Council Meeting Minutes

October 18-19, 2018

Attendees:

Council Members: Bobray Bordelon, Lisa Cook, Jane Fry, Michael Jones-Correa (Chair), Elizabeth Groff, Verna Keith, Lindsey Malcom-Piqueux, Chandra Muller, Esther Wilder, and Katherine Wallman

Guests: Mark Hansen, and Matthew Woollard

ICPSR Staff: Dharma Akmon, Trent Alexander, JD Alford, Johanna Bleckman, Dieter Burrell, Stephanie Carpenter, Alina Conn, Edward Czilli, Linda Detterman, Libby Hemphill, Lynette Hoelter, Stuart Hutchings, Dory Knight-Ingram, Abay Israel, Susan Jekielek, Lisa Kelly, Kilsang Kim, Kathryn Lavender, Daphne Lin, Susan Leonard, Maggie Levenstein, Jared Lyle, James McNally, John Marcotte, Tom Murphy, Justin Noble, Asmat Noori, Michelle Overholser, Amy Pienta, Jukka Savolainen, Saundra Schneider, Chelsea Sample-Steele, Annalee Shelton, Michael Shove, Fillippo Stargell, David Thomas, Harsha Ummerpillai, and Vanessa Unkeless-Perez

Approval of Minutes

ICPSR Council Chair called for approval of the minutes of the June 2018 meetings. Council voted unanimously to approve.

Director's update

Maggie Levenstein gave a brief overview of what we have been doing at ICPSR since June 2018.

Maggie indicated that our membership totals have risen recently, and we are up 11 members for a total of 782 member institutions. The ICPSR Summer Program in 2018 had 952 participants. We have recruited more curators and more acquisitions staff bringing our staff numbers to 112. The ICPSR Data Fair had 26 online sessions, 243 member institutions represented, 40,000 social media impressions, and 15 countries represented. All of the videos are now available on YouTube and are closed captioned.

We have had an organizational paradigm shift with a new organizational structure which consists of eight units. There is more communication across the organization. We are still working on achieving the goals in the Strategic Plan and are treating the Strategic Plan as a living document. We will continue to meet with Ruth Bardenstein on how the units are doing with their goals.

We have had successful renewals for the topical archives The National Archive of Computerized Data on Aging (NACDA), Health and Medical Care Archive (HMCA) and the National Archive on Criminal Justice Data (NACJD). We have more renewals coming up this year for NACJD, National Addiction & HIV Data Archive Program (NAHDAP), Patient Centered Outcome Research Institute (PCORI), and the Data Sharing for Demographic Research (DSDR). The Research Connections archive has become the Early Childhood and Education Archive. The Office of Planning Research and Evaluation (OPRE) decided to end support for the Research Connections project, based at Columbia University; however, they asked for Susan Jekielek to apply for a sole source five-year contract for Early Care and Education.

We have a number of education related projects: College and Beyond II (Mellon), Gates Foundation (Measures of Effective Teaching and Millennium Scholars), American Education Research Association (hosting the data for its journals and providing data stewardship support for NSF fellowship recipients), and registry for education research projects (joint with Society for Research on Educational Effectiveness (SREE) funded by the Institute of Education Sciences (IES)). IES is supporting an archive started by SREE and moved to ICPSR; we will be archiving registrations of research strategies.

Maggie updated the Council on non-archival projects: LinkageLibrary, Research Passport and Open Badges, Geospatial VDE, IRB and consent statements (joint with Qualitative Data Repository), Experiments in Crowdsourcing Metadata, Recovery and linkage of 20th Century Tax Records and Census Data, and Training Grants (Rigor & Transparency to Enhance Reproducibility and the Summer Program Diversity Initiative). We are also working on various other projects such as Census collaboration, State of Michigan collaboration, Metadata enhancements, and Using AI for identifying data-related publications.

ICPSR goals for Archonnex this year include moving curation, Researcher Passport, MyData, RADIUS, LinkageLibrary, StatSnap (our online analysis system), Summer Program Portal, and the deposit system enhancements.

Professor Mark Hansen, Director, David and Helen Gurley Brown Institute for Media Innovation & Professor of Journalism, Columbia University gave a presentation on "Data Journalism." Mark spoke of the basics of journalism being someone with a notepad and a curiosity, and how the relationship between journalism and data is just as critical as ever. Journalists of course use data to learn about goings on and to further back up their stories, or in the least, there is narrative potential in data. However, the investigation must also be applied to the data itself - how was it gathered? Is it accurate? Who controls it? Can they themselves gather it? Thus their investigations must not only be about the topic the data is being collected on, but the data and data collection itself.

Roundtable discussion of data and the media with Matthew Woollard, Director, UK Data Archive. Matthew discussed how the UK Data archive engages with the media. For the last 9 years, UKDA has run a secure data enclave. It provides personal data to researchers. The key risk is a data breach. There is a trust relationship between the data archive and the government which provides the personal data. Most breaches are caused by researchers who make accidental mistakes, e.g., a researcher who sends personal information in an email. They are banned for 3

months, and then retrained. Matthew's interaction with the media is based on a fear: What happens in the event of a breach—'crisis management.' That's not an ideal relationship to have with the media. It's important that data archives and services have a closer relationship w/ the media, not only as data suppliers, but trying to understand the data.

Education Committee

Council: Verna Keith (Chair), Lisa Cook, Jane Fry, Michael Jones-Correa, Lindsey Malcom-Piqueux, and Esther Wilder

Staff: Dieter Burrell, Scott Campbell, Stephanie Carpenter, Margaret Levenstein, Saundra Schneider, Anna Shelton, Fillippo Stargell, and David Thomas

I.   Education in the ICPSR Strategic Plan

Maggie Levenstein reported on the establishment of an Education Strategic Planning committee (Libby Hemphill, lead, Sandy Schneider, and Lynette Hoelter) to review the role of educational activities, writ large, among all ICPSR units.

First Step: staff has already completed an in-house survey of all education and training activities within all ICPSR units.

Second Step: the Education Strategic Committee is currently using the survey to develop strategic education goals for ICPSR. Among the issues being considered are:

- which audiences ICPSR should serve
- what activities benefit which audiences
- how best to establish priorities for education or training activities

Maggie also reported on the establishment of a 'Conferences Strategic Plan' for ICPSR, which would potentially help to coordinate conferences activities across the organization. Staff noted that that sponsored archives pay for their conference attendance, and that Membership attends major conferences as decided annually, and some on a rotating basis.

Lisa Cook suggested that staff investigate attending the United Negro College Fund (UNCF) conference, and to see if UNCF itself will sponsor ICPSR attendance.

II.   Undergraduate Training

Maggie Levenstein reported on a June 2018 meeting at ICPSR sponsored by the Sloan Foundation. The Sloan Empirical Reasoning project brought together ICPSR staff, Council members, and several scholars from U.S. universities with the aim to foster the development of new materials for training undergraduates in empirical reasoning. Among the ideas discussed was the notion of 'training the trainers.' Esther Wilder then gave an overview of the new grant

from NSF that CUNY has received to train and encourage faculty to include empirical analysis in their own classrooms.

III. Summer Program

A. 2018 Summer Program: Sandy Schneider reported on the status of the 2018 Summer Program. The Program offered 84 courses, taught by 110 instructors. A total of 925 participants from over 30 countries and over 30 different disciplines attended in 2018. In 2018, of participants who reported, around 60% were female and 40% male. Of those who reported, approximately 51% were white, with 49% in other categories. Of those who reported, approximately 61% were U.S citizens while 49% were international participants.

2018 Course Offerings:

- Regular Four-Week Sessions: 28 statistical workshops, 1 substantive workshop, 10 lectures.
- Short (Two- to Five-Day) Workshops; 40 statistical and 5 sponsored workshops, held at 13 locations around the U.S. and internationally.

Other highlights from the 2018 Summer Program:

- a new Middle East and North Africa (MENA) initiative in which American Political Science Association (APSA) sponsored three scholars from the Middle East and North Africa to attend the 4-week sessions
- the large number of Summer Program participants who attended the ISR Fair in July

B. 2019 Summer Program Course Schedule: Sandy also reported that she is already working on the Course Schedule for the 2019 Program. Among innovations planned for next year are new offerings on machine-learning, on Python, and a second (advanced) workshop on Race, Ethnicity, and Quantitative Methods.

C. Diversity, Equity & Inclusion Initiative (DEI) in the Summer Program: Staff and Council members then discussed DEI efforts by the Summer Program, including the diversity of the Summer Program faculty. The discussion included general issues surrounding faculty recruitment for the Program including salaries and housing issues. Sandy provided information on the Program's diversity efforts and accomplishments including co-leadership of ISR's DEI education workgroup and the recruitment of a more diverse TA cohort. She provided additional information on the ISR/UM Provost DEI scholarships that we awarded the year: 15 Summer Program scholarships were awarded to incoming graduate students from underrepresented groups entering programs at the University of Michigan, Michigan State University, the University of Florida, the University of Chicago, and CUNY.

D. More DEI in the Summer Program: Sandy announced the receipt of a Gates Foundation award of almost $150,000 to the Summer Program to provide scholarships for graduate student from underrepresented groups to attend the Program.

E. Visa Issues: Staff and Council members discussed the visa issues that affected Chinese participants last summer, as well as potential visa issues for next summer.

F. Summer Program Finances: Council and staff members then discussed Summer Program finances, and staff noted the loss of what had been a large expected payment from Chinese students who were unable to attend in 2018 due to the visa problems. Council suggested that staff investigate whether Open Society is amenable to sponsoring participants from Eastern Europe.

G. Archonnex and the Summer Program: Staff reported on the goal to move the Summer Program's registration and administration process to the Archonnex platform.

Technology Committee Meeting

Council Members: Bobray Bordelon, Elizabeth Groff, Chandra Muller, Katherine Wallman

Guest: Matthew Woollard

Staff: Trent Alexander, Johanna Bleckman, Linda Detterman, Lynette Hoelter, Stuart Hutchings, Abay Israel, Susan Leonard, Daphne Lin, Jared Lyle, Elizabeth Moss, Tom Murphy, Asmat Noori, Mike Shove, Harsha Ummerpillai, and Vanessa Unkeless-Perez

Archonnex Marketization

Tom Murphy presented what he referred to as the beginnings of a vision for marketing ICPSR's technology stack, Archonnex, with the intent of getting Council's guidance on this topic. This was developed by Trent Alexander, Linda Detterman, and Tom Murphy and based on discussions that occurred at the June 2018 Council meetings.

Tom shared that we had been approached by the University of Iowa about Archonnex as part of a market analysis they were performing to address a need they have for a hosted archive. Their archive is of non-social science data, but what we provide meets many of their needs. This inquiry about our 'product' demonstrates a few questions for which ICPSR needs direction from Council:

- Should ICPSR entertain these inquiries and or seek similar opportunities
- What is Council's expectations of how we would respond? Any requirements?

Tom provided an overview of Archonnex from the perspective of capabilities provided to ICPSR and staff roles evolved with defining, developing and supporting the associated applications and services. The two main categories of services ICPSR utilizes are Self Publishing Repositories and Curated Archives. Tom described a collection of components we deliver as a 'service bundle' and provided some examples, such as reporting and an institutional repository. As we continue to develop the overall Archonnex platform, ICPSR could possibly make available these service bundles for others to use.

Tom noted a number of questions we have been asking ourselves:

- What is Council's perspective on this topic (similar products/offerings they've seen or are familiar with)?
- What additional services should ICPSR be offering?
- Should ICPSR provide support for and or take on archives of non-social/political science data?

A challenge ICPSR is trying to solve is that of our financial sustainability. We would like a flow of funds for ongoing support and development, and technology (staff, equipment, licensing).

The discussion began with the observation from Council that ICPSR has an 'edge' due to the completeness of our offering, technology, and our subject matter experts. It is felt that ICPSR does not leverage the Subject Matter Expert (SME)'s skill set. Suggestion was made for ICPSR to look into tagging variables as data is ingested and use common/shared tags link to other studies and areas of interest in our catalog. If we could add additional value such as this tagging to our self-publishing offering, it may lead people to ICPSR and the membership. It was also suggested we could build off of this to automate adding items like 'characteristics of the study,' things the producers of the data should know to do and are easy for our curation staff.

Tom indicated that what was being described by Council is similar to a model Microsoft uses (home, enterprise, etc.), offering a path from basic to the more full service ICPSR curation. Trent expressed concern that the self-publishing repositories have a lot of competition, often are essentially free. If ICPSR were to proceed further on that track we would need a very clear strategy. He agreed the SMEs are a strength of ICPSR.

Tom clarified his purpose was to obtain guidance from Council on defining a "market plan," so ICPSR can focus our technology development plans, establish better costing for sustainable budgeting, and determine if and how we could offer our technology to potential buyers.

Several members of the committee asked additional questions about what University of Iowa was requesting from ICPSR and whether they were looking to define their own institutional repository. Tom reported it is a homegrown archive for which they are considering having us curate. How or where it would be hosted was not determined, which led to the question of was

Council open to our hosting a non-social science archive that may or may not reside at with our collection. From a technology standpoint, it is something we could make work.

Concerns were expressed. Insufficient vetting of sourcing could expose ICPSR to lawsuits. Non-social science data may involve metadata we do not have expertise with. And scale of the data was also discussed. Tom responded questions related to data outside our current SME area(s) of knowledge would require us to adapt and that our technology can reduce some of the risks associated to this. On a separate note, he reminded the committee that currently ICPSR provides ongoing support for a number of services for which we are no longer funded.

Discussion continued on the ongoing challenge of how we maintain our archives (curation, technology, enhancements). Grants are provided to 'standup' an archive but ongoing operation of archive still generates cost. Over time, long running funded archives ended up with different levels of funding. Council members asked if ICPSR was moving to be a job shop or how this advances the mission/niche. Tom stated that the main goal is to figure out a sustainable mechanism for funding our technology. Trent communicated that he and Maggie share the concerns shared by council about getting away from our principal focus and mission.

Technology Roadmap

Abay gave a review of projects to be implemented by end of 2018 and look ahead for 2019. He communicated that conflicting priorities has been an issue in the past and to correct that ICPSR leadership defined the four core priority projects: Curation, Linkage Library, StatSnap, and Researcher Passport.

Curation

ICPSR's methods and procedures for taking submitted "raw" data sets and preparing them in multiple formats for use by the broader research community.

First phase of this project was a deep dive of user story mapping to fully understand existing practices, processes and 'what Curation is.' Once application development began, the project team worked iteratively with Curation staff to clarify understanding. ICPSR is moving to a standardized workflow for all curation projects and our directory. Much of the development work is focused on eliminating manually driven processes, to help improve efficiency and change control (accuracy, veracity). Trent used this topic to help connect the financial and budgeting challenges Tom mentioned.

Recent updates:
- Deposit Tracker (help with workflow of curation requests) implemented Fall 2019
- MetaData Editor Feb 2019, Curation staff have access to a demonstration version to allow them to provide feedback with its development.
- Significant overall of Curation framework is scheduled to be completed in June 2019

StatSnap

ICPSR's online statistical analysis system; using ICPSR hosted data without requiring download. Somewhat comparable to ICPSR's legacy SDA offering, one difference is that with SDA there was additional work needed to prepare the studies for use whereas with StatSnap the standard curation process will prepare the data. Goal of StatSnap is to provide an analysis tool that can be used by users of all skill levels.

Recent updates:
- StatSnap Phase 1 provides support for frequencies, crosstabs, and univariate statistical analysis will be available in December 2018

LinkageLibrary

Effort by ICPSR to begin to provide support for user contributed content, such as: data; code; and linkages that connect different data sets.

Council members inquired whether it was ICPSR's intent to test the data/code that users provide to verify it works as claimed. Trent clarified that it is unlikely ever going to be a feature of this product; its purpose is to provide a canonical linkage of data sets. A launch strategy for this offering will be determined in 2019. First iteration available in December 2018, a "soft launch" with testing by folks at Texas A&M.

ResearcherPassport

ICPSR's effort to develop a researcher identity to help ease the process of, and administration of requests for, accessing restricted-use data.

Initial site is up and provides ability to apply for a passport and additional details on the offering. Next step include developing the ability to share the passport and create a visual representation. This first iteration provides no vetting of information people submit. We intend to join this effort to a separate project underway that will introduce OpenBadges and begin to provide that capability.

ICPSR's effort to refine project development and management process
Abay gave overview of what is involved in the effort to define, prioritize, develop, and test new features and enhancements. He provided an in depth explanation of how our development approach has evolved and how we are managing coordination between each of our product owners. Many projects are on our list and resources to develop, test and implement are limited. Meeting wrapped up with demonstration of the products discussed.