

Appendix A: Data Sharing Plan

The Census Bureau desires that data collected as part of this project be made available to the researchers and the public at no (or minimal) cost. Access to the data is an important aspect of transparency at all stages of the panel, which will foster trust in the legitimacy of the panel and the results it generates. The Longitudinal Internet Studies for the Social sciences (LISS) Panel, run by CentER Data in the Netherlands, is a good example of this approach; their probability panel is similar to the Ask U.S. Panel, and all background and topical survey datasets are available for download by registered users (<https://www.lissdata.nl/>).

At the same time, we must weigh the benefits of data release against the risks to the privacy of panel members. Linking together data from the baseline survey with data from several topical surveys may enable reidentification of members, especially when panel data are merged with other data sources. Census researchers are certainly aware of this risk.^{67,68} For this reason, we must carefully consider what data can be released publicly, and we have a responsibility to inform panel members about what data will be released, when, how, and to whom. During the Year 1 design phase, we will work with Census to create policies about data release, and we will review and update these policies throughout the life of the panel.

A.1 Data Collected

While building and maintaining the Ask U.S. Panel, we will collect names, addresses, email addresses and phone numbers from panel participants as well as demographic characteristics, socioeconomic status, opinions, attitudes, and behaviors. The information captured in the topical surveys will depend on the questions posed by investigators. The panel will also generate paradata, such as the number and timing of survey invitations sent, which surveys each case responds to, incentives received, and response times in completed surveys.

A.2 Data Storage and Access

During the project's period of performance, data will be stored on secure RTI servers. Systems are available and monitored 24 hours per day, 7 days per week. Significant investments have been made to ensure that the systems maintain an extremely high level of availability and reliability in support of RTI's campus and global networks. This includes, but is not limited to, redundant power systems, VPN and redundant networking technologies, load-balancing devices, clustering, and multiple data centers. RTI maintains two data centers on its Research Triangle Park (NC) campus in support of administrative and project-dedicated systems. Both data centers are protected by uninterruptible power supplies capable of sustaining data center operations until emergency backup generators come online.

To ensure compliance with all applicable information security laws, statutes, and agency directives, RTI has voluntarily and enthusiastically implemented the IT security guidelines and principles published by the National Institute of Standards and Technology (NIST). In accordance with the Federal Information Processing Standards (FIPS) 199 Standards for Security Categorization of Federal Information and Information Systems Assessment (FIPS 199 Assessment), RTI has implemented continuous monitoring capabilities to ensure that all security controls are regularly monitored and reported on. These monitoring capabilities include but are not limited to regular vulnerability scanning, automated network monitoring, intrusion detection and prevention measures, and periodic control auditing.

We will store all personally identifiable information (PII), such as panelists' names, addresses and email addresses, in a separate location and database that only project team members with a need-to-know can access.

A.3 Data Documentation

The data we collect and release from the Ask U.S. panel will be well documented, so researchers understand how the questions were asked and how to interpret the results. There are several standards available for documenting the data, metadata, survey questions, and response options used in each survey. We believe the Data Documentation Initiative (DDI) standard is best suited for our use. The DDI standard is used by the Inter-university Consortium for Political and Social Research (ICPSR) and by the LISS panel, among others. The Bureau of Labor Statistics uses the DDI to document and track changes in the Consumer Expenditure Interview Survey questionnaire.^{69,70} From our work with the Energy Information Administration on the Residential Energy Consumption Survey, we know that they are planning to use the DDI to document that survey as well.

There are several tools that assist in implementing the DDI standard. The Bureau of Labor Statistics uses the Colectica tool, which integrates well with Blaise, simplifying the documentation process. RTI has used the Colectica tool on previous Blaise projects, such as a recent project for the Centers for Disease Control and Prevention, Reducing Disparities in Vaccination Coverage.

A.4 Data Release

In making the data available, we must balance the benefits of data access with our responsibilities to protect the privacy of our panel members, as discussed above. We must also balance the benefits of data availability with investigators' interest in analyzing responses before others can access them. We will work with Census to develop standards for timely data release that may include an embargo period of approximately 1 year after data are collected before they are released to the public.

Performing data linkage may require access to PII by the topical survey investigator, which should be disclosed in the linkage consent question. During Year 1 and in close collaboration with the Census, we will carefully think through whether and under what conditions we could provide access to PII to facilitate data linkage.

RTI is experienced with the production of public use data files and documentation aimed at advancing research for the common good. Examples include:

- Yearly datasets and codebooks for the National Survey on Drug Use and Health (NSDUH) for the Substance Abuse and Mental Health Services Administration. See, for example, the 2017 NSDUH Final Analytic File Codebook.⁷¹
- The High School Longitudinal Study of 2009 (HSL:09) Base-Year to Second Follow-Up⁷² for the National Center for Education Statistics.
- As part of a recent Robert Wood Johnson Foundation grant, RTI created publicly available datasets by linking clinical trial data with comparable cancer survivors in the Medical Expenditure Panel Survey. This effort produced seven publicly available datasets.⁷³
- In 2016, RTI collected and released data from the 2016 Law Enforcement Management and Administrative Statistics Body-Worn Camera Supplement. The data are available at the ICPSR repository.