Social Science Data Archives:

A Historical Social Network Analysis

by Kristin R. Eschenfelder, Morgaine Gilchrist Scott, Kalpana Shankar, Greg Downey[1]

Abstract
This paper reports findings about inter-organizational influence and collaboration relationships among social science data archives over time, focusing on activities of institutions affiliated with the journal International Association of Social Science Information Services and Technology Quarterly (IASSIST Quarterly). We examine how archives interacted from 1976-2014 by tracing relationships described in articles published in IASSIST Quarterly.

Introduction
This paper reports findings about inter-organizational influence and collaboration relationships among social science data archives over time, focusing on activities of institutions affiliated with the journal International Association of Social Science Information Services and Technology Quarterly (IASSIST Quarterly). We examine how archives interacted from 1976-2014 by tracing relationships described in articles published in IASSIST Quarterly. .

Keywords
Social science data archives, history, IASSIST Quarterly, social network analysis, collaboration

Introduction
Research disciplines increasingly rely on data archives and repositories to share results, advance their work, and support large-scale collaboration. While there have been numerous studies that examine the technologies and practices of how research fields and researchers develop and use data archiving and archives, there has been less attention paid to how data archives themselves as information institutions have adapted over time to evolving research trends, institutional changes, and funding models. Social Science Data Archives (SSDA) are exemplars of long-lived information infrastructures (broadly defined as the computing and technological resources and their supporting institutions that are designed to advance scientific inquiry) that have successfully adapted to such changes (Heim, 1980; O'Neill Adams, 2006). Understanding inter-organizational relationships among SSDA, funders and partner institutions over time is essential to understanding how SSDA have evolved to serve their user communities.

In this paper, we examine how such relationships have evolved, focusing on activities of institutions that appear in articles published in IASSIST Quarterly (IQ) from 1976-2014. Our larger goal is to better understand how SSDA have cooperated and competed to achieve their goals and draw out lessons learned that can be applied to the development and maintenance of contemporary cyberinfrastructures for research in the social sciences. We hope that lessons learned by SSDA may be useful to similar infrastructure projects in other fields.

This paper explores the following research questions:

1 Which institutions are most influential as depicted in IQ articles?
2 Which institutions collaborate the most as depicted in IQ articles?
3 To what degree are international relationships represented in IQ articles?
4 To what extent are highly collaborative institutions collaborating with each other?

Methods
Social network analysis (SNA) is a data collection and analysis approach useful for examining patterns in connections between people or social institutions. The results of SNA analysis are often depicted as a web of 'nodes' (people or institutions) and relationships or 'links' which represent the links between nodes. As Hansen, Schneiderman and Smith (2011) describe, SNA analysis may trace:

a. The number of unique links connected to a node. Nodes that have more links connected them to other nodes may be more important or influential.

b. Changes in the patterns of connections between nodes.
c. Variation in the types of links or between nodes.

To answer our research questions, we use SNA to analyze links between organizations or institutions (the nodes or our social network) as represented in IASSIST Quarterly (IQ) articles from 1976 through the end of 2014. We obtained back-issues of IQ from the IASSIST website, from archive.org or from the University of Wisconsin-Madison library. We did not include papers from the annual IASSIST conference proceedings because we only had access to the full text of conference presentations after 2000 (http://www.iassistdata.org/conferences). We first identified all articles by issue according to the volume number and date listed on the bottom of the article. For each paper we identified the following nodes in each paper:

- Institutional home of author or co-author of IQ article,
- Institutions mentioned in the paper as influencers or collaborators (more on this below) by IQ authors
- Funders of projects described by IQ authors.

We identified the following major types of relationships, or links, as explicitly described in the papers:

1 Influencing relationship: When one node mentioned another node as being influential, including co-authorship when the authors were at different institutions. Social network analysis typically refers to nodes that have more attached links as having a higher 'degree of influence' because a link between institutions represent opportunities for influence or evidence of influence. We examined the IASSIST networks to see which institutions were the most connected to other institutions. Arguably these linkages represent the potential for one institution to influence the practice of the other institution through sharing of knowledge or resources.
2 Collaborating institution relationship: When one node described a collaboration with another node. (see below)
3 Funding relationship: When a node provided funding (explicitly stated in the article). Funding relationships are shown in green on network graphs.
4 International relationship: We specifically noted data provider or collaborating institution relationships that crossed national boundaries. Because the Consortium of European Social Science Data Archives (CESSDA) is a pan-national organization, any relationship with CESSDA was marked as international.

We tracked influencing, collaborating, funding and international relationships among nodes in IQ using the tool NodeXL, an open source template for conducting SNA with data from Microsoft Excel (NodeXL 2015). We also tracked change over time of all of the four types of relationships. In order to show change over time, it is a common practice in longitudinal social network analysis papers to analyze data in multiyear sections rather than year by year. Data within a year often do not provide sufficient nodes and links to show a network of relationships. At the same time, analyzing the data as one large set (1976-2014) cannot show change. We determined that five year sections were sufficient to show both network relationships and change over time.

Operationalizing Influence and Collaboration

We coded the IQ articles in two ways: first broadly for influence, and then more narrowly for collaboration. Collaboration relationships are a narrower subset of influence relationships.

Influence: First, we coded broadly for relationships indicating influence among data archives. For the purpose of this project influence between nodes was included by one node's reference to another node as:

- an inspiration or model,
- part of a larger grouping of affiliated organizations,
- part of a collaboration,
- a provider of data,
- a provider of resources (staff, software), or
- a co-author on an IQ paper from a different institution

As we describe below, and as we depicted in our (Eschenfelder et al., 2015) IASSIST 2015 conference paper, the influence analysis resulted in a large loose network of relationships in the IASSIST community. We argue that influence patterns show which data archives other archives talk about, or which data archives are most influential in the larger community.

Collaboration: In this narrower analysis, we examined a subset of influence relationships describing collaboration between data archives. To do so, we first developed a definition of collaboration and coding rules by conducting comparative analysis of three random samples of approximately ten IQ articles. From the analysis, the team identified the following types of collaboration relationships:

- **New Collection/Service:** Institution Y gets data from institution X when institution Y creates a new data collection or service. Or X allows Y to make X's data accessible through a portal.
- **New Entity/Project**: Institutions X and Y create a new entity 'project Z' related to data.
- **Cross National Surveys:** Institution X describes participating in a cross national survey. Or a cross national survey project describes getting participation from nations X, Y and Z.
- **Software/Processing:** Institution X gets software from institution Y and creates a new data product or service. Or, institution X uses institution Y's processing power to manipulate or manage data to provide a service.

- **Collaborate to Collect Data**: Institutions Y and Y collaborate to obtain data from individuals (PIs' projects or from research subjects themselves).
- **Research about Practice:** Institution Y collects data from other data archives in order to compare practices and reports results.
- **Mergers:** Reports on acquisition of another data archive or its collection. Taking over a previous archive.
- **Learning Materials:** Institutions X, Y and Z collaborate to create learning materials.

In order for the description of an influence relationship in IQ to count as a collaboration, the depiction also needed to meet significance criteria. To count as significant, a description of a collaboration had to: have a header, have its own paragraph, be visually separated from other text, be mentioned multiple times within the article, or have at least 4 lines of text devoted to it.

IQ Articles included many descriptions of influence relationships that we did not count as collaboration including: X and Y co-authoring the IQ paper; providing curated links to data that lives at other institutions; compiling, indexing and coding data created through surveys run by others; simple descriptions of data deposit; descriptions of institution X getting data from institution Y and X writing a paper; mere descriptions of data available at institution X (even if the data is from other institutions); collaboration of subunits within the same larger organization; aspirational or planned collaborative activities; X describing how they contract out services to Y; and descriptions of teaching activities.

Coding Data for Influence and Collaboration

The research team first coded the relevant articles to identify influence relationships. A team member read through each article included in IQ, identified the names of nodes, and entered information about co-author, influence and funder relationships between nodes into an Excel template. Any given node typically had several links (influence relationships) and each link was listed in a separate row.

To code articles for the collaboration code, three members of the research team co-read and co-coded five separate samples of IQ influence articles from across different time periods in order to ensure we had agreement on the application of the coding rules. One team member then coded the remainder of the IQ articles and entered information about collaborations into a new Excel template. Each collaboration relationship was listed in a separate row.

Institutions and Relationships: 1976-2014

Figure 1 and Table 1 summarize data on the types of relationships reported in IQ articles from 1976-2014 including influence, collaborator, funder and international relationships. The number of nodes, or SSDA appearing in relationships grew over time from 12 in the 1976-1980 time period to 111 in the 2011-2014 period. The number of all types of relationships also grew over time, but not steadily. In particular, the number of funder relationships has been very up and down over the years. Moreover, the number of collaborator relationships reported in IQ grew steadily, but then dropped in the 2006-2010 reporting period.

Most nodes (or SSDA) in the data have a very small number of relationships. A small number of SSDA have a higher number of relationships. In this paper we focus on those SSDA with the higher number of relationships.

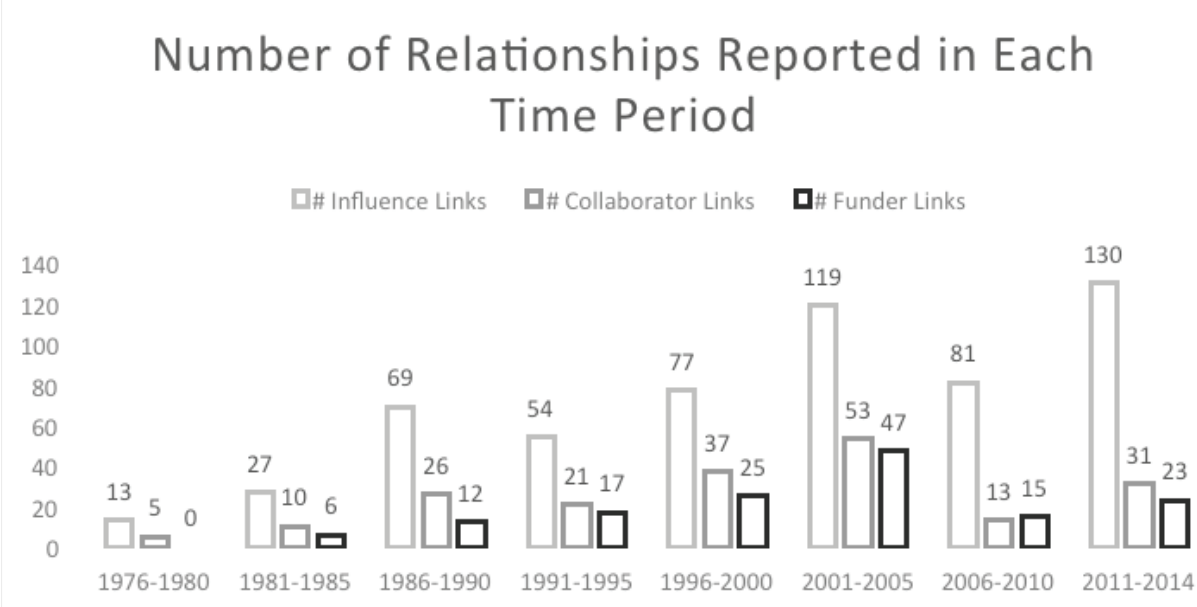

Figure 1: Types of Relationships Over Time

Time Period	Nodes	# Influence Links	# Collaborator Links	# Funder Links
Overall 1976-2014	653	285	196	145
1976-1980	12	13	5	0
1981-1985	37	27	10	6
1986-1990	81	69	26	12
1991-1995	76	54	21	17
1996-2000	108	77	37	25
2001-2005	130	119	53	47
2006-2010	98	81	13	15
2011-2014	111	130	31	23

Table 1: Relationships over time

Which SSDA has had the most influence?

We examined the data for which SSDA had the most influence. Influence links included when one node referred to another node in an IQ article as an influence or inspiration, as part of a larger grouping of affiliated organizations, as part of a collaboration, as a provider of data, or a provider of resources (staff, software). One problem is that many influence relationships in our data stem from large projects with many partners. We call these instances 'projects.' These projects are depicted in only one or two IQ articles, but include a high number of relationships.

Overall 1976-2014: SSDA With The Highest Number Of Influence Links To Other SSDA	# of Influence Links	# Articles in Which the Institution is Coded	Influence Ratio: Articles/ Links	Links/articles
Inter-University Consortium for Political and Social Research (ICPSR)	30	22	0.73	1.36
Zentralarchiv fur Empirische Sozialforschung (ZA)	16	9	0.56	1.78
UKDA	32	17	0.53	1.88
University of Edinburgh	13	6	0.46	2.17
International Social Survey Program (ISSP)	21	7	0.33	3
US Federal Reserve	14	4	0.29	3.5
University of Minnesota	15	4	0.27	3.75
Pennsylvania State University	19	2	0.11	9.5
The Pew Forum	19	2	0.11	9.5

Table 2: 1976-2014 Nodes with Most Influence Links

We created a measure called 'influence ratio' to depict those institutions with the most influence. This measure helped us identify institutions with many influence relationships over time, in addition to those with many influence relationships. The influence ratio captures this measure of an institution's influence across different articles. The ratio divides the number of total articles in which institution X appears by the total number of influence links between institution X and others. Table 2 reports the institutions with the

highest influence ratio from 1976-2004. Large data archives appear at the top with ICPSR leading with a ratio of .73, followed by the ZA and UKDA with ratios of .56 and .53 respectively.

Which SSDA has collaborated the most?

We examined the data to see which SSDA has the most collaborative relationships. Collaborative relationships were a narrower subset of influence relationships that involved a specific set of relationship types including: creating new collections or services, new projects, cross national surveys, sharing of software or processing, new data collections, comparisons of practices or mergers.

Because we were interested in reporting on institutions with many collaborative relationships over time, we focus our analysis on those institutions depicted in more than one IQ article. Table 3 lists the institutions with the most collaborative relationships in terms of the same article/links ratio used above. Again, large data archives appear at the top of the list with ICPSR having a collaboration ratio of 1 and UKDA having a ratio of .85. The US Census Bureau had the next highest ratio with .43.

Overall 1976-2014: Institution with the Highest Number Of Links To Other Nodes (c)	# of Collaboration Links (Coding 2)	Number of IQ articles in which institution is mentioned	Collaboration Ratio (articles/ links)
ICPSR	6	6	1
UKDA	13	11	.85
US Census Bureau	7	3	.43
International Social Survey Program (ISSP)	7	2	.29
Integrated Library and Survey-data Extraction Service (ILSES)	11	2	.19
East Asian Business and Development (EABAD) archive	7	2	.29

Table 3: Institutions with the Most Collaborating Relationships across Multiple IQ Articles 1976-2014

Who has funded whom?

We examined funding relationships and nodes to determine which funders had the most relationships from 1976-2014. Table 4 below shows the funding nodes with the most relationships in the 1976-2014 period. JISC was the top funding node with 18 reported relationships, followed by NSF and ESCR with 12 relationships each.

Funder	Number of Funding Links 1976-2014	Number of IQ articles in which funder is mentioned
Joint Information Systems Committee (JISC)	18	14
National Science Foundation	12	7
Economic and Social Research Council (ESRC)	12	6
Higher Education Support Project (HESP)	3	2
National Institute on Aging	3	2
Library of California	6	1
National Institute of Child Health and Human Development	4	1
California Department of Finance	3	1

California Digital Library	3	1
Juan March Institute, Spain	3	1
US Bureau of the Census	3	1

Table 4: Top Funding Nodes

Table 4 shows several instances of projects with many funding links that are only reported in one article (e.g., Library of California). These instances represent IQ articles that describe projects in which a funder funded a project with multiple partner nodes.
In the next section, we continue by providing more detailed data on those nodes with the most influence and collaboration relationships during each of eight five year time periods of our study. We also report on international relationships during each period.

Period 1: 1976-1980

The early time period of 1976-1980 saw the lowest number of nodes (N=12) and the lowest number of influence and collaboration links (N=13, N=5). At this stage in IQ's history, most articles tended to simply describe activities at an author's institution, and few described cooperative activities.

Influence: Roper had the highest number of influence relationships (6), followed by University of Iowa (4) and then Yale, Williams College and the University of Connecticut (all of which were affiliated with Roper Center – 3 each).

Collaboration: In the narrower collaboration measure, only University of Iowa and the Roper Center showed more than one collaboration link during this period.

1976-1980: Institutions with the Highest Number of Relationships	Influence	Collaboration
Most Relationships	Roper (6)	(2 each) Roper Center;
2nd Most Relationships	University of Iowa (4)	
3rd Most Relationships	Yale University; Williams College; University of Connecticut (3)	

Table 5: 1976-1980 Number and Type of Relationships

International: IQ articles did not depict any international collaborations during the 1976-1989 period.

Period 2: 1981-1985

The second time period saw a growth in nodes (N=37) and influence and collaboration relationships (N=27, N=10). This period also saw a rise in international relationships.

Influence: In this period, ICPSR and the US Census Bureau had the highest number of influence links (6) followed by Rand Corporation (4).

Collaborations: IQ articles still reported very few collaboration relationships. Four institutions reported two collaborations each: Center for Human Resource Research, Indian Council of Social Science Research (ICSSR), the International Federation of Data Organizations for the Social Sciences (IFDO), Norwegian Social Science Data Services, Zentralarchiv fur empirische Sozialforschung (ZA).

1981-1985: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	ICPSR; US Census Bureau (6 each)	(2 each) ICSSR, IFDO, NSD, ZA
2nd Most Relationships	Rand (4)	
3rd Most Relationships	(all with 3) (NORC); NSD, NARA; US DoE; Bonneville	

Table 6: 1981-1985 Number and Type of Relationships

International Collaborations:

This period saw three international collaborations. Both involved partners in Europe or partnerships with international organizations such as IFDO.

Year	Collaboration Relationship
1981	Norwegian Social Science Data Service and European Consortium on Political Research (ECPR)
1985	Two projects between Zentralarchiv fur empirische Sozialforschung and the International Federation of Data Organizations for the Social Sciences (IFDO)

Table 7: 1981-1985 International Collaborations

Period 3: 1986-1990

The third period showed growing interactivity among SSDA. Nodes increased to 81. Influence relationships increased to 69 and collaboration relationships grew to 26.

Influence: The International Social Survey Programme (ISSP) had the highest number of influence links (13) followed by Australian National University (6).

Collaboration: The East Asian Business and Development (EABAD) archive had the most collaboration links (5) stemming from a large multi-institution project. The US Census had 4 reported collaborations in this period.

1986-1990: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	International Social Survey Program (ISSP) (13)	East Asian Business and Development (EABAD) (5)
2nd Most Relationships	Australian National University (6)	US Census Bureau (4)
3rd Most Relationships	(5 each)National Opinion Research Center (NORC); East Asian Business and Development Research Archive (EABAD)	(3 each) UKDA; TARKI (Hungarian Social Science Information Center)
4th Most Relationships	(4 each) ICPSR; University of Amsterdam; University of Alberta; Office of Population Census and Surveys UK; Hunter College CUNY; University of Mannheim	

Table 8: 1986-1990 Number and Type of Relationships

International:

This period's IQ articles described an international collaboration between CELADE in Chile and the ICRC in Canada in 1989. In 1990 the East Asian Business and Development (EABAD) Archive at UC Davis reported on a collaborative project involving numerous partners in Asian nations. The ECPR, an international scholarly political science association located in Essex UK reported collaborations with two European universities.

Year	International Collaboration Relationship
1989	United Nations Latin American Demographic Center(CELADE) in Chile and International Development Research Center of Canada (ICRC)

1990	East Asian Business and Development (EABAD) Archive at UC Davis and a series of partners including University of Hong Kong, National University of Singapore, Tunghai University in Taiwan and the China Credit Information Service ECPR (UK) and both University of Mannheim and University of Amsterdam

Table 9: 1986-1990 International Collaborations

Period 4: 1991-1995

The fourth time period saw a decline in reported activity among SSDA. The number of nodes fell from 81 to 76, the number of influence links fell from 69 to 54, and the number of collaborations fell from 26 to 21.

Influence: University of Manchester and the Manchester Computing Center had the highest number of links in this period (10).

Collaboration: University of California Davis had the most collaboration links in IQ (5).

1991-1995: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	University of Manchester and Manchester Computing Center (10)	UC Davis (5)
2nd Most Relationships	University of Wisconsin-Madison (8);	(2 each) Bringham University SUNY; East Asian Business and Development (EABAD); ICPSR; Lehman College; ROADS; US Census Bureau
3rd Most Relationships	University of Illinois Urbana Campaign (7)	
4th Most Relationships	(5 each) US Census Bureau; University of California Davis	
5th Most Relationships	University of Missouri St Louis (4)	

Table 10: 1991-1995 Number and Type of Relationships

International: This period's IQ articles described only one international collaboration between the University of Ulster and the United Nations University (a think tank and post graduate education institution associated with the United Nations and located in Japan).

Year	International Collaboration Relationship
1995	University of Ulster and the United Nations University (Japan)

Table 11: 1991-1995 International Collaborations

Period 5: 1996-2000

Activity grew again during the 1996 to 2000 time period. Nodes grew to 108, influence links grew to 77 and collaborations grew to 37.

Influence: In this period, the UKDA had 11 influence links and the ZA had 7 links.

Collaborations: The Integrated Library and Survey-data Extraction Service (ILSES) appeared as the largest collaborator during this period with eleven links. The Data Liberation Initiative followed with seven links.

1996-2000: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	UKDA (11) ILSES (11)	Integrated Library and Survey-data Extraction Service (ILSES) (11)
2nd Most Relationships	Zentralarchiv fur Emprische Sozialforschung (ZA) (7)	Data Liberation Initiative (7)
3rd Most Relationships	(all with 6) University of Minnesota; Conference of Rectors and Principals of Quebec Universities	ISSP (5)
4th Most Relationships	(5 each) University of Quebec; Sociometrics Corporation	Royal Statistical Society (4)

Table 12: 1996-2000 Number and Type of Relationships

International : International collaborative relationships described in this period include three descriptions of Integrated Library and Survey-data Extraction Service projects with multiple European partners in 1997, 1998 and 2000

Year	International Collaboration Relationship
1997	Integrated Library and Survey-data Extraction Service (ILSES) and partners including ProGAMMA, Groningen, SWIDOC in Amsterdam, University of Amsterdam, Zentralarchiv fur Empirische Sozialforschung (ZA), Trinity College and CIDSP in Grenoble
1998	Integrated Library and Survey-data Extraction Service (ILSES) and partners including ProGAMMA, Groningen, Zentralarchiv fur Empirische Sozialforschung (ZA), University of Amsterdam, BSDP, Trinity College
2000	International Social Survey Program (ISSP) and partners including National Opingion Research Center (NORC), Zentrum fur Umffragen, Methoden und Analysen (ZUMA), Zentralarchiv fur Empirische Sozialforschung (ZA), National Center for Social Research, Research School of Social Sciences

Table 13: 1996-2000 International Collaborations

Period 6: 2001-2005

The sixth period saw growth in reported activity as nodes rose to 130, influence relationships rose to 119, and collaborations rose to 53.

Influence: Penn State and Pew had the highest number of influence links with 18 each.

Collaboration: The Association of Religious Data Archives had 20 reported collaborations in this period.

2001-2005: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	(18 each) Pennsylvania State University; the Pew Forum; also Association of Religion Data Archives (ARDA) (20)	Association of Religion Data Archives (ARDA) (20)

2nd Most Relationships	UKDA (11)	(7 each) Sociological Data Archive (SDA) the Institute of Sociology in Prague; UKDA
3rd Most Relationships	(10 each) Deakin University; University of Ballarat	Collection of Census Data and Resources (UK) (5)
4th Most Relationships	Czech Academy of the Sciences (9)	
5th Most Relationships	(7 each) University of Manchester; Princeton University	

Table 14: 2001-2005 Number and Type of Relationships

International: IQ articles in this period described more examples of international collaborations than previous periods. They described two cross-Atlantic collaborations involving archives in the US and European partners.

Year	International Collaboration Relationship
2001	Sociological Data Archive (SDA) at the Institute of Sociology in Prague and nine other European collaborators. Institute for Quality of Life Research Bucharest (IQLR), UKDA and ZA.
2002	Collection of Census Data and Resources (UK) and the Frauhofer Institut Autonome Intellligente Systeme (Germany)
2003	Finish Social Science Data Archive (Finland) and ICPSR (USA)
2005	Association of Religion Data Archives (USA) and 20 other partners, mostly based in US but some European

Table 15: 2001-2005 International Collaborations

Period 7: 2006-2010

The seventh time period was another period of decline. Articles only mentioned 98 nodes, and 81 influence links. And we saw a large decline in collaboration links described from 53 in the earlier period to merely 13 in this period.

Influence: The Finnish Social Science Data Archive and the University of Ljubjana showed the most influence links with 10 and 9 links respectively.

Collaboration: University of Edinburgh, University of Oxford, University of South Hampton and London School of Economics all showed three collaborations each.

2006-2010: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	Finnish Social Science Data Archive (10)	(3 each) University of Edinburgh; University of Oxford; University of South Hampton; London School of Economics
2nd Most Relationships	University of Ljubjana (9)	
3rd Most Relationships	University of Minnesota (7)	
4th Most Relationships	(6 each) University of Windsor; University of Albany; University of Cape Town	

5th Most Relationships	(5 each) UKDA; US Federal Reserve; MIT	

Table 16: 2006-2010 Number and Type of Relationships

International: The number of international collaborations described in IQ also dropped in 2006-2010. We found only one: between the African Association of Statistical Data Archivists and the International Household Survey Network in 2007

Year	International Collaboration Relationship
2007	African Association of Statistical Data Archivists (AASDA) and the International Household Survey Network (IHSN)

Table 17: 2006-2010 International Collaborations

Period 8: 2011-2014

Growth increased in the final analysis period with 111 nodes, 130 influence relationships and 31 collaborations.

Influence: The Lithuanian Center for Social Research and Vilnius University scored highest in terms of number of influence links with 24 each.

Collaboration: The node with the highest number of collaborators was the Nestor Memorandum of Understanding (13). This stems from one large project with many European collaborators.

2010-2014: Institutions with the Highest Number of Relationships	Influence	Collaboration
Top Most Relationships	(24 each) Lithuanian Center for Social Research; Vilnius University	Nestor Memorandum of Understanding (13)
2nd Most Relationships	(11 each) Grottingen State University; Cologne University of Applied Sciences	(3 each) American National Election Study (ANES); Lithuanian Humanities and Social Science Data Archive (LiDA); VOICES Project/EQUALAN; Wienmer Institute for Social Science Data Documentation and Methods (WISDOM)
3rd Most Relationships	UKDA (9)	
4th Most Relationships	(8 each) US Federal Reserve; Wienmer Institute for Social Science Data Documentation and Methods (WISDOM)	
5th Most Relationships	Council of European Social Science Data Archives (CESSDA) (7)	

Table 18: 2011-2014 Number and Type of Relationships

International: This period's IQ articles described three major international collaborations. Two were largely European, but one stretched between Taiwan and Norway.

Year	International Collaboration Relationship
2011	VOICES Project/EQUALAN and 3 European partners, Lithuanian Humanities and Social Science Data Archive (LIDA) and 2 European partners
2012	Center for Survey Research, Taiwan (SRDA) and the Norwegian Social Science Data Services Nestor Memorandum of Understanding among 13 European partners
Table 19: 2011-2014 International Collaborations	

Summary Discussion

In this section we answer the paper's research questions by examining our data over time.

Research Question 1: Who is the most influential SSDA as depicted in IQ?

The most influential SSDA (according to our measure of influence) are ICPSR, the ZA, UKDA, and the University of Edinburgh. The ISSP project was also very influential but not at the level of the other four. In other words, within the IQ environment, these nodes were most often mentioned, across different articles, as having an influence via collaboration, data sharing, or just serving as an example for others.

Overall 1976-2014: SSDA With The Highest Number Of Influence Links To Other SSDA	Influence Ratio: Articles/Links
Inter-University Consortium for Political and Social Research (ICPSR)	0.73
Zentralarchiv fur Empirische Sozialforschung (ZA)	0.56
UKDA	0.53
University of Edinburgh	0.46
International Social Survey Program (ISSP)	0.33
US Federal Reserve	0.29
University of Minnesota	0.27
Pennsylvania State University	0.11
The Pew Forum	0.11
Table 20: Nodes with the Most Influence Relationships	

Research Question 2: Who collaborates the most as depicted in IQ?

The SSDA with the most collaborative relationships according to our measure is ICPSR, with UKDA close behind it. The US Census Bureau also appears prominently. This means that as depicted in IQ articles, ICPSR and UKDA had the most relationships that met our criteria for collaboration described earlier.

Overall 1976-2014: Institution with the Highest Number Of Links To Other Nodes (c)	Collaboration Ratio (articles/links)
ICPSR	1

UKDA	.85
US Census Bureau	.43
International Social Survey Program (ISSP)	.29
Integrated Library and Survey-data Extraction Service (ILSES)	.19
East Asian Business and Development (EABAD) archive	.29

Table 21: Most Collaborative SSDA

Research Question 3: To what degree are international collaborations represented in IQ articles?

As shown in Table 22, the number of international collaborative links varies widely across the time periods of the study. International collaborative relationships have grown over time but seen two periods of decline (1991-1995 and 2006-2010).

Time Period	# All Collaboration Links	# International Collaboration links	Ratio Intl/All
1976-1980	5	0	0
1981-1985	10	3	.3
1986-1990	26	8	.31
1991-1995	21	4	.19
1996-2000	37	16	.43
2001-2005	53	30	.57
2006-2010	13	1	.07
2011-2014	31	18	.58

Table 22: Collaborations and International Collaborations Over Time

To examine whether the proportion of international collaborative relationships to overall collaborative relationships has grown in IQ articles, we plotted a ratio of international collaborations over all collaborations by time period.

Collaboration downturns: The results shown in Figure 2 suggest that the proportion of international collaborations is growing over time, with periods of downturn (e.g., 1991-1995, 2006-2010). This pattern of growth and dips in growth parallel the overall pattern of collaboration activity shown in Figure 1 and Table 22 above. Other kinds of analysis might yield more insight into whether these collaboration downturns were due to financial pressures or other factors; for example as economies go into recession, or governments cut funding on education fewer resources may be available to support travel to collaborate with other SSDA. Alternatively, SSDA staff may be more under resourced and have less time to write up articles describing collaborations to IQ.

Research Question 4: To what extent are highly linked institutions interlinked with each other?

We took the institutions with the most relationships from Table 21 and looked to see to what degree they had collaboration relationships with other 'most collaborative' nodes versus collaborative relationships with other nodes. We found that the 'most collaborative' nodes did not collaborate with each other; rather, they collaborated with other nodes with lower collaboration scores. This suggests that highly connected nodes are mostly serving as hubs, linking outwards toward less connected nodes.

Conclusions

The data generated by this analysis provide one limited view of the relationships among SSDA and related institutions. There was partial overlap between most influential and most collaborative SSDA. Our analysis show that the most influential SSDA include ICPSR, ZA, UKDA and University of Edinburgh. The most collaborative SSDA were ICPSR and the UKDA followed by the US Census Bureau. Our analysis also found that the top ranked collaborative SSDA did not collaborate with each other but with other institutions. Top

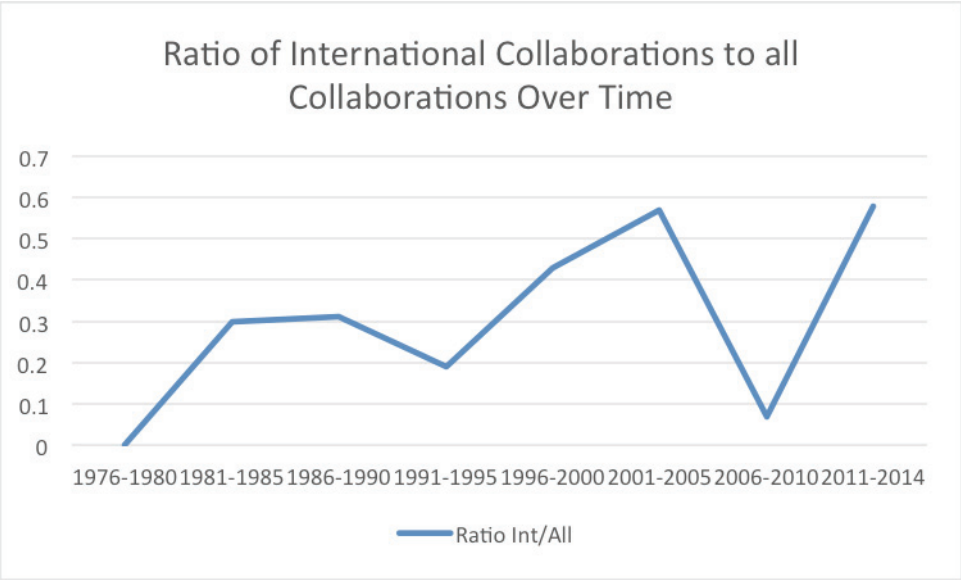

Figure 2: Collaborations and International Collaborations Over Time

funders including JISC, SNF and ESRC. Depiction of influence relationships grew most steadily in IQ articles. Collaboration and funding relationships grew over time, but showed periods decline in 1991-1995 and 2006-2010.

Because the data set stems from articles published in IQ, the data cannot represent the many collaborations that were never described in IQ. For example, renewing a contract with a major partner may be very important, but because it is not a new project, it may not merit a write-up in IQ. Because IQ is the publication of record of the IASSIST organization, one can argue that the most important and innovative relationships would likely be included because IASSIST members would seek to share information about their projects with other members. Secondly, while the data set of relationships is incomplete, it is still representative of the most innovative and noteworthy relationships in the study period. Future research could enrich understanding of influence and collaboration relationships between SSDA by adding data from IASSIST annual meeting conference proceedings. At this time, only data from 2000 and later is available via the IASSIST website. It is much more difficult to draw conclusions about lack of collaboration or competition from this data sets or reasons why relationships might have ended. To learn more about these issues, we plan to rely more on historical documentation from case studies of individual SSDA.

SSDA are complex institutions with a long history (Heim, 1980; O'Neill Adams, 2006). This IQ analysis provides one view of how SSDA have developed and maintained relationships with their funders, and each other over time to innovate, serve their user bases, and grow their products and services. This paper has described and summarized inter organizational relationships among SSDA and funders from 1976-2014 as depicted in articles published in IASSIST Quarterly. As one component of a larger study about the history of social science data archives (SSDA) and the field of social science data archiving, the data provides an important broad view of relationships among data archives in the shaping of a research discipline. We would suggest that this study, even though it is partial in scope, represents the kinds of institutional analyses that could yield insight for other kinds of data repositories and institutions in how they are developing and leveraging their own institutional networks and the implications such networks have for growth and maintenance.

References

Eschenfelder, K.R.; Gilchrist Scott, M.; Shankar, K.; LeClere, E.; Lin, R.; Downey, G. 2015. Social Science Data Archives: A Historical Social Network Analysis' IASSIST Conference 2015, Minneapolis MN.

Hansen, Derek L.; Shneiderman, B.; Smith, Marc A., 2011. Analyzing social media networks with NodeXL insights from a connected world. Burlington, Mass: Morgan Kaufmann.

Heim (McCook), K., 1980. Social Science Data Archives: A User Study. Unpublished Dissertation, University of Wisconsin-Madison.

NodeXL, 2015. NodeXL: Network Overview, Discovery and Exploration for Excel [online] Available at: < http://nodexl.codeplex.com/> [Accessed May 2015].

O'Neill Adams, M., 2006. The origins and early years of IASSIST. IASSIST Quarterly, Vol 30 Fall Issue, pp 5-14.

Acknowledgements

This project has been supported by the Alfred P. Sloan Foundation, the University of Wisconsin Alumni Research Foundation, University College Dublin Seed Funding Scheme, and the American Society for Information Science and Technology History Fund.

Notes

1. Corresponding author is Professor Kristin R. Eschenfelder, School of Library and Information Studies, University of Wisconsin-Madison, eschenfelder@wisc.edu
2. As explained by IQ editors, the year dates of publication for an article are not always precise because publishing is sometimes behind schedule. Analysis by five year periods may therefore be a better depiction of trends than analysis by specific year.
3. Where there were more than two authors, all authors' institutions were listed as collaborating with one-another. When one author was associated with multiple institutions, each institution was listed as having all the relationships referenced in the article. However, they were not listed as collaborating with each other.