

Making the case for data archiving: The changing “value proposition” of social science data archives

Kristin R Eschenfelder

*University of Wisconsin-Madison, USA.
eschenfelder@wisc.edu*

Rachel D. Williams

*Simmons College, USA.
rachel.williams@simmons.edu*

Kalpana Shankar

*University College Dublin, Ireland.
kalpana.shankar@ucd.ie*

ABSTRACT

In this paper, we analyse how three social science data archives (SSDA) have adapted to provide value to different stakeholders over time. Drawing on historical administrative documents and interviews with current and former staff at three long standing SSDA, we examine how these archives have provided value, to whom, and how they have situated themselves vis-à-vis alternative archives over 20-40 year time spans. Although data archives have been in operation for decades, how they have sustained themselves over time by continuing to provide value to stakeholders in changing conditions is less well understood. Studies of value have tended to focus on a snapshot in time rather than providing a view that emphasizes change over time. We conclude with a comparative analysis of changes in value across archives and suggestions for future work

KEYWORDS

Data archives, sustainability, organizational change, business models, value proposition

INTRODUCTION

Repositories for the curation and management of data are an essential infrastructure for research across the disciplines. Although they have been in operation for decades, how they have sustained themselves over time (and can continue to do so) has only been explored in recent years. As these repositories have grown in scale and scope, they have been dependent upon heterogeneous sources of funding and have made structural and programmatic changes to accommodate shifts in user needs, technologies, and institutional relationships.

This paper explores the changes in the perceived value of data archives over time. The role of data archives has been well documented for their role in spurring research activity in not-for-profit and commercial enterprises and reducing the cost of secondary research (Beagrie and Pink 2012). However, most attempts to demonstrate the value of data archives have been at a specific point in time, a snapshot, whereas data have changing value: Some argue that the longer the investment, the more valuable data collections become (OECD 2017). Nevertheless, making the case for value and impact over time can be challenging. As Maron and Loy (2011) note, the transition from grant funding to a longer-term plan for ongoing growth and development of a repository or archive may seem just like a financial problem, but actually runs deeper since projects must justify their ongoing existence to depositors, users, hosts, funders, and others. Without a value proposition, digital resources, many of them representing the outlay of significant resources (financial and institutional), are likely to become siloed or unusable.

One of many challenges archives face is the complex, fluctuating, and ill-defined organizational concept of “sustainability.” Some definitions of organizational sustainability emphasize structural components, referring to “an economically viable infrastructure, both social and technical...” or “business, social, technological, and policy mechanisms.” (Berman & Lavoie 2011; Bradley 2007) Others focus on resources, such as the ability “to generate or gain access to the resources-financial or otherwise.” Other definitions emphasize sustainability as a culture or set of organizational competencies and skills that include ongoing engagement with user communities and other stakeholders, being prepared for an environment of constant change, demonstrating accountability, and managing risk (Guthrie, Griffiths, and Maron, 2008). Further, while definitions allude to data-related services, they vary with respect to the set of services they assume ought to be provided sustainably and how.

Another challenge is the complex, fluctuating and ill-defined nature of “value.” What is the relationship of “value” to creating and managing a sustainable data archive and more importantly, how does “value” change over time? Further how do perceptions of value vary between different stakeholders or users? The idea that value underpins any business model, and the understanding that different stakeholders might value different aspects of a program are well accepted (OECD 2017); however, the evolution of value and what stakeholders’ value over time has been less examined. In this paper we first summarize key ideas from the literature of business models about the role of value in business models. We then present the results of an empirical study of three large social science data archives, and we describe major changes to their value for different users and stakeholders over a 40 year

period. This analysis of change in value over time is one part of a larger historical comparative case study of social science data archives (Eschenfelder & Shankar, 2017) and draws on and extend socio-technical approaches that emphasize how different stakeholders may have very different perceptions of the success or failure of technologies (Bijker, 1995).

VALUE PROPOSITIONS, BUSINESS MODELS AND DATA ARCHIVES

There is little agreement on what a business model actually consists of, and the exact role of value in a business model (DaSilva and Trkman 2014). Some argue that this stems from the newness of the term “business model” which only developed as the internet increased the number of options an organization could choose to organize itself and offer services (DaSilva and Trkman, 2014). In a structured review of the literature of business models, Zott, Amit, and Massa (2011) also found there is no agreement on what a business model is, and that researchers tend to adopt or create a definition instrumental to their particular research needs. One similarity that they did find across the literature reviewed is that business models are routinely invoked to understand/explain how business create value for stakeholders.

The intersection of business models and the creation of value in cultural and scientific organizations (which have different mandates than for-profit firms), has recently received increased interest. But most current work on business models of data organizations has focused on revenue generation and diversifying sources of income. The concepts of “value” or the “value proposition” offered by of data archives and data sharing are more lightly discussed. For example, the Research Data Alliance/World Data System (RDA-WDS) Cost Recovery Interest Group explored current and potential future alternative income streams via structured phone interviews with representatives from 22 mostly science data repositories globally (Dillo, Hodson, & Waard 2016). The European Commission Horizon 2020 Advisory Group on European Research Infrastructure report emphasizes how “flexible business models” are “essential to keep [research infrastructure] sustainable in the long run,” calling for creation of innovative and flexible business models. (European Commission 2015). In 2017, the Organization for Economic Co-operation and Development (OECD) released a report on the business models of 47 data archives currently in operation across eighteen countries and numerous disciplines. While mostly reporting on income streams, the OECD’s report recommended that all repositories develop a clearly articulated business model and that the model should change over time for many reasons, including the changing needs of stakeholders. The OECD suggests that value-added mixes of data and services should evolve over time in reaction to changes in the “market” for those data and services, and numerous other cost drivers (change in user base and expectations, to give one example). They write, “Engaging and maintaining structural, institutional, philanthropic, or other funders depends on their understanding of the value proposition, and ensuring such engagement may involve repositories undertaking detailed benefit/cost, value, and impact analyses....Sustainability depends, inter alia, on a clearly articulated value proposition and the development of a ‘business model’ (p. 9).

Another gap in the value discussion is the lack of empirical research on different stakeholders’ perceptions of how value is created and maintained in data archives and how those perceptions might change over time. Ember and Hanisch (2013) take as given that data archives are “valuable” to their constituent communities and that the involvement of the community serviced (traditional disciplines or sub disciplines, networks of researchers, among others) is essential to the ongoing functioning of a data archive. This stands to reason; if the community values the resource, it will engage in developing and sustaining it (Michener et al 2012). However, they are forced to surmise that other ways that archives could add value, such as certification or commercial partnerships) have not proven to be as useful to the long-term success.

To explore changes in perceptions of value over time, we draw on an approach by Osterwalder. Osterwalder (2005) defines a business model as “a set of elements formed by a logical plan and a purpose to create and drive value.” His popular Strategyzer “Business Model Canvas” addresses nine components including value propositions. (Osterwalder et al 2018) In this paper we focus on the “value propositions” which are defined as bundles of goods or services that create value for a given customer segment. Important to our paper, Osterwalder emphasizes that different customer segments may have different value propositions because they are trying to achieve different goals, or as Osterwalder describes “get different things done in their work or in their life,” they have different “pains,” or things that annoy or thwart them, and they have different “gains,” or desired outcomes or benefits. (Osterwalder 2018). We also draw on work by Morris et al (2005) that analyzed 30 definitions of business models and identified three emphases: (1) economic aspects of an organization and how it creates revenue, (2) operational aspects of an organization and how it arranges work processes, and (3) strategic issues such as positioning in relation to competitors. Finally, we draw on socio-technical approaches that emphasize how different stakeholders may have very different perceptions of the success or failure of technologies, and the processes through which those perceptions of success or failure may come to change over time (Bijker, 1995). We combine these ideas to ask a series of questions about how data archives have created value for the different sets of stakeholders in relation to alternatives over decades.

Question 1. How has the data archive created value? What products or services has the data archive offered? Does the data archive provide access to other’s data, its own data? Through what channels has the archive distribute its products services (direct distribution/indirect distribution)?

Question 2. For whom has the data archive create value? Who are the different types of direct and indirect users of the archive or stakeholders?

Question 3. How has the archive competitively positioned itself vis-a-vis alternatives over time? For example this might include the image the organization creates of itself in relation to alternatives, including alternatives of free data or no data, in terms of costs, quality or features available.

METHODOLOGY

This paper is part of a larger study on the sustainability of social science data archives (SSDA). We selected SSDA that have been in continuous operation for decades in order to examine change in the organizations, their data collections, their services and their stakeholders over time. We used purposive sampling to select long-lived archives with contrasting revenue models and whose organizational documents would be in the English language. Both ICPSR and LIS collect membership fees, while UKDA have received multi-year grants from a UK national research funder. The authors limited analysis from inception of each archive to 2002. We chose 2002 as an artificial if reasonable ‘historical’ cut off although in many cases we have later public data from annual reports and other posted materials. Our decision to end data collection in the early 2000s was based on a need to assure participants we would not be evaluating current archive practices, and a desire to focus on analysis of change over time rather than get caught up in the excitement of contemporary events.

ICPSR, or the Inter-university Consortium of Political and Social Research has been located at the University of Michigan – Ann Arbor. We collected data on operations from its founding in the early 1960s to the mid-2000s. 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 has operated as a membership organization and its fees were modelled off journal pricing with larger institutions paying a larger membership fee to provide access to all faculty, students and staff. But, since the 1990s, ICPSR has generated the majority of its revenue from serving as a contract data platform for US government agency offices. Further, it applies for external grants to fund special projects.

The UK Data Archive (UKDA) is now part of a cross-institutional UK Data Service. The UKDA was founded in 1967 at the University of Essex. For the purposes of this 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. We collected data on UKDA operations from its founding in 1967 to the mid-2000s. The UKDA has operated with the support of regular large multi-year grants from the major UK funder, the Economic and Social Research Council (ESRC). As part of this arrangement, the UKDA has provided access to its data to all qualified UK educational institutions. The UKDA supplements the ESRC funding with additional grants for special projects and contract work. In the 2000s the ESRC modified its support to fund UKDA and several other UK data services in a more coordinated fashion as the UK Data Service.

The LIS Cross National Data Center (formerly known as the Luxembourg Income Study) began in 1979 in Luxembourg. It has a more thematically focused data collection: LIS harmonizes and provides access to cross-national microdata on individual and family financial data from 50 different countries. We have data on LIS operations from the mid-1980s (its inception) to the early 2000s. LIS is a membership fee driven organization, however LIS memberships are at the national rather than the institutional level. One organization pays a membership fee for a country and then all academic institutions in that country can use LIS data. In some cases the national membership fee is paid for by a governmental science funding agency (e.g., NSF, ESRC), but in other cases it may be paid for by a government agency or an individual university on behalf of others.

In addition to gathering documents, we conducted semi-structured interviews with former and current staff with roles including directors, collections managers, membership services coordinators and technology managers. Interviews took approximately 45 minutes and were both in-person and phone based. A preset list of topics guided analysis of both documents and interviews, but we remained open to new themes. Because our data stem from sources internal to the archives, readers should remember that findings about value or changes in value stem from documents predominantly written by archive staff or governance bodies and thus represent their perception of the perceived value of the archive in the eyes of stakeholders. In some cases we have external review reports and other materials that provide an external view of value. In the findings below we highlight snapshots of changes that each of our three sites enacted related to our three value questions.

FINDINGS

Based on our socio-technical approach that emphasizes how different stakeholders may have very different perceptions of the success or failure of technologies (Bijker, 1995), we begin by exploring how our case study sites believe they have created value for different stakeholders over time. We then describe the different stakeholders for whom our case study data archives perceived they might create value. Finally, we look at how our case sites believed they positioned themselves against a variety of competitors for their stakeholders’ resources of money, usage, or enthusiasm. Because of space constraints, we include a few examples in each area, and we primarily chose examples that illustrate change in value.

Question 1. How has the data archive created value?

ICPSR

Since its inception, ICPSR has maintained that its value stems from stewardship and curation of collections of high-quality social science data. The proposition is that ICPSR institutional members will choose to pay membership fees to access this aggregated, indexed, cleaned and highly-described collection. ICPSR has also positioned itself as providing value to researchers looking to deposit their data to ensure responsible access by others. ICPSR promises that for each deposit, "... a data archivist reviews the data and documentation, builds a study description, enhances the documentation...and archives the data for long-term preservation." (ICPSR, 2018). ICPSR places the data within a larger search and indexing structure, and will also modify the data to protect the confidentiality of participants ICPSR. Through these services, it arguably adds value to researchers "raw" data that they could otherwise simply post in a more isolated fashion to the Web.

In its early pre-internet days, ICPSR promoted its value as a distribution mechanism. By depositing data with ICPSR, a researcher could rest assured that peer scholars could get access to the data via punch card and then tape drive distribution systems. Further, researchers could order copies of any data in the ICPSR collection for delivery via card, tape and then CD-ROM. But the advent of computer networking and internet sharing protocols challenged all SSDA's distribution value. By 1994, staff expressed concern that scholars found FTP is "more efficient" than ICPSR's traditional distribution systems (ICPSR 1994). Countering concerns about FTP in its 1995 Annual Report, ICPSR argued that "providing FTP access is not the same as preserving it for posterity or making [data] ready for ...use" and it urged members to continue to support the shared ICPSR collection (ICPSR 1995b pg 5).

While data quality has long been part of ICPSR's value claim, at times stakeholders have pushed for faster and more open access to data than curation would permit. This pressure for fast access to "as is" lower-quality data created complications with ICPSR's membership model which reserved access to data to paying members. To soften this tension and create value for different sets of stakeholders, ICPSR began to tier data, doing different levels of work on different data, and providing different types of access to the different tiers. Calls for open data however continue to draw the value proposition of access to a high quality collection into question. In interviews staff members worried that stakeholders might not no longer appreciate the work that ICPSR does and just use lower quality open alternatives (personal communications).

ICPSR also promotes itself as an R&D organization whose products provide value to other data archival organizations by furthering standards development, technologies and best practices in curation. For example, staff developed early statistical analysis software called OSIRIS, an early desktop statistics application called ABC, and ICPSR was a major player in the development of the SSDA community's development of the DDI metadata standard and markup language (ICPSR 1995).

UKDA

In the first few years the archive's founding, UKDA staff were primarily interested in exploring technical questions related to the feasibility of data archiving mostly of interest to other data professionals (Farlie 1971). But by 1971, after negative feedback from funders, it began to work to change the perception of the archive "an unresponsive self-centred organization" to one more "anxious and able to serve clients' needs" (Annual Report, 1982). UKDA leadership sought to create value by acquiring data that would be otherwise difficult to obtain, prioritizing data of "importance" rather than quantity (UKDA 1971). Also central to UKDA's value claims was high quality processing, and creation of extensive documentation that made data more accessible to users, and creation of collections of related surveys for temporal trend analysis. Further, by the mid-1980s, UKDA leaders portrayed the archive as a key player in facilitating data exchange between different countries, pitching the value of the archive as a data broker with other archives (Annual Report, 1983).

Like other archives such as ICPSR and LIS, UKDA did provide some training programs and seminars to gain more users and increase understanding of users' needs, but its ability to create value was more centered on other areas (UKDA 1982). By the 1980's UKDA was examining how to expand value through identifying new types of data to acquire or entertaining the possibility of commercial users. For example, UKDA began preserving data for time series and trend analysis (UKDA 1984). Like ICPSR, UKDA also created value by being responsive to technology trends and sought to increase value by changing how data was made available to users. The introduction of new technologies in the 1990s saw data usage double in the span of just three years.

LIS

LIS had a more narrow and niche data collection. To create more potential users, it developed extensive training programs to introduce new researchers to their data, generate demand for their data, and increase its perceived value within the scholarly community examining household wealth. LIS has had a stated goal of introducing "young" scholars to existing research, promoting new lines of inquiry that could make use of the data, and facilitating teaching using the LIS datasets (LIS 1988).

Like the other cases, LIS has also generated value through its curatorial, custodial and documentary activities that add value to data originally produced by national statistical agencies. The curatorial value propositions of the LIS Data Archive has three

elements: microdata, cross-national and harmonization. First, while there are many sources of national-level economic data, fewer focus on household-level economic microdata such as household income, debt, spending etc. Second, LIS gathers microdata from many different national sources, saving the researcher a great deal of effort. In many cases these data are controlled by national government statistical agencies and it takes negotiation, relationships of trust, and the development of a secure access system in order to get the data out of its home country and provide remote access. “This time-consuming process usually requires negotiating national privacy restrictions, and often entails years of building and maintaining relationships” (LIS 2018). Thus, while country Z would have copies of its internal national data, and a researcher from country Z could get access to Z data, that researcher would not have access to similar data from other nations. A manager explained why national agencies will release microdata to LIS but not individual researchers, “... they can’t because data providers won’t release data to individuals. It has to be behind our protective firewall.” (LIS 2004) LIS’s value proposition has been to develop a consolidated collection, including data from multiple countries, on a secure platform, which ensures data use comports with the expectations of the data providers. Finally, data from different nations requires “harmonization” or the creation of common variables and normalized measures across data sets so that researchers can compare similar phenomena across nations. As LIS describes, “this process currently takes anywhere from one to three months per country dataset.” (LIS 2018).

Question 2. For whom has the archive created value?

ICPSR

ICPSR initially conceptualized itself as creating value for scholars in U.S. political science departments by collecting voting, poll and other U.S. election-related data (ICPSR 1963). This collection of cross-state elections data was more valuable in aggregate; further, staff increased the value by cleaning the data and formatting it for the statistical processing software of the day. But by 1971 ICPSR documents discuss the need to broaden the archive’s user audience and the range of relevant granting agencies. Staff propose changing the archives name to refer to Political *and* Social Research instead of just political research (ICPSR 1971). The eventual formal name change in 1975 was preceded by broadening collection development and grant activity in the early 1970s to curate data on a broader range of social science issues including aging, law and criminal justice, victimization, public health, demography, and history.

Additionally, throughout the 1970s and 1980s ICPSR expanded the audience for whom it created value by offering data archiving and distribution services to government agencies. This provided an important and significant new source of revenue, but it also created identity confusion. Meeting minutes describe how by 1996, some felt that agency revenue created a “confused organizational identity” and conflicted with ICPSR’s original mission to provide data to (paying) members. Meeting notes express concern that ICPSR might be more responsive to agencies because of the financial importance of grants vs membership fees (ICPSR 1996). Our analysis of budgets over time shows that by the early 2000s, contracts and grants overshadowed membership fees as a revenue source.

As ICPSR sought to retain membership revenue during periods of campus budget cuts, ICPSR actively sought to cultivate a new audience of deans and other budget decision makers in order to convince them to continue paying (ICPSR 1994). ICPSR had initially saw “organizational representatives” (OR) -- often librarians or staff -- as their primary value advocates on campus. But given falling memberships, minutes express concerns about OR’s unwillingness to “argue with deans” to keep memberships. By 1990 ICPSR created an advisory group to figure out how to better communicate ICPSR’s value to campus budgetary decision makers.

UKDA

Due to UKDA’s funding through the ESRC, the archive has sought to create value for the ESRC as defined by the terms of its funding agreements which specified a balance between educational users, government and commercial users. Its primary focus has been on UK academic users: Like ICPSR, it provided access to secondary data for research and education. It accepted deposit from individual researchers, and sought to capture data collected from studies funded by ESRC. UKDA’s funding applications provide a window on its changing value proposition: UKDA has prepared grant applications every 5 years that make a case for its value for ESRC; further, ESRC has an external review panel assess the UKDA grant and provide feedback on reviewers’ perceptions of UKDA’s value.

UKDA worked to develop close relationships with government agencies that create and statistical and other data. From its inception, the UKDA was conceived of, in part, as “centre for storing data derived from social surveys sponsored by government departments...” (UKDA 1967) Getting government agencies to deposit data however was not always easy because of the need to build confidence with depositors, concerns about confidentiality of data, and the difficulty of transferring data. Early ESRC grant assessments critiqued the archive for low overall use, and especially low use by government agencies and potential commercial users (UKDA 1970). But UKDA argued that by providing a central means of disseminating government data, it relieved agencies from that work, saved money from “public purse,” and made it easier for data seekers to locate data

from different agencies (UKDA, 1980). UKDA continued to curry governmental and commercial users who would be interested in using the archive's data, particularly Census data and other housing and income datasets (UKDA, 1982).

Also like ICPSR, UKDA saw part of its value as stemming from research and development, and the archive worked on new retrieval techniques and data analysis software (UKDA 1978). UKDA also facilitated access to data for users overseas (ESRC 1988). The archive thus saw itself as providing access to a variety of users, primarily academic, but with a broad range of disciplines and interests.

LIS

LIS differed from ICPSR and UKDA archives in that it never offered a means for individual researchers to deposit and distribute their data. Its goals have been to gather a very specific type of data from national statistical and research offices. Further, while it is similar to ICPSR and UKDA in that it solicits data from government agencies -- albeit a much narrower set -- it does not market itself as a long term data preservation service for those agencies. Instead, it markets the value of its harmonized cross-national comparative data to both researchers and government agencies who might use the data to conduct comparative policy analysis (LIS 2017).

For the organizations that provide the national-level membership fees, LIS claims that it creates value through facilitation of better policy analysis about the effects of different kinds of national programs on poverty, income adequacy, and economic well-being generally. But, this value has been contested: for example, LIS records show how in the mid to late 1990s, the US National Science Foundation complained that not enough US economists were using LIS data, drawing NSF's continued payment of the US membership fee into doubt (LIS 1997). In another example, in interviews LIS staff described how in many cases the payment of annual fees for LIS is contingent on personal relationships with budget decision makers in government agencies. Change of personnel can disrupt the revenue as new decision makers may not be aware of, or value the LIS data collection (personal communications). As one described, new decision makers would ask "what is LIS? I just took over this job and I have this portfolio and I'm supposed to be giving you money for some data organization in Luxembourg? I don't know who you are." (Personal communication) LIS archive managers described spending a great deal of time cultivating and maintaining relationships with budget decision makers in order to ensure they appreciated the value of LIS membership and use of LIS data by policy analysts in their country.

LIS has always sought to expand the pool of policy analysts for whom it creates value by increasing the range of national data sets it contains. This draws new users to the resource, but also expands the breadth of comparative analysis possible. LIS has also expanded its membership pool to include international organizations such as the OECD. Finally, LIS has also received inquiries about whether they would consider serving as a data platform for other groups, and it is considering providing this service to research programs in intellectually aligned areas as a means of both diversifying revenue and supporting financial microdata research (personal communication).

Question 3. How has the archive competitively positioned itself vis-a-vis alternatives over time?

ICPSR

ICPSR and other social science data archives in the US purposefully attempted to distinguish themselves from one another through collection specialization with a set of early US data archives planning to divide up specific topical areas (Converse P. 1963). Among archives that charged membership fees, like ICPSR and the Roper Center for Public Opinion, disputes sometimes arose when one archive copied and began offering access to another's data.

In another example of positioning, as ICPSR sought to retain memberships and influence campus budget decision makers, it actively sought to reframe itself less as a club with membership fees and more as 'similar to a journal subscription.' It did this because staff believed that budget decision makers were already familiar with journal subscription fees and would more readily continue payment to ICPSR under this mental model (ICPSR, 1974). As part of this, ICPSR suggested that member campuses move membership fees from academic department budgets (e.g., political science) to the library because while the membership fee seems big in the context of a department budget, it is small in the context of a library budget (ICPSR, 2000)

In a final example, from the late 1990s onwards, some stakeholders have pressed for faster access to key data sets, and more open access to data. These pressures conflicted with the long-standing quality value proposition provided by the curation and documentation activities. The open access to data conflicted with ICPSR's members-only restrictions. In response to these pressures, in 1994 ICPSR developed its "FastTrack" FTP service that provided "as is" access to high-demand data to members. That same year, in response to pressure from a funder and PIs who wanted immediate open access, ICPSR provided non-member anonymous FTP access to certain federally funded data sets. Documents justify the change as necessary to avoid those data "being disassociated with us" and as necessary to avoid having PIs distribute their own data (Rockwell, 1994). ICPSR has provided both high quality curated data and low quality "as is" data to provide value in comparison to alternatives.

UKDA

For UKDA, being competitive over time involved having a thorough understanding in how its competitor archives operated and ensuring UKDA maintained its niche among competitors. Early after its creation, UKDA conducted a comparative analysis of the international landscape of existing data archives, and how UKDA could fit into that landscape (Land, 1970). For example, one of the ways in which, at least initially, UKDA was distinct in particular from other American archives was in its emphasis on data use by researchers, rather than students. It later changed that approach by bringing datasets into the classroom and providing additional training opportunities for students.

Because UKDA serves primarily UK educational institutions and government agencies, its primary alternatives are other UK data archives. It risks losing users, and ESRC support, to alternative UK data archives. Over time, ESRC grant review panels noted pros and cons of ESRC investment in a more centralized Essex operation, versus spreading out investment amongst several data archives in a more federated approach. Grant reviews in the 1980's describe how ESRC chose Essex as a location for a data archive over Manchester because of the computer processing facilities at Essex. (UKDA 1980) But by 1994, grant reviewers expressed concern that reduction or loss of ESRC funding was a possible threat. They describing how the structure of the ESRC funding as a "research center competition" encouraged other UK campuses to bid to create alternative data archives (UKDA 1994).

LIS

LIS does not tend to have the "open data" alternative because in most cases the data sets they receive from national agencies are restricted, and also because of the harmonization they provide. But they perceive that some stakeholders may dismiss LIS because the stakeholders' assume that some open data alternatives must exist. As one manager explained in an interview "We used to get [that question from X], and it made us want to throw ourselves out a window "Why can't individual researchers just do what you do?" Take 50 data sets and harmonize them? <Incredulous laughter>." (Personal communication). Another challenge is that some potential funders do not adequately understand the differences between microdata and other types of economic data. Many other archive projects exist that include other types of economic data, and funders may erroneously believe that redundancy in service exists between LIS and these other projects.

From the perspective of national agencies that provide data to LIS, an alternative is simply to not provide the data. Review of LIS meeting notes show how historically some government statistical agencies have perceived submission of data LIS as a risk. Many concerns have centred on privacy protections (LIS 2004). In order to satisfy data contributors' privacy concerns regarding remote access to digital data, LIS built a sophisticated access and use control system (LISSY) that would provide a method to allow researchers to use data under restrictions named by data contributing nations. But technology has not completely solved the problem, and in recent years LIS has lost former data contributors due to changing data privacy regulations and new interpretations of existing regulations. Agencies that formerly provided data will now not contribute data to LIS despite the robust LISSY infrastructure (personal communication).

DISCUSSION

In the business literature, it is taken for granted that adapting business models to changing circumstances is essential to the success of an organization (Wirtz, Pistoia, Ullrich & Göttel, 2016). While the literature presents business models as having varied constitutive elements, all business model frameworks include an element of "value." Further, some scholars like Zott, Amit, and Massa (2011) make value a central organizing principle for business model planning. To explore changes in value perceptions of data archives over time we asked three questions about value across three case sites including: How does the data archive create value? For whom does it create value? How does the data archives position itself in comparison to alternatives over time? In this section we present the questions in a different order because identification of the stakeholder for whom the archive is creating value drives answers to the other two questions. Drawing on socio-technical theorizing (e.g., Bijker, 1995) about stakeholder variance in the perceived success or failure of technologies, we draw out how different stakeholders may have very different expectations for, or perceptions of, the value of an infrastructure like data archives. The value is relational and based on the stakeholder.

For whom has the archive created value?

All of our data archives have attempted to create value for a similar set of stakeholders including scholars who create data, scholars who use data, social science methodology students and teachers, other organizations like government agencies that need curation services, practitioners in the field of data archiving, and funders. While each archive was developed with a core set of users and uses in mind, all of them expanded that user set over time by adding data in completely different topic areas (ICPSR & UKDA) or from more nations (LIS). Further, all our data archives interacted with government agencies as stakeholders, but in different ways. For ICPSR, agencies were a new source of revenue or a new service. UKDA served government agencies from its inception and as part of its core ESRC funding. For both UKDA and ICPSR, in-nation government agencies were data providers, but LIS had the additional challenge of trying to move data across national boundaries. Funding agencies

(especially ESRC and NSF) were stakeholders for whom our data archive created value. But, that value was often in question and data archives risk being passed over for funding or having funding reduced. Another commonality we saw was that each archive has sought to justify its value to budgetary decision makers that might not be end users of the data. For example, ICPSR worked to increase name recognition among campus budget decision makers, LIS works constantly to maintain and renew relationships with organizations that pay national membership fees. UKDA must convince its major funder of its impact on research and education every five years. Data archives must prove value to these stakeholders even if they are not direct users of the data.

How have the data archives created value over time?

Stakeholders have varied expectations about, and perceptions of, the potential value of the data archive. The question of how the archive creates value is a function of the specific stakeholder in question. Our data archives adopted some common strategies for creating value including developing data collections to attract academic users including ensuring data quality, data documentation, educational endeavors to promote data use, adopting preferred technology and access methods, developing new software services, providing search and indexing tools, and managing privacy and security issues. Further, some value strategies were unique to particular archives. For example while all archives grouped related data such as time series, panels etc., the LIS data archive was unique in harmonization work it provided to facilitate cross-national analysis. Further, all our case sites sought to create value for government agencies, but in very different ways. UKDA sought to relieve agencies of the work of curating data as part of their ESRC remit. ICPSR also sought to relieve agencies of curation work, but for a contractual fee. LIS did not provide curation services for agencies, but through inclusion of a country's data in the comparative database, it provided that country's policy analysts with potential new comparative analyses.

How does the data archives position itself in comparison to alternatives over time?

Our case archives faced a number of different alternatives or competitors for resources over time including other data archives. What counts as an alternative is also driven by the perspective of the stakeholder and the potential value the archive might create for that stakeholder. One very important alternative facing all our cases was “free data.” All our archives have distinguished themselves from the free data alternative by pointing to the quality of their data product, their documentation and their data deposit services to ensure privacy and distribution. Further, the archives will argue that they provide more data management and preservation services than free data alternatives. Another very important data alternative to deposit of data in our case study archives was “no data.” By this we mean that potential depositors (e.g., PIs or government agencies) may simply forgo data contributions. Or, budget decision makers paying membership fees may choose not to pay, and not have access to members-only data.

CONCLUSION

In this paper, we explored three historical case studies of how data archives create value over time, for which stakeholders they create value, and how they position themselves against alternatives such as free data and no data. We used a documentary analytic approach from three social science data archives that have been curating data for decades to explore how perceptions of their “value” changed over 40 years.

This paper provides entry into two fruitful arenas of research. One is temporality and the long-view. Data archives and similar organizations have been creating and articulating value through the evolution of their business models over time; the literature has focused to date on how this happens at discrete points in time instead of examining them over time (OECD 2017). We would suggest that there is now a sizeable body of work upon which to build research that takes the long view. Secondly, this paper suggests that a focus on business models as an aggregation of resources, institutions, data, content, and technologies (among other factors) would provide one means of extending research on information infrastructures beyond the singular cyberinfrastructure/e-science platform case study.

We conclude with a provocation. In the case of sustainability in digital repositories, the assumption persists that infinite sustainability is a universal good (and thus is a “value” in and of itself – perhaps the greatest one?). This assumption places limitations on a discourse that should be more critically considered. Because digital technologies have made it possible to archive almost all content and because there are always costs associated with maintaining any repository, information professionals are in the position not just to advocate for the creation of infinitely sustainable archives but to also consider the needs of repository users recognizing that some repositories may not need to be infinitely sustained. Just because repositories CAN last, should they? To have substantive and thoughtful discussions about what and how long particular kinds of digital content should be retained, we cannot rely on a shared assumption that longer is better. Instead, we must consider how long is good. And sometimes, that what is good might not necessarily be forever.

ACKNOWLEDGMENTS

This study was funded by the Alfred P. Sloan Foundation G-2014-14521. Special thanks to Allison Langham for assistance with data analysis.

REFERENCES

- Bauin, S., & Rothman, H. (1992). "Impact" of journals as proxies for citation counts. In P. Weingart, R. Scharinger, & M. Winterhager (Eds.), *Representations of science and technology* (pp. 225-239). Leiden: DSWO Press.
- Beagrie, N., & Pink, C. (2012). "Benefits from research data management in universities for industry and not-for-profit research partners", White Paper, Retrieved from <http://opus.bath.ac.uk/32509/>
- Bijker, W. (1995) *Of Bicycles, Bakelites and Bulbs: Towards a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
- Berman, F. & Lavoie, B. (2011). Economics and digital preservation: Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. July 2011. Retrieved from: <http://www.digitalpreservation.gov/meetings/documents/ndiipp11/Berman.pdf>
- Bradley, K. (2007). Defining digital sustainability. *Library Trends*, 56(1), 148-163.
- DaSilva, C.M.; Trkman, P. (2014). "Business model: What it is and what it is not", *Long Range Planning*, 47(6), 379–389.
- Dillo, I., Hodson, S., & de Waard, A. (2016). *Income Streams for Data Repositories*, White Paper, February 28, 2016. Retrieved from <https://zenodo.org/record/46693#.WsSj59PwaV4>
- Ember, C. and Hanisch, R. (2013), "Sustaining domain repositories for digital data", White Paper, ICPSR, University of Michigan-Ann Arbor, Michigan, 11 December 2013. Retrieved from: https://rd-alliance.org/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf
- Eschenfelder, K. Shankar, K. (2017) *Organizational Resilience in Data Archives: Three Case Studies in Social Science Data Archives*. *Data Science Journal*, 16, p 12. DOI: <http://doi.org/10.5334/dsj-2017-012>
- Gambardella, A. and McGahan, A. M. (2010). Business-model innovation: General purpose technologies and their implications for industry structure. *Long range planning*, 43(2-3), 262-271.
- Guthrie, K., Griffiths, R., & Maron, N. (2008). Sustainability and Revenue Models for Online Academic Resources. *Strategic Content Alliance*, (May), 66.
- LIS Cross National Data Archive (2018) "Services". Retrieved April 1, 2018 from <http://www.lisdatacenter.org/about-lis/services/>
- LIS Cross National Data Archive (2018) "About LIS". Retrieved April 1, 2018 from www.lisdatacenter.org/about-lis/.
- ICPSR (2018) "Start Sharing Data" Retrieved April 3, 2018 from <https://www.icpsr.umich.edu/icpsrweb/deposit/>
- Maron, N. L., & Loy, M. (2011). Revenue, Recession, Reliance: Revisiting the SCA/Ithaca S+R Case Studies in Sustainability.
- Michener, W. K., Allard, S., Budden, A., Cook, R. B., Douglass, K., Frame, M., ... & Vieglais, D. A. (2012). Participatory design of DataONE—enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*, 11, 5-15.
- Morris, M.; Schindehutte, M.; Allen, J. (2005). "The entrepreneur's business model: toward a unified perspective", *Journal of Business Research*, 58, 726–735.
- OECD Global Science Forum. (2017) "Business models for sustainable research data repositories." *OECD Science Technology and Industry Policy Paper No.47*. Retrieved from: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/STP/GSF\(2017\)1/FINAL&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/STP/GSF(2017)1/FINAL&docLanguage=En)
- Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). "Clarifying business models: Origins, present, and future of the concept". *Communications of the association for Information Systems*, 16(1), 1.
- Osterwalder, A.; Pigneur, Y.; Papadacos, T.; Bernarda, G.; Smith, A.; Smith, A. (2018) "Value proposition design : How to make stuff people want". John Wiley & Sons: Hoboken, NJ
- Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Shankar, K.; Eschenfelder, K.R.; Downey, G. et. al. (2014) "Social Science Data Archives: Case Studies in Data Sustainability" Poster at the 2014 iConference.
- Trimi, S., & Berbegal-Mirabent, J. (2012). Business model innovation in entrepreneurship. *International Entrepreneurship and Management Journal*, 8(4), 449-465.

- Wirtz, B. W., Pistoia, A., Ullrich, S., & Göttel, V. (2016). "Business models: Origin, development and future research perspectives", *Long Range Planning*, 49(1), 36-54.
- Zott, C.; Amit, R.; Massa, L. (2011). "The business model: Recent developments and future research", *Journal of Management*, 37(4), 1019–1042.
- REFERENCES – archival sources
- Converse, Philip E. (1963) Memo to Committee of 8. [Minutes of Meeting Ann-Arbor Dec 1-2, 1962] , 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- ESRC (1988). 1988-1989 ESRC Data archive annual report [planning document]. UKDA Administrative files, University of Essex UK.
- Farlie, D. J. C. (1971). Memo on review of the present and future of the SSRC Data Bank. UKDA Administrative Files. University of Essex UK.
- ICPSR (1963) Annual Report 1962-1963. 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- ICPSR (1971, October) Council Meeting Minutes [minutes] 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- ICPSR (1974, March) Council Meeting Minutes [minutes] 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan.
- ICPSR (1994, October) Council Meeting Minutes [minutes]. 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- ICPSR (1995) Journey to the 21st Century: Towards a Strategic Plan for ICPSR [manuscript] Ann Arbor, MI: ICPSR.
- ICPSR (1995b) Annual Report. 1994-1995. 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- ICPSR (1996) The Future of ICPSR [Long Range Planning Subcommittee Memo] 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan.
- ICPSR (2000, June) Council Meeting Minutes [minutes] 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan
- Land, F. (1970). 1970 Report to the SSRC Data Council.
- LIS (1997, Jan 22) Minutes for Sixth LIS Biennial Meeting 21 & 22 July 1997, LIS Headquarters, Differdange, Luxembourg.
- LIS (1998, July) Luxembourg Income Survey. (1)2, 6. UKDA Administrative files, University of Essex UK.
- Rockwell, R. (1994, Nov 17) Distribution of AHEAD Data. [email to ICPSR Council] 0071 Bimu 2, ICPSR publications, Bentley Historical Library University of Michigan.
- LIS (2004, May 28) Tim Smeeding to Carolyn Kennedy [Letter]. LIS Administrative Files.
- UKDA (1967) Social Science Research Council Databank. [Grant to the ESRC], UKDA Administrative Files, University of Essex UK.
- UKDA (1970, March) Minutes. [Minutes SSRC Database Council] UKDA Administrative Files, University of Essex UK.
- UKDA (1978). SSRC Data archive: Report and proposal from the director [Report]. UKDA Administrative Files, University of Essex UK.
- UKDA (1980) Report on the Review Panel on the SSRC Archive [Report], UKDA Administrative Files, University of Essex UK.
- UKDA (1982, October). UKDA Annual report [Report], UKDA Administrative Files, University of Essex UK.
- UKDA (1971, Jan 15). Summary of major points arising from Databank meeting [minutes], UKDA Administrative Files, University of Essex UK.
- UKDA (1983, October). UKDA Annual report [report], UKDA Administrative Files, University of Essex UK.
- UKDA (1984). Director's report [report]. UKDA Administrative Files, University of Essex UK.
- UKDA (1994, November) Review of the ESRC Data Archive [Report]. UKDA Administrative Files, University of Essex UK.