DOI: 10.1002/asi.24691

Revised: 21 May 2022

### RESEARCH ARTICLE



# The financial maintenance of social science data archives: Four case studies of long-term infrastructure work

<sup>1</sup>Associate Dean for Computer, Data & Information Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>2</sup>School of Information and Communication Studies, University College Dublin, Dublin, Ireland

<sup>3</sup>School of Journalism and Mass Communication and the Information School, University of Wisconsin-Madison, Madison, Wisconsin, USA

### Correspondence

Kristin R. Eschenfelder, Associate Dean for Computer, Data & Information Sciences, University of Wisconsin-Madison, Madison, WI, USA. Email: eschenfelder@wisc.edu

**Funding information** Alfred P. Sloan Foundation, Grant/Award Number: G 2014-14521

Kristin R. Eschenfelder<sup>1</sup> Kalpana Shankar<sup>2</sup> Greg Downey<sup>3</sup>

### Abstract

Contributing to the literature on knowledge infrastructure maintenance, this article describes a historical longitudinal analysis of revenue streams employed by four social science data organizations: the Roper Center for Public Opinion, the Inter-university Consortium for Political and Social Research (ICPSR), the UK Data Archive (UKDA), and the LIS Cross-National Data Center in Luxembourg (LIS). Drawing on archival documentation and interviews, we describe founders' assumptions about revenue, changes to revenue streams over the long term, practices for developing and maintaining revenue streams, the importance of financial support from host organizations, and how the context of each data organization shaped revenue possibilities. We extend conversations about knowledge infrastructure revenue streams by showing the types of change that have occurred over time and how it occurs. We provide examples of the types of flexibility needed for data organizations to remain sustainable over 40-60 years of revenue changes. We distinguish between Type A flexibilities, or development of new products and services, and Type B flexibilities, or continuous smaller adjustments to existing revenue streams. We argue that Type B flexibilities are as important as Type A, although they are easily overlooked. Our results are relevant to knowledge infrastructure managers and stakeholders facing similar revenue challenges.

#### INTRODUCTION 1

Over the past two decades, many scientific fields have come to rely upon shared collections of managed data to support computationally intensive research (National Science Foundation [NSF], 2016). Creating and sustaining the infrastructure for data-driven science "raises challenges around long-term storage, preservation, and curation" of the data (National Academies, 2016, p. 13). Data organizations (DOs) (i.e., data archives, repositories, etc.) assemble and manage large collections of data, foster reproducibility, facilitate public access, and promote data reuse and integration-thus facilitating the asking of new questions and supporting advancement of research (Borgman, 2015; Borgman et al., 2019). The day-to-day data practices involved in managing research data vary widely across fields, among different DOs, and between projects within DOs (Borgman, 2015; Mayernik, 2015), but all require expenditure of resources to acquire, prepare, describe, preserve, and disseminate data.

Many have raised concerns about how infrastructural organizations, like DO, might organize themselves to

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. © 2022 The Authors. Journal of the Association for Information Science and Technology published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

<sup>2</sup> WILEY JASIST

increase their long-term sustainability given increasing costs and tightening financial support (e.g., Knowledge Exchange, 2014; National Academy of Sciences, 2014). Flexibility and adaptation are important, and one key area for flexibility is revenue streams. Revenue and finances are important elements of infrastructure sustainability (Borgman & Wofford, 2021; Fidler & Russell, 2018; Russell & Vinsel, 2018). Sustainable funding of data infrastructure like DOs is a key component of Open Science and Responsible Research and Innovation (Horizon 2020 Expert Advisory Group, 2015; Leonelli, 2013; Leonelli & Tempini, 2020; OECD, 2017). Yet, sustainable financing remains a challenge (Borgman et al., 2020), and changes in funding mechanisms, resource availability and evolutions in what stakeholders' value require infrastructure managers to adapt (Ribes & Polk, 2014). While there is general agreement that flexibility to adapt to changing revenue circumstances is important, there is less understanding of what that flexibility looks like in terms of changes to work practices, organizational structures, data products and services offered etc. To help KI be more flexible, especially regarding finances, greater information about what the needed flexibility might consist of and how it might be achieved is needed (Ribes & Polk, 2014).

This paper contributes to the above goal by illuminating a set of data practices that have received less attention, the work of developing and maintaining revenue streams to support ongoing data operations. Leading analysis with exploration of revenue sources provides a unique perspective on the assemblages of people, artifacts and institutions that comprise knowledge infrastructure (like DOs).

We explore the historical revenue streams of a particular type of DO, the social science data archive (SSDA). First developed in the post-World War II period with the public opinion survey boom of the 1930s and 1940s, social science DOs grew across the globe alongside quantitative data methodologies in the social sciences throughout the 20th century. Because of their long history, their global presence, and their varied research audiences and purposes, SSDAs provide a unique and foundational lens for the study of data infrastructure financial design, maintenance, and sustainability.

We take a historical, longitudinal, process view to examine how these DOs have paid for acquiring, preserving, and disseminating data (Langley et al., 2013). Using the case study method, we describe how the revenue streams of four prominent SSDAs developed and evolved over the last decades of the twentieth century. For each case, we describe the types of revenue streams the DO founders imagined and the changes in revenue sources over time, and we discuss how particular situations and contexts shaped revenue streams. We compare and

contrast the cases and conclude with observations for furthering the study of finances and maintenance in infrastructure studies.

Prior studies have pointed to revenue as both a resource and a challenge that shapes the development of infrastructures. Variable durations of funding influence staffing and strategic planning. Fluctuations in government funding may pressure infrastructure organizations look for more commercialized solutions to (Borgman, 2015). Over time, the drive to develop new support sources may encourage infrastructures to take up new types of data, new methodologies, and new user communities (Bietz et al., 2013; Borgman, 2015; Edwards et al., 2013; Ribes & Finholt, 2009; Ribes & Polk, 2014; Russell & Vinsel, 2018).

Scholars have also identified many costs of data infrastructure maintenance. Ribes and Polk (2014) point to changes in topics deemed important, the nature of data, instrumentation, the investigators who could use the infrastructure, and how collaborations and work coordination occurs. Borgman et al. (2016) notes that some data are more expensive to curate because they are more complex or less "processed"; sometimes the data have a broader audience with more diverse needs. High value data sets may be processed faster and to a higher level of quality, requiring more resources. Alternatively, they may apply almost no work at all to some data sets, posting them caveat emptor (OECD, 2017). However even data with little or no processing still entails infrastructure costs (Curation Costs Exchange, 2021). And spending resources to process a data set not only reflects an initial value judgment about that data, but also increases the likelihood that the data will be used (increasing its value and maintenance costs) over time.

Several reports provide snapshots of contemporary revenue sources for data and digital culture infrastructure (Crow, 2013; Dillo et al., 2016; Kitchin et al., 2015; Maron, 2014; OECD, 2017). Revenue may include "structural funding" such as annual government budgets or multivear grants from a government entity. Other revenue streams include project grants, deposit contracts or fees, host institution support, access fees, licensing, philanthropy, consulting, and/or corporate sponsorship. Each possibility has advantages and disadvantages. A 2016 survey of 22 DOs found that over half relied on structural funding for most of their income, with a quarter of the overall sample receiving all their income from structural funding. A smaller number of DOs reported significant data access or deposit revenue (Dillo et al., 2016). A 2017 OECD survey of 47 repositories rated structural funding sources as most important, but most respondents were also exploring new revenue sources as their structural funding was not sufficient

(OECD, 2017). Repositories reported an average of two revenue streams each, but 16 reported reliance on one source. Industry players also introduce new models, like fee-based intermediary services (Plantin et al., 2018).

Prior work provides an excellent snapshot in time, but we have little understanding of how revenue sources change, and how particular contexts shape revenue possibilities. To examine change over time, we employed an approach that assumes continuous ongoing change in organizations. Tsoukas and Chia (2002) use the analogy of standing on a balance beam. While someone is seemingly standing still, their muscles constantly twitch in order for them to keep their balance. In periods of seeming stability, many small changes are occurring to maintain that stability.

Applying this lens to the revenue streams of social science data organizations, we seek to identify both big flashy revenue changes and smaller "muscle twitches," done to maintain financial stability. Both the large changes and the small and easily overlooked changes are important parts of DO maintenance and qualify as institutional work, or "the purposive action of individuals and organizations aimed at creating, maintaining and disrupting organizations." The work may be dramatic and highly visible or "invisible and mundane" consisting of small "adjustments, adaptations and compromises" (Lawrence et al., 2009, p. 1).

### 2 | METHODOLOGY

Initial conversations with SSDA began in 2012 and data collection began in 2014; the bulk of the case studies and analysis was conducted from September 2014 to August 2018. With institutional permission, the authors traveled to and gathered administrative documents from four English-language SSDAs:

- The Roper Center for Public Opinion (Roper) located at Cornell University in Ithaca New York, USA.
- Inter-university Consortium for Political and Social Research (ICPSR) located at the University of Michigan, Ann Arbor, Michigan, USA.
- The UK Data Archive (UKDA) which is now part of the UK Data Service located at the University of Essex, Essex, UK.
- The LIS Cross-National Data Center in Luxembourg (LIS), located in Luxembourg and with US offices located at the City University of New York Graduate Center.<sup>1</sup>

We chose our cases purposefully to compare social science data archives dependent on structural funding to

those that instead relied on membership or subscription fees. ICPSR, Roper and LIS employ some type of subscription as a source of revenue; however, the UKDA is more reliant on structural funding, or what it calls "core funding" from UK national funding agencies. This initial dichotomy of membership revenue versus national core funding proved to be overly simplistic, as we discuss later.

We were granted access to archival material and permission to interview staff, provided we were conducting a "historical" study. A more recent history was deemed too sensitive, so we limited our analysis to pre-2002. We chose 2002 as a cutoff as preliminary documentary analysis suggested that 2001/2002 heralded a new set of standards and technologies (the Web in particular), and most importantly, our participants were comfortable with that date.

Documents were our primary source of data and we spent 2-3 days at each case site acquiring them. ICPSR's documents were formally archived with finding aids at the Bentley Library, but in the other cases old documents were in informal storage (often the back of closets or old file cabinets). We explored the files, identified potentially relevant documents, and made digital copies. Choosing which of the hundreds of documents to copy constituted a first stage of analysis. We targeted founding documents, governance rules, annual reports, grant applications and reports, strategic plans, external reviews, budgets, and governance agendas or meeting minutes. We collected 106 documents at the Roper Center, 274 at ICSPR, 212 at the UKDA, and 115 at LIS. Most of these documents were multiple pages in length, so our analysis included thousands of pages. Documents spanned a time frame from the 1950s to the early 2000s. They included formally published brochures, typewritten letters, dot-matrix printer minutes and hand-written notes. We saved the documents as PDF images, but OCR scanning was not possible given the variation in text/scanning quality.

For initial organization and preliminary analysis, we used Mendeley to tag and sort the PDFs. A team member read each document, entered metadata on document type, author, and date of initial creation. The team member also applied a set of nine broad codes (derived from an earlier literature review) to sections of the documents (Eschenfelder et al., 2019). Any one section could be coded multiple times. We used the nine codes as a means of identifying documents that addressed topics of interest and prioritizing text for closer analysis. This initial deductive coding was determined by when the team obtained documents from each archive. It began in fall of 2015 and extended through 2016. In 2016 and early 2017, the team created comparative grids for each of the nine codes identifying commonalities and unique features.

▲ WILEY JASST

For this paper, analysis drew on text identified with the "financial issues" code but also includes the related codes of "memberships," "markets/marketing," "mission/values," and "collection development."<sup>2</sup> Analysis explored the unique contexts of each case and identified similarities and differences across the cases by decade. The first author reread all the relevant tagged documents, taking notes in analytical memos, and creating financial analysis comparison grids organized by DO, by financial sub themes and by decade. At this stage we also pulled out pricing, membership numbers and other financial data that we tracked longitudinally in Excel. The authors discussed the financial analytical memos and tables, guided both by the lower-level themes and other conceptual frameworks explored in earlier related work (Eschenfelder & Shankar, 2017; Shankar & Eschenfelder, 2015). Writing this paper proved to be a further level of analysis as we corroborated our findings with other previously gathered historical data from decades of journal and newsletter literature on social science DOs, and debated which financially-related themes and findings would be of most interest to JASIST readers and whether to organize the paper by DO, by time, or by cross-cutting themes.

Documentary analysis was supplemented by conversational interviews of 45 min to 1.5 hr in length with 21 total DO staff across the four sites. At each site, we interviewed several current staff and at least one former staff member. While we started with a list of standard interview questions, as we became more familiar with the history and current context of each DO, we typically modified the questions to be more specific.<sup>3</sup>

As described in other publications (Downey et al., 2019; Eschenfelder & Shankar, 2020), the authors also interviewed leaders of, and examined the archival papers for, major social science data archive organizations.<sup>4</sup> This related work provided us with a deeper understanding of the field of social science data archiving and the nature of the relationships among individual data archives.

Through prolonged engagement with the field, the authors developed a rich understanding of field-level concerns and challenges. From approximately 2016 to 2019, the authors participated in field conferences including presenting papers and posters at the IASSIST conferences (2015, 2017), SciDataCon (2016), RDAP (2016), the International Conference on Digital Curation (2020) and various cyberinfrastructure workshops (2019). These presentations provided a form of "member checking." In addition, the authors also presented early outputs of their work, and collected feedback from staff at ICSPR and the Data Archiving and Network Services (DANS) in the Netherlands.

# **CASE DESCRIPTIONS**

In the sections that follow, we draw special attention to:

- a. Founders' assumptions about revenue streams;
- b. Common changes to revenue stream over the long term (30-60 years);
- c. The importance of financial support from host organizations; and
- d. How the unique context of each DO shaped revenue possibilities.

#### 3.1 Roper

3 

The Roper Center for Public Opinion, founded in 1947 at Williams College, opened for general use in 1957. Roper's collection includes public opinion data provided by over 40 polling firms, media organizations, government agencies and academic researchers. Its collection includes over 23,000 data sets and survey questions. It has two main products: the main data archive, and the iPoll collection of polling questions linked to data sources. Roper is the second smallest DO in our study in terms of staff with seven staff members listed on its website in 2019. Most of Roper's annual reports (1950s to present) did not contain regular budgetary information, thus our analysis relies on fragmented financial information included in other documentation.

Roper's primary source of revenue has been university subscriptions, but Roper also has a long history of both corporate sponsorships and subscriptions drawing on founder Elmo Roper's media industry connections. Research grants have not been as prominent a source of revenue. In early 2019 its website listed 264 subscribing members including 42 commercial organizations, government agencies or think tanks.

#### 3.1.1 Subscription revenue

From inception, Roper has employed a subscription revenue model, but it struggled with the instability of subscription revenue and continuously tinkered with the categories of subscriptions fees. Unlike our other cases, Roper always had both academic and commercial subscribers due to its strong relationship with the media and US polling firms. For example, of 108 subscribers to IPOLL in 1997, 31 were academic libraries/campuses, 34 were pollsters, 17 were commercial press organizations and 26 were other organizations including government agencies. Further, Roper has long provided "metered access" or different amounts of data for different fee levels.

Roper's initial subscription model was aimed primarily at academic libraries. The original fees of \$1,000 the first year, and \$500 each succeeding year provided members with permission to copy and store a set number of datasets as well as limited data processing labor. Roper experimented with subscription categories over time. As shown in Figure 1, in 1978 Roper created three levels of membership at different subscription price points which provided metered access to different numbers of data sets.

Roper also provided data sets on a pay-per-use basis, charging nonmembers 50% more for data access. In 1997 it offered two categories, "limited" and "full," with different prices and access to data for each category. Around 2003, Roper began providing an "all you can eat" unlimited access subscription option for members.

Documentation shows that membership did not grow as fast as hoped, fluctuated year to year, and required investments in marketing staff. Late 1980s reports describe a leveling off of general membership at around 50 members: reports from the 1990s show membership subscriptions going up and down between the mid-40s and mid-60s. The 1999–2000 annual report describes general membership as having "remained constant" and membership revenue as having only grown only slightly in that decade. The 2005–2006 annual report indicated 85 general members.

### 3.1.2 | New revenue

To attract new subscribers when the obvious sources were tapped, Roper began developing new data products. In 1983, Roper developed IPOLL, a database of poll questions, which Roper hoped in its 1983–1984 Annual Report would become a "key factor in the center's financial health and intellectual role." By the end of the 1980s, subscriptions to IPOLL surpassed regular subscriptions. But the 1990–1991 report complained of flattening IPOLL growth and the trend continued. As shown in Figure 2, revenue stemming from subscriptions to IPOLL began to decline after 1997–1998.

In the early 1990s Roper began to offer pay-perminute access to data through resellers like DIALOG and LEXIS/NEXIS, both commonly used by commercial organizations. The 1991–1992 Annual Report describes how new growth had come from government and industry through these services.

But not all new product experiments succeeded in developing revenue. In the 1990s, Roper sought to diversify revenue by offering fee-based analytical services. As part of this fee for service effort, it created a bimonthly magazine, *Public Perspective*, to advertise Roper's analytical expertise. But Public Perspective never covered its own costs and publication ended in 2003. We did not find evidence that Roper generated significant analysis services revenues.

### 3.1.3 | Host support

The Roper Center changed hosts four times; financial support was an important factor in each move. Roper's original location at Williams stemmed from the Roper family's personal connections with and support for the college. But the small liberal arts nature of Williams made getting resources for advanced computing difficult, so Roper had to rely on industry donations of computing equipment. In 1977, Roper announced a move to the University of Connecticut and formation of a governing cooperative including Connecticut, Yale, and Williams (see Figure 3). The vision was that support would be shared. But notes show the partners believed that Roper should wean itself from host contributions and become self-sufficient. By 1986, the Roper bylaws no longer mention Yale or Williams as governing and financial partners.



FIGURE 2 Roper IPOLL revenue 2000 annual report

	Category	Annual Fee (1978-1979)	Number of Data Sets Furnished*	Cost**
	Participating Member	\$ 900	15	\$60/data set
FIGURE 1 Roper subscription levels 1978–1979	Associate Research Member	\$2,500	50	\$50/data set
	Senior Research Member	\$5,000	120	\$42/data set



partnership

The 2000 Annual Report describes Roper as being in debt to University of Connecticut, and in 2005 the university announced a plan to reduce its support for Roper by \$150,000 a year. In 2014 Connecticut announced the Roper Center would be leaving campus. This was a time of severe budget cuts at state universities, and the university's budget was largely flat in the face of increasing costs. In 2015, Cornell announced that Roper would join its campus.

#### 3.1.4 Industry revenue

Roper was unique among our DOs for its cultivation and development of industry gifts as a revenue stream. Sporadic data and financial corporate gifts had been part of the Roper budget from inception, but a formal corporate sponsorship program began in 1990. Annual reports tracked growth from one sponsor in 1989 to 55 sponsors in 2000. But many of the sponsorships were often small and unstable, varying with larger economic cycles.

### 3.2 Inter-University Consortium for Political and Social Research

ICPSR, founded in 1962 at the University of Michigan, began with an emphasis on political science, but quickly broadened to encompass an array of social science data.

In summer of 2019, it listed over 11,000 data sets. ICPSR is the largest DO in our study in terms of staff, membership size with 778 institutional memberships, and total revenues of \$18.4 million (2017-2018 annual report). Collections include its original membership archive, open thematic collections developed from multiyear curation contracts with US federal agencies, and an open access archive of non-curated data (Open ICPSR) that complies with federal agency open data regulations. ICPSR was our most fully documented case site.<sup>5</sup>

#### Subscription revenue 3.2.1

While ICPSR founders imagined it as a scholarly cooperative funded by annual membership subscriptions (which have remained an important revenue stream), ICPSR has continuously experimented with its subscription categories and pricing to attract new members. For example, in the 1960's, ICPSR had just three categories, and in the late 1970s it had five. But more categories complicated management as members argued that they ought to be in a less expensive category. Experiments with new categories did not always draw in new members and some were abandoned. By 2002, category management had grown so burdensome, the board voted to "outsource" categorization by using the Carnegie Classification of Institutions of Higher Education to classify its members.

ICPSR also created the federated membership category in the 1970s that would allow smaller campuses to share a membership. ICPSR assumed that a hub campus would handle data requests from federation members and reduce service costs for ICPSR. But the federation subscription fee became a complex formula that took up staff time to administer. Furthermore, other developments made existing revenue arrangements less successful. For example, the rise of FTP made it even more likely that federation members with service requests would bypass their hub institution and seek to download data directly from ICPSR, requiring new resources to service these requests. In 1994, the ICPSR board voted to phase out federations by declining any new federation members and increasing fees for existing ones.<sup>6</sup>

There were also changes in who paid for subscriptions at member universities. Founders assumed academic departments would pay subscriptions as they were the target audience of the data collections. But academic budget decision makers' lack of knowledge about ICPSR, and the relatively small size of departmental budgets, created vulnerabilities. ICPSR came to believe that academic libraries were less likely to cancel subscriptions. Beginning in the mid-1970s ICPSR reframed its services as analogous to a journal subscription and recommended

that university libraries (with their larger budgets) pay the subscriptions. By the 1980s, ICPSR believed most membership subscriptions were paid from libraries.

While membership subscriptions remained an important revenue stream, *the amount of revenue* produced was unstable, especially during larger economic downturns. Maintaining the subscription revenue stream throughout financial ups and downs required spending on marketing and member relations. The 1970s and 1980s saw an overall net growth in membership, with more schools joining than dropping out. In comparison, the 1990s and 2000s saw flat membership numbers except for federations.

### 3.2.2 | Federal grants and contracts

Founding documents from the early 1960s assumed that subscriptions would be sufficient to cover archival costs, but by 1966 staff were discussing the importance of project grants to provide revenue. Figure 4 shows that project grants quickly overtook membership revenue as the largest source of income during the 1960s.

But federal project grant funding also underwent periods of contraction. Figure 5 shows the decline in research grant revenue in comparison to membership in the 1970s.

Another limitation on the grant revenue stream was the US NSF's declaration that it would only provide grant funding for research projects and not ongoing archival activities (NSF was one of ICPSR's biggest grant funders). Across the decades, ICPSR tried to persuade NSF to fund ongoing curation, or to commit to a structural funding revenue stream. Position papers and reports express a hope that the US federal government would be persuaded to treat ICPSR as a national institution, analogous to the Library of Congress. Then ICPSR would be able to abandon the subscription revenue stream and open its general



FIGURE 4 ICPSR income by source 1960s—Created by authors

archive to serve everyone. This hope has not yet come to pass.

### 3.2.3 | Agency contracts

Beginning in 1975, ICPSR began developing a new revenue source which grew into a dominant stream: renewable contracts with government agencies to curate and provide access to agency data. By the early 2000s grants and contracts combined overtook membership subscriptions (Figure 6). Proponents argued that it was no longer realistic to depend on membership subscription revenue; contracts provided new access to government data. Critics argued that contract revenue confused ICPSR's organizational identity as a scholarly cooperative and reduced attention to the needs of subscribed members.



FIGURE 5 ICPSR income by source 1970s—Created by authors

**REVENUE FROM MEMBERSHIPS AND GRANTS** 

\$6,000,000 \$5,000,000 \$4,000,000 \$3,000,000 \$2,000,000 \$1,000,000 \$1,000,000 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 MEMBERSHIP GRANTS AND CONTRACTS

**FIGURE 6** Revenue from membership and grants and contracts from 2002 to 2003 ICPSR annual report

\* WILEY JASST

However, the development of the contracts revenue stream required new staffing and skills. Curation costs increased because agencies often provided larger and more complex datasets in addition to some poorer quality data. Developing new contractual relationships required staff to cultivate relationships and create demonstration projects. Maintaining the relationships also entailed costs: agencies might require unusual billing procedures or have highly restrictive travel rules for contractors. Agencies could open contracts for rebidding or demand lowered fees that did not leave room for cost changes. In some cases, the transfer of funds to pay the contract was not smooth, leaving gaps in expected revenues.

The contracts revenue stream and increasing agency demands led ICPSR to develop different tiers of data curation and open access to some agency funded data. As early as 1974, government funders were beginning to require that investigators put agency funded data into the public domain. In response to early demands ICPSR agreed to distribute agency data "as is," a quality standard that came to be known as "Class V" data. Importantly, NSF was not willing to pay ICPSR to store the "Class V" data nor to "mandate ICPSR as a host [for archived data]." NSF argued that subscription revenue should cover the costs of archiving, or that ICPSR charge a deposit fee. But the arrangement still entailed costs: for postal transfer, then anonymous FTP, and then the open access ICPSR Direct archive. The use of different tiers of data curation became an important strategy for ICPSR to maintain its subscription revenue stream. Paying members had access to the high-quality curated versions of data, while the Class V data was accessible to all.

#### 3.2.4 Host relationships

Host support from the University of Michigan has been a smaller but important revenue stream for ICPSR. ICPSR brought in a large amount of overhead funds to the University of Michigan through its grants and contracts. Overheads in theory are distributed by the host to support the infrastructures used by many different projects. However, conflict sometimes arose between ICPSR and the host about how much of the overhead should be returned to ICPSR. Tensions flared in the late 1980s as ICPSR wished to recapture more of its overhead and argued that the overhead it generated was worth more than the services provided by the host. The overhead question flared again in 1995. In both cases, negotiations gave ICPSR a greater cut of overhead, but they had to take on some costs previously covered by the host.

#### **UK Data Archive** 3.3

The UKDA was founded in 1967 at the University of Essex. Originally a stand-alone archive, in the early 2000s the UKDA was brought into collaboration with several other UK DOs to form the UK Data Archive. Currently the UK Data Archive hosts over 6,000 data sets and approximately 60 staff members. The UKDA did not produce detailed and regular budget reports, which limits our analysis.<sup>7</sup>

#### Subscription revenue 3.3.1 1

Initial funding for the UKDA in 1965 stemmed from the University of Essex and 1967 seed grants from the Economic and Social Research Council (ESRC).<sup>8</sup> Stakeholders hoped this funding would quickly be replaced by annual subscription fees. Founders envisioned UKDA as the Roper-like home for UK polling and public opinion data. But the subscription revenue stream was not successful as hoped, and UKDA suspended the subscription fees in 1972 due to concerns that the fee was deterring use. A 1975 ESRC midterm grant review lamented that hopes raised in the original proposal for the archive to be financially self-sufficient (i.e., not grant funded) were not fulfilled. After 1972, academic users enjoyed free access funded by ESRC grants. While membership subscriptions were abandoned, fee-based services for commercial and governmental organizations continued.

#### Structural funding 3.3.2

For much of the study period, UKDA's primary revenue source was renewable structural or "core" funding from the ESRC and then also JISC, both UK national funders. But renewal of funding was often unpredictable, and some stakeholders believed that the ESRC should not fund UKDA at all. As early as 1971, some argued that the ESRC, as a research organization, should not support a continuous service with a (partial) educational mission. In the early 1970s the discussion became heated enough that UKDA developed a contingency plan for loss of ESRC support. Looking for alternative solutions, ESRC stakeholders suggested a variety of alternative government agency funders. In 1996, JISC took on funding educational portions of UKDA services, but the ESRC has continued to provide structural funding to support research.9

Renewal periods for ESRC funding fluctuated between 6-, 5-, 2- and 1-year extensions, making planning, management, and reporting difficult. These fluctuations were driven in part by larger political and economic issues. One manager speculated that if a particular grant renewal had happened 1 year later under a different administration, UKDA likely would have received increased funding. Political changes sometimes brought increased demands for competition. For example, in 1998, an ESRC review called into question the wisdom of "the rolling basis of the contract" and suggested an open call for proposals for the DO functions.

The importance of the structural funding led the DO to reorganize itself and its work, both within its own staff and in relation to other UK DO. In 2002, the ESRC, seeking to coordinate funded data services at separate funded DOs, released a bid for a national social science data service. This funder-driven reorganization changed staffing and the nature of work at the UKDA at Essex, leading to the creation of a virtual organization that combined the staff of several DO into what is now known as the UK Data Service.<sup>10</sup> While restructuring did not change the sources of the structural funding (the ESRC and JISC), it motivated reorganization of work within and between DO.

### 3.3.3 | Fee-based services

Fee-based services were set up at founding for commercial and nonprofit users and have remained a small revenue stream. In certain periods, the political climate or UKDA's desire for additional revenue sources encouraged greater experimentation with fee services. Fees fit well with the ethos of the 1980s Thatcher era, so ESRC stakeholders encouraged the introduction of fee structures. For example, a 1981 ESRC review led to the reinstatement of a data handling charge for all users. In 1982 the ESRC governing board encouraged the DO to reimplement usage fees to recover some costs. Further, a 1984 memo described how "the current climate of opinion in government favors commercial exploitation" to "develop the commercial potential of services." Finally, a 1987 review committee "congratulated" the archive for a proposed plan to develop fee-based services. The interest in fees continued with the Conservative government under John Major. A 1994 ESRC review committee recommended that UKDA hire a consultant to develop even more commercial possibilities.

Internal considerations also motivated the desire for more fee revenue. Unpredictable ESRC grant renewals (discussed earlier), combined with the archives' desire to develop new services, led staff to support experimenting with fees. Fee-based services shaped staffing and the arrangement of work in the DO as complex fee schedules required more customer service work to complete transactions. Charging for data required renegotiation of deposit agreements that precluded commercial use. Further, in the 1990s, some UK government agencies began charging for commercial use of their data, requiring UKDA staff to negotiate between agency charging policies and the DO's new fees.

The importance of fees as a revenue source ebbed and flowed but overall, fee-based revenue generation was modest. A 1990 report suggests that fees were not successful because the retrospective nature of the archive's collections limited commercial demand, and also because of UKDA's lack of marketing resources.

### 3.3.4 | Government data broker

As part of an effort to develop increased financial support from government stakeholders, UKDA staff began to offer data services to government agencies, "brokering" the relationship between the agency and public data users. This new relationship was ideally funded through additional ESRC structural funding. But the duration and static nature of the structural funding renewals created challenges. Meeting notes describe how costs of ingest and curation of government data were often higher than expected due to poor data and metadata quality, but UKDA could not obtain additional funds to cover the costs because of the large multivear nature of the structural funding grants. Another complication was competition from other UK DO. Competing DO hosted at other universities sometimes won data brokering contracts the UKDA desired. Moreover, with the rise of FTP, groups could just post their own data, leading to concern about the future of DOs as brokers.

### 3.3.5 | European project revenue

Starting in 1993, the United Kingdom's membership within the European Union facilitated access to new sources of project revenue. In the late 1990s, UKDA began several collaborative projects with European DO that were funded by the European Commission. While the UKDA had long collaborated with European DO, the emergence of new European funding sources further encouraged UKDA to organize itself to submit proposals for European-scale projects. This led to several large, funded projects on cross-border data usage, international standards and shared tools. For example, in 1995 UKDA submitted a Large-Scale Facilities Grant with German and Norwegian DOs.

Because UKDA documents did not regularly report on all sources of revenue, we cannot determine what



FIGURE 7 UKDA project funding from 1998 report

⊥WILEY\_ **JASIS**T

10

percent of revenue stemmed from project-based grants. Evidence suggests that it was a much smaller proportion than the ongoing structural grants. A 1998 report shows that "other" funding, which would include project grants (but also host support and fees), provided the smallest proportion of the DO's funding (Figure 7).

# 3.4 | LIS Cross-National Data Center in Luxembourg

LIS was organized in 1982 in partnership with the Luxembourg government and US academic institutions to collect, harmonize, and provide secure remote access to international income and wealth microdata (World Bank, 2021). Since microdata describe individual people or households and are typically produced by national statistical agencies there are special confidentiality concerns. LIS founders envisioned the DO as an infrastructure to facilitate comparative analysis of outcomes of different nations' social programs. Data include individual or household income from employment, investments, pensions, social programs, assets and debt, market and government income, household characteristics, and labor market outcomes. Typically, government agencies generate the data and transfer it to LIS for harmonization. Links between LIS and government agency staff are key because agency staffers often serve as advocates for LIS subscriptions, and because the DO is dependent on contributions of national microdata by agency staff for collections growth.11

LIS staff host data in Luxembourg to take advantage of Luxembourg's strong privacy laws. LIS lists 13 total staff. A 2006 budget listed a net worth of  $\notin$ 464,108.

In a 2013 NSF grant, LIS described its revenue streams as subscriptions or what it calls "country contributions," host support, project grants and





FIGURE 8 LIS founding revenue plan

philanthropy. The national subscriptions are the dominant revenue source.

### 3.4.1 | Subscription revenue

The LIS founders' original plan (Figure 8 below) imagined that revenue would transition from 1983 to 1993 from start-up grants and foundation bridge funding to 5-year "permanent consortium funding," consisting of subscription pledges from national funding agencies.

Yearly payment of the subscription would allow every scholar in the nation access to the DO. The multiyear subscriptions have remained the dominant source of revenue in LIS, but the source and payments have changed over time and thus revenue has been unpredictable. Funders have included funding agencies, universities, national banks, government agencies, and foundations. Fees were sometimes split between two or more funders.

Recruiting and retaining funders was expensive. Courting new funders could take years of relationshipbuilding. For example, one partnership first broached in 1992 did not provide a subscription payment until 2001. To begin recruiting a new funder for a nation, LIS would give the first year's access for free, add that country's data to the system, and gather usage statistics to demonstrate to potential national funders how researchers in the nation would use LIS. Ideally, the recruited funder would start paying after that first year, but this could take more time due to lack of understanding of LIS, budget constraints, and sponsors' rules about funding and expenditures.

Nonpayment of subscriptions has been an ongoing challenge. Sometimes, the funder sometimes does not pay-often due to national economic downturns-and the nonpayment requires DO staff to recruit a new funder for that nation. LIS might accept a delay in payment while recruiting a new funder. According to LIS rules, nonpayment of subscription fees would result in cutting off access. But a strict cutoff could interfere with negotiations for renewals and damage personal relationships important to keeping national memberships (and data contributions) long term. Paying members resented fee increases necessitated by nonpayment from other nations.

Different financial supporters, including government agencies, granting agencies and foundations, each provided special challenges. For example, funding relationships with government agencies were complicated by turnover of staff due to changes in governments or career progressions. The new staff person with budget oversight typically had no knowledge of LIS and would question the subscription payment. The "courting process" would need to begin again to reestablish the relationship. In cases where LIS applied to national funding agencies for sponsorship through routinized grant systems, success was not assured. LIS had to solicit letters of support from prestigious users in each nation and provide nationspecific impact data to convince grantors. For example, as part of an application to the US NSF, the funder warned that LIS needed to demonstrate to grant reviewers that top US economists were using the data. Finally, foundations typically expected that payment of

the subscription would be a one-time expenditure and were not willing to renew, requiring recruitment of a longer-term funder.

LIS offered different levels of subscription costs to attract members and over time increased the membership subscriptions to cover costs. While LIS advertised a base membership fee, actual fees were based on usage, data donations, and in-kind labor contributions. Membership and contribution of data were more important than maximizing membership fees. But as LIS sought to expand the collection to include more nations' data, many target nations could not afford to pay the base rate. In 1993, the LIS board gave the Director discretion to accept a smaller base fee from lower income countries. LIS had four fee increases during our study period (Figure 9).

Tension between the desire to expand access and problems of free ridership grew. For example, in the early 2000s LIS Directors argued for free access for researchers in two nations they were trying to recruit as members, while advising against free access for researchers in another country, whom they perceived as trying to avoid subscription. In another example, LIS increased nonmember usage fees because the Board felt some nations were avoiding subscriptions by just paying pay-per-use fees. In 2003, the Council required that the Director seek its approval for lower subscription payments.

LIS created a new membership category to attract a new type of member-supranational/intergovernmental organizations such as the European Commission, OECD, the World Bank and NATO. While these organizations sometimes paid for fee-based nonmember access, none had joined as membership subscribers. Media sources had also expressed interest in paying fees to use LIS data, but the rules of the organization did not allow them to use the data under any model.



FIGURE 9 LIS fee base country contribution increase over time-Created by authors



LIS dollars revenue from membership by year



**FIGURE 10** LIS revenue from country contributions/ membership by year—Created by authors

### 3.4.2 | Project grants

While founders hoped that national subscriptions would pay for all DO costs, they quickly learned they would not be sufficient. Within 2 years of founding, leaders changed the LIS Rules of Organization to recognize that the collections would be funded by both subscriptions and project grants. While subscription revenue grew over time, the percent of costs it covered fell as the costs of the DO increased (Figure 10). Budget reports show that in 1995 subscriptions paid for 75% of expenses, in 1996 they paid for 60%. In 1999 subscription revenue grew to cover 90% percent. But 2003 budgets report subscriptions covered only 58% of core costs.

### 3.4.3 | Host relationships

LIS has received direct and indirect revenue from two hosts: the Luxembourg government and the City University of New York Graduate Center (CUNY). The national research agency of Luxembourg was one of the original founding funders of LIS, so LIS became part of a Luxembourg government science agency. Funding channeled through the research agency provided staff salaries, computing, building space, and administrative services. Some years LIS stored their sensitive data for free in highly secure government data. The director of the research agency aided LIS in obtaining national project grants in addition to the regular funding.

LIS shifted hosts within the Luxembourg government in the early 2000s to the Ministry of Culture, Higher Education and Research and to the Centre Universitaire. As part of this move, LIS became an independent nonprofit institution and staff had to arrange both the transition in organizational form and a physical move to different offices. In the United States, originally LIS offices and staff were funded from NSF grants, but the LIS Board advised developing different funding sources. By 1996, LIS obtained a host support commitment from the CUNY for offices, computers, student assistants and a subsidy for staff salaries.

### 4 | DISCUSSION

We began this paper with an explanation of concerns about financial sustainability in the knowledge infrastructures (KI) field and specifically among data organizations (DO). We reviewed calls for explication of the specific practices through which infrastructure managers ought to enact flexibility and adapt to changing conditions (Ribes & Polk, 2014). We contribute to this literature by detailing instances of flexibility in creation, change and maintenance of revenue streams. In doing so, we begin to describe what flexibility looks like and how it is achieved. We make four important points. First, our cases illustrate that flexibility in revenue practices can only be observed by looking across longer time spans. Second, we distinguish between two types of revenue flexibility: Type A-developing entirely new products and services, and Type B-making continuous smaller adjustments to existing revenue streams to accommodate changing demand or new opportunities. We describe how both types of revenue flexibility require relationship work-either courting new relationships or bolstering existing relationships to maintain subscribers and supporters.

Table 1 provides a summary of the larger revenue changes the DOs underwent in the 30-to-60-year periods from their founding to approximately 2002.

The stories of these revenue stream changes raise several points relevant to managers of knowledge infrastructures, funders, and other supporters.

First, large scale dramatic changes in revenue sources do occur where staff may develop entirely new products and services, or where a major source of support (such as a host institution) changes. Some Type A dramatic changes succeed in becoming new revenue sources and others do not. Examples of Type A flexibilities in our longitudinal data include:

- UKDA's experiments with university membership fees (later dropped);
- Movement of some UKDA renewable funding to JISC;
- Roper's development of IPOLL, which became a successful new product;
- Roper's experiments with a periodical and related analysis services (later dropped);



<b>TABLE 1</b> Major revenue source change	s over time
--	-------------

Data organization	Founding revenue assumptions/sources	Changes to revenue sources	Ongoing revenue sources after 1999
Roper Center for Public Opinion	<ul> <li>1957 NSF/Ford Foundation start up grants</li> <li>Host: Funding from Williams College</li> <li>Industry donations of money and data</li> </ul>	<ul> <li>1964: Created membership subscriptions (library oriented)</li> <li>Host: Support switched to UCONN/Yale, then just UCONN</li> <li>New product data: iPOLL</li> <li>Project grants</li> <li>1989: Industry sponsorship program started</li> </ul>	<ul> <li>Greater emphasis on membership development</li> <li>Host: Move to Cornell University</li> </ul>
ICPSR	<ul> <li>1962 NSF and foundation founding grants</li> <li>Membership subscriptions</li> <li>Small project grants</li> <li>Host support the University of Michigan</li> </ul>	<ul> <li>1966: Recognition of need for project grants for revenue</li> <li>1975: First government agency contract</li> <li>Summer training courses</li> </ul>	<ul> <li>Post 2001: Rapid expansion of government contracts as revenue source</li> <li>Mid 1990s Tensions with the University of Michigan regarding overhead</li> </ul>
UKDA	<ul> <li>1967 start-up grants: The ESRC (assumed temporary)</li> <li>Host support—University of Essex</li> <li>University subscription fee (dropped by 1972)</li> <li>Fees: Commercial deposit, analysis, data provision</li> <li>Project grants, the ESRC and others</li> </ul>	<ul> <li>Recognition of dependence on ongoing ESRC grants</li> <li>1974: First government agency agreement for deposit/curation services</li> <li>Encouragement for fee-based services</li> <li>Summer training courses</li> </ul>	<ul> <li>1994 JISC (educational agency) takes over parts of sustaining funding</li> <li>2002 the ESRC national social science data services bid released</li> <li>Early 2000s: EC new source of project funding</li> </ul>
LIS cross National Data Archive	<ul> <li>1982 start-up grants</li> <li>Multiyear subscriptions paid by national science agencies</li> <li>Use fees for occasional users</li> <li>Host: Science government agency Luxembourg</li> </ul>	<ul> <li>1989: Recognition that project grants will need to supplement subscriptions</li> <li>Grants fund training workshops and conferences</li> </ul>	<ul> <li>Host: 2006 CUNY host support added</li> <li>Host: Switch to educational government agency Luxembourg</li> </ul>

- ICPSR's development of government agency contracts, which became a dominant revenue stream;
- Decision that project grants will need to supplement other revenue (all cases); and
- Change in host institutions and types of support provided (Roper and LIS).

Type A flexibilities may involve significant changes in how the DO organizes its staffing and services. In our case studies, staff had to enact Type A flexibilities infrequently. Further, these changes often took many years to fully come about; sometimes they existed in beta form for many years. While Type A flexibilities are exciting, our data suggest that Type B flexibilities are equally, if not more important.

Second, well-functioning revenue streams require continual tinkering to remain well-functioning. This is our "Type B" flexibility where infrastructure managers must make continuous smaller adjustments to existing revenue streams to accommodate changing demand or shifts in opportunities. Our data contain many stories of the small-scale, nonglamorous work done by DO staff to keep revenue streams functioning. For example, subscription revenue has remained important for ICPSR, LIS and Roper (Table 1). But to keep it functioning as a revenue stream, staff have had to regularly adjust both the membership category definitions and fee amounts associated with the membership categories. In making these adjustments, the DO staff must balance the tension between the need for higher subscription revenue and discouraging resubscribers or new members due to fee increases.

Related to this point, while a revenue source for a DO may remain the same over time, the package of data products and services required to maintain that revenue source typically changes. Adjusting the products and services associated with revenue stream also falls under the "Type B" flexibility. In one example, grant-based infrastructural funding remained a dominant venue source for

⊥WILEY\_ **JASIST** 

the UKDA during most of the study period. But our data illustrate how UKDA adjusted the bundle of data and services promised in return for that grant-based revenue. Infrastructure managers must devise what is attractive to funders given societal, political and technological changes, and relationships with the stakeholders must be renewed in order to ensure continued funding (Horizon 2020, 2015; Ribes & Polk, 2014). This Type B flexibility (or Tsoukas and Chia's (2002) twitches) can easily be overlooked and undervalued (Bowker et al., 2010). DO supporters and funders may not realize the labor inherent in Type B flexibilities.

Our results also highlight the practical importance of host support. Host support might get overlooked as a source of revenue, but it was the most common source of revenue across our four sites, and it is a common contemporary revenue source (OECD, 2017). The proportion of a DO budget supplied by host support may be small. For example, Dillo et al.'s (2016) survey found that most DO did not receive substantial funding from their hosts, and many complained of host's claims on DO grant overhead monies. Despite the small scale of the revenue, it is important. Our case studies include numerous examples of DO staff enacting "Type B" flexibilities to maintain good relationships with hosts. This might require periodic reporting, or the more delicate work of renegotiating overhead agreements. Host goals did not always align with DO goals. As one interviewee explained, "I have the University as a master, who has particular demands ... But I also have [my] funder, and the two strategic goals and the vision of both organizations are not the same." Our cases also include two examples of "Type A" flexibilities where data organizations had to develop a new host relationship because the host organization changed (LIS and Roper). Examples of Type A flexibilities required due to changes in host support included:

- LIS applied for nonprofit status and moved from one government agency host to another government agency.
- Over time, Roper migrated between three different university hosts.
- UKDA had to integrate its staff and reorganize processes when it became part of the larger UK Data Services, which included multiple data organizations in the United Kingdom. This reorganization was funder driven.

Another lesson is that structural funding does not mean you do not have to do financial maintenance work. The 2017 OECD study presents structural funding as a key path to sustainability (OECD, 2017). Thus, we began this project with the intention of contrasting DOs that enjoyed structural funding (UKDA) with those that did not (ICPSR, Roper, LIS). We assumed that a structurally funded DO would have an easier go of it, not having to constantly hunt for revenue. But exploring UKDA's structural funding history highlighted numerous frustrations including uncertain renewal cycles, inability to account for higher-than-expected costs and managing the expectations of funding agency stakeholders. All of these frustrations required Type B flexibility to adapt and keep the structural funding revenue stream going. At the same time, UKDA experimented with other revenue sources to supplement its structural funding (see Table 1).

We found that renewable funding, while it is renewable, is still unpredictable; and the unpredictability requires Type A or Type B flexibilities. Some of our cases involved grant-based infrastructural funding, where the DO received periodic national grants supporting infrastructural services or data access (e.g., UKDA). DOs funded by the government infrastructural funding grants must still reprove their value every grant cycle and shorter-term cycles do not match well with longer term goals (Borgman, 2015; Borgman et al., 2019). In the case of the UKDA, this uncertainty manifested itself as staff coping with occasional threats to not renew funding, shorter 1- or 2-year awards instead of the desired 5-year awards, or increased scrutiny for funding renewals. It also led to movement of part of the UKDA's funding from the ESRC to JISC (Type A). Another challenge in many of our cases was the mission mismatch where government funders hesitate to provide long-term funding for shared infrastructure because of its nature as infrastructure, as opposed to research (Borgman, 2015). As one DO staffer explained, "One ongoing issue of getting grant funding [from research agencies], for archives, is that people don't like funding general resources. They like funding research." These uncertainties required both Type A flexibilities (given a new funder) and Type B flexibilities to assuage existing funders. Staff had to scramble to meet grant agencies' changed expectations or adapt internal processes to unexpected changes in funding.

A related point is that changes in the larger economic situation, or turn over in which political parties control government, impacts all government sources of revenue (i.e., grants, contracts, renewable grants, fees). For example, because LIS depended on funding from many different government funders (via national subscriptions), it had to cope with changes across numerous governments (all with different election cycles) and with the loss of key contacts/supporters at administrative agencies as staff moved on. In some instances, that led to delays in the renewals of funding, and it required extensive relationship-rebuilding labor (Type B flexibility). In another example, during certain periods aligned with UK conservative governments, DO stakeholders strongly encouraged the UKDA to further develop its fee-based revenue streams as fee for service fit well with the general political approach to government services. In the United States, periods of economic downturn or government spending cutbacks reduced the amount of funds available to research funding bodies and contracts with government agencies came under renewal pressure as agency clients also looked to cut costs. Our longitudinal analysis illustrates how government-based revenue streams are subject to the vicissitudes of economic cycles and changes in government and DOs must continually adjust merit and impact strategies to these changing conditions (Ribes & Polk, 2014). Recent calls for studies of knowledge infrastructure in times of scarcity might employ a historical approach to analyze past periods of scarcity (Borgman et al., 2020).

Importantly, the mission of the DOs shape possible revenue sources, particularly the possibility of earning revenue from providing services to for-profit commercial organizations. Both Roper and UKDA have a long history of services to industry (Table 1). In contrast, the missions of both ICPSR and LIS have precluded developing industry revenue streams. Review of organizational documents 15

shows ongoing debates about the potential of commercial revenue sources; further, DOs have struggled to redevelop service boundaries in light of revenue opportunities from organizations that seem to blur boundaries, such as "research-oriented but commercial" organizations. Boundaries may shift over time as governments emphasize public–private partnerships for support of infrastructure (Horizon 2020, 2015). But many (including some of our participants), are concerned about negative consequences stemming from more porous boundaries between commercial interests and knowledge infrastructures (Borgman et al., 2020).

Many scholars have pointed to the dependencies of data infrastructures on their larger environments, and our data provide examples of how environment and revenue strategies interact. For example, the evolution of computing technologies, such as the introduction of FTP, made ICPSR's federated membership subscription model less viable. In a different direction, the establishment of the European Union and the emergence of the European Commission opened new project funding opportunities for the UKDA. In all our cases, global economic cycles expanded and contracted government funding, making sustaining funding, project grants and contracts more tenuous (Table 2).

Revenue source	ICPSR	Roper	UKDA	LIS
Host institution support	Since inception	Inception—but host changed three times	Since inception	Inception—with one host change
Fees for data access and/or data analysis work	Since inception	Since inception	Since inception	Since inception
Limited term grants	Adopted soon after inception	Became more common over time	Adopted soon after inception	Since inception
Subscriptions/ memberships	Since inception	Shortly after inception to today	Dropped	Since inception
Sustaining funding from national government agency			Inception to today (renewable)	For some nations since inception (renewable)
Contracts with government or nonprofit institutions	Grew to become dominant		Yes, but funded through sustaining funding package above	
Contracts for work with commercial organizations		Since inception	Since inception	
Third party distribution	CD ROM sales	Distribution relationships started with DIALOG (1990), NEXIS (1994)	CD ROM sales	
Training courses	Summer program		Summer program	Grant funded workshops and conferences

⊥WILEY\_ **JASIS**T

### 5 | CONCLUSION

The scientific data landscape has changed dramatically in the 20 years since our 2002 analysis cut off. We have seen an explosion in the volume of research data, the development of data-intensive research, growth of cyberinfrastructures, mandates for data management, technological advances for preservation and dissemination and culture change with respect to open data and data sharing (Borgman, 2015; Edwards et al., 2013). Despite these changes, the challenge of ensuring ongoing revenue for data infrastructures remains (Borgman et al., 2020). Developing better strategies for sustainable funding of infrastructure like DOs is a key component of contemporary projects like Open Science and Responsible Research and Innovation (Horizon 2020 Expert Advisory Group, 2015; Leonelli, 2013; Leonelli & Tempini, 2020; OECD, 2017). Staff in knowledge infrastructures (like DOs) need to be flexible and adapt to changing circumstances to keep the organization functional and relevant (Ribes & Polk, 2014).

Our paper extends previous work on the organization of knowledge infrastructures by illuminating the data practices involved in developing and maintaining revenue streams. We begin to explicate what flexibility looks like and how it might be achieved by different DOs. We introduce two categories of revenue labor flexibilities: Type A flexibilities are required to develop entirely new products and services or adjust to major changes in funders or host institutions. Type B flexibilities require making continuous smaller adjustments to existing revenue streams to accommodate changing demand or new opportunities. By distinguishing between types, we draw attention to the less glamorous, small tweaks regularly made to existing revenue streams to keep them functioning, and we provide a vocabulary to talk about the overlooked and undervalued labor related to revenue maintenance (Mattern, 2018). To use Tsoukas and Chia's (2002) analogy, our work highlights the muscle twitches work necessary to stay on the balance beam rather than just focusing on the dramatic moves. We also emphasize how both Type A and Type B flexibilities require relationship work-either courting new relationships or bolstering existing relationships to maintain subscribers and supporters.

Our study also adds a long view perspective to current snapshot-in-time analyses of DO revenue streams. While prior research has typically described how DOs currently finance themselves, this paper describes how DO *had been* financing themselves over a 30–60-year period. Further, our cases demonstrate how the unique context of each DO shapes and constrains revenue possibilities as technologies, patterns of government funding and research fashions change over time.

As our prior work on social science DOs has suggested (Downey et al., 2019), there is a great deal of norming and learning that occurs through consultation and collaboration between and across DOs in a particular scientific field, through the organizing efforts of data infrastructure professionals themselves. And, as noted earlier, the DO and KI practice communities have responded with a wealth of reports on revenue stream options. In addition, the field is developing frameworks to help infrastructure managers think about how to enact flexibilities (Type A or B) through both small and larger scale changes to infrastructural services (Knowledge Exchange, 2014; Osterwalder et al., 2013). More work is needed to prepare infrastructural managers to enact flexibilities to be sustainable and adjust to changing revenue opportunities and constraints over the long term.

While this paper focused on historical revenue sources used by DOs in the social sciences—one of the longestrunning examples of global DOs created for collaborative, computational scientific study—all knowledge infrastructures will face similar revenue challenges over the long term and infrastructure leaders will need to employ similar data practices to generate and maintain revenue streams.

### ACKNOWLEDGMENTS

This work was supported by grant G 2014-14521 from the Alfred P. Sloan Foundation from 2014 to 2019. The authors would also like to thank the Irish Research Council, the University College Dublin College of Social Sciences and Law Seed Fund, the Wisconsin Alumni Research Foundation, the University of Wisconsin Sarah M. Pritchard Open Access Support Fund, and the ASIS&T History and Foundations of Information Science Fund for their support for this project. We would also like to acknowledge Rachel Williams and Allison Langham for their vital assistance in organizing and coding documents.

### DATA AVAILABILITY STATEMENT

Documents from ICPSR are held by the Bentley Historical Library at the University of Michigan. Documents from Roper, the UKDA and LIS are the property of those data organizations. Documents were made available to the authors on the condition that they not be further shared. Interested researchers should contact the data organization for access.

### ORCID

Kristin R. Eschenfelder <sup>b</sup> https://orcid.org/0000-0003-4033-9837 Kalpana Shankar <sup>b</sup> https://orcid.org/0000-0001-8788-466X

### ENDNOTES

- <sup>1</sup> Our need to read, code and analyze historical organizational records limited us to English-language DOs. The large number of organizational records involved, and the authors' need to make sense of subtle differences in financial and management terminology across cases made relying on translations unwise.
- <sup>2</sup> Topical subcodes in "financial issues" included membership revenue, extramural funding, contract funding, host institution financial support, other financial inputs, overhead, administrative resource allocations, prices, costs, business model change, and free access.
- <sup>3</sup> Our larger project has been the creation of a broad history of social science data organizations (Shankar et al., 2016) and for this reason interviews covered a range of topics. The topics most relevant to this paper included financial sustainability concerns, relationships with host institutions, past current and future users and uses of collections, and relationships with funders.
- <sup>4</sup> The other organizations included, the Consortium of European Social Science Data Archives (CESSDA), the International Federation of Data Organization (IFDO) and the International Association for Social Science Information Service and Technology (IASSIST).
- <sup>5</sup> ICPSR regularly produced and archived annual reporting documents that included budget information as well as discussion of ongoing challenges. This regularity allows us to show trends in the financial history not possible with the other cases.
- <sup>6</sup> ICPSR's founding vision saw only universities as members, but over time ICPSR expanded access to permit pay-per-use, and then subscription access, by research-oriented nonprofits and government agencies. Use by commercial organizations was and remains complicated due to data depositor intentions.
- <sup>7</sup> Annual reports did not always include budgets; and when they did, information provided varied. The ESRC and JISC grant applications and reports were focused on expenditures of grant monies and did not address other sources of revenue.
- <sup>8</sup> Note that both the UKDA and the ESRC changed names several times during our study period (e.g., Data Archive at Essex, SSRC Data Archive, Social Science Research Council) We refer to both organizations by their 2019 names throughout this paper. Further, the UKDA is the lead management organization for the larger UK Data Service. In most cases we refer to the UKDA separately. When we refer to the larger UKDS, we use the term UKDS.
- <sup>9</sup> JISC was a new nonprofit representing the joint educational infrastructure interests of the UK higher education funding councils.
- <sup>10</sup> The coordinated service was originally known as the Economic and Social Data Service before being folded into the UK Data Service in 2012 with some additional partners including Edinburgh.
- <sup>11</sup> LIS has produced regular annual budget reports that allowed us to show financial trends.

### REFERENCES

Bietz, M. J., Paine, D., & Lee, C. P. (2013). The work of developing cyberinfrastructure middleware projects. In *Proceedings of the* 

## JASIST \_\_WILEY 17

2013 Conference on Computer Supported Cooperative Work, San Antonio Texas, 23–27 February (pp. 1527–1538). ACM.

- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world.* The MIT Press.
- Borgman, C. L., Darch, P. T., Pasquetto, I. V., & Wofford, M. F. (2020). Our knowledge of knowledge infrastructures: Lessons learned and future directions: Report of knowledge infrastructures workshop. eScholarship. https://escholarship.org/uc/item/ 9rm6b7d4
- Borgman, C. L., Sands, A., Darch, P. T., & Golshan, M. S. (2016). The durability and fragility of knowledge infrastructures: Lessons learned from astronomy. In *Proceedings of the Association for Information Science and Technology*, 53(1), Copenhagen Denmark, 14–18 October (pp. 1–10). Wiley. https://doi.org/10.1002/pra2.2016.14505301057
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science & Technology*, 70(8), 888–904.
- Borgman, C. L., & Wofford, M. F. (2021). From data processes to data products: Knowledge infrastructures in astronomy. *Har*vard Data Science Review, 3(3). https://doi.org/10.1162/ 99608f92.4e792052
- Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in the networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97–117). Springer. https://doi.org/10.1007/978-1-4020-9789-8
- Crow, R. (2013). Sustainability of open access services report. Phase 3: The collective provision of open access resources. Scholarly Publishing and Academic Resources Coalition (SPARC).
- Curation Costs Exchange. (2021). Understand costs. https://www. curationexchange.org/
- Dillo, I., Hodson, S., & de Waard, A. (2016). Income streams for data repositories, RDA-WDS cost recovery interest group. Research Data Alliance. https://doi.org/10.5281/zenodo.46693
- Downey, G., Eschenfelder, K., & Shankar, K. (2019). Talking about metadata labor: Social science data archives, professional data librarians, and the founding of IASSIST. In W. Aspray (Ed.), *Historical studies in computing, information, and society.* Springer.
- Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). *Knowledge infrastructures: Intellectual frameworks and research challenges*. Deep Blue. http://hdl.handle.net/2027.42/97552
- Eschenfelder, K., & Shankar, K. (2017). Organizational resilience in data archives: Three case studies in social science data archives. *Data Science Journal*, 16, 12. https://doi.org/10.5334/dsj-2017-012
- Eschenfelder, K. R., & Shankar, K. (2020). Of seamlessness and frictions: Transborder data flows of European and US social science data. In A. Sundqvist, G. Berget, J. Nolin, & K. Skjerdingstad (Eds.), Sustainable digital communities. iConference 2020. Lecture notes in computer science (Vol. 12051). Springer. https://doi.org/10.1007/978-3-030-43687-2\_59
- Eschenfelder, K. R., Shankar, K., Williams, R., Langham, A., Zhang, M., & Salo, D. (2019). A nine-dimensional framework for digital cultural heritage organizational sustainability: A

WILEY\_ JASIST

content analysis of the LIS literature (2000–2015). Online Information Review, 43(2), 182–196.

- Fidler, B., & Russell, A. L. (2018). Financial and administrative infrastructure for the early internet: Network maintenance at the defense information systems agency. *Technology and Culture*, 59(4), 899–924.
- Horizon 2020 Expert Advisory Group. (2015). European research infrastructures including e-infrastructures: Key recommendations. European Commission.
- Kitchin, R., Collins, S., & Frost, D. (2015). Funding models for Open Access digital data repositories. Online Information Review, 39(5), 664–681. https://doi.org/10.1108/OIR-01-2015-0031
- Knowledge Exchange Project. (2014). Sustainability index. Knowledge Exchange. http://www.knowledge-exchange.info/projects/ project/open-access/sustainability-oa-services
- Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, 56(1), 1–13.
- Lawrence, T. B., Suddaby, R., & Leca, B. (Eds.). (2009). Institutional work: Actors and agency in institutional studies of organization. Cambridge University Press.
- Leonelli, S. (2013). Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties*, *8*(4), 449–465.
- Leonelli, S., & Tempini, N. (2020). Data journeys in the sciences. Springer Open. https://www.springer.com/gp/book/ 9783030371760
- Maron, N. (2014). A guide to the best revenue models and funding sources for your digital resources. Ithaka Consulting.
- Mattern, S. (2018). Maintenance and care. Places Journal. https:// doi.org/10.22269/181120
- Mayernik, M. S. (2015). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, 67(4), 973–993.
- National Academies of Sciences, Engineering, and Medicine (2016). Future directions for NSF advanced computing infrastructure to support. In U.S. science and engineering in 2017–2020. The National Academies Press. https://doi.org/10.17226/21886
- National Science Foundation. (2016). *Ten big ideas*. https://www. nsf.gov/news/special\_reports/big\_ideas/index.jsp

- OECD (2017). Business models for sustainable research data repositories. In OECD science, technology and industry policy papers (No. 47). OECD Publishing. https://doi.org/10.1787/302b12bb-en
- Osterwalder, A., Pigneur, Y., & Clark, T. (2013). Business model generation: A handbook for visionaries, game changers, and challengers. Wiley.
- Plantin, J. C., Lagoze, C., & Edwards, P. N. (2018). Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data & Society*, 5(1), 2053951718756683.
- Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, 10(5), 375–398.
- Ribes, D., & Polk, J. B. (2014). Flexibility relative to what? Change to research infrastructure. *Journal of the Association for Information Systems*, 15(5), 287–305.
- Russell, A. L., & Vinsel, L. (2018). After innovation, turn to maintenance. *Technology and Culture*, 59(1), 1–25.
- Shankar, K., & Eschenfelder, K. (2015). Sustaining data archives over time: Lessons from the organizational studies literature. *New Review of Information Networking*, 20(1–2), 248–254. https://doi.org/10.1080/13614576.2015.1111699
- Shankar, K., Eschenfelder, K. E., & Downey, G. (2016). Social science data archives: A trans-Atlantic history of field-level knowledge infrastructures from 1960–2004. *Science & Technology Studies*, 29(2), 62–73. https://doi.org/10.23987/sts.55691
- Tsoukas, H., & Chia, R. (2002). On organizational becoming: Rethinking organizational change. Organization Science, 13(5), 567–582.
- World Bank. (2021). What do we mean by microdata? https:// datahelpdesk.worldbank.org/knowledgebase/articles/228873what-do-we-mean-by-microdata

How to cite this article: Eschenfelder, K. R., Shankar, K., & Downey, G. (2022). The financial maintenance of social science data archives: Four case studies of long-term infrastructure work. *Journal of the Association for Information Science and Technology*, 1–18. <u>https://doi.org/10.1002/asi.</u> 24691