Council Meeting Minutes
October 6 – 7, 2022

Attendees:

Council Members (In-Person): David Armstrong (Chair), Bobray Bordelon, Jon Cawthorne, James Doiron, Kristin Eschenfelder, Susan Frazier-Kouassi, Trevon Logan, and Gisela Sin

Council Members (Remote): Michael Cafarella, Mark Hansen, Ken Smith, and Katherine Wallman

ICPSR Staff (In-Person): Trent Alexander, JD Alford, Anuvesh Anugu, Scott Campbell, Tessa Coleman, Alina Conn, Linda Detterman, Aalap Doshi, Allyson Flaster, Robert Franzese, Robert Gessner, Libby Hemphill, Lynette Hoelter, Jeannette Jackson, Joy Jang, Ann Jastrzembowski, Jeff Jones, John Lemmer, Maggie Levenstein, Jared Lyle, John Marcotte, Trisha Kunst Martinez, Michelle Overholser, Amy Pienta, Amy Pienta, Darleen Poisson, Kyrani Reneau, Annalee Shelton, Fillippo Stargell, David Thomas, and Rujuta Umarji

ICPSR Staff Remote: Gauri Ajri, Annahita Akbarifard, Dharma Akmon, Ambyr Amen-Ra, Homeyra Banaeefar, Annie Beaubien, Sandeep Bedadala, Sage Belrose, Zachary Bennett, David Bleckley, Stephanie Carpenter, Megan Chenoweth, Robert Choate, Rebekah Chu, Lara Cooper, Gin Corden, Evan Cosby, Elyssa Daniel, Austin De Spirito, Dom DiFranco, Amanda Draft, Benjamin Dreyer, Julie Eady, Allyson Flaster, Robert Franzese, Irmarie Fraticelli-Rodriguez, Aubrey Garman, Chelsea Goforth, Nishit Gopani, David Thomas, Hassan Hodges, Sal Holguin, Rachel Huang, Stuart Hutchings, Dory Knight-Ingram, Abay Israel, Meghan Jacobs, Euijin Jung, Kevin Kapalla, Jennifer Koski, John Kubale, Kathryn Lavender, Susan Leonard, Jenny Li, Scott Liening, Daphne Lin, Dale Luff, Lynn MacLeod, Vrinda Mahishi, James W McNally, Erin Meyer, A.J. Million, Elizabeth Moss, Narmadha Rajendran, Justin Noble, Kelly Ogden-Schuette, Sam Orosz, Michelle Overholser, Alieysa Patel, Shelly Petrinko, Katey Pillars, Felix Ramirez, Kelly Doonan-Reed, Kyrani Reneau, Sarah Rush, Chelsea Samples-Steele, Mike Shallcross, Michael Shove, Sam Imbody, Brenae Smith, Michael Staggs, Cory Steiner, Alison Sweet, Sharvetta Sylvester, Sandra Tang, Michael Traugott, jay winkler, Melissa Wlodkowski, Sandy Zalmout, and LingLing Zhang

Council Member Absent: Randall Akee

Approval of the Minutes

David Armstrong called the meeting to order and called for approval of the April 2022 Council. The April 2022 minutes were approved unanimously.

Director's Report

Membership Update
ICPSR membership is at an all-time high of 812 institutions. Membership revenue is at $4.5 million in FY2022. This reflects several things: 1, the enormous hard work of our Membership and Communications team, a special shout out to Anna Shelton and the entire Membership and Communications team. 2, the value that the organization creates for our membership we make it possible. 3, data ubiquity is increasingly recognized and reflected in the policy environment; this is much bigger than just ICPSR but makes ICPSR an extremely important player in this broader context. Institutions want to be connected to us because of our broad leadership in this area.

Examples of the broader data policy context

- Foundations of Evidence-Based Policymaking Act of 2018 aka Evidence Act or Open Data Act
- Advisory Committee on Data for Evidence Building
- 2023 NIH Data Management and Plan Policy
- White House Office of Science and Technology Policy
- CHIPS Act

ICPSR's advocates for sustainable, Findable, Accessible, Interoperable, Reusable (FAIR) data so that research is transparent and reproducible, data that is as open as possible, as closed as necessary. Calls for "open" data often miss the importance of sustainability, FAIRness, and confidentiality protection.

ICPSR's innovative sponsored projects

- Research Data Ecosystem
- ResearchDataGov (RDG): supported development of new restricted data application system, ReDMAS (as well as building the single application portal for the federal statistical system mandated by the Evidence Act)
- Millennium Challenge Corporation (MCC): supported development of new data model that allows us flexibility to manage different data types and connect different objects

ICPSR Values
We have also grown considerably since March 2020, and we have tried to do this while leaning into the core values for our organization. These include:

- Intentional communication
- Innovative learning organization
- Collaboration
- Connection with ICPSR and its mission

Diversity, Equity and Inclusion
The University of Michigan had a DEI 5-year strategic plan; we are in the process of drafting the next five years (DEI2.0). ISR plans a town hall called *Be Heard* on November 1, 2022, which is an opportunity for everyone in ISR to make their voices heard about what they think we should prioritize in that plan. At ICPSR, we are beginning the process of identifying our goals, for inclusion in ISR's plan.

Living with COVID
We are largely back in person; on any given day about 40% of ICPSR is in the building. That varies from day to day except on Tuesdays, which is the ISR is our focal day. Our goal is to provide opportunities for people to come together as well as the flexibility people desire.
We want to ensure that the Summer Program thrives in this new hybrid world. Last summer's hybrid program was successful, but not as successful as a program as it could have been had there been a larger number of in-person participants. We are planning changes that will enable and enable more in-person participation.

Discussion and Questions:
Bordelon: Given the labor market situation, what are plans for staff retention?

- Competitive salaries, consistent with UM policy
- Flexibility in remote work, when consistent with job requirements

Action Items/Next steps none at this time

Director's Report Slides.

ICPSR Innovation: Social Medical Archive (SOMAR)

Presented by Libby Hemphill, Research Associate Professor, ICPSR; Associate Professor, School of Information; Director, Resource Center for Minority Data; Director, Social Media Archive; Associate Director, Center for Social Media Responsibility.

SOMAR goals:

- Enable research using data from social media platforms
    - Data access and protection
    - Timeliness and preservation
    - Training
    - Linkage to other data resources
- Honor social media user's preferences and expectations when possible.

Challenge: conflict between perspectives of researchers and social media users.

- Researchers: data has been generated, so it should be available for my research
- Users: they do not owe researchers their data. They do not owe anyone insight into their regular lives, even if their data is public by some definition of public.

SOMAR endeavors to honor the preferences and expectations that users were generating this data have about their data when we can or try to balance those with society's needs to understand ourselves.

Research goals
1. Replication (for verification or extension)
2. Verify
3. Extend

Primary threats
- Data decay
- Difficulty in verifying sources and results
- Lack of standards for capturing provenance, metadata
- Scale of work for individual researchers to recreate data

Discussion and Questions:
Alexander: Is it a core goal of the project to provide a computation environment?
Answer: Yes, for two reasons: 1, because of scale, costs, and regulation, data doesn't move. 2. One of SOMAR's major goals is to support researchers in less well-resourced institutions who would not otherwise be able to participate in this research program.
Hansen: Do we know what computational analysis we will be supporting?
Answer: R, Stata, and Python. Hiring a data science person to provide technical user support, especially for working with data in sequel.

Action Items/Next steps: None at this time

SOMAR Updates Slides

ICPSR Innovation: ResearchDataGov (RDG)

Presented by J. Trent Alexander, Research Professor, ICPSR and Population Studies Center; Associate Director, ICPSR

Research Data Gov (RDG) is the single portal for applications for restricted federal statistical data. RDG covers the 16 principal agencies of the Federal statistical system. All those agencies provide restricted data to external researchers, and each, until very recently, had their own unique application process with different requirements, different timelines, different metadata. The Evidence-based Policy Making Act, co-sponsored by Congressman Paul Ryan and Senator Patty Murray, established a commission that recommended harmonization of these processes, and that was mandated by the Evidence Act passed in 2019. In August 2019 ICPSR was awarded a contract to pilot this system, using our existing infrastructure (IDARS) with many necessary tweaks. In 2020, NSF awarded us a 5-year contract to make a production version of this system that would support all 16 principles statistical agencies and that portal will be used for all restricted data applications for federal data beginning in December 2022.

Discussion and Questions:
Alexander demonstrated the site.
Smith: How will this work for proposals that want to link data sets? Are personal identification keys that facilitate linkage already in place? If users request data from multiple agencies, do they get separate reviews from each agency involved or a synthesized response?
Trent's response: If the data have been PIKed, the metadata in RDG will indicate that, so that it is possible to link them. If proposals are reviewed by multiple agencies, the user can see each agencies response separately. When there are multiple agencies, the agencies have a longer period to submit their responses.

Action Items/Next steps: None at this time

RDG Slides

Budget and Finance Committee Report

Jon Cawthorne, Committee Chair, reviewed the ICPSR budget for FY22. Jon thanked the members of the committee for their time and effort in reviewing and bringing the documents together for the October Council meeting. Jon also thanked Membership and Communication for the revenue raised.

ICPSR's three sources of revenue: membership fees, sponsored projects, and the summer program. ICPSR's overhead costs include: ISR building fund and the I/O split. The FY22Budget had an approved deficit of $319K; the Year-End of FY22 deficit was $1.9M. The primary reason for the large deficit was the delay in the launch of RDE. The FY23 budget projects a deficit of $258K.

FY 2023 revenue: membership revenue no change from 2022
FY 2023 operating expense: non-salary expenses have fallen due to limited travel during the pandemic. Provost tax expected to increase from 454K to 746K, as these are tied to our (increasing) direct expenditures.

Impact of RDE: ICPSR has increased staff from 101 FTE in January 2020 to 168 FTE in September 2022.

Discussion and Questions:
Bordelon: Are staffing levels currently where they should be? Are we hiring in curation?
Umarji: Curation staffing levels now match work needs.
Eschenfelder: Is ICPSR increasing FTEs in HR or are those ISR employees?
Lemmer: ICPSR hired one HR staff person (Ann Jastrzembowski), but most HR employees are at ISR and paid for through the I/O split.
Bordelon: Shoutout to former chair, Lisa Cook, who changed Finance presentations to make them more accessible to Council. Thanks to the ICPSR staff, to Jon Cawthorne and Lisa Cook and all the others, (Susan Frazier-Kouassi and Trevon Logan).

Cawthorne: Shoutout to John Lemmer, JD Alford, Darleen Poisson, and Fillippo Stargell, as well as new committee member Susan Frazier-Kouassi.
Smith: How is admin doing given the large growth in the organization, but relatively small growth in admin?
Lemmer: Admin is a good group of people, and we turn to ISR when we need additional help.

Action Items/Next Steps: None at this time.

Budget and Finance slides

Membership and Outreach Committee Report

Katherine Wallman, committee chair, congratulated Linda Detterman and her team. ICPSR membership has increased to over 800 members, an all-time high for its membership. COVID has not adversely affected membership numbers as had been feared. The Membership There were a few cases where ICPSR temporarily reduced member fees due to the pandemic. The anticipated decline in membership revenues did not emerge.

Other activities

- Data Fair, led by Annalee Shelton, deserves Council attention for its success. More than 800 individuals registered for this year's data fair. The virtual fair does not have the ambience of an in-person even, but does allow many, many more people to participate.
- Modernized branding to be completed by 2025, with some changes as early as 2023.

Council has continuing concern about how all the work on all the shiny new things may affect or interact with what are kind of the core functions of ICPSR. Challenge is to take advantage of the shiny new developments so that they bring value to ICPSR member institutions and the research community.

Doiron: echoes the kudos to the Membership unit.
Frazier-Kouassi: Described her NSF proposal targeting 4 HBCUs .
Bordelon: What would be the cost to subsidize MSIs that cannot afford membership.
Detterman: We have ongoing outreach that offers reduced membership fees but receive relatively little response. The critical issue is identifying an advocate who can serve as OR. The OR can provide training and awareness, as free access alone does not increase utilization.
Bordelon suggested that many ORs would volunteer to work with ORs at MSIs.

Action Items/Next Steps: None at this time

NO SLIDES

Summer Program Committee Report

David Armstrong, committee chair, reviewed the Summer Program's progress for 2022. Attendance was down, although not as much as we had feared. The short courses had the biggest attendance drop. Attendance was much better in the 4-week courses. Online participation was the modal engagement mode, somewhere in the neighborhood of 80% online in the 4-week courses, split roughly evenly between synchronous and asynchronous engagement. Many of the short courses went to an entirely online format. Our data collection from participants found that the following were the principal reasons that they opted for remote access.

- Cost: 39% (housing or travel or some combination)
- Family obligation (24%)
- Work obligations (12%)
- Travel logistics (10%)

The Summer Program team is working to facilitate in-person participation, understanding that it is likely that the hybrid program will remain in place, because remote access reaches people who cannot participate otherwise. The Summer Program recognizes that housing is a challenge for instructors and participants alike. Changes in the Ann Arbor rental market have made it harder to sublet in the summer, decreasing the traditional options.

In terms of longer-term planning, we recognize that we are trying to achieve multiple goals. The quality of the experience is higher when in-person because of the informal networking that is supported, but in-person participation has higher access barriers, both in terms of time and money. The online program lets us deliver these courses to people who simply can't come in person. The pandemic has opened the program up to that audience, and we all recognize that it's important that we don't abandon that audience now. We need to offer access to these courses in some way or another, as much as we can, while also trying to incentivize in-person participation by reducing costs. We also continue to think about the nature of the relationship with our off-site locations. Franzese's overall goal is to increase the reach of the Summer Program to currently underserved populations.

Discussion and Questions:
Franzese noted that the short courses ended up almost entirely virtual. He is considering the creation of an intercession in between three-week core sessions, as offers a bit more flexibility for shorter courses.
Frazier-Kouassi asked if the proposal to move to 3-week sessions was based on feedback from instructors.
Franzese said that the proposal was responsive to feedback from instructors, but that the idea came from the SP leadership.
Frazier-Kouassi asked about housing options and the rental market.
Franzese responded that they are under negotiations to secure dormitory housing.

Summer Program Slides

Technology and Standards Report

Kristin Eschenfeld, committee chair, presented the Technology and Standards report.

- Ongoing human resource challenges related to both remote work and pay. The possibility of remote work means that ICPSR now faces competition from employers in other regions.
- Research Data.Gov release date is November 1, 2022. It is currently in a testing environment and getting a lot of feedback. More generally ICPSR is building processes, including a more complex and robust testing environment.
- Curation group has been working on changes to curation levels, with the goal of trying to increase the efficiency of the curation process
- Using staff survey from 2021 to improve operations
- Bibliography team is working to automate bibliography maintenance. The team described a project to use natural language processing to identify data citations in a document; committee members had helpful suggestions for the team.
- Metadata team has various projects underway to ensure that metadata remains standardized with the new, more flexible data model. One is a metadata portal, an internal tool for staff to view and select metadata options. They have created a metadata governance scheme, so that proposed changes to metadata are not made on an ad hoc basis but rather use the approved process.
- Metadata team is working to build an API to allow external parties to download ICPSR metadata to DCAT or other metadata formats.

Discussion and Questions:
Smith asked if we should curate open ICPSR.

Eschenfelder added that are two related issues:

1. Should the staff spend time working to improve the metadata submitted by people to open ICPSR
2. Should we improve metadata that ICPSR to better describe the relationship between datasets.

Bordelon noted that many of the OpenICPSR studies have provocative titles, but no documentation, so they are not useful.

Levenstein said that ICPSR will not curate all OpenICPSR. OpenICPSR should be used for replication data sets. High value data in should be curated and included in the membership collection. We have permission to curate OpenICPSR deposits and include them in the membership collection, but it does create confusion as there are two active, but not connected, versions of the data in our catalog.

General Questions and Discussion:
Bordelon asked to discuss the Meta donation to ICPSR and whether anyone was consulted about ICPSR's willingness to accept this gift. Levenstein noted that this gift was thoroughly vetted at the ISR and UM levels.

Action Items/Next Steps: None at this time

NO SLIDES

Council Discussion and Votes

October 7, 2022

Voting and Discussion

Summer Program
Motion: The Summer Program 2023 schedule will consist of two, three-week sessions, and a one week intersession. This proposal passed unanimously.

Reconfiguring the Membership and Education Committee to include data acquisitions and user support

Armstrong noted that in April 2022, ICPSR staff suggested that the Membership & Education Committee expand to include Acquisitions, since the current committee structure does not provide a forum to review and advise on data acquisitions strategies or challenges.

The purpose of the Membership, Acquisitions, and Education Outreach Committee would be to:

- Review membership health status including member counts, revenue collection, and set annual membership fees; address concerns regarding current/future membership in ICPSR and assist in strategy development
- Review Acquisitions health (deposits and releases of data, with a focus on member data); assist in acquisitions strategy development and address concerns; review and approve collection development policy as required
- Discuss updates to agency data sharing policies (NIH/NSF, etc.) and training initiatives
- Review and advise education outreach & communications and on member data tools and services development, including user support
- Advise on teaching and learning initiatives

The expanded committee would begin with the April 2023 Council meeting; the Acquisitions and Membership & Communications Unit leads would construct the committee agenda and invite relevant staff to pre-Council committee meetings.

DISCUSSION:
Councilmembers commented that this would make the committees more balanced in their responsibilities and more completely cover ICPSR activities.
Motion: the scope of the Membership and Education Committee will be expanded to cover acquisitions and user support.
This passed unanimously.

Open discussion with ICPSR Staff

Anna Shelton asked Council: What is it that matters to you about ICPSR?

Bordelon: The people. A trusted repository that has for 60 years led the way with social science data making sure it's there in the future, findable, seeing the impact through citations. SOMAR. High quality education.

Armstrong: The Summer Program in 2001 changed my life. Took a class on measurement and scaling, took those skills back and those skills directly turned into a publication which got me a postdoc which got me a job which put me in touch with a professor which led to the job that I'm doing now. The connections I've made through the Summer Program have smoothed a path through my career I've made in the last 15 years. The people I met in the first couple years are still some of my best friends. I met my wife here. It's not an exaggeration to say that the Summer Program changed the trajectory of my life.

Doiron: The people, the people. The relationships. That's a testament to the staff, the people here that make up this place. I was a student who knew I had to carve my path in research. I remember my mentors talking about ICPSR in this way, that it was a force. That it had utility to create culture change. ICPSR was like the shining beacon of light, what we aspired to. When you have naysayers who say you can't do these hard things, I would point to ICPSR and say yes, you can, they do and that means we can. I think about my time on Council - you get more than you give. I bring things back to my institution.

Frazier-Kouassi: Both the Summer Program and access to the data, it's key in helping faculty and grad students accelerate their careers. When I went to Prairie View where I am the OR, I said, we need to belong to ICPSR. I said "This will be life changing. If you take a course or access the data, it can change the trajectory of your career." Look at what's available to you to accelerate your research agenda.

Sin: I knew about ICPSR when I was an undergrad in Argentina. I didn't' know about Michigan, but I did know about ICPSR. ICPSR has such a wide and far reach. This is me down in Argentina when internet wasn't even so available. I knew about data and the Summer Program.

Eschenfelder: I do research on data science data infrastructure in the United States. We need to remain a leader and always pushing the boundaries.

Wallman: I'm an outlier. Had I known what I could get at ICPSR I would have come a long time ago? To me, ICPSR is an opportunity. It may be well known to those in the academy but may not be well known to those in other places like the government or the private sector.
Logan: I don't think I've published anything at my career that doesn't include data at ICPSR. 1) History of the institute is the intellectual history of people who were novel data creators in earlier generations. There are many diamonds at ICPSR. There are nuggets in surveys that were done decades ago that can answer today's questions. 2) ICPSR can prevent us from reinventing the wheel. The privatization of data that was the practice of earlier generations has now been democratized. That allows us to make better science. The number of things we can do as social scientists has expanded exponentially. If you think about the world where ICPSR does not exist, you would probably take away at least half of our empirical knowledge base in social science. I think about the things that come across my desk the things I read, at least half would not be there if there was no ICPSR.

Cawthorne: I learned about ICPSR through the provost at Wayne State. I was told by him that "I've gone to this meeting a couple of times, and I think these are my people." This isn't different than anybody else who is trying to come together to do new stuff. It's hard stuff, its good stuff. To be a leader, you need to be an example. Keep struggling, keep addressing these issues. You'll keep doing amazing things.

Appendix

I. Directors Report

II. Social Medical Archive (SOMAR)

III. ResearchDataGov (RDG)

IV. Budget and Finance Committee Report

V. Summer Program Committee Report

ICPSR
The Data Consortium

Director's report

October 6, 2022

ICPSR

ICPSR Birthday Reception

Today!
5:00-7:00 pm
Perry Atrium

ICPSR

2

Membership update

- ICPSR membership is at an all time high of 812 institutions
- Membership revenue also at all time high
  - Collected over $4.5 million in FY 2022 (ending June 30, 2022)

- Reflects
  - Hard work of our Membership & Communications team
  - Value that the whole organization delivers to our members
  - Data ubiquity
    - Importance to understanding the world we live in and empowering us to act upon it

ICPSR

Data Policy Environment

- Foundations of Evidence-Based Policymaking Act of 2018
  - AKA Evidence Act or Open Data Act
  - Advisory Committee on Data for Evidence Building
- NIH Data Sharing Policy
  - Covers data from projects with applications submitted after January 25, 2023
  - ICPSR guidance to NIH research community
    - https://www.icpsr.umich.edu/web/deposit/cms/4254
    - Guidelines for data management and sharing plans
    - Where to appropriately share your data
- White House Office of Science and Technology Policy
  - Memorandum issued August 25, 2022
- CHIPS Act
  - Section 10344. Research Reproducibility and Replicability

ICPSR

Data Policy Messaging

- Data must be FAIR
  - Findable, Accessible, Interoperable, Reusable
- Research must be transparent and reproducible
- For “open” data to support reproducible research, we need
  - Confidentiality
    - As open as possible, as closed as necessary
  - Sustainability
    - Accessibility has to include sustainability

ICPSR

5

RDE is ICPSR’s answer

- New data types, larger scale, frequently updated, geographic specificity, linked to other data products, no documentation

- RDE funding technological transformation to support new capabilities
  - It’s hard, it requires a lot of people, it requires a lot of patience
    - Goal is integrated system, so we are having to update the underlying platform
    - New ways of working, new development environments
  - We are making progress

ICPSR

6

RDE Goals

- Focus on End-users
  - Provide straightforward pathways for users to do their work
- Modernized Technology
  - Take advantage of latest advances in technology
- Access to Research Materials
  - Easier to find and apply for access
- Connections between Research Materials
  - Single stream, integrating self-publishing and curated data, research output
- Trust across ICPSR and the Research Community
  - Supporting transparency and reproducibility over data lifecycle

RDG & MCC supported first steps

- MCC supported development of new data model that allows us the flexibility to manage different data types and connect different objects
  - Didn't create all the software that makes use of newer schemas
  - Metadata and Preservation has been working hard at defining schemas and appropriate workflows
- RDG supported development of new restricted data application system, ReDMAS
  - In beta testing for restricted federal statistical data users now
    - https://www.researchdatagov.org/
  - Has pulled us into struggles between the 16 federal agencies that are required by law to use RDG (which they call SAP)

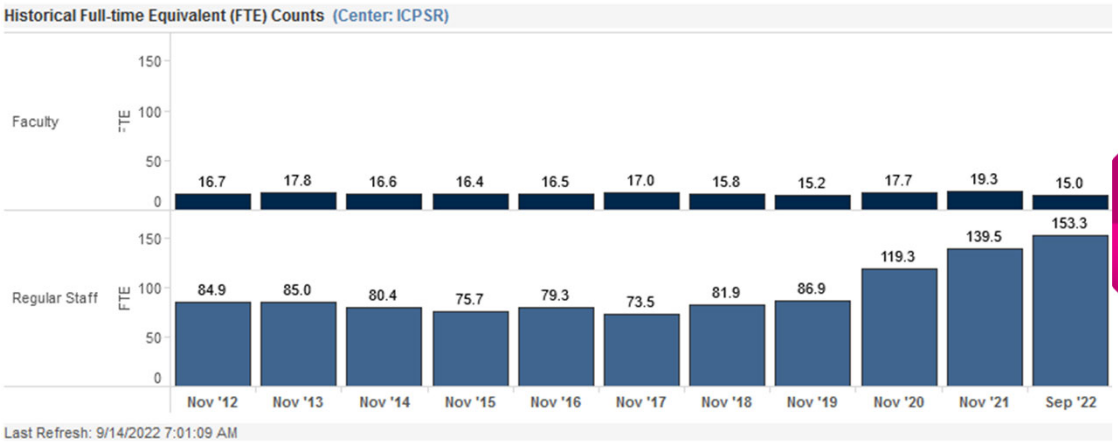
ICPSR values

- Communication
  - Intentionality
  - Assuming good intent
- Innovation
  - Learning organization
  - Collaboration: no one can know or do it all
- Connection with the organization
  - We have grown and changed a lot, fast

ICPSR

9

ICPSR staff growth



ICPSR

10

We are hiring!

- Let your friends, family, networks know
  - Project managers
  - Data analysts
  - Librarian
  - Research administration
  - Software engineers
  - Devops engineers
  - Data engineer

ICPSR

11

We need to hire

- We have a lot of work to do
- Spending money helps us generate money
- RDE pays for new technology, but tech is only part of it
- Growing security team
  - Privacy Officer with three direct reports, plus security position in CNS
  - Disclosure review board
- Growing bibliography team

ICPSR

12

Hiring creates opportunity

- DEI 2.0 five-year strategic planning under way
  - Be Heard all-ISR town hall, November 1
  - ICPSR process
    - Our mission: advance and expand social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations
    - Our DEI statement says that we believe our commitment to DEI is fundamental to our achieving our mission
      - What are we doing well that we should continue?
      - Identify unmet goals from DEI 1.0
        - Why didn't we meet them?
        - What can we do differently?
      - What new goals are we missing, and how can we achieve them?
    - Period of rapid growth is opportunity to further diversify our faculty and staff

ICPSR

13

Living with COVID

- Recreating in-person Summer Program experience
  - Shortening program in order to reduce housing costs, fit within university calendar to access housing
  - Will maintain hybrid option
- ICSPR has successfully transitioned to hybrid, about 50% in office on any given day
  - Need to have events when people who are 100% hybrid are physically present
    - Maybe at organization level, maybe at unit level

ICPSR

14

Data Jeff wants to party with you!

ICPSR Birthday Reception

October 6, 2022 5:00-7:00 pm
Perry Atrium

15

Questions?

ICPSR

60 years

ICPSR

SHARING DATA TO ADVANCE SCIENCE

SOMAR Updates

Libby Hemphill

ICPSR INSTITUTE FOR SOCIAL RESEARCH UNIVERSITY OF MICHIGAN

SOMAR
Social Media Archive @ ICPSR

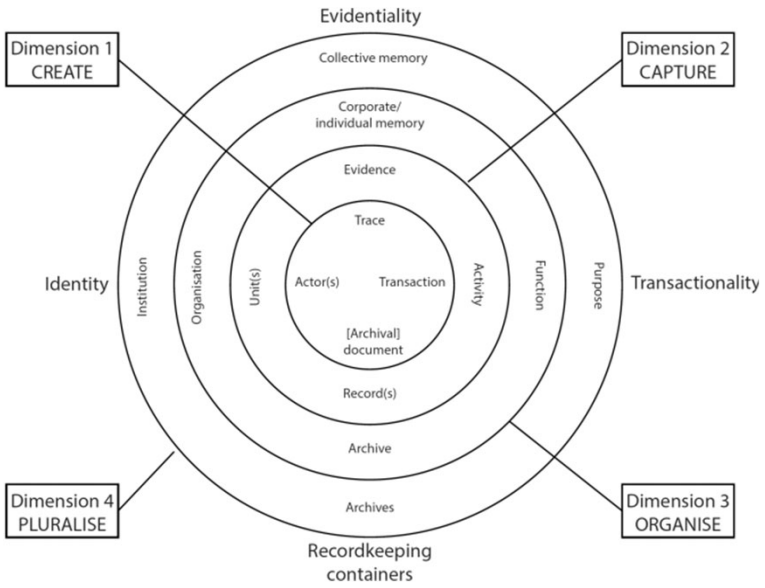
Goals

- Connect researchers with data from social media platforms, now and in the future
- Honor social media users preferences and expectations when possible
- Facilitate research that relies on social media data

Research Goals and Threats - Why is SOMAR necessary?

Research Goal	Primary Threat
Replicate	Data decay
Verify	Data documentation
Extend	Data accessibility

Archival Theory: Records Continuum Model



Upward, F., Reed, B., Oliver, G., & Evans, J. (2018). *Recordkeeping informatics for a networked age*. Monash University Publishing.

What the Records Continuum Model Means for SOMAR

Stage/Transformation	Example Implication for SOMAR
Create	Attend to the individual decisions that lead to data creation (e.g., why did Libby retweet a Twitter thread about bike lanes yesterday?)
Capture	Close to tradition “appraisal” or “collection development policy” - why created data should we capture?
Organize	Metadata - how do we adapt and extend provenance and original order to meet new networked models of connecting individual datum to one another
Pluralize	Who should have access to which data and under which conditions?

SOMAR’s Activities

Stage/Transformation	Activity
Capture	- Sampling social media streams and collecting data
Organize	- Experimenting with PostgreSQL, ElasticSearch, and Parquet data storage and indexing - Work with metadata team to create new documentation guidelines
Pluralize	- Cloud-based, scalable virtual data enclaves - Agreements with data providers to review applications for access to restricted use data - Agreements with platforms to provide proxied access to their APIs - Partner with other archives and platforms to federate data

SOMAR Went Rogue

- Buy
  - InfoReady
- Build
  - Wix
- Partner
  - Meta, GESIS

SOMAR's Challenges

- Computing costs
  - Data storage: $180/TB/yr
  - VDE for the first year: $168,000 - $1.3M
- Staffing
  - Hiring data science person right now!
  - Will be hiring engineers
- Legal agreements
  - Both sides are slow and cautious

Get updates!

Join the mailing list: http://socialmediaarchive.org

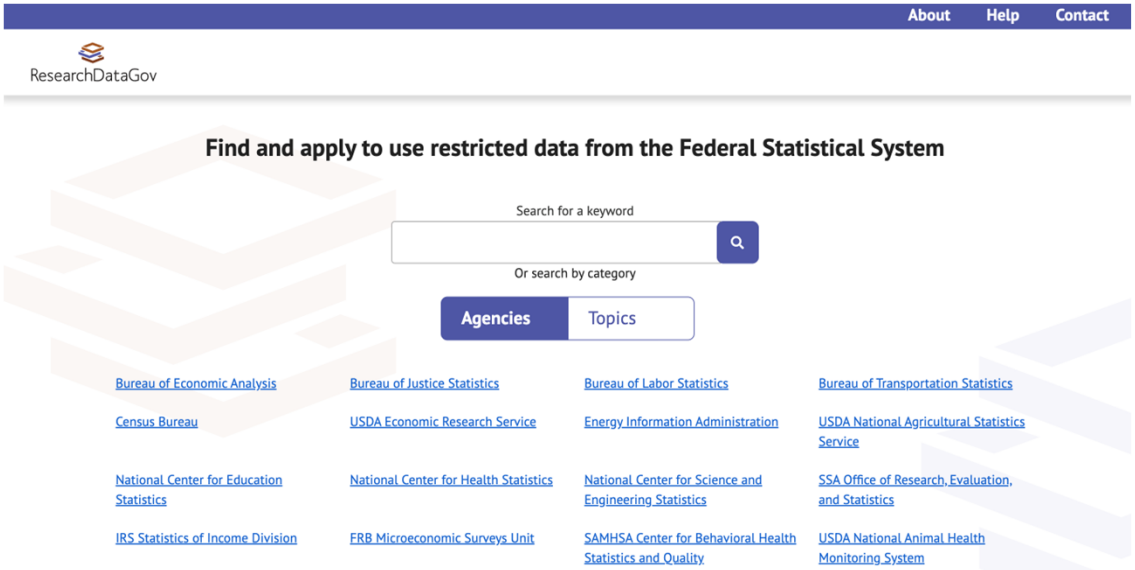
Contact the team: somar-icpsr@umich.edu

*Planned data releases

- Late 2022 through 2023
  - Data containing information on impact of social media on key political attitudes and behaviors during the U.S. 2020 elections

SOMAR Advisory Board Members!

Name	Affiliation	Term
Rebekah Tromble	George Washington University	2022 - 2024
Nate Persily	Stanford University	2022 - 2024
Kristin Eschenfelder	University of Wisconsin - Madison	2022 - 2024
Alex Hanna	Distributed AI Research Institute (DAIR)	2022 - 2024
Meredith Clark	Northeastern University	2022 - 2024
danah boyd	Microsoft Research (MSR)	2023 - 2024



Why we built it

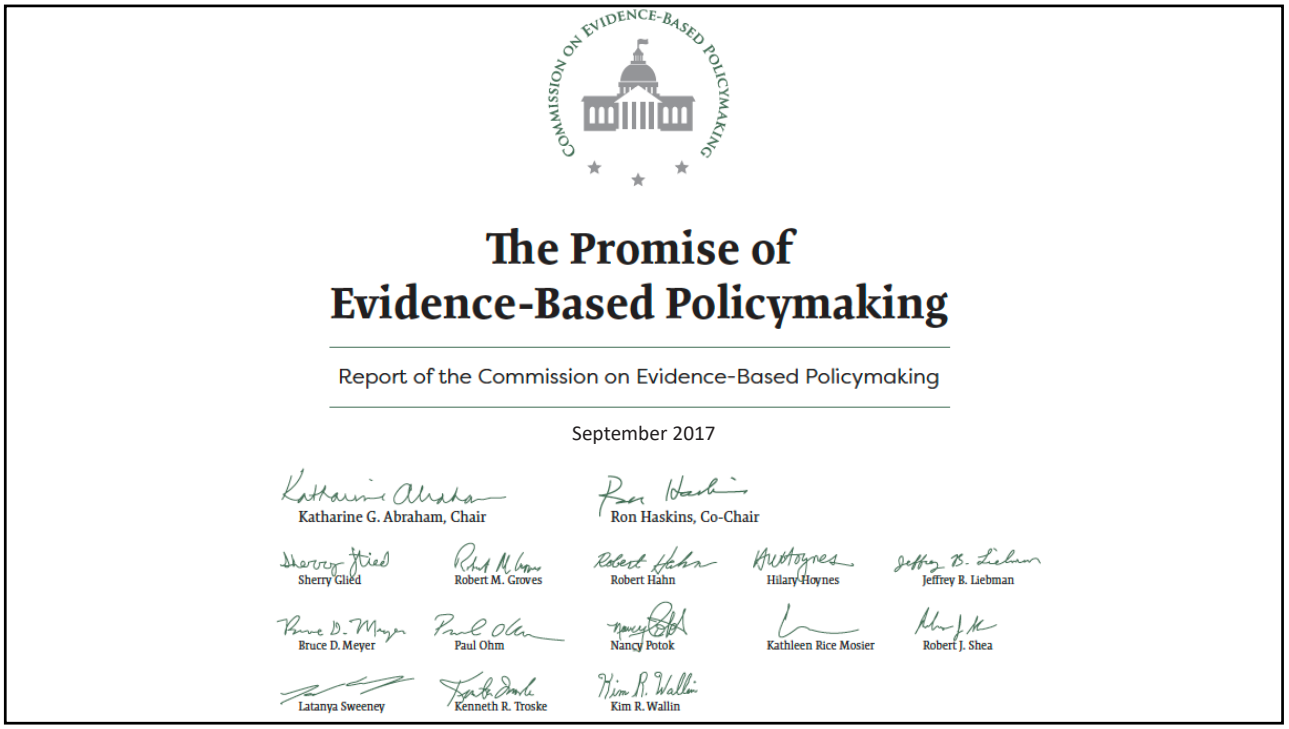
How it works

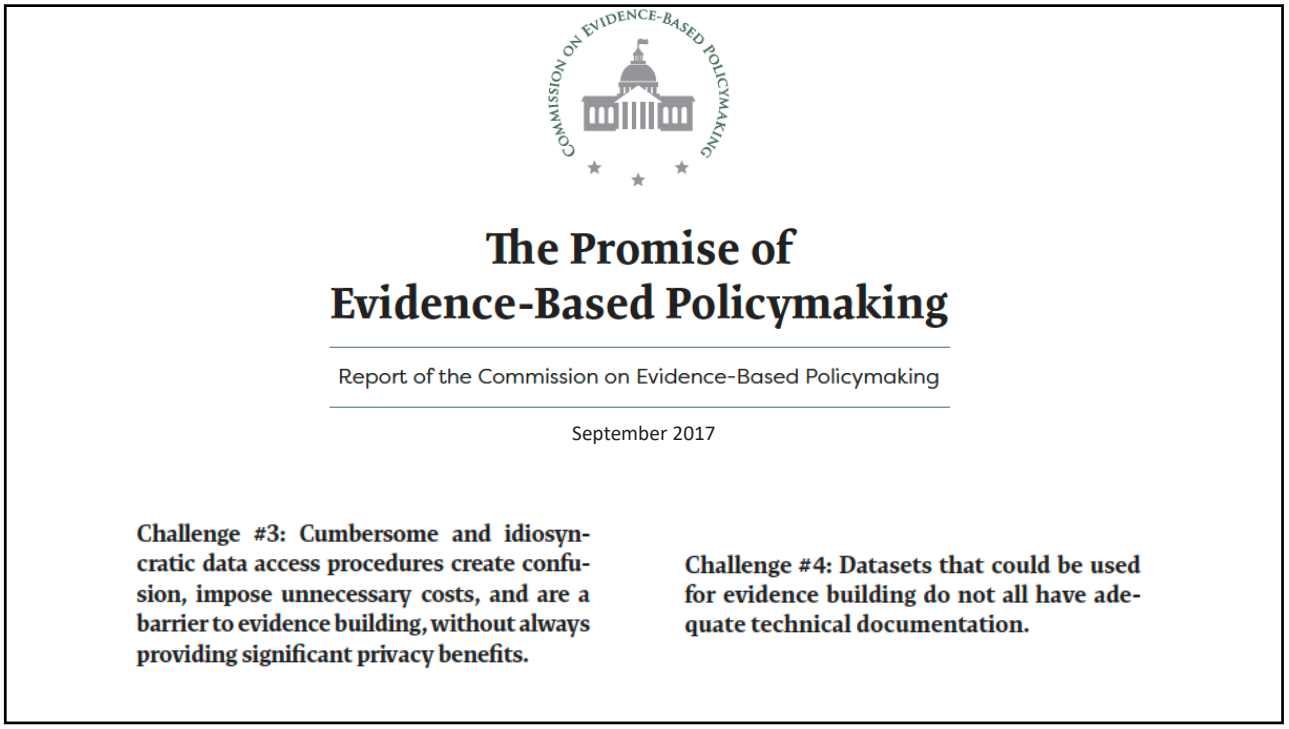
1

Evidence-Based Policymaking Commission Act

Public Law 114-140, 2016

2


COMMISSION ON EVIDENCE-BASED POLICYMAKING
The Promise of
Evidence-Based Policymaking
Report of the Commission on Evidence-Based Policymaking
September 2017
Katharine G. Abraham, Chair
Ron Haskins, Co-Chair
Sherry Glied
Robert M. Groves
Robert Hahn
Hilary Hoynes
Jeffrey B. Liebman
Bruce D. Meyer
Paul Ohm
Nancy Potok
Kathleen Rice Mosier
Robert J. Shea
Latanya Sweeney
Kenneth R. Troske
Kim R. Wallin


COMMISSION ON EVIDENCE-BASED POLICYMAKING
The Promise of
Evidence-Based Policymaking
Report of the Commission on Evidence-Based Policymaking
September 2017
Challenge #3: Cumbersome and idiosyncratic data access procedures create confusion, impose unnecessary costs, and are a barrier to evidence building, without always providing significant privacy benefits.
Challenge #4: Datasets that could be used for evidence building do not all have adequate technical documentation.

Foundations of Evidence-Based Policymaking Act of 2018

Public Law 115–435
115th Congress

An Act

To amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes.

Jan. 14, 2019
[H.R. 4174]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Foundations for Evidence-Based Policymaking Act of 2018".

Foundations for Evidence-Based Policymaking Act of 2018.
5 USC 101 note.

5

"§ 3583. Application to access data assets for developing evidence

"(a) STANDARD APPLICATION PROCESS.—The Director shall establish a process through which agencies, the Congressional Budget Office, State, local, and Tribal governments, researchers, and other individuals, as appropriate, may apply to access the data assets accessed or acquired under this subchapter by a statistical agency or unit for purposes of developing evidence. The process shall include the following:

"(1) Sufficient detail to ensure that each statistical agency or unit establishes an identical process.

"(2) A common application form.

"(3) Criteria for statistical agencies and units to determine whether to grant an applicant access to a data asset.

"(4) Timeframes for prompt determinations by each statistical agency or unit.

"(5) An appeals process for adverse decisions and noncompliance with the process established under this subsection.

"(6) Standards for transparency, including requirements to make the following information publicly available:

"(A) Each application received.

"(B) The status of each application.

"(C) The determination made for each application.

"(D) Any other information, as appropriate, to ensure full transparency of the process established under this subsection.

"(b) CONSULTATION.—In establishing the process required under subsection (a), the Director shall consult with stakeholders, including the public, agencies, State and local governments, and representatives of non-governmental researchers.

"(c) IMPLEMENTATION.—The head of each statistical agency or unit shall implement the process established under subsection (a).".

6

Timeline of work at ICPSR

2019 - August	Census Bureau/OMB awards ICPSR contract to pilot an inventory & application system
- November	ICPSR releases ResearchDataGov (version 1)

Timeline of work at ICPSR

2019 - August	Census Bureau/OMB awards ICPSR contract to pilot an inventory & application system
- November	ICPSR releases ResearchDataGov (version 1)
2020 - December	NSF awards ICPSR contract to make production version of portal
2021 - September	NSF and ICPSR complete initial requirements gathering
2022 - February	ICPSR releases ResearchDataGov (version 2), inventory only

Timeline of work at ICPSR

2019 - August	Census Bureau/OMB awards ICPSR contract to pilot an inventory & application system
- November	ICPSR releases ResearchDataGov (version 1)
2020 - December	NSF awards ICPSR contract to make production version of portal
2021 - September	NSF and ICPSR complete initial requirements gathering
2022 - February	ICPSR releases ResearchDataGov (version 2), inventory only
- Sept/Oct	Extensive user testing and feedback period
- Nov/Dec	Will release ResearchDataGov (version 2), complete system

9

ResearchDataGov project team

Principal Investigators: Trent Alexander and Lynette Hoelter

Project Manager: David Thomas

Product Managers: Stuart Hutchings, Abay Israel, Hassan Hodges, Kelly Doonan-Reed, and Wendy Umbriac

Research administration: Lisa Kelley and Darleen Poisson

Directors of CNS and Product Management: Trisha Martinez and Dharma Akmon

Technical leads: Bing She, Veena Kannan, Karun Seelam, and Raghu Ravi

Development, operations, testing
Chinedu Amadi-Ndukwe, Mei Shen, Grant Schonhoff, Dan McKay, Vasudha Doppalapudi, Lia Chen, Sam Orosz, Salvador Holguin, Narmadha Rajendran, Nishit Gopani, Ryan Lacek, Swetha Chigurupati, Anvesh Anugu, Greg Farris, Dan Pritts, Jaime Magiera, Jason Weirauch, and Eduardo Beltran

Design
Deepti Pandey, Euijin Jung, Jenny Li, Gauri Ajri, Sandeep Bedadala, Lynn Macleod, Aalap Doshi, and Wendi Fornoff

ICPSR INSTITUTE FOR SOCIAL RESEARCH UNIVERSITY OF MICHIGAN

Demo!

How researchers find data and documentation

How agencies get documentation in the site

How researchers apply to use data

How agencies review applications

11

ICPSR Operating Budget and Year-End
Fiscal Year 2022 Budget and Year-End
Fiscal Year 2023 Budget

*** This display does not include sponsored direct revenue or expenses ***

	FY22 Budget	FY22 Year-End	FY23 Budget
Total Revenue	**13,409,605**	**12,547,548**	**14,754,000**
Unit Expenses	11,828,349	12,458,680	12,859,000
ICPSR Overhead	1,900,351	1,962,730	2,153,000
Total Expenses	**13,728,700**	**14,421,410**	**15,012,000**
Operation Surplus (Deficit)	**(319,096)**	**(1,873,862)**	**(258,000)**

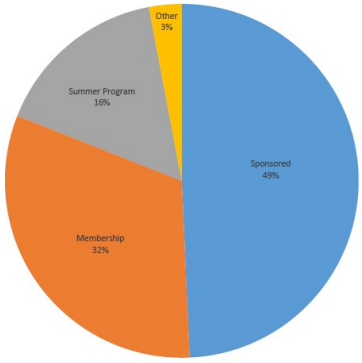
ICPSR
Statement of Fund Balance and Reserves
as of June 30, 2022

Fund Balance as of June 30, 2022		**9,974,739**
Restricted Funds:		
Endowment (Restricted Gifts) [1]	479,960	
Sponsored Fund Balance [2]	2,259,957	
Plant Fund (Restricted for Plant & Equipment)	249,237	
Unrestricted Funds:		
Quasi-Endowment (Invested Unrestricted Reserves)	2,630,898	
Cash on Hand	4,354,687	
Total Reserves		9,974,739
Less Restricted Reserves		(2,989,154)
Unrestricted Reserves		6,985,585
Less: Minimum unrestricted funds per ICPSR policy [3]		(4,292,462)
Excess unrestricted reserves		**2,693,123**

ICPSR Operating Revenue by Category

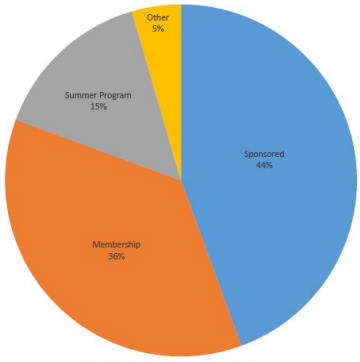
*** This display does not include sponsored direct revenue ***

FY2022 Budget



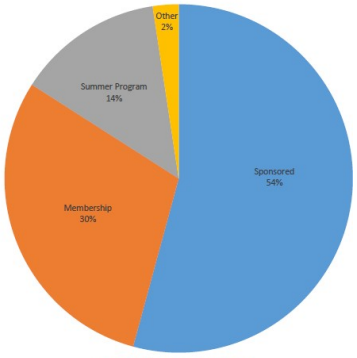
Total: $13,409,605

FY2022 Year-End



Total: $12,547,548

FY2023 Budget

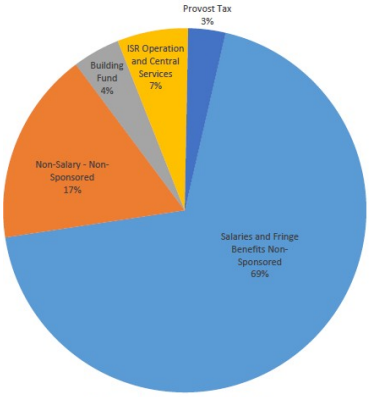


Total: $14,754,000

ICPSR Operating Expense by Category

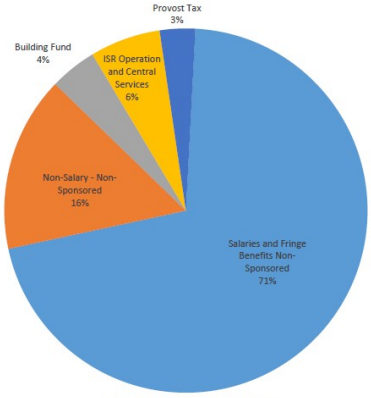
*** This display does not include sponsored direct expenses ***

FY2022 Budget



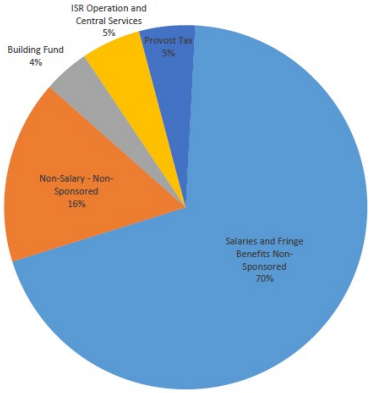
Total: $13,728,700

FY2022 Year-End

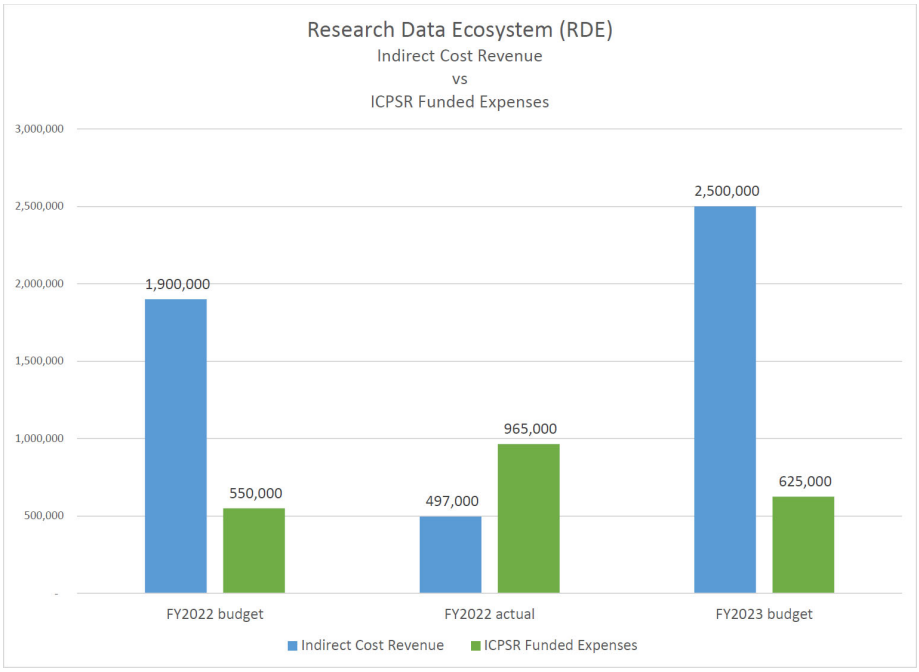


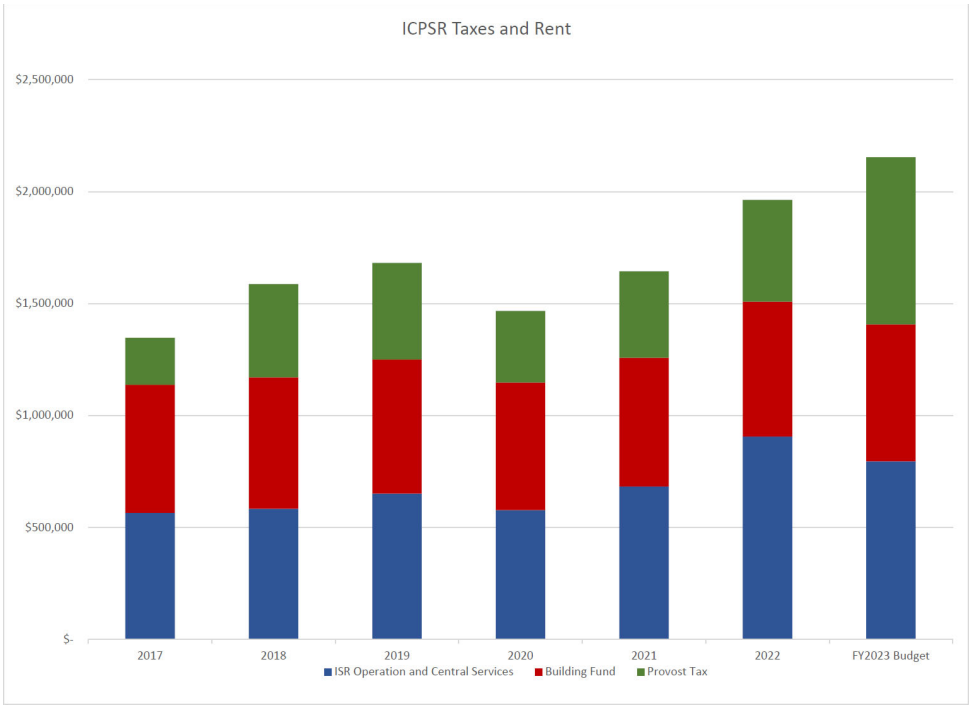
Total: $14,421,410

FY2023 Budget



Total: $15,015,000


Research Data Ecosystem (RDE)
Indirect Cost Revenue
vs
ICPSR Funded Expenses
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
-
1,900,000
550,000
497,000
965,000
2,500,000
625,000
FY2022 budget
FY2022 actual
FY2023 budget
Indirect Cost Revenue
ICPSR Funded Expenses


ICPSR Taxes and Rent
$2,500,000
$2,000,000
$1,500,000
$1,000,000
$500,000
$-
2017
2018
2019
2020
2021
2022
FY2023 Budget
ISR Operation and Central Services
Building Fund
Provost Tax

ICPSR Staffing Growth
Full Time Equivalent

Functional Unit	1/1/2020	1/1/2021	1/1/2022	9/30/2022	FY2023 Approved Budget		
					Sponsored	Non-Sponsored	Total
Administration	7.0	7.0	7.0	7.0	0.4	7.6	8.0
Business and Collection Development	7.3	10.8	10.3	11.3	8.7	2.1	10.8
Computing and Network Services	15.0	24.0	43.0	48.9	22.0	29.9	51.9
Curation	28.0	43.6	40.9	35.9	30.0	12.9	42.9
Director's Office	8.4	8.1	6.0	9.0	3.1	5.9	9.0
Education	0.5	0.8	0.8	1.0	0.0	0.8	0.8
Membership and Communications	10.4	9.5	11.5	9.6	1.3	7.4	8.7
Metadata and Preservation	6.0	7.0	8.0	9.0	4.8	5.2	10.0
Product Owner	0.0	0.0	5.0	7.0	4.3	2.8	7.0
Project Management and User Support	18.7	27.7	29.7	29.7	23.5	7.0	30.5
Total FTE's	**101.3**	**138.4**	**162.1**	**168.3**	**98.0**	**81.5**	**179.6**

Summer Program Committee Report

ICPSR Council Meeting
October 6-7, 2022

2022 Program Recap

- Attendance Down (esp. short courses)
- Online participation was the modal mode
  - ~80% online in 4-week courses (40% synchronous, 40% asynchronous)
  - >95% online participation in short courses (more asynchronous participation)
- Cost was main impediment to in-person attendance
  - Responses: Cost (39%), Family (24%), Work (12%) Travel Logistics (10%)
- Most instructors willing to return for hybrid program (~20% Unsure)
- Housing was a big issue for everyone (participants and instructors)

Proposed 2023 Program Details

- 3 Sessions: 3 weeks (6/19-7/7), 1 week (7/10-14), 3 weeks (7/18-8/4)
  - Sync with potential on-campus housing opportunity for participants
- Hybrid format as in 2022
- Maintain roughly 40 instruction hours per workshop
- Offer at least as many courses as in recent Summers
  - Classroom space is not a limitation

A vote of support was requested on the format of the 2023 Program

Longer-term Planning/Concerns

- Optimize on two dimensions – Quality and Access
  - Concerns about the long-term viability of a hybrid program
- Incentivize in-person attendance
  - Continue focus on reducing costs
  - Change the balance of incentives to be in favor of in-person participation
- Nature of relationships with off-site locations
- Increasing reach of SP to currently underserved populations