



OMB Control Number: 0925-0648
Expiration Date: May 31, 2021

Public reporting burden for this collection of information is estimated to average 15 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a current valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0648). Do not return the completed form to this address.

* 1. In which of the following NIH Institutes and Centers do you work? Please select only one.

- CC (NIH Clinical Center)
- CIT (Center for Information Technology)
- CSR (Center for Scientific Review)
- FIC (Fogarty International Center)
- NCATS (National Center for Advancing Translational Sciences)
- NCCIH (National Center for Complementary and Integrative Health)
- NCI (National Cancer Institute)
- NEI (National Eye Institute)
- NHGRI (National Human Genome Research Institute)
- NHLBI (National Heart, Lung, and Blood Institute)
- NIA (National Institute on Aging)
- NIAAA (National Institute on Alcohol Abuse and Alcoholism)
- NIAID (National Institute of Allergy and Infectious Diseases)
- NIAMS (National Institute of Arthritis and Musculoskeletal and Skin Diseases)
- NIBIB (National Institute of Biomedical Imaging and Bioengineering)
- NICHD (Eunice Kennedy Shriver National Institute of Child Health and Human Development)
- NIDA (National Institute on Drug Abuse)
- NIDCD (National Institute on Deafness and Other Communication Disorders)
- NIDCR (National Institute of Dental and Craniofacial Research)
- NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases)
- NIEHS (National Institute of Environmental Health Sciences)
- NIGMS (National Institute of General Medical Sciences)
- NIMH (National Institute of Mental Health)
- NIMHD (National Institute on Minority Health and Health Disparities)
- NINDS (National Institute of Neurological Disorders and Stroke)
- NINR (National Institute of Nursing Research)
- NLM (National Library of Medicine)
- OD (Office of the Director)

* 2. Which of your degrees and/or certifications are applicable to your role? (e.g. PhD, RN)

* 3. What role(s) do you play in data science research? Please select all that apply.

- Clinical
- Research
- Policy
- Other (please specify)
- Quality
- Technical

* 4. In which category(ies) do you conduct research? Please select all that apply.

- Clinical
- Laboratory
- Other (please specify)
- Population



* 5. In which category(ies) do you conduct research? Please select all that apply.

| | Very Low | Low | Medium | High | Very High | N/A |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Capture and create metadata (descriptive information about your data, how it was collected, and other contextualizing information) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Use common data elements, ontologies (formal models of concepts within a domain and their relationships), or other predefined terms for describing your data or variables | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Organize, tag, and track data so multiple team members can work on the same dataset | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

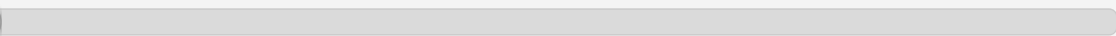
| | Very Low | Low | Medium | High | Very High | N/A |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Conduct research through data mining (using computational methods to discover patterns in large datasets) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Locate and obtain other researchers' shared data to use in your research, and clean or process it to meet your research needs | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Demonstrate, analyze, or communicate your research results through data visualization | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Create a plan for long-term storage and retention of your data | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Publish and deposit data in a repository suited to your research field | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Write a formal data management plan, including selecting file formats, choosing a standard for data description, and planning for storage and preservation | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

* 6. Please indicate which of the following sources of data are currently or potentially relevant to your research:

| | No interest | Potentially helpful to my research | Essential to current research | Essential to future research-currently don't have acces |
|---|-----------------------|------------------------------------|-------------------------------|---|
| electronic health records (EHRs) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| electronic Patient Reported Outcomes (ePRO) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| patient registries | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| clinical trial summary (aggregate) data | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| clinical trial individual patient data | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| insurance medical claims | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

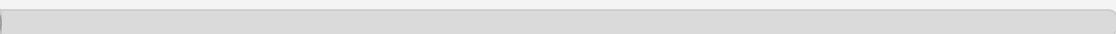
* 7. How valuable for your research would it be to have access to a resource that integrated data from insurance billing claims, electronic health records (EHRs), and patient registries? (0=not useful, 5=essential)

0 (not useful) 5 (essential)



* 8. How important is it to you for the data sources you use to be open and have limited restrictions of use?

0 (not useful) 5 (essential)



* 9. Please provide a rough estimate of the percentage of time – expressed as a whole number between 0 and 100 – you spend per week preparing data? (Preparing data is defined as a process of normalizing and features filtering in order to prepare the data for analysis; Source: *A data analysis framework for biomedical big data: Application on mesoderm differentiation of human pluripotent stem cells*: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5487013/>)

* 10. Of these, which types of data do you use? Please select all that apply.

- | | |
|--|-------------------------------------|
| <input type="checkbox"/> Structured | <input type="checkbox"/> Real-World |
| <input type="checkbox"/> Unstructured | <input type="checkbox"/> Public |
| <input type="checkbox"/> Simulated/Synthetic | <input type="checkbox"/> Private |

* 11. What methods do you use for analyzing your data? Please select all that apply.

- | | |
|---|--|
| <input type="checkbox"/> Machine Learning | <input type="checkbox"/> Text Analytics |
| <input type="checkbox"/> Linear Regression | <input type="checkbox"/> Bayesian Methods |
| <input type="checkbox"/> Nonlinear Regression | <input type="checkbox"/> Simulation |
| <input type="checkbox"/> Classification | <input type="checkbox"/> Prescriptive Analysis |
| <input type="checkbox"/> Data Mining | <input type="checkbox"/> Natural Language Processing |
| <input type="checkbox"/> Other (please specify) | |

* 12. How do you define an acceptable data quality level?

* 13. When working with data, what barriers do you face? Please select all that apply.

- | | |
|---|--|
| <input type="checkbox"/> Inconsistent Data | <input type="checkbox"/> Lack of domain expert input |
| <input type="checkbox"/> Incomplete Data | <input type="checkbox"/> Multiple ad-hoc environments |
| <input type="checkbox"/> Duplicative Data | <input type="checkbox"/> Limitations of tools |
| <input type="checkbox"/> Lack of strong data analytics skills | <input type="checkbox"/> Coordination with IT |
| <input type="checkbox"/> Lack of management support for use of data and tools | <input type="checkbox"/> Difficult to explain data science to others |
| <input type="checkbox"/> Lack of financial support for use of data and tools | <input type="checkbox"/> Issues with privacy |
| <input type="checkbox"/> Data unavailable or difficult to access | <input type="checkbox"/> Unclear expectations for project impact |
| <input type="checkbox"/> Results not used by decision makers | <input type="checkbox"/> Integrating findings into decisions |

* 14. What is the primary objective for use of data in your research?

* 15. Please provide 1-2 examples of a typical research question for you or your team that would require collection and/or analysis of data? (e.g. "What microbiome characteristics of children exposed to antibiotics and different diets influence long-term health outcomes?" or "What are the genetic, epigenetic, and environmental factors of endometriosis, and how can they help us identify its subtypes?")

* 16. In which areas, if any, do you need help with access to bioinformatics resources for your research program? Please select all that apply.

| | None at all | A little | A lot |
|--|-----------------------|-----------------------|-----------------------|
| Experimental design/planning (replicates, randomization) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Data analysis | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Data management | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Statistics | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Software/Software planning | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Storage | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Security | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Computers/Equipment | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Training | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Standards | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sharing/Collaborating | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Licensing | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

* 17. Select the data resources you currently use and are interested in using. Please select all that apply.

| | Currently use | Interested in using |
|---|-----------------------|-----------------------|
| Agency Core Support Services, e.g. NIH CREx (Collaborative Research Exchange) | <input type="radio"/> | <input type="radio"/> |
| NIH Library | <input type="radio"/> | <input type="radio"/> |
| Common Data Elements (CDEs) – use, recommend, contribute to a repository | <input type="radio"/> | <input type="radio"/> |
| Data Cleansing, e.g. Drake, OpenRefine | <input type="radio"/> | <input type="radio"/> |
| Data Dictionary | <input type="radio"/> | <input type="radio"/> |
| Data Capture tools, e.g. REDCap | <input type="radio"/> | <input type="radio"/> |
| Data Extraction tools, e.g. Octoparse, Talend, Informatica | <input type="radio"/> | <input type="radio"/> |
| Data Repositories, e.g. NIH Data Sharing Repositories | <input type="radio"/> | <input type="radio"/> |
| Data Visualization tools, e.g. Tableau, Datawrapper, Spotfire | <input type="radio"/> | <input type="radio"/> |
| Formal Vocabularies, Terminologies, Ontologies, or Coding Systems | <input type="radio"/> | <input type="radio"/> |
| Open Source Data tools, e.g. RapidMiner, Hadoop | <input type="radio"/> | <input type="radio"/> |
| Sentiment tools, e.g. Opentext, Semantria | <input type="radio"/> | <input type="radio"/> |
| Statistical Programming/Analysis tools, e.g. R, SAS, MiniTab | <input type="radio"/> | <input type="radio"/> |

* 18. What types of training or resources did you take in order to use those tools?

19. Please share any final thoughts regarding the current and future needs for your research not captured through the survey. We value your insight.