

NIH ODSS STRIDES Initiative Baseline Survey

Survey Introduction

Thank you for participating in the National Institutes of Health (NIH) Office of Data Strategy Science (ODSS) Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative baseline survey. The survey will take maximum 20 minutes and is designed to help NIH better understand the culture of cloud usage and broader challenges among the scientific community. Additionally, the survey will help NIH identify, segment and prioritize potential STRIDES participants. Your responses will be used to assess the programs' impact on NIH's business processes and how research is performed. Please identify only facilitators and barriers that have had a significant impact on your scientific research using big data. Thank you for your time!

OMB#: 0925-0648 Exp. Date: 05/31/2021

Burden Disclosure: Public reporting burden for this collection of information is estimated to average 20 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 currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection 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. Please enter your email address.

Researcher Characteristics

The information collected within this section will help NIH identify, segment and prioritize current and potential STRIDES participants and their unique needs. For example, Question #6 asks if respondent works with sensitive data. We ask this because we want to be sure that we are connecting them with the resources they need to architect their environment properly.

2. Please provide your academic institution, if applicable.

3. Please select your NIH Institute and Center (IC), if applicable.

- NCI
 - NEI
 - NHLBI
 - NHGRI
 - NIA
 - NIAAA
 - NIAID
 - NIAMS
 - NIBIB
 - NICHD
 - NIDCD
 - NIDCR
 - NIDDK
 - NIDA
 - NIEHS
 - NIGMS
 - NIMH
 - NIMHD
 - NINDS
 - NINR
 - NLM
 - CC
 - CIT
 - CSR
 - FIC
 - NCATS
 - NCCIH
 - Not applicable
-

4. What is your current role at your NIH Institute and Center (IC) or the big data/cloud-based project you support?

- Senior Investigator
- Investigator
- Senior Clinician
- Senior Scientist
- Assistant Clinical Investigator
- Staff Clinician
- Staff Scientist
- Senior Research Assistant/Research Assistant
- Scientist Emeritus
- Postdoctoral Fellow
- Pre-Doctoral Fellow
- Post-Baccalaureate Fellow
- Technical Fellow
- ORISE Senior Fellow
- Senior Clinical Fellow
- Clinical Fellow
- Senior Research Fellow
- Research Fellow
- Other - Write In (Required)

5. How would you best describe your research? (e.g., clinical research)

6. Do you work with sensitive data? If so, what kinds?

- I don't work with sensitive data
- Human subject data
- Non-human subject data

7. Select the data categories that you generate, use, process or analyze in your research. Check all that apply.

- Genomic
- Proteomic
- Metabolomic
- Micro/Macro Imaging
- Clinical Observations
- Clinical Chemistry
- Clinical Reports
- Social
- Other - Write In (Required)

*

8. How large is the total footprint of your data?

- Between 0 and 1 TB
- Between 1 TB and 500 TB
- Greater than 500 TB
- I don't know the total footprint of my data

9. How rapidly does your data grow?

- Less than double per month
- Greater than or equal to double per month
- I don't know how rapidly my data grows

Experience with Cloud Tools or Capabilities

LOGIC Show/hide trigger exists.

10. Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply.

- Yes; I use commercial cloud tools or capabilities to access data sets or other digital objects
- Yes; I use commercial cloud tools or capabilities to store data sets or other digital objects
- Yes; I use commercial cloud tools or capabilities to manipulate, visualize, or evaluate data sets or other digital objects
- Yes; I use commercial cloud tools or capabilities to share data sets or other digital objects
- No; I don't currently use commercial cloud tools or capabilities when conducting scientific research, but am potentially interested in using them to better fulfill my research needs
- No; I don't currently use commercial cloud tools or capabilities when conducting scientific research and do not believe they are required to fulfill my research needs

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("Yes; I use commercial cloud tools or capabilities to access data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to store data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to manipulate, visualize, or evaluate data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to share data sets or other digital objects")

11. What percentage of your data is currently stored in the cloud?

- 0%-1%
- 2%-25%
- 26%-50%
- 51%-100%

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("Yes; I use commercial cloud tools or capabilities to access data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to store data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to manipulate, visualize, or evaluate data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to share data sets or other digital objects")

12. How would you rate your ability to conduct scientific research using cloud tools and capabilities?

- Beginner
 - Novice
 - Intermediate
 - Advanced
 - Expert
-

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("Yes; I use commercial cloud tools or capabilities to access data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to store data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to manipulate, visualize, or evaluate data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to share data sets or other digital objects")

13. Please select the commercial cloud platform(s), if any, you have used while conducting scientific research. Check all that apply.

- Google Cloud Platform
- IBM
- Amazon Web Services
- Microsoft Azure
- Other - Write In (Required)

*

Experience with Cloud Tools or Capabilities

PIPING Piped Values From Question 13. (Please select the commercial cloud platform(s), if any, you have used while conducting scientific research. Check all that apply.)

14. Please select the vendor, if any, associated with the cloud service provider (CSP) from which you have obtained cloud services from while conducting scientific research. Check all that apply.

Carahsoft

Onix

REAN

Four Points

Figshare

OmniBond

SevenBridges

Not applicable

Enter another option

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("Yes; I use commercial cloud tools or capabilities to access data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to store data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to manipulate, visualize, or evaluate data sets or other digital objects", "Yes; I use commercial cloud tools or capabilities to share data sets or other digital objects", "No; I don't currently use commercial cloud tools or capabilities when conducting scientific research, but am potentially interested in using them to better fulfill my research needs")

15. Select one or more cloud services that have helped, or could help, accelerate your research. Check all that apply.

- Bulk data transfer technologies
- Storage
- Compute (i.e. virtual machines and container management techniques)
- Data delivery
- SQL/no SQL technologies
- Big data technologies
- Analytics
- Artificial intelligence (e.g., machine learning)
- Domain-specific applications or tools
- Internet of things (IoT)
- Other - Write In (Required)

*

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("No; I don't currently use commercial cloud tools or capabilities when conducting scientific research, but am potentially interested in using them to better fulfill my research needs", "No; I don't currently use commercial cloud tools or capabilities when conducting scientific research and do not believe they are required to fulfill my research needs")

16. What has prevented you from utilizing commercial cloud services or tools while conducting scientific research? Check all that apply.

- Concerns related to data privacy or security
- Lack of skills or training
- Lack of funding or resources
- General hesitancy
- Cloud tools/capabilities are not user friendly
- Cloud tools/capabilities do not support research methods performed (e.g., deep learning, big data analytics)
- Research requires computational performance currently only available through high performance computing (HPC)
- Research uses well-defined computational approaches on a scale that does not justify migrating to the cloud to exploit its dynamic, elastic and scalable capability
- Available publicly-funded cloud services, such as National Institute of Standards and Technology (NIST), obviate the need for any commercial cloud services
- Traditional computational support locally available at low/no cost
- Ingress/egress cost waivers (e.g., those available with Internet 2 membership)
- Ability to leverage the same computational hardware investments for more than one research project
- Cloud is not part of your longer-term data management strategy
- Collaboration goals for your program/project does not require sharing data broadly (e.g., team is geographically dispersed)
- Do not anticipate variability (i.e., spikes and valleys) in terms of computational and storage needs over the course of your program/project
- No one on your team who can support the administration, provisioning, and management of data in the cloud, along with serving as the accountable party to monitor usage and ensure proper information security controls are in place
- Other - Write In (Required)

*

LOGIC Hidden unless: #10 Question "Do you typically utilize commercial cloud tools or capabilities while conducting scientific research? (e.g., AWS, GCP, Azure or other cloud-based software solutions) Check all that apply." is one of the following answers ("No; I don't currently use commercial cloud tools or capabilities when conducting scientific research, but am potentially interested in using them to better fulfill my research needs", "No; I don't currently use commercial cloud tools or capabilities when conducting scientific research

and do not believe they are required to fulfill my research needs")

17. What would you suggest NIH do to support or encourage the utilization of cloud tools and capabilities when conducting scientific research? Check all that apply.

- Offer training
- Offer cost incentives
- Minimize paperwork required to acquire cloud resources
- Promote culture of using cloud
- Provide easier access or change data access/sharing policies
- Provide funding
- Provide education on how cloud could benefit or provide long-term value to your research
- Demonstrate how to make tools "cloud-ready" (i.e., configured or configurable) to work effectively in the cloud
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Show/hide trigger exists.

18. Please select when, if at all, you encounter barriers or challenges. Check all that apply.

- Using commercial cloud service(s) or tool(s)
- Accessing data sets
- Storing data sets
- Manipulating, aligning, visualizing or evaluating data sets
- Sharing data sets
- I do not encounter barriers or challenges related to the response options provided

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Using commercial cloud service(s) or tool(s)")

19. What significant barriers have you encountered when using commercial cloud services or tools? Check all that apply.

- Governance or control
- Cloud vendor reliability or performance
- Research lead may feel apprehensive using certain cloud tools or capabilities
- Lack of knowledge or expertise
- Cloud tools or capabilities do not meet research needs
- Cloud vendor pricing structure
- Lack of desired data sets available on the cloud
- Lack of funding or resources
- Data migration such as extensive troubleshooting or slow data migration
- Data security
- Agreement issues or legal considerations
- Onboarding or account setup
- Controlled Access
- Timely routes to access data concerning a single scientific focus area
- Lack of ability to perform analysis across multiple data sets
- No capability that easily enables collaboration
- Type of data is not readily available via cloud resources
- Ability to manage cloud spend
- Personal preference
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Using commercial cloud service(s) or tool(s)")

PIPING Piped From Question 19. (What significant barriers have you encountered when using commercial cloud services or tools? Check all that apply.)

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Using commercial cloud service(s) or

tool(s)")

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Using commercial cloud service(s) or tool(s)")

PIPING Piped From Question 19. (What significant barriers have you encountered when using commercial cloud services or tools? Check all that apply.)

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Accessing data sets")

20. What significant barriers have you encountered when accessing data sets? Check all that apply.

- Data type necessary for meeting scientific research objectives is not available because the cultural or social norms within community does not promote sharing that type of data
- Research lead may feel apprehensive using certain cloud tools or capabilities
- Lack of understanding or implementation of FAIR criteria
- Willingness to give others data set ownership
- Unreliable data or corrupt data files
- Complex, or lack of, data access resources
- Lengthy data access approval process
- Complex data access policies
- System errors
- File size and bandwidth
- Metadata is not sufficiently detailed enough to understand the provenance of the data
- Data is not readily available via cloud resources
- Lack of knowledge or experience accessing large data sets using NIH publicly available software tools/APIs/etc.
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Accessing data sets")

PIPING Piped From Question 20. (What significant barriers have you encountered when accessing data sets? Check all that apply.)

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Accessing data sets")

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Accessing data sets")

PIPING Piped From Question 20. (What significant barriers have you encountered when accessing data sets? Check all that apply.)

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Storing data sets")

21. What significant barriers have you encountered when storing data sets? Check all that apply.

- Research lead may feel apprehensive using certain cloud tools or capabilities
- Lack of infrastructure that can support data at scale
- Lack of funding for cloud resources
- Concerns regarding data security
- Moving data to the cloud isn't feasible
- Lack of experience or knowledge
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Storing data sets")

PIPING Piped From Question 21. (What significant barriers have you encountered when storing data sets? Check all that apply.)

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Storing data sets")

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Storing data sets")

PIPING Piped From Question 21. (What significant barriers have you encountered when storing data sets? Check all that apply.)

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Manipulating, aligning, visualizing or evaluating data sets")

22. What significant barriers have you encountered when manipulating, aligning, visualizing or evaluating? Check all that apply.

- Research lead may feel apprehensive using certain cloud tools or capabilities
- Lack of infrastructure that can support data at scale
- Lack of available resources or guidance
- Unclean or unformatted data
- Lengthy processing time
- Lack of usability (the adequacy of metadata associated with data)
- Publicly available software libraries or data analysis tools do not meet scientific research needs
- Insufficient skill in research or data science methods necessary to meet scientific research objectives
- Insufficient skill in software libraries or data analysis tools
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Manipulating, aligning, visualizing or evaluating data sets")

PIPING Piped From Question 22. (What significant barriers have you encountered when manipulating, aligning, visualizing or evaluating? Check all that apply.)

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Manipulating, aligning, visualizing or evaluating data sets")

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Manipulating, aligning, visualizing or evaluating data sets")

PIPING Piped From Question 22. (What significant barriers have you encountered when manipulating, aligning, visualizing or evaluating? Check all that apply.)

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Sharing data sets")

23. What significant barriers have you encountered when sharing data sets? Check all that apply.

- Type of data you want to share does not align with the cultural or social norms within your research community, which does not promote sharing that type of data
- Research lead may feel apprehensive using certain cloud tools or capabilities
- Lack of data sharing resources or guidance
- Lengthy IRB review/approval process
- Lengthy data sharing process
- Meeting NIH data quality standards (e.g., format, metadata)
- Privacy of research participants (e.g., data includes identifiers that would permit linkages to individual research participants and variables that could lead to deductive disclosure of the identity of individual subjects)
- Lack of understanding FAIR Guideline application
- Lack of understanding data sharing policies or process
- Other - Write In (Required)

*

Barriers or Challenges

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Sharing data sets")

PIPING Piped From Question 23. (What significant barriers have you encountered when sharing data sets? Check all that apply.)

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Sharing data sets")

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or

challenges. Check all that apply." is one of the following answers ("Sharing data sets")

PIPING Piped From Question 23. (What significant barriers have you encountered when sharing data sets? Check all that apply.)

Additional Input on Barriers, Facilitators, or Solutions

LOGIC Hidden unless: #18 Question "Please select when, if at all, you encounter barriers or challenges. Check all that apply." is one of the following answers ("Using commercial cloud service(s) or tool(s)", "Accessing data sets", "Storing data sets", "Manipulating, aligning, visualizing or evaluating data sets", "Sharing data sets")

24. How can STRIDES help you address the barrier(s) you identified and described earlier? *

25. Are there significant facilitators (e.g., tools, people, cloud capabilities, programs) that you rely on when accessing, storing, using, and sharing data for scientific research?

26. Would you be interested in participating in a future focus group or Townhall to further discuss your responses?

Thank You!

Thank you for taking our survey. Your response is very important to us.
