Contribution ID: **52** Type: **Presentation**

Research Data Ecosystem: Innovating Infrastructure for the Social Sciences in the 21st Century through Building a Modernized Software Platform, Data Description Framework, and Tools for the Research Data Community

Thursday 16 October 2025 11:44 (11 minutes)

ICPSR, one of the world's largest social science data archives, located at the Institute for Social Research at the University of Michigan, is leading a $38M National Science Foundation project, the Research Data Ecosystem (RDE), to modernize research data infrastructure and support efficient, cutting-edge, and reproducible data-driven science.

In this presentation, we will briefly discuss the current state of data infrastructure, outline the infrastructure needed to support 21st century social science, and show how RDE will help close the gap between the two states. The presentation will be 60 minutes allowing 30 minutes for Q&A.

The extant data infrastructure cannot adequately support 21st century social science. Diverse types of data enable path-breaking analyses into human behavior but also present challenges of scale, sensitivity, and structure, requiring new research approaches. There is an urgent need for new modes of access, confidentiality protection, methodological approaches, and tools so that research using a variety of data types meets accepted scientific standards of transparency, reproducibility, efficiency, and ethics. Current barriers include multiple incompatible standards for data, lack of interoperability, and the inherent difficulty of managing big and often-sensitive data.

Data types driving research across the social science disciplines include social media, administrative, commercial transaction, streaming, audio, video and photo, satellite imaging, and biological. These data enable path-breaking analyses into human behavior as innovative projects combine data from different sources to allow for timely analysis with unprecedented granularity. New data types present challenges of scale, sensitivity, and structure, requiring new approaches to collection, privacy, analysis, storage, and preservation. A consensus exists in the scientific community on the urgent need for new modes of access, confidentiality protection, methodological approaches, and tools so that research using a variety of data types meets accepted scientific and ethical standards.

Frontier social science, relying on new categories of data, needs convergent standards and interoperable research infrastructure for producing, managing, and analyzing research data that includes non-designed data and "big"data. All stages of the data lifecycle need infrastructural support. RDE is building infrastructure that spans across all stages of the research lifecycle (ensuring research data are FAIR), with standards and methods making the tools for using the infrastructure interoperable. RDE, encompassing the full data life cycle, will make scientific analyses using that data more rigorous, transparent, and reproducible.

We will share how the RDE infrastructure is being built to enable social and data scientists across disciplines to conduct their work more efficiently and to create, organize, archive, access, and analyze data in ways that they cannot with existing infrastructure through the specific components outlined below:

Research Data Description Framework: a flexible system of metadata standards
Research Document Registry: a registry for digital documents including pre-registered research designs and hypotheses, data management plans, participant consent statements, and data use agreements
Tubocurator: software for harmonizing data and generating appropriate metadata to assure FAIRness.Turbocurator facilitates the curation and sharing of high quality, discoverable, and re-usable data by reducing the cost of preparing data and metadata. It will help harmonize data across studies, make it easy for researchers to attach

appropriate metadata, maintain provenance, prepare data for re-use and re-discovery, and check for confidentiality issues.
Explore Data: interactive tools to preview, explore, and discover data
Video Data Tools: tools for facilitating data discovery and integration of video data
Geospatial Data Tools: tools for facilitating data discovery and integration, including confidential and geospatial data
Researcher Passport: credentialing system for researcher access to confidential data
COBRE: cloud-based platforms for analyzing confidential or large, complex, social science data

This session will be relevant to any person or organization involved in modernizing their cyberinfrastructure across the research data lifecycle to ensure FAIR data.

The speakers will be Dr. Maggie Levenstein and Aalap Doshi.

Levenstein is Director of ICPSR, Professor in the School of Information, Research Professor, Institute for Social Research, at the University of Michigan. She is the Principal Investigator of the NSF infrastructure project, RDE, and the NIH's Social, Behavioral, and Economic COVID-19 Consortium Coordinating Center. She is Co-Director of the Michigan Federal Statistical Research Data Center. She serves on the boards of the Social Science Research Council; World Data System; the Data Documentation Initiative (DDI); National Internet Observatory; Data Archiving and Access Requirements Working Group (DAARWG) of the NOAA Science Advisory Board; Criminal Justice Administrative Records System (CJARS); and the Wealth and Mobility Study, Stone Center for Inequality Dynamics. She received her PhD in economics from Yale University and BA in economics from Barnard College, Columbia University. She is the author of Accounting for Growth: Information Systems and the Creation of the Large Corporation and a fellow of the American Association for the Advancement of Science. Her research examines the production, dissemination, and confidentiality protection of novel, non-designed data for social and economic measurement.

Aalap Doshi is the Director of Technology for ICPSR at the Institute for Social Research and Lecturer in the School of Information at the University of Michigan where he teaches "Navigating Ambiguity in User Experience." He holds a Master of Science in Information (Human-Computer Interaction) from the University of Michigan, and has over two decades of experience at the intersection of design, technology, and strategy. Previously, he helped establish and scale human-centered design and innovation at Michigan Medicine, where he led the design of UMHealthResearch, an award-winning health research recruitment platform. He is also the co-founder of Findcare, a nonprofit connecting low-income communities to affordable healthcare.

Primary authors: Mr DOSHI, Aalap (ICPSR University of Michigan); Dr LEVENSTEIN, Margaret (ICPSR University of Michigan)

Presenters: Mr DOSHI, Aalap (ICPSR University of Michigan); Dr LEVENSTEIN, Margaret (ICPSR University of Michigan)

Session Classification: Presentations Session 10: Infrastructures to Support Data-Intensive Research - Local to Global

Track Classification: SciDataCon2025 Specific Themes: Infrastructures to Support Data-Intensive Research - Local to Global