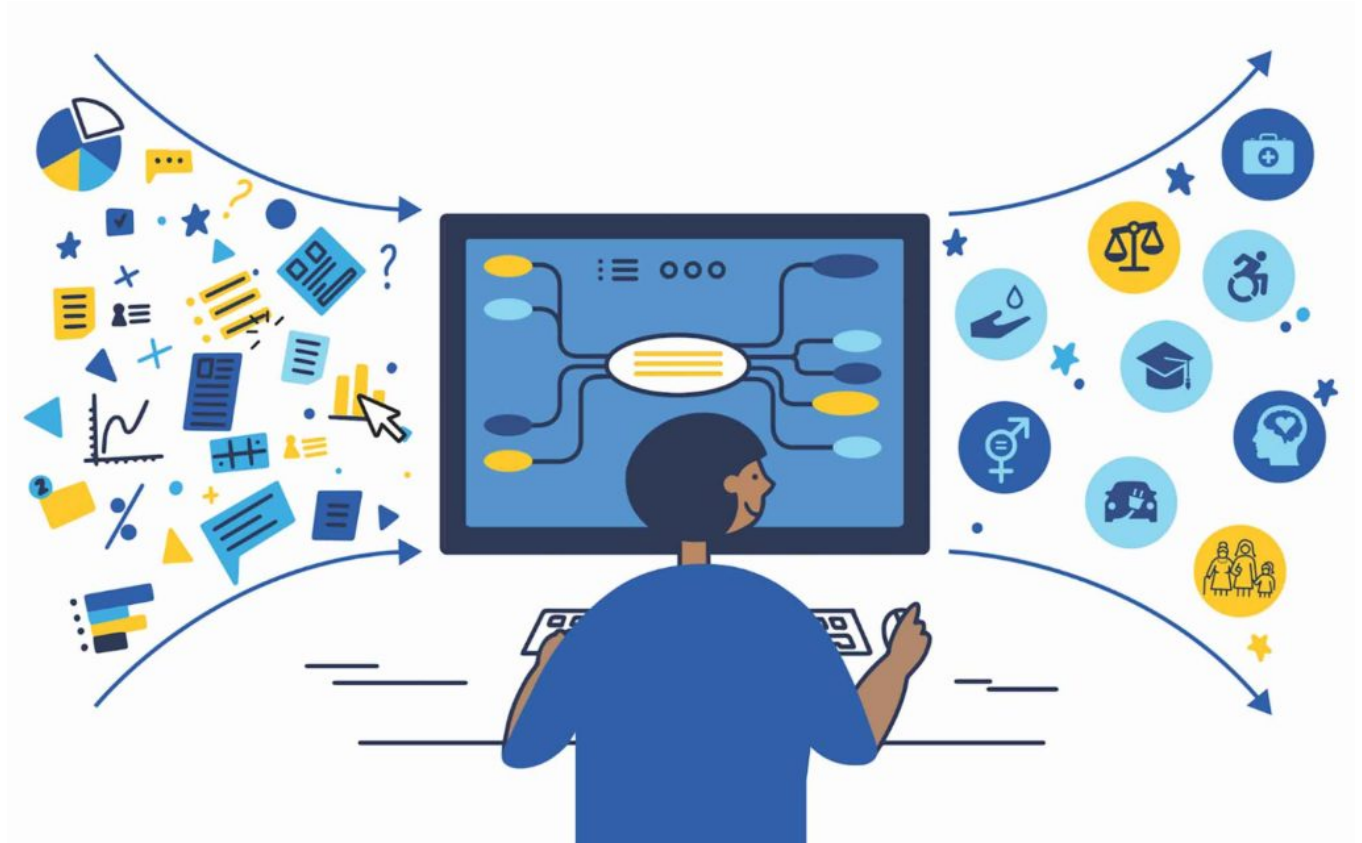


# Inside Grants: National Science Foundation Research Data Ecosystems (RDE)

October 31, 2022



Research on human behavior can help improve a wealth of societal issues, such as predicting and preventing infectious-disease outbreaks, building resilience to disasters, and improving health outcomes. When data sets are incomplete, contain errors, or are missing information, researchers face obstacles to conducting research and providing insights. To combat this, the National Science Foundation has invested \$38 million to establish the [Research Data Ecosystem: A National Resource for Reproducible, Robust, and Transparent Social Sciences in the 21<sup>st</sup> Century](#). Leading the creation of the data archives and software is [The University of Michigan Institute for Social Research](#) (UMich ISR). This project will modernize data management and collection to improve scientific research.

APS's Kekoa Erber spoke with the UMich ISR team for this special look at the Research Data Ecosystems project.

## Research Data Ecosystem (RDE) Infrastructure Project

- **Country:** United States
- **Organization:** National Science Foundation
- **Grant mechanism:** Mid-scale Research Infrastructure-2

- **Amount:** \$38,357,018

### **Tell us about the origins of this project. What spurred this new initiative?**

The Institute for Social Research (ISR) at the University of Michigan initiated the Research Data Ecosystem (RDE) infrastructure project because ISR recognized the need to provide better support throughout the research lifecycle for researchers using novel data to engage in cutting-edge social science. The Inter-university Consortium for Political and Social Research (ICPSR), the world's largest social-science archive specializing in curated data, began constructing digital archives for social-science data in the 1960s to preserve and disseminate the novel data that ISR researchers were creating. At that time, each data set was created with its own bespoke framework, permissions, metadata, etc. Advances in our ability to collect data have led to a massive influx of different data types that, theoretically, can be linked to inform research within the social sciences, thus requiring a modernized software platform.

The RDE is a transformative infrastructure project that will modernize the ICPSR software platform and develop an integrated suite of software tools to advance research in the social and behavioral sciences with a focus on the democratization of data. The RDE will enable:

1. **Interoperability:** an integrated system for the entire research data lifecycle, so that work done early in the data lifecycle is useful at later stages, making it possible to integrate data from different sources;
2. **Reproducibility:** making it easier to reproduce and build on prior research results by being able to find and re-use data and code;
3. **Transparency:** providing information about provenance, including source, code, method of collection, etc. for research data;
4. **Increased efficiency of data sharing:** reducing the burden on data producers in sharing data and ensuring that shared data are FAIR (Findable, Accessible, Interoperable, Reusable); and
5. **Confidentiality protection:** protecting confidentiality while increasing research access.

To achieve these goals, the project will also develop the Research Data Description Framework—a metadata specification similar to the Resource Description Framework—for describing different research data lifecycle events. The RDE will include standalone functional components for each stage of the research lifecycle that will be interoperable with one another and with key existing research infrastructure. The platform will support social and behavioral science researchers using traditional (e.g., survey and experimental) and novel (e.g., digital tracing, imaging) types of data over the entire research lifecycle, from data collection to analysis to sharing to rediscovery and reanalysis.

This infrastructure will improve the quality, integrity, and safety of data while increasing accessibility to data and collaboration between users across social-science and behavioral-science disciplines, and it will do so with a user interface designed to make data more accessible across the board.

**When people think of research infrastructure, opportunities in behavioral and social sciences might not come to mind. What is the importance of research infrastructure to our fields? And how can our fields continue to contribute to this area?**

The RDE will modernize the management of data to enable a new era of interconnected research for the social and behavioral sciences and beyond. The platform will improve the quality of data-driven social and behavioral science research over the entire data lifecycle. The RDE will enable researchers across disciplines to conduct their work more efficiently and to create, organize, archive, access, and analyze data in ways that they cannot with the existing infrastructure.

Imagine you would like to study a particular ZIP code that is known to have specific adverse health conditions and that you are able to come to ICPSR and safely and securely identify all sorts of studies with data related to this ZIP code (EEG data, survey data, video data, geospatial data, criminal-justice data, educational data, etc.) and conduct research in a way that was never before possible.

The RDE, once built and in conjunction with the work being done at ICPSR to curate data, will enable you to do just that.

The fields of behavioral and social sciences can accelerate the ability to develop new knowledge by considering the data lifecycle at the very beginning of a research project—with the creation of machine actionable research plans, data-management plans, preregistrations, human-subject approvals, consent statements, and data-use agreements that can be used to manage and share data during and after a research project. Too often, the consideration of what will happen to the data set (where it will be stored, how it can continue to be accessed, what are the right metadata to apply to it so it can be connected with other data, etc.) comes as an afterthought or once the paper is published. Often data ends up “in the attic” and unusable, which is an enormous waste of valuable data resources.

Psychological scientists can contribute to the development of the RDE infrastructure by collaborating with us by volunteering as a user of research data, participating in interviews and testing, providing feedback, and using the tools in their own research and organization.

**Can you speak to the importance of this effort for the field of psychological science?**

The RDE will make it easier for psychological scientists to find, access, explore, and connect data in ways they never have before that will lead to new discoveries. This will allow, for example, aggregation of small experimental data collections so that psychological scientists can draw inferences for populations. We will make it possible to easily navigate social-science archives to add to and enhance data sets. We experience so many barriers to working with existing data, especially data that haven't been curated and appropriately tagged with metadata. We are working to provide a delightful, safe, and smooth pathway to access, explore, and analyze social-science data, both restricted and unrestricted, in the cloud. The RDE has the potential to revolutionize how scientists work with social-science data to create new knowledge.

[\*See all Inside Grants columns\*](#)

## **What was the grant process like? What advice would you have for researchers considering applying for NSF infrastructure funding?**

Be patient, persistent, and responsive. Applying for an infrastructure project is a big lift, especially for academic software builders. Be clear about why your infrastructure will enable scientists to do research in ways they have never done before. Identify project-management expertise and invest time in building a high-performing team to work on this project. NSF infrastructure people are collaborative and exacting, and we are confident that you, like us, will find your organization benefits from working with them.

## **How can APS members get involved in the activities of the RDE?**

APS members can sign up for updates about the RDE [here](#). If you have any questions about this particular infrastructure grant, please contact Jeannette Jackson, the RDE managing director, at [jackjean@umich.edu](mailto:jackjean@umich.edu). The code for this project, once developed, will be open source.

*Feedback on this article? Email [apsobserver@psychologicalscience.org](mailto:apsobserver@psychologicalscience.org) or login to comment.  
Interested in writing for us? Read our [contributor guidelines](#).*