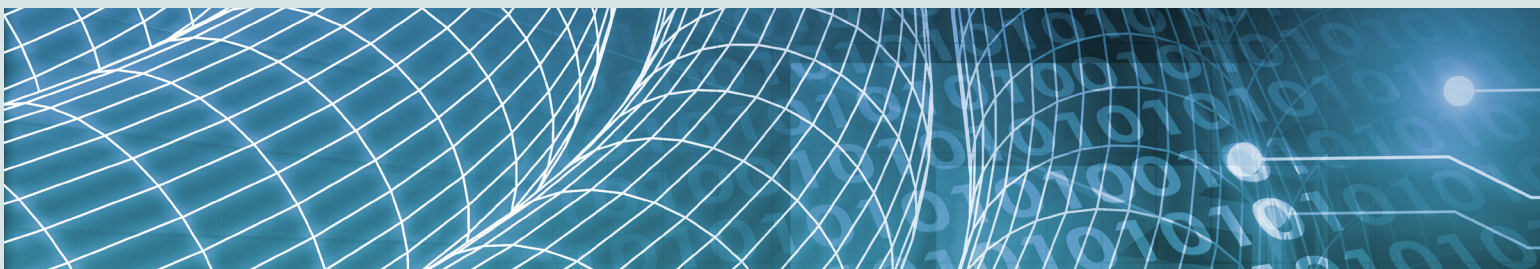
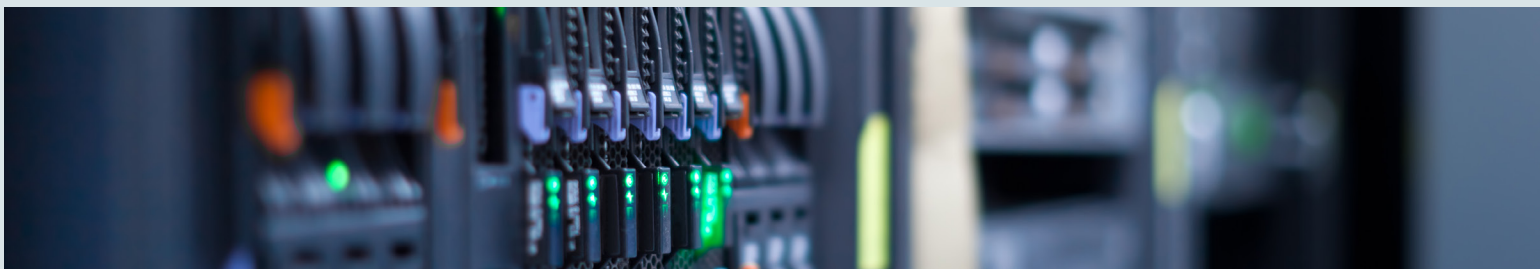




Technical Report: **Cyberinfrastructure for Sustained Scientific Innovation**

ANALYSIS OF COMMUNITY RESPONSE TO NSF CSSI PROGRAM AND
SOLICITATION NSF 19-548

JUNE 2020



Report Authorship

Ross Brindle, Nexight Group

Fred Hansen, Nexight Group

Robert Foy, Nexight Group

Rachel Lanspa, Nexight Group

Changwon Suh, Nexight Group

Author for correspondence:

Changwon Suh, email: csuh@nexightgroup.com

8405 Colesville Rd

4th Floor

Silver Spring, MD 20910

Please cite as:

Brindle, R., F. Hansen, R. Foy, R. Lanspa, C. Suh. *Technical Report: Cyberinfrastructure for Sustained Scientific Innovation: Analysis of Community Response to NSF CSSI Program and Solicitation NSF 19-548*, Nexight Group, June 2020.

Disclaimer

Any opinions expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF) or any of the organizations that are partner to the Cyberinfrastructure for Sustained Scientific Innovation (CSSI) solicitation.

Acknowledgements

The authors would like to thank the NSF staff who led this effort and provided inputs on the survey design and development, including Micah Beck, Vipin Chaudhary, Amy Friedlander, Alexis Lewis, Manish Parashar, Stefan Robila, Amy Walton, and Joseph Whitmeyer. Most particularly, the authors thank the hundreds of Principal Investigators (PIs), co-PIs, and other stakeholders who took time to complete the survey.

table of contents

Introduction.....1

Demographic Breakdown of Respondents3

SECTION I: SOLICITATION4-12

Section 1 Introduction 4

Aspects for Change..... 5

 1. *Types of Awards*..... 6

 2. *Combined Software Data/Umbrella* 7

 3. *Presentation of Domain-Specific Interests* 8

 4. *Solicitation-Specific Criteria* 9

 5. *Solicitation Logistics*.....10

 6. *Guidance Included in the Solicitation*11

 7. *Support Beyond the Written Solicitation*12

SECTION II: PROGRAM13-17

Section II Introduction13

 1. *Near-Future Assumptions*.....14

 2. *Success in Objectives*.....15

 3. *Success in Guiding Principles*16

 4. *Near-Future Suggestions*17

 5. *Near-Future Suggestions Funding*17

Appendix A: Acronym List18

Appendix B: Near-Future Suggestions Open Responses..... 20

Appendix C: Select Multi-Step Analysis.....28

introduction

This technical report describes the methodology and results of a survey of Principal Investigators (PIs), co-PIs, and others in the Cyberinfrastructure (CI) research community. The survey had two primary purposes:

- To inform decisions about changes to be made to the National Science Foundation (NSF) Cyberinfrastructure for Sustained Scientific Innovation (CSSI) solicitation.
- To inform decisions about the future direction and focus of the NSF CSSI umbrella program.

This survey-based assessment was carried out by Nexight Group under the Office of Advanced Cyberinfrastructure (OAC) Award (#1930025). The grant was intended to benefit the scientific research community by providing insights that enhance CSSI's support of CI for scientific research.

OAC and the CSSI Program

The OAC supports and coordinates the development, acquisition, and provision of state-of-the-art CI resources, tools, and services essential to S&E research. Advances in CI help to drive scientific discovery and innovation by enabling new research capabilities and new pathways for experimentation and theory.

The NSF OAC CSSI umbrella program ([NSF 19-548](#)) seeks to enable funding opportunities that are flexible and responsive to evolving and emerging needs in CI. The goal of the CSSI program is to create a CI ecosystem that spans all levels of the data and software stack and scales from individual or small groups of innovators to large community resources. The CSSI program targets services that address all aspects of CI, from embedded sensor systems and instruments, to desktops and high-end data and computing systems, to major instruments and facilities.

Importance of Input from the CI Research Community

In developing investment priorities, the CSSI program considers the needs of the science & engineering community. However, the program also recognizes that the advancement of CI research is best served by engaging not only the capabilities of the CI researchers, but their creativity and energy as well. In other words, research priorities should also be informed by PIs' passions and interests. In addition, feedback from the CI research community helps NSF ensure that PIs are able to write effective proposals and have the support to successfully carry out their projects.

Accordingly, ongoing feedback from and dialogue with PIs from the CI research community is critical. Surveys are an efficient and effective mechanism by which NSF directorates stay connected to the community and maintain a dialogue. Recognizing that surveys place a burden on the time and effort of PIs, OAC established a design imperative to minimize the time and effort required to complete this survey. The average survey completion time was less than ten minutes.

Survey Development

Nexight Group, a small business technical and management consultancy in Silver Spring, MD, worked with NSF to establish a framework for survey design that began with an understanding of the decisions to be informed by the survey. As noted in the previous page, the survey was designed to gather input that would inform decisions about the CSSI solicitation and decisions about the program's direction and focus. Nexight worked with OAC decision makers to specify the structure of the decisions, the inputs to the decisions, and the impact of the decisions. Our three phases of survey design are:

- Determine the nature of the decisions to be made
- Determine the information needed to inform the decisions (who needs to say what about what?)
- Write survey items

The structure for the decisions describes how the decisions are made. The decisions follow a two-stage decision-making process. The **first stage** is focused solely on determining whether change is needed – it represents a simple yes/no decision point. If the determination is **no**, then the decision-making process ends. If the determination is **yes**, then the decision-making process moves to the second stage. The **second stage** is focused on determining how to change. Since the options for change were limited by organizational and practical constraints, the second stage reduced to a choice among options. This decision-making structure is illustrated on the right.

The input to the decisions describes the specific information (e.g., judgments, observations, preferences) that Pls can provide to inform the decisions. In order to keep the survey short, only questions that directly tied to the decisions were included in the survey. Questions that were of interest to the decision makers but not directly tied to a decision were excluded.

The impact of the decisions describes why the decisions need to be made and what difference they make. Nexight developed a logic model that described how the CSSI program has impact and achieves its goals. This helped to ensure that the decisions informed by the survey were meaningful and significant. The logic model also provided a conceptual context for framing the survey questions.

Nexight Group developed and finalized both the survey and a survey landing page with information on the CSSI umbrella program, the survey project and outcomes, and a link to complete the survey. The survey and landing page were created using QuestionPro survey software (<https://www.questionpro.com>).

The survey was launched on December 2, 2019 and closed on January 10, 2020.

Survey Respondents

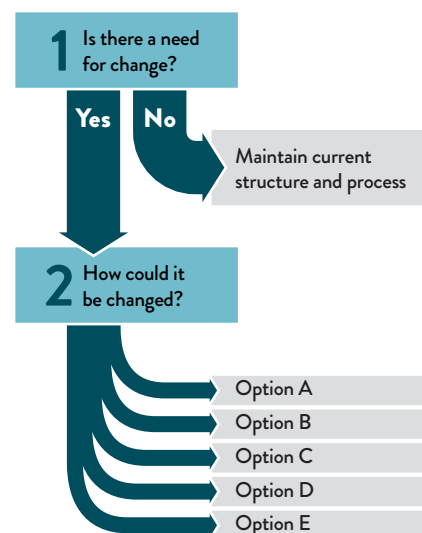
A total of 251 people provided responses to the survey, 31 of whom supplied answers for solely Section 1: Solicitation. The demographics indicated that the respondents were representative of the target population and comparable to the respondents of other NSF surveys.

97% interact with NSF as Pls or co-Pls. 55% had served as a proposal reviewer in the past year. 98% had previously submitted a proposal to NSF, 57% of whom received NSF funding. 90% listed universities as their primary organization. Survey responses were received from 42 U.S. states.

A more detailed demographic breakdown of respondents is shown on the following page, and in Appendix C (p. 28).

Nature of Decision

Decision makers used a two-stage decision-making process.



Demographic Breakdown of Respondents

1 Survey respondents primarily interact with NSF as Principal Investigators or co-Principal Investigators.

96.8% Principal Investigator (PI) / co-Principal Investigator (co-PI)

- 1.1% Sponsored Projects Office (SPO) staff
- 1.1% Authorized Organization Representative (AOR)
- 0.5% Research Administrator
- 0.5% Other

2 Half of survey respondents have served as reviewers in the past year.

55% Have served as a reviewer in the past year, reviewing an average of 10.4 proposals

- 45% Have not served as a reviewer in the past year.

3 The majority of respondents describe themselves as PIs and professors (some respondents selected multiple categories).

49.4% Principal Investigator

- 42.2% Professor
- 19.1% Faculty/Educator
- 12.4% Staff Scientist/Researcher
- 4.8% Institutional Administrator
- 2.8% Scientific Administrator
- 2.4% Laboratory Manager
- 0.8% Laboratory Technician
- 0.4% Post-Doctoral Fellow
- 2.8% Other

4 Nearly all respondents have previously submitted a proposal to NSF, 73% of whom received funding.

98% Have submitted

- 2% Have not submitted

Of those who submitted:

- 57% Funded
- 18% Under review
- 5% Rejected
- 19% No response

5 The majority of respondents are most related to the Directorate for Computer and Information Science and Engineering (CISE), followed by the Directorate for Mathematical and Physical Sciences (MPS), then the Directorate for Engineering (ENG), and the Directorate for Geosciences (GEO).

34% CISE Computer and Information Science and Engineering

- 20% MPS Mathematical and Physical Sciences
- 10% ENG Engineering
- 8% GEO Geosciences
- 3% BIO Biological Sciences
- 1% SBE Social, Behavioral, and Economic Sciences
- 1% EHR Education and Human Resources
- 2% Other

6 90% of respondents list universities as their primary organization, followed by non-profit research organizations at 5%.

90.6% University

- 5.2% Non-profit Research Organization
- 3.1% Federally Funded Research and Development Center
- 0.5% Associate's College
- 0.5% No Organizational Affiliation / Independent Researcher

7 Large organizations (1001 or more) form the majority of respondents.

74.7% 1001 or more

- 13.2% 501 to 1000
- 6.8% 101 to 500
- 3.2% 25 or fewer
- 2.1% 26 to 100

8 Most respondents work in research groups, as individuals or as part of research projects.

86% Work in a research group

- 14% Do not work in a research group

Of those who work in a research group:

- 36% Research project
- 34% Individual researcher
- 19% Laboratory
- 10% Other
- 1% Did not indicate

For details and select multi-step analysis, refer to Appendix C (p. 28)

section I: solicitation

THE SURVEY RESULTS ARE PRESENTED IN TWO SECTIONS.

Section I (pages 4-12), described below, presents the survey results intended to inform changes to the CSSI solicitation.

Section II (pages 13-17), described on page 13, presents the survey results intended to inform decisions regarding the future direction and focus of CSSI program investments.

Section I

The survey questions focused on the solicitation were organized around the two-part decision-making process – the **Need for Change** was assessed separately from **How to Change**. Accordingly, the results for Section I are organized around this two-part decision-making process, with the Need for Change results presented first.

CSSI decision makers identified seven aspects of the solicitation for which meaningful change was both allowed and feasible (see the next page of the report for the list of aspects). Only these aspects were included in the survey. Survey respondents were asked to identify three of the seven areas where improvement is most needed and could make the most difference.

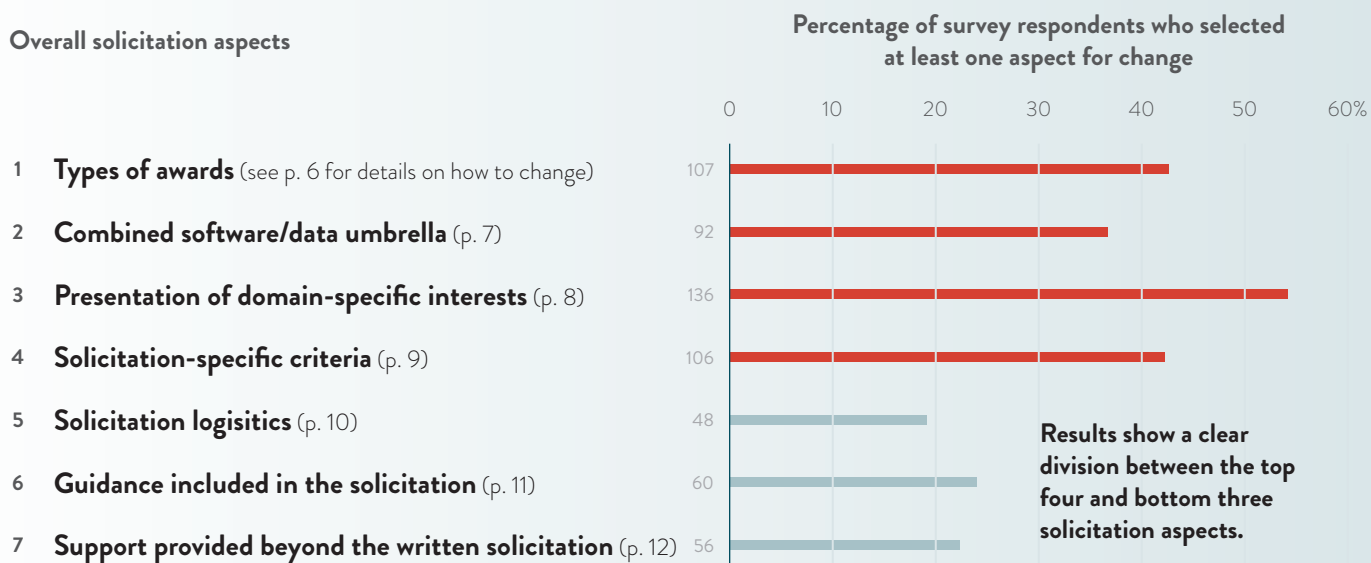
The Need for Change results are displayed in a bar chart that indicates the percentage of respondents who chose that aspect (note: the percentages do not add to 100% because people could select more than one option). The Key Takeaway section summarizes the main results displayed in the bar chart and indicates the conclusions that may be drawn from those results.

The How to Change results are presented second. For each aspect they checked as needing change, respondents were asked to indicate the How to Change options that they believed should be made. The How to Change results for each aspect are presented on a separate page. Again, the results are displayed in a bar chart followed by a Key Takeaway section summarizes the main results displayed in the bar chart, which indicates the conclusions that may be drawn from those results.

The How to Change results also include a demographic analysis that breaks the results down in terms of reviewers versus non reviewers. This breakdown is warranted given the fairly even split between the two subpopulations (55% reviewers and 45% non-reviewers). The comment on the demographic analyses indicates whether and how the takeaways change when the reviewers are separated from non-reviewers.

is there a need for change?

Participants were asked to **select three of seven** indicated aspects of the current solicitation where improvement is most needed or could make the most difference. **Multi-step analysis based on this question can be found in Appendix C (p 28).**



KEY TAKEAWAY **Top four aspects of the solicitation where change is needed and would make the most difference:**

Types of awards

the two classes of CSSI awards (elements and framework implementations) and the size, duration, and timing of the funding associated with each type

Combined software/data umbrella

the inclusion of both software and data perspectives under a single umbrella scope

Presentation of domain-specific interests

the number of areas, the clarity of the descriptions, and the level of specificity and amount of detail provided

Solicitation-specific criteria

the appropriateness and number of additional solicitation-specific review criteria and the clarity and detail with which they are described

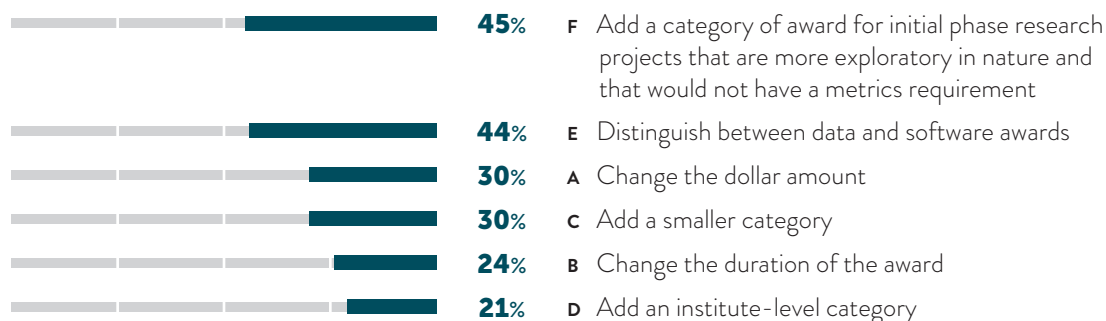
SOLICITATION: HOW TO CHANGE

1 TYPES OF AWARDS

This refers to the two classes of CSSI awards (i.e., elements versus framework implementations), and the size, duration, and timing of the funding associated with each type. [Solicitation NSF19548, pp. 6-7]

Respondents selected the changes that they thought should be made to the types of awards offered in the solicitation.

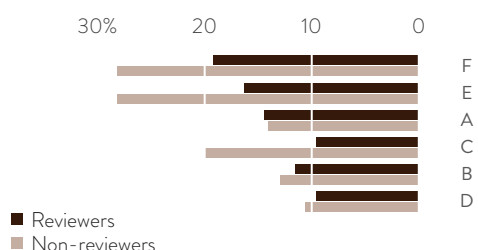
% of those who chose Aspect 1



KEY TAKEAWAY Though all options were selected by a significant portion of the population, although the following two stand out as being the most selected options:

- F **Adding an award category for exploratory research**
- E **Distinguishing between data and software awards**

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



The breakdown analysis does not change the Key Takeaway regarding the top two How to Change options. However, among non-reviewers, **1C: Add a smaller category** would be added to the list of top change options.

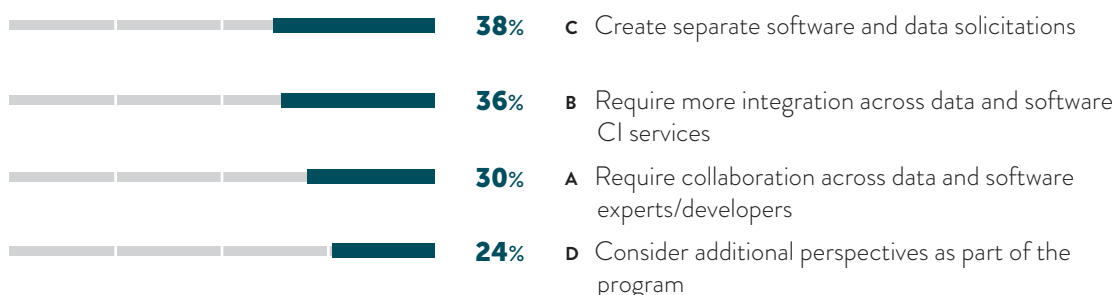
SOLICITATION: HOW TO CHANGE

2 COMBINED SOFTWARE / DATA UMBRELLA

This refers to the program's inclusion of both software and data perspectives under a single umbrella scope. [Solicitation NSF19548, pp. 1-2]

Respondents selected the changes that they thought should be made to the combined software/data umbrella.

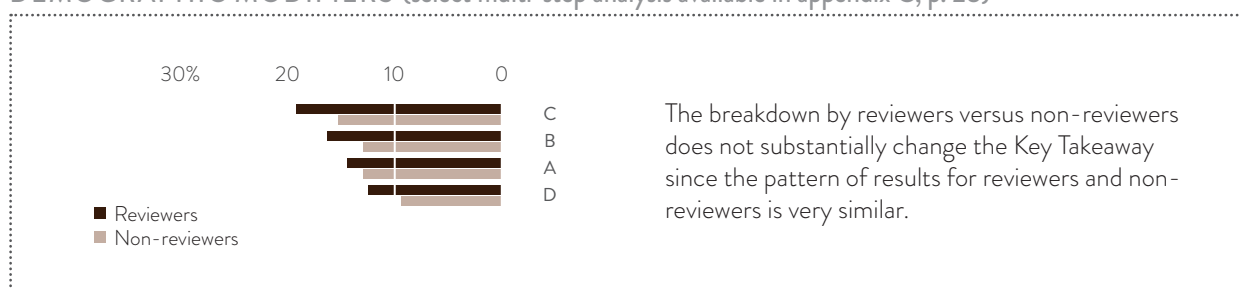
% of those who chose Aspect 2



KEY TAKEAWAY Though all options were selected by a significant portion of the population, **“Create separate software and data solicitations” was the most selected.**

Although respondents recommend separating the solicitations (echoing the key takeaway of “1: Type of Awards” to distinguish between software and data awards), they also advocate for more **integration across data software services** and more **collaboration across data and software developers.**

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



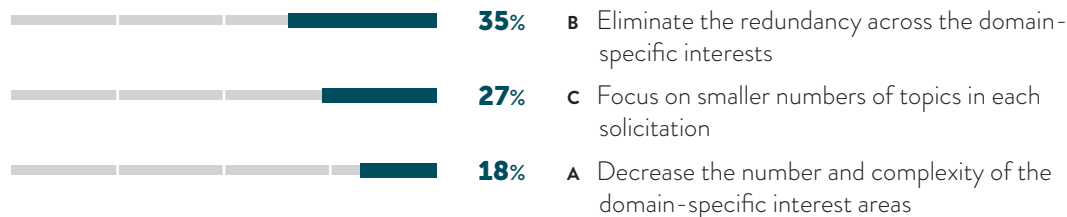
SOLICITATION: HOW TO CHANGE

3 PRESENTATION OF DOMAIN-SPECIFIC INTERESTS

This refers to the way in which the specific interests of domains (i.e., BIO, CISE, EHR, ENG, GEO, MPS, or SBE) are presented, including the clarity of the descriptions, the level of specificity and amount of detail provided, and the number of interest areas listed. [Solicitation NSF19548, pp. 7-10]

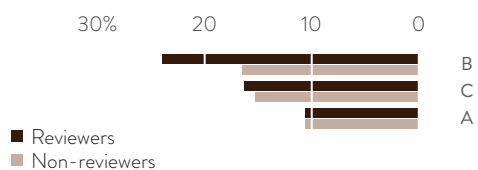
Respondents selected the changes that they thought should be made to the presentation of domain-specific interests.

% of those who chose Aspect 3



KEY TAKEAWAY All options were selected by a significant portion of the population, indicating that **respondents generally advocate for increasing clarity and simplicity in the presentation of domain-specific interests.**

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



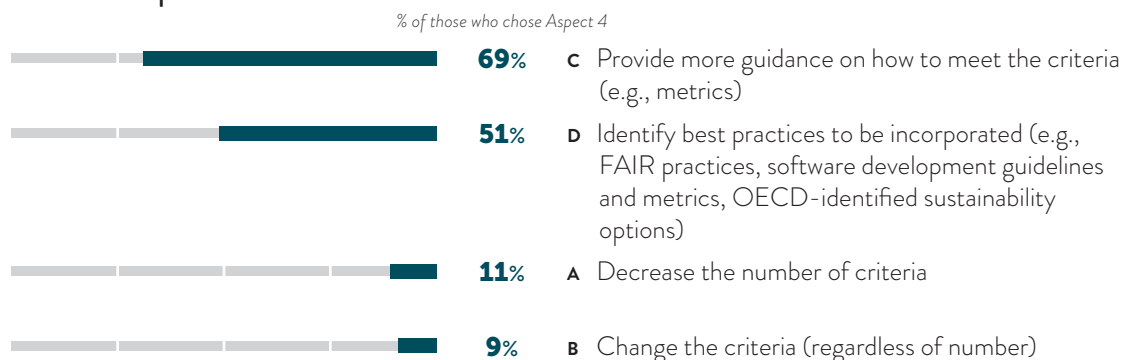
The breakdown by reviewers versus non-reviewers does not substantially change the Key Takeaway, although reviewers were more likely to recommend eliminating the redundancy across domain-specific interests.

SOLICITATION: HOW TO CHANGE

4 SOLICITATION-SPECIFIC CRITERIA

This refers to the appropriateness and number of additional solicitation-specific review criteria (e.g. science-driven, innovation, metrics, sustainable impacts), as well as the clarity and detail with which they are described. [Solicitation NSF19548, pp. 16-17]

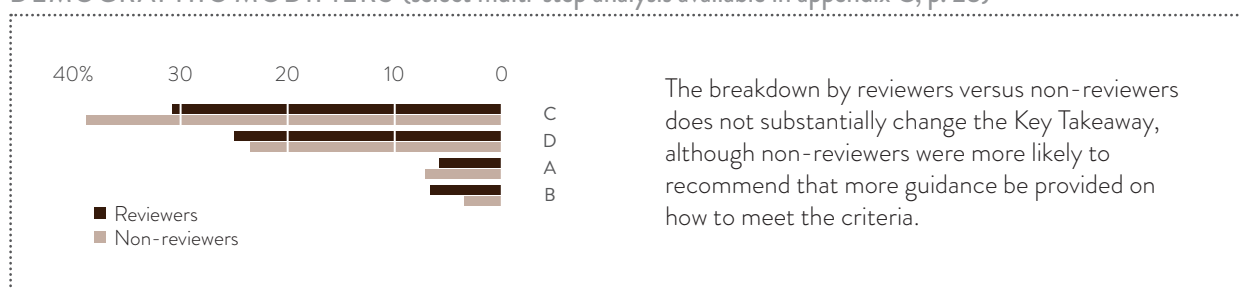
Respondents selected the changes that they thought should be made to the solicitation-specific criteria.



KEY TAKEAWAY Respondents are advocating for more help on how to meet the solicitation criteria. In particular, they think that the following two changes should be made:

- C **Provide more guidance on how to meet the criteria**
- D **Identify best practices to be incorporated into a proposal**

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)

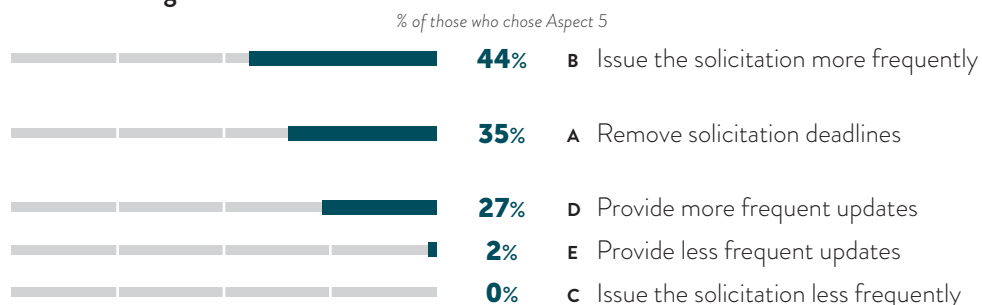


SOLICITATION: HOW TO CHANGE

5 SOLICITATION LOGISTICS

This refers to the proposal deadlines, frequency of solicitation, and the overall submission timeline. [Solicitation NSF19548, pp. 1, 14-15]

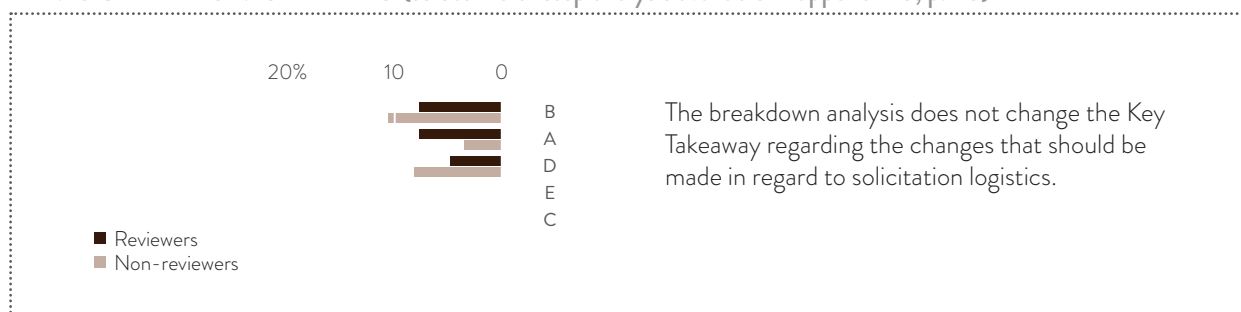
Respondents selected the changes that they thought should be made to the solicitation logistics.



KEY TAKEAWAY Solicitation logistics was not one of the top areas where a Need for Change was indicated so caution should be exercised in drawing conclusions from the results.

Nevertheless, it is clear that **respondents would prefer that the solicitation deadlines be removed** and that the **issuance of the solicitation and solicitation updates should occur more frequently.**

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



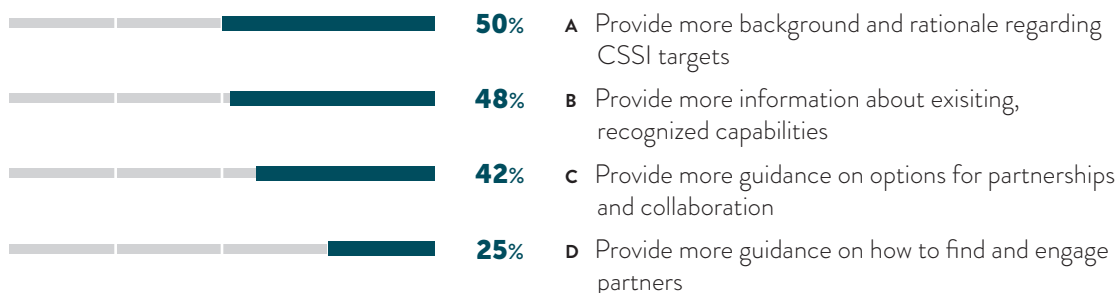
6 SOLICITATION: HOW TO CHANGE GUIDANCE INCLUDED IN THE SOLICITATION

This refers to the background information, instructions, and guidance included in the solicitation document regarding how to prepare a project and write a proposal.

[Solicitation NSF19548, pp. 11-14]

Respondents selected the changes that they thought should be made to the guidance included in the solicitation.

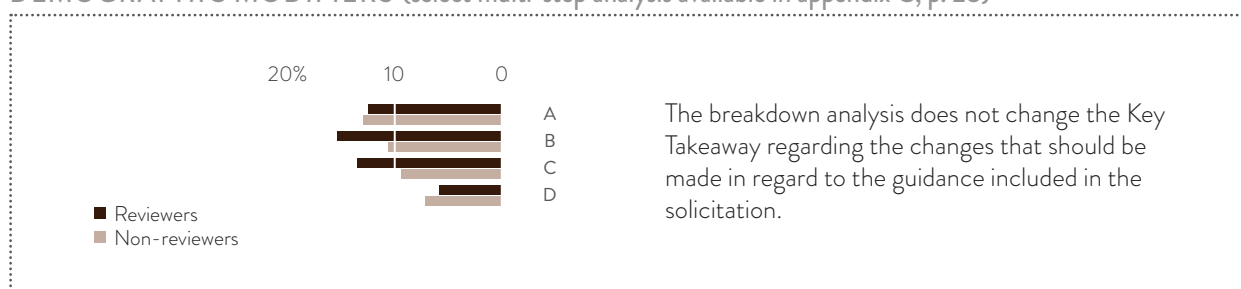
% of those who chose Aspect 6



KEY TAKEAWAY Guidance included in the solicitation was not one of the top areas where a Need for Change was indicated so caution should be exercised in drawing conclusions from the results.

Nevertheless, it is clear that **respondents would like more background information, instructions, and guidance** to be included in the solicitation document.

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



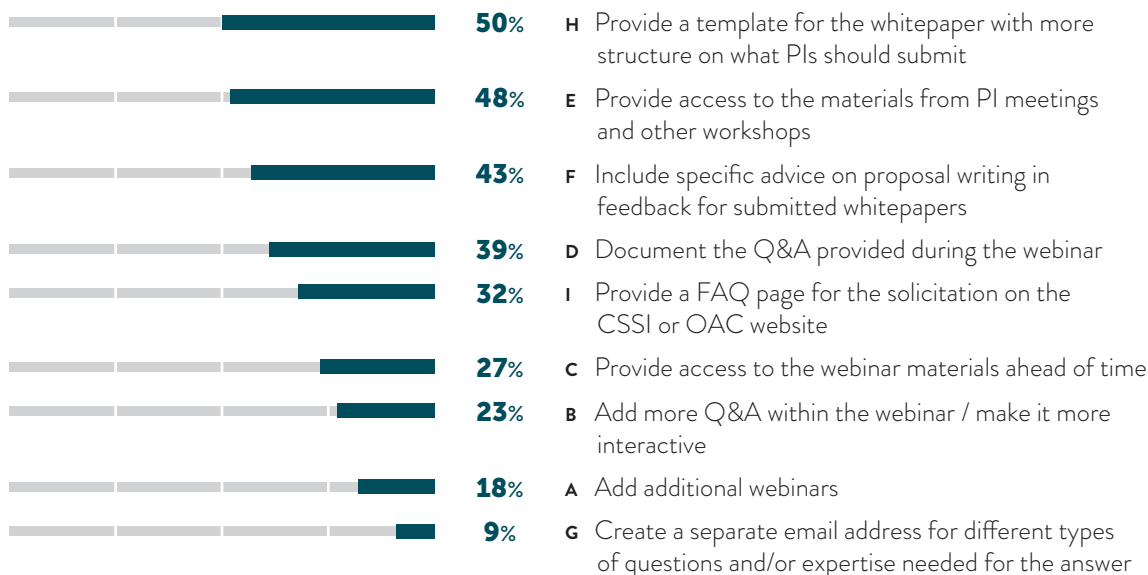
SOLICITATION: HOW TO CHANGE

7 SUPPORT BEYOND THE WRITTEN SOLICITATION

This refers to the support and guidance provided through the PI webinars, the feedback provided on submitted whitepapers, the common email for submitting questions, and the answers provided.

Respondents selected the changes that they thought should be made to the guidance included in the solicitation.

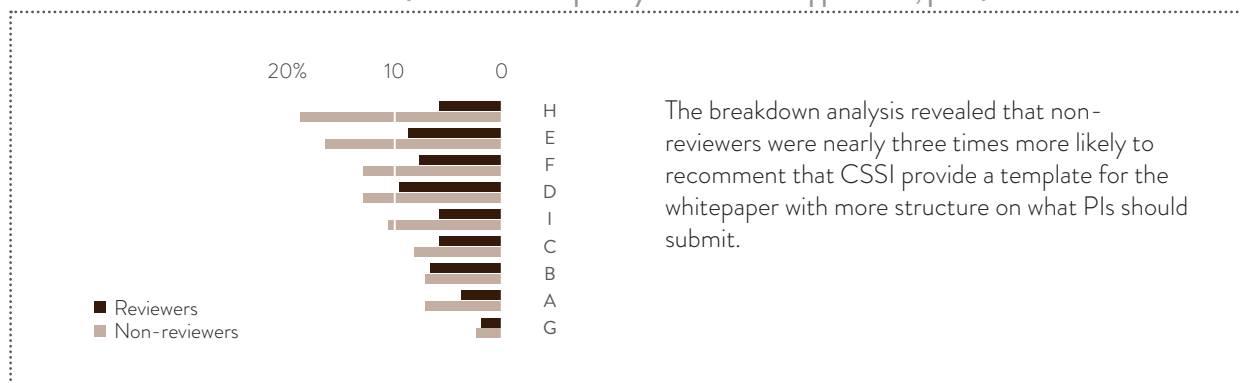
% of those who chose Aspect 7



KEY TAKEAWAY Support beyond the written solicitation was not one of the top areas where a Need for Change was indicated so caution should be exercised in drawing conclusions from the results.

Nevertheless, it is clear that **respondents would advocate for all of the How to Change options** except for creating a separate email address for different types of questions/expertise.

DEMOGRAPHIC MODIFIERS (select multi-step analysis available in appendix C, p. 28)



section II: CSSI program

Section II

Section II presents survey results intended to inform decisions about the direction and focus of the CSSI program and its investment priorities going forward. The survey questions were organized around the two-part decision-making process – the **Need for Change** was assessed separately from **How to Change**. Accordingly, the results for Section II are also organized around this two-part decision-making process, with the Need for Change results presented first.

The Need for Change was assessed from multiple perspectives. First, respondents were asked to indicate whether they thought the assumptions of the CSSI program will still apply in the near future. A need for change would be indicated to the extent that respondents believe that those assumptions are no longer valid.

Next, respondents were asked to rate the degree to which the CSSI program has been successful in achieving its objectives. A lack of effectiveness would suggest a need for change at the program level.

The final Need for Change question asked respondents to rate the degree to which the CSSI program has been successful in implementing its guiding principles. Again, a lack of effectiveness would suggest a need for change at the program level.

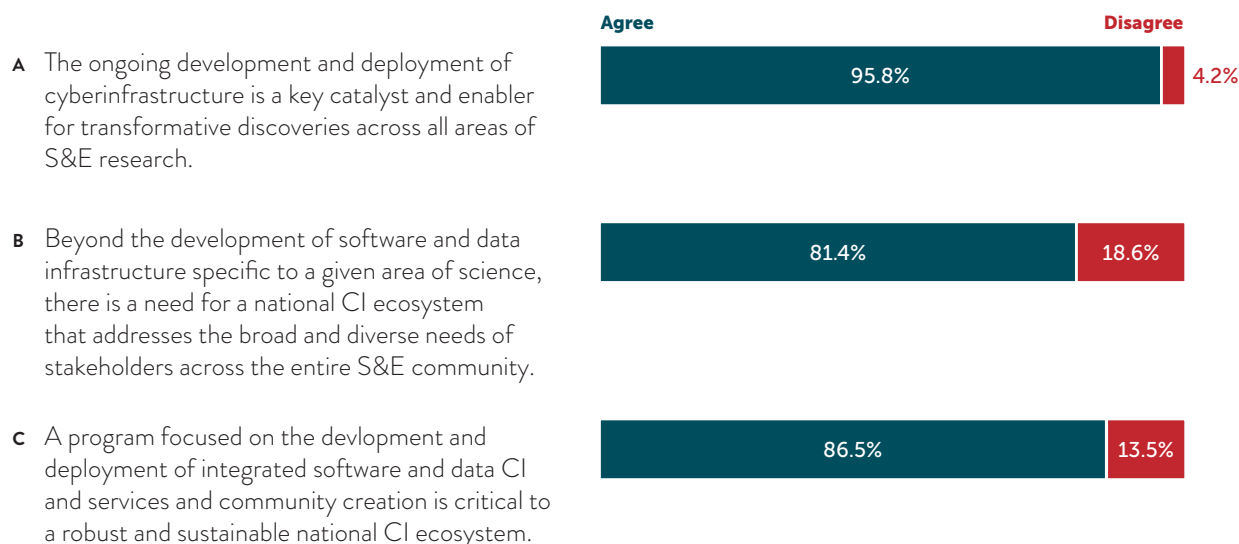
The How to Change questions focused on the allocation of funding over the next 1-5 years. Respondents were first asked to identify three investment priorities for the next 1-5 years. They were then asked to indicate the allocation of funds across those priorities.

CSSI PROGRAM: NEED FOR CHANGE

1 NEAR-FUTURE ASSUMPTIONS

NSF's CSSI program is based on certain assumptions about how infrastructure (i.e. software and data infrastructure) can support science and engineering (S&E) research and enables new thinking and transformative discoveries.

Respondents indicated the extent to which they agree with the following assumptions as they apply to the near future (1-5 years) for the CSSI program.



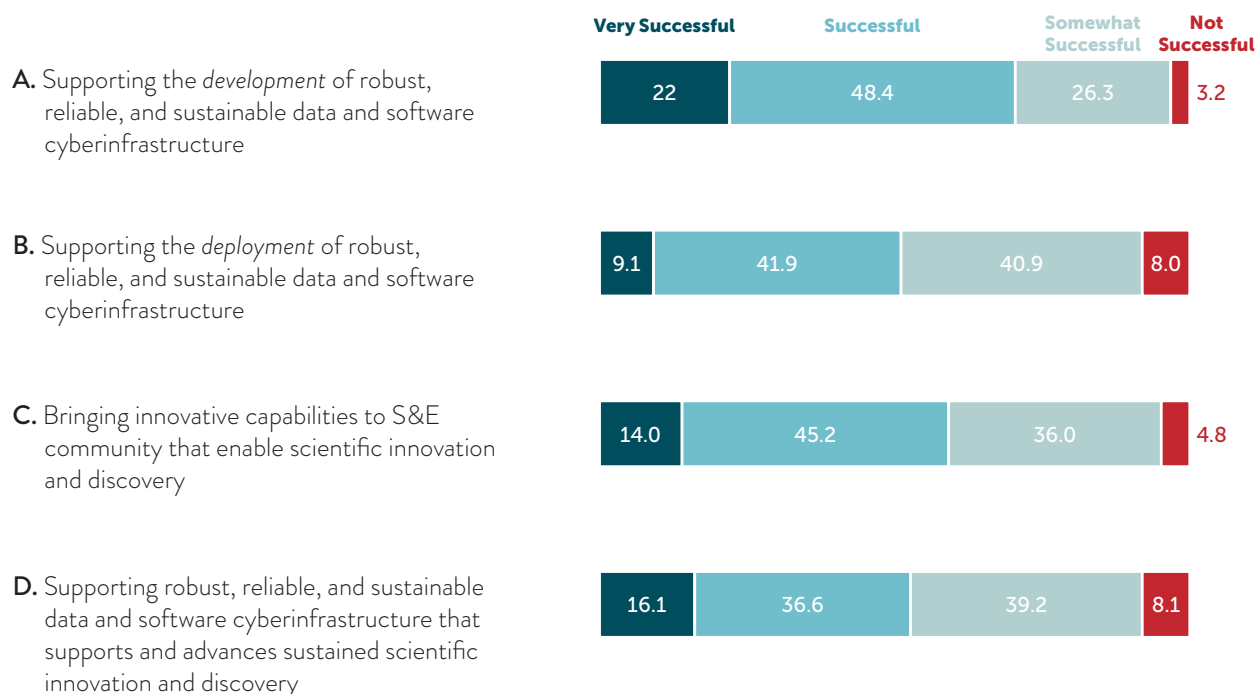
KEY TAKEAWAY There was a high level of agreement for each of the three assumptions on which the CSSI program is based. This indicates that **the rationale for the CSSI program is still valid, and that there is no need for change in regard to the overall direction and focus of the program.**

CSSI PROGRAM: NEED FOR CHANGE

2 SUCCESS IN OBJECTIVES

The CSSI program aims to support CI development and deployment to support scientific innovation and discovery.

Percentage of respondents who would classify the CSSI program's achievement of the following objectives as...



KEY TAKEAWAY The majority of respondents indicated that the CSSI program has been successful in achieving each of the four objectives. Respondents rated CSSI's success higher on supporting the development of CI relative to the deployment of CI.

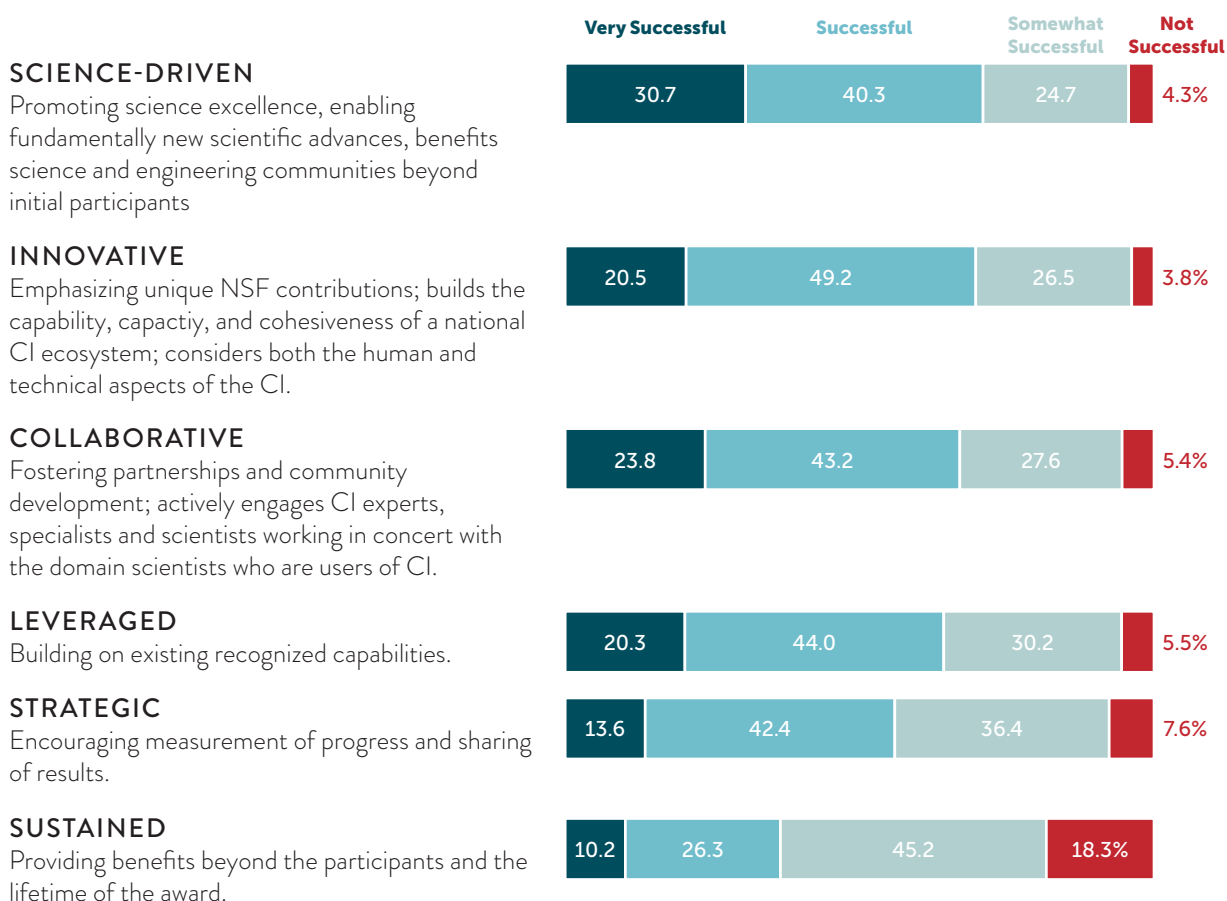
There is also a slight indication of a need for change given that a significant percentage of respondents indicated that **CSSI has only been somewhat successful**. In particular, the results suggest that **CSSI may want to adopt an increased focus on deployment and supporting scientific innovation**.

CSSI PROGRAM: NEED FOR CHANGE

3 SUCCESS IN GUIDING PRINCIPLES

The CSSI program is based on six guiding principles. A lack of effectiveness would suggest a need for change at the program level.

Percentage of respondents who would classify the CSSI program's implementation of its guiding principles as...



KEY TAKEAWAY The majority of respondents indicated that the CSSI program has been successful in implementing all of its guiding principles except for **SUSTAINED**. Less success was also indicated in regard to the principle of **STRATEGIC**. In contrast, a large majority of respondents indicated success in implementing the principles of **SCIENCE-DRIVEN**, **INNOVATIVE**, and **COLLABORATIVE**. Only a very small percentage indicated that CSSI has not been successful in implementing its guiding principles.

Again, there is some indication of a need for change, especially in regard to implementing the principles of **SUSTAINED** and **STRATEGIC**. It is worth noting that these two principles are the most difficult to implement.

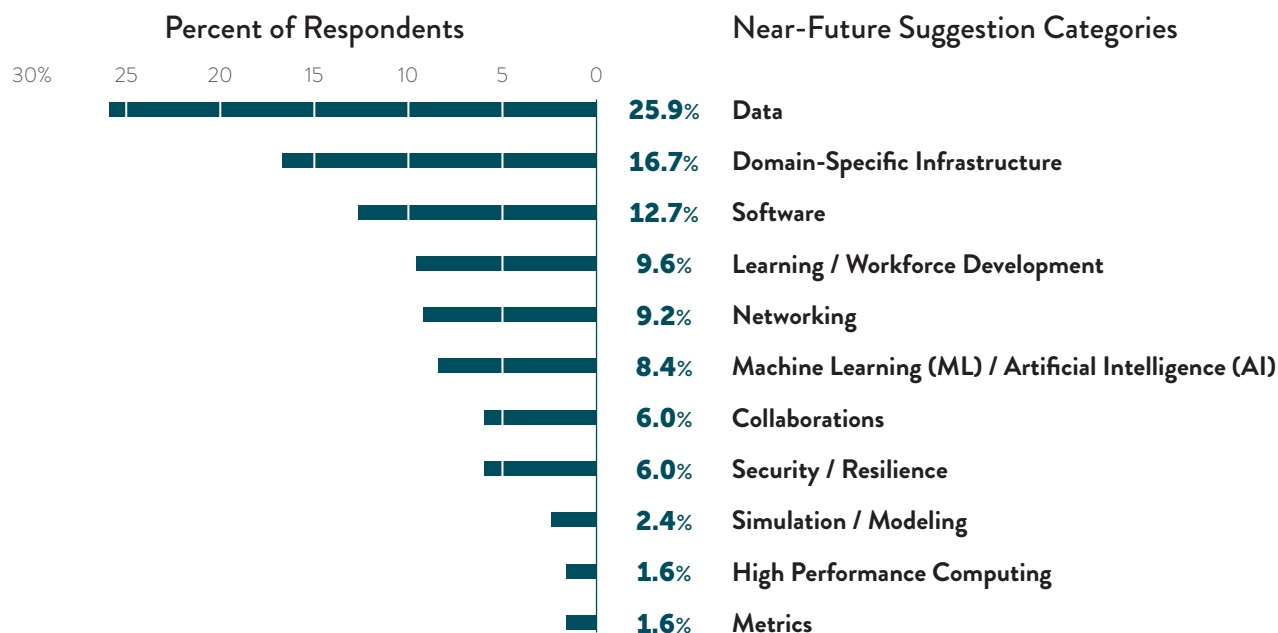
CSSI PROGRAM: HOW TO CHANGE

4 NEAR-FUTURE SUGGESTIONS

Respondents indicated up to three areas in open response where they thought CSSI funding should be directed in the next 1-5 years in order to have the most impact or best serve the S&E community.

Responses were assigned a category according to CSSI investment priorities, charted below.

For a full list of open response answers, refer to Appendix B (p. 20).



5 NEAR-FUTURE SUGGESTIONS FUNDING

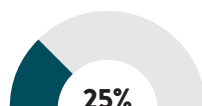
Respondents also indicated the amount of total funding (out of 100%) that NSF should allocate for each of their near-future suggestions.

Percentage of respondents who recommended allocating X amount of total funding to their near-future suggestions:

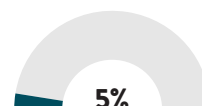
25% of total funding



50% of total funding



75% of total funding



KEY TAKEAWAY Few respondents recommended allocating large percentages of funds to their near-future suggestions. Rather, **the large majority of respondents favor broader funding, at lower allocation levels.**

list of acronyms

AI	Artificial Intelligence
AOR	Authorized Organization Representative
API	Application Programming Interface
BIO	NSF Directorate for Biological Sciences
BLAS	Basic Linear Algebra Subprograms
CI	Cyberinfrastructure
CS	Computer Science
CISE	NSF Directorate for Computer and Information Science and Engineering
CSSI	Cyberinfrastructure for Sustained Scientific Innovation
DFT	Density Functional Theory
DIBBs	Data Infrastructure Building Blocks
EHR	NSF Directorate for Education and Human Resources
ENG	NSF Directorate for Engineering
FAIR	Findable, Accessible, Interoperable, and Reusable
FFT	Fastest Fourier Transform
FFRDC	Federally Funded Research and Development Center
GB	Gigabytes
GEO	NSF Directorate for Geosciences
GIS	Geographic Information System
HBCU	Historically Black Colleges and Universities
HPC	High-Performance Computing
HTC	High-Throughput Computing
HW	Hardware
IT	Information Technology
IoT	Internet of Things
LAPACK	Linear Algebra Package
MPS	NSF Directorate for Mathematical and Physical Sciences
MSI	Minority-Serving Institution

List of Acronyms (cont.)

NIH	National Institutes of Health
NSF	National Science Foundation
OAC	Office of Advanced Cyberinfrastructure
OAU	Other Authorized User
OECD	Organization for Economic Co-operation and Development
OSS	Open Source Systems
PI	Principal Investigator
RSEs	Research Software Engineers
S&E	Science and Engineering
SBE	NSF Directorate for Social, Behavioral, and Economic Sciences
SGCI	Science Gateways Community Institute
SI2	Software Infrastructure for Sustained Innovation
SPO	Sponsored Projects Office
SW	Software
TCP-IP	Transmission Control Protocol/Internet Protocol
THz	Terahertz
USDA	United States Department of Agriculture
XSEDE	Extreme Science and Engineering Discovery Environment

near-future suggestions

Open responses to Section II, Question 4: Near-Future Suggestions, organized by categories chosen by Nexight to map with CSSI goals and investment priorities

CATEGORY

RESPONSE

Data

Adopting big data technologies

Algorithms for quantum computing

Algorithms for scalable computing

Any innovative approach to data generations, analysis, etc. as long as it will be open source

Big data processing

Big omics data

Blueprinting a secure cyber infrastructure which can easily share relevant data and best practices

Broad open source infrastructure - data analysis

Broad open source infrastructure - data visualization

Broad open source infrastructure - numeric computation

Broadly applicable storage and curation approaches to diverse data sets

Building data repositories for standardized benchmarking of physics software (e.g. Phase transitions in lattice models)

Cloud data storage

Collocated data and tools to re-use data

Creating and publicizing a resource library that contains existing CSSI projects - to help other researchers solve CI problems

Cyberinfrastructure that integrates databases and makes them more accessible/usable

Cyberinfrastructure to support completely open data for large projects (e.g. In astrophysics)

Data Access, storage and management

Data analytics

Data and AI

Data archiving

Data assimilation

Data curation

Data infrastructure hardware for institutions

Data management for the long term

Data mining

Data quality

Data repositories are being funded with high \$\$ amount, but those are just storage with a few tools developed at a very high cost, without knowing they are useful for the domain. Big projects, big funding really need better user requirements collection processes up front before an investment is made in something the computer scientists 'think' needs to be done.

CATEGORY

RESPONSE

Data (cont.)

Data repository access and development.

Data science

Data services

Data standardization: Formats and APIs

Developing CI to assure high quality data and services

Developing CI to provide quality data and services to the community

Developing infrastructure to make data available and useful on a large scale

Developing methods and tools to evaluate data and services quality

Enable data-driven initiatives

Ensuring cost effective access to massive scale digital data assets across all domains and to all members of the global research ecosystem

Frictionless data packages and data preservation in a machine-readable and interoperable way

High-quality data generation, curation, and dissemination

Infrastructure and tools for data sharing and interoperability.

Infrastructure for data sharing that include 'found' data

Integration of FAIR guidelines in CI projects

Integration of mechanisms to verify and collect metadata on input data and computation

Libraries of tools

Linked data

Long term support (i.e. NSF-supported) for high-value data sets and services across science & engineering

Maintenance of existing, well-used tools (not adding novel capabilities, but rather user support, bug fixes, refactoring, etc.)

Management, curation, and reuse of data (which is gathered at high cost) need to be emphasized (viz. Fair principles)

Math libraries (e.g., BLAS, LAPACK, FFT, Sparse matrix, etc.) For the modern computer architectures.

Methods to more easily integrate variable data types into larger analysis frameworks

New AI and big data algorithm that does not required centralized data

Numerical libraries

Open Data / Encouraging Release of Data

Questions of data permanence/how do we not lose data in old and legacy formats?

Science data management infrastructure

Science-driven data alignment and integration

Secure storage technology and infrastructure (compliance oriented)

Semantic interoperability

Standards development for digital object representation of scientific data

Support CI that enables FAIR principles

Visualization Tools

Data science DATA

Development of a federation of data repositories in various disciplines

CATEGORY

RESPONSE

Data (cont.)

Creating a solicitation geared towards creating community-specific data standards in all S&E communities

Domain-Specific Cyberinfrastructure

CI for a specific domain (e.g., Sociology, Chemistry, etc.)

Domain-specific cyberinfrastructure

Earth sciences

Agriculture

Smart cities

Social sciences

Sociology

Research infrastructure for particular scientific communities (like Astropy)

Renewable energy

Quantum chemistry

Particle physics

Nuclear physics

Nanotechnology

Polar science

Materials discovery

Materials research

Materials science

Materials Science and Nanotechnology

Interfaces with research in material science and soft matter in particular

Infrastructure development for biology-specific disciplines, e.g., Neuroscience

Highly efficient implementation of new algorithms for game changing accuracy in predicting fast (THz) response of materials to electromagnetic pulses

Health

Geography/GIS

Geophysical domain

Global change

Discipline focus

Crystal structure prediction

Computational biology

Computational materials science

Computational materials science and quantum chemistry tools

Computational mechanics

Computational physics

Computer Infrastructure and methodological innovation for next-generation research in life sciences

Climate change

Climate variability and adaptation to climate change

Astronomical surveys

Atomic and molecular physics, including strong-field interaction and relativistic theory of heavy atoms

Biology

CATEGORY

RESPONSE

**Domain-Specific
Cyberinfrastructure
(cont.)**

Biophysics
 Biosciences
 Electric power systems
 Energy efficiency
 Power and energy grid edges
 Building data repositories for standardized benchmarking of physics software (e.g. Phase transitions in lattice models)

Software

Adding functionality to third-party open source software packages
 Common software frameworks
 Community-driven open source software
 Defining a software lifecycle and what success in sustainability looks like
 Develop better tools to broaden access to hero type scientific computing simulations
 Development of large-scale multiprocessor software
 Early-stage software/algorithm development
 Exascale domain-specific software
 Hardware/software co-design for scientific and commercial advancement
 Highly efficient implementation of new algorithms for game changing accuracy in predicting drug-protein binding site and energy
 Highly efficient implementation of new algorithms for game changing accuracy in predicting electrocatalytic conversion of solar energy to fuels and CO₂ to chemicals
 Hybridized approaches to software which blend web-hosted applications with local software
 Long term sustainability of large-scale software systems supporting research
 Maintenance of existing, well-used tools (not adding novel capabilities, but rather user support, bug fixes, refactoring, etc.)
 Middleware for domain applications
 More on the relationship between hardware and software, i.e. Support algorithms optimized for specific hardware and task.
 More support for existing, widely used software; less on speculative small awards for PIs with no history of providing CI
 OSS (open-source software) for S&E Problems that do not have current OSS solutions.
 Program level activity to establish and require standards and protocols across funded components such that the ecosystem can be realized independent of specific implementations of software (analogous to the Internet with TCP/IP, HTTP, HTTP and the fact that any browser can be used/any web server can be used and it all still works).
 Program level activity to increase exposure and dissemination of CI components. As discussed in the DIBBS '18 PI meeting the best way for someone to find information about the past 50 DIBBS activities is to search NSF award search site for the 50 abstracts. This is not sufficient. There should be a central index with required basic information (e.g. Documentation, code, demonstration instance, and activity information indicating whether or not the effort is still active as time goes on). This goes beyond the SGCI effort which is too narrow, and beyond the Big Data Hubs which doesn't have the resources for this. This likely belongs as part

CATEGORY

RESPONSE

Software (cont.)

of the next iteration of XSEDE.

Scalable software

Shared software development

Software development for next-generation high performance computers

Software engineering within disciplines falling behind

Software infrastructure for all of science (like NumPy and SciPy)

Software maintenance

Software support for new hardware

Software sustainability efforts (i.e., improvements such as refactoring, writing documentation, and maintenance of existing code)

Sustainable and sustained scientific software development, as broadly as possible and/or as driven by proposed research. This continues to be as critical as ever

Tools that support research on decision making in complex systems

Usability and accessibility of scientific software

User-friendly and accessible resources to use software tools

Learning/Workforce Development

Preparing the future CI workforce by supporting new knowledge and training programs.

CI education infrastructure?

Developing CI professionals that are fully committed to fully open source development, open collaboration and modern, evolving practices

Educational domain

Educational tutorial documentations and videos.

EHR

Encouraging the sharing of human resources between R1 and smaller institutions.

Environment and education

Funding exploratory work and young investigators will ensure a healthy pipeline of fruitful research and CI development for the future

HBCUs (Historically black colleges and universities) and MSIs (Minority-serving institution) centric solicitation

Human resource development in ways that keep S&E professionals available to academia and science in general

Human resource development that helps cross the divide between 'domain scientists' and IT and/or CS experts

Integration of education for all sectors of the science community including professors

Knowledgeable workforce that can assist in the development of software located anywhere in the country

Non-solicited workshop/training proposals

Outreach, communication and training to help those outside of CI leverage new capabilities developed in CSSI

Reaching out to communities that are not currently using or are less likely to use/understand CI as a resource and set of tools

Small and medium size institutional assistance

Software/database infrastructure specialists to develop APIs, manage servers, etc.

Support more young researchers

Support personnel

CATEGORY

RESPONSE

Learning/Workforce Development (cont.)

Training scholars and students to use more advanced ecostructure

Transdisciplinary training with respect to data science and software engineering targeting domain scientists

With crossover into classroom efforts, student and faculty development

Workforce development

Networking

Domestic and international efforts to build data transfer capability and computer networks capacity.

100Gb connectivity

Cloud services

Community microgrids and networked microgrids

Data science is changing the way we do research using what Jim Gray stated as the 4th paradigm. Emphasize building ecosystems to meet this grand challenge in the way we are doing new science

Dynamic cloud computing

Elastic compute service (like those available in the commercial cloud) for S&E

Extending existing successful CI in significant ways

How to develop mid- to large sized community owned software capabilities

Intelligent infrastructures to optimize services

IoT

Leveraging the public cloud for scientific innovation

Long term infrastructure sustainability

Long-term, sustained support for gateways and networks

Networking

Portable, cloud-native data and code

Promote widespread utilization of 1GB/sec data transfer between institutions, facilities and science endpoints (last mile CI)

Propagating good practices learned in some part of the ecosphere to others

Redundant and accessible storage (with management) for S&E in a cloud model

Secure and private cloud computing

Should focus on sustenance of services and deployments

Sustainability through containers

Sustaining and advancing existing CI

Use of Cloud services and funding models

Machine Learning (ML) / Artificial Intelligence (AI)

Super AI computers for local communities

Super AI computers

Robust and Reliable Machine Learning

Real-time machine learning e.g. For complex instrument control solutions

Physics driven machine learning

Machine learning

Machine learning deployment cyberinfrastructure

Integrated AI Algorithms and Sciences

Infrastructure embedding AI algorithms for chemical and bio/material discovery and design

Frameworks for ML & AI methods development and deployment over data and computing resources that could serve interdisciplinary communities

CATEGORY

RESPONSE

Machine Learning (ML) / Artificial Intelligence (AI) (cont.)

Elements for Machine Learning & AI methods development that could serve a broad interdisciplinary community

Development and deployment of machine learning models

Comprehensive knowledge representation of scientific data to enable AI and ML

Combination of NVIDIA and Intel AI chips

Advancing applications of AI/ML for scientific innovation

AI

AI & machine learning

AI for good

AI/Machine Learning and Adversarial Learning

Application of AI and Machine Learning innovations to orchestrate and manage CI effectively and efficiently

Application of artificial intelligence in science and engineering that can generate new knowledges

Focus on innovative high-risk high-rewards software and data analysis development, such as large-scale realistic supercomputer simulations and machine-learning analysis

Collaborations

Interdisciplinary research involving computer scientists and domain scientists

Science / engineering collaborations

Interagency research topics, such as NSF+NIH or NSF+USDA

Integrated frameworks that bring capabilities together

Incentive funds for collaborations with existing centers, institutes, and successful projects or for other behaviors that exemplify other NSF goals than scientific excellence

Foster more collaboration with domain scientists and computer scientists

Create an environment to strength collaboration and no competition

Consider joint events and teams between academic CI researchers and industry CI researchers, e.g., University CI team and Amazon AWS team working jointly towards academic CI for science in cost-effective and sustainable fashion

Collaboration between domain scientists, computer scientists, data scientists, and software engineers.

Collaboration between software/hardware developers, data producers and data consumers

Collaboration. Encouraging and enabling multidisciplinary access to and use of CI capabilities at an institution, and between institutions

Collaborations with industry to bring in industry-led advances to the science CI

Collaborative programs between large and smaller institutions

Brand new innovative, science driven, collaborative CI that leverages new technologies/advances.

Harmonization of different CI communities of practice

Security/Resilience

Consider investment into cyberinfrastructure that supports aging scientific equipment

Cybersecurity

Data science / AI in identifying and predicting cyber threats.

CATEGORY

RESPONSE

**Security/Resilience
(cont.)**

Documentation

Long-term persistent infrastructure for the computing and storage

Networking and Security

Phishing

Secure and private cloud computing

Secure storage technology and infrastructure (compliance oriented)

Secure tools

Natural hazard Security/Resilience

Hazards Security

Cyber physical resilience in microgrids and urban power distribution grids

Cybersecurity

Extreme events – security/resilience

Disaster Recovery and Resilience Security

Simulation/Modeling

Predictive and deterministic simulations

Methods to ensure reproducibility of simulations

Multiphase flow simulation

Multiphysics and multiscale modeling

Cyber infrastructure for multiscale simulation in chemistry, physics, materials and bioscience, especially from nanometers (excluding quantum) to microstructure

Coupling existing codes to enable multi-physics simulation

Modeling for additive manufacturing

Focus on innovative high-risk high-rewards software and data analysis development, such as large-scale realistic supercomputer simulations and machine-learning analysis

**High Performance
Computing (HPC)**

Tools and software for HPC (High Performance Computing)

High performance computing HPC

High-performance computing infrastructure, increased capacity, shorter wait times HPC

Domain agnostic enabling technologies, especially enabling modern tooling for high performance computing

Metrics

Recognizing value by not just download count but by criticality of end-use Metrics

Demonstrating or measuring impact of CI

Automatic metrics of effective deployment: what can we measure to determine if a system is current successful and in wide use by a scientific community?

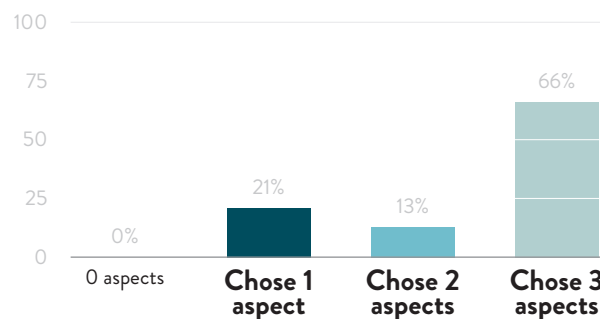
Provide best practice metrics to show improvement. Perhaps allow research into this.

Developing methods and tools to evaluate data and services quality

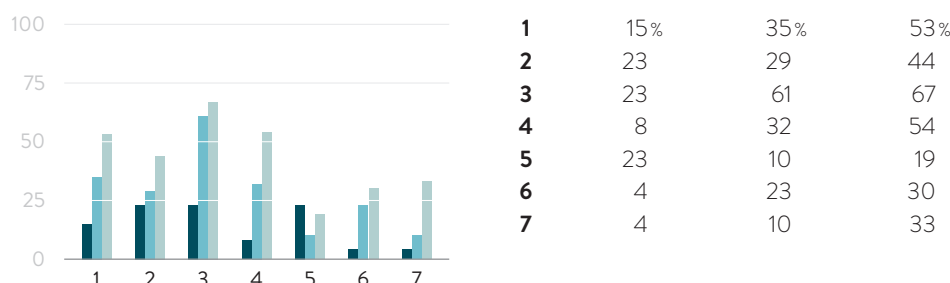
multi-step analysis

A Answers to the survey's solicitation and selected demographic questions, sorted by groups determined by how many aspects for change respondents voted for.

Participants were asked to **select three of seven** indicated aspects of the current solicitation where improvement is most needed or could make the most difference.



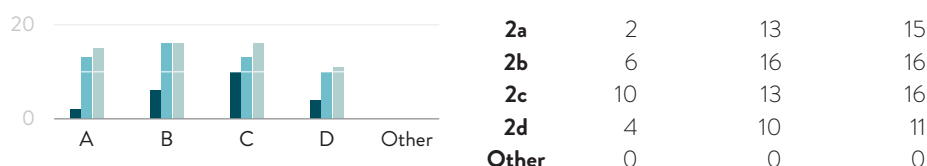
Overall solicitation aspects (p. 5)



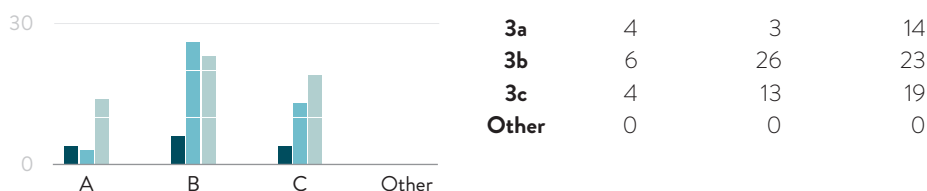
1 Types of awards (p. 6)

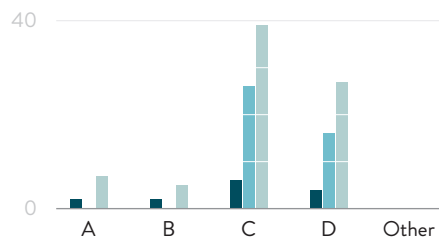


2 Combined software/data umbrella (p. 7)



3 Presentation of domain-specific interests (p. 8)

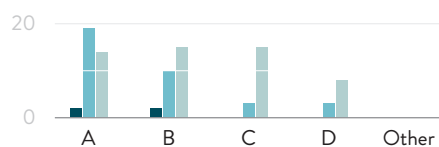


4 Solicitation-specific criteria (p. 9)

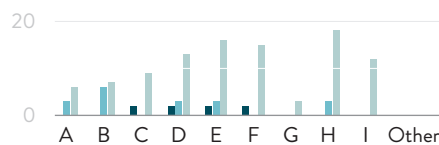
	Chose 1 aspect	Chose 2 aspects	Chose 3 aspects
4a	2	0	7
4b	2	0	5
4c	6	26	39
4d	4	16	27
Other	0	0	0

5 Solicitation logistics (p. 10)

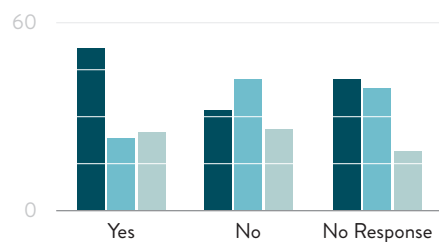
5a	13	6	5
5b	10	6	8
5c	0	0	0
5d	2	0	7
5e	0	0	0
Other	0	0	0

6 Guidance included in the solicitation (p. 11)

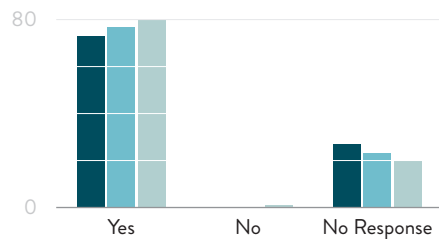
6a	2	19	14
6b	2	10	15
6c	0	3	15
6d	0	3	8
Other	0	0	0

7 Support provided beyond the written solicitation (p. 12)

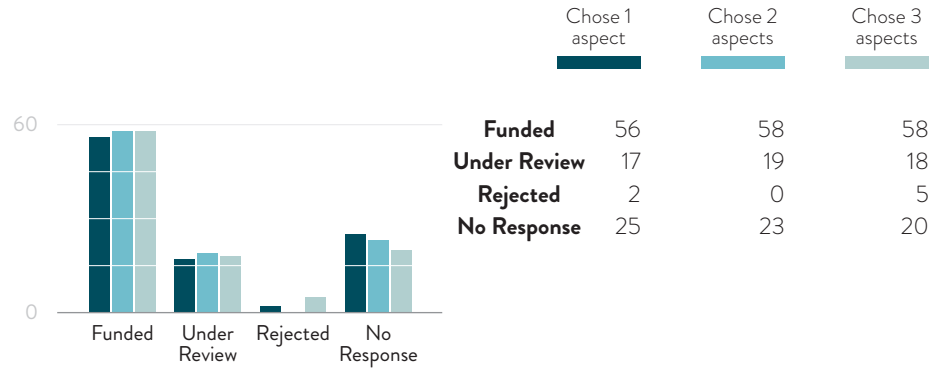
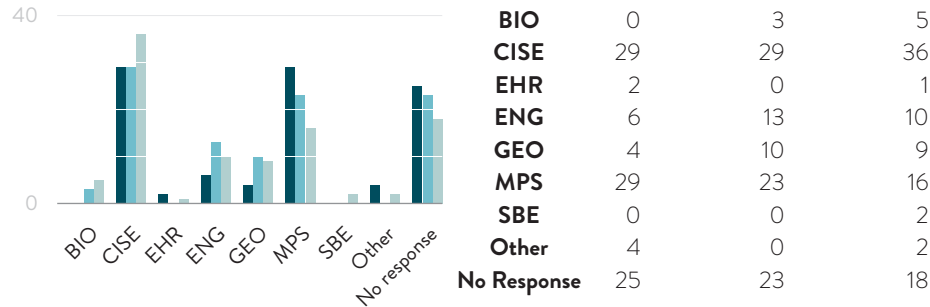
7a	0	3	6
7b	0	6	7
7c	2	0	9
7d	2	3	13
7e	2	3	16
7f	2	0	15
7g	0	0	3
7h	0	3	18
7i	0	0	12
Other	0	0	0

D2 Served as Reviewer (p. 3)

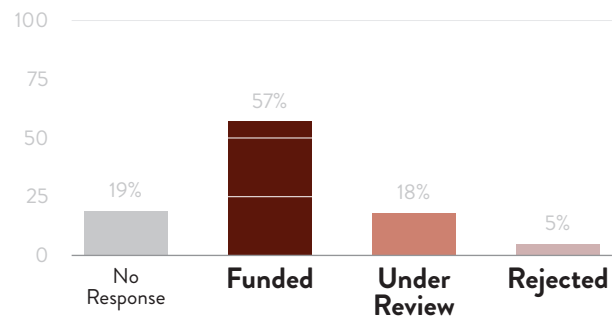
Yes	52	32	42
No	23	42	39
No Response	25	26	19

D4a Have you submitted a proposal to NSF? (p. 3)

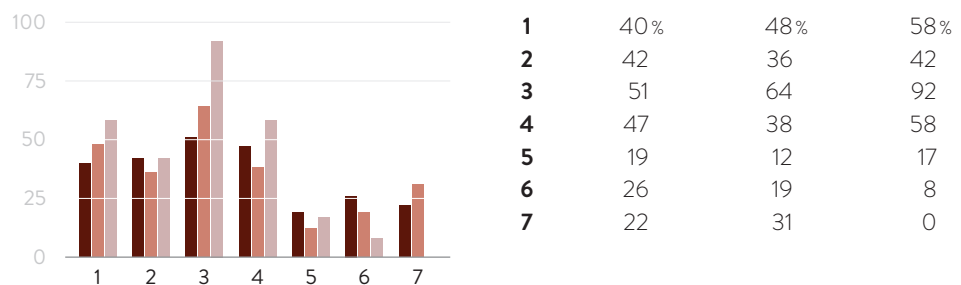
Yes	73	77	80
No	0	0	1
No Response	27	23	20

D4b Status of proposal (p. 3)**D5 Directorate** (p. 3)**B** Answers to the survey's solicitation and selected demographic questions, sorted by groups determined by the status of proposals submitted by respondents (demographic question 4b).

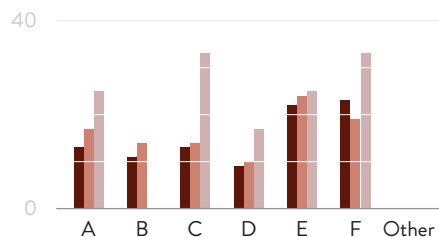
Respondents were asked to indicate whether or not they had previously submitted a proposal to NSF, and if yes, then asked to indicate the status of their proposal.



Participants were asked to **select three of seven** indicated aspects of the current solicitation where improvement is most needed or could make the most difference.

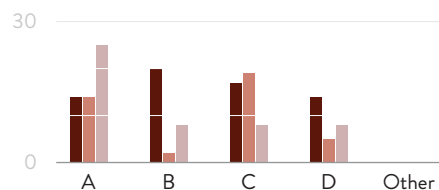


1 Types of awards (p. 6)



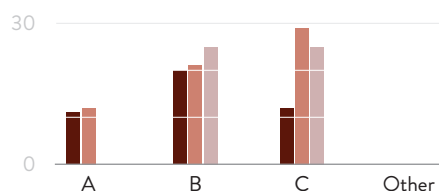
	Proposal Funded	Under Review	Proposal Rejected
1a	13	17	25
1b	11	14	0
1c	13	14	33
1d	9	10	17
1d	22	24	25
1d	23	19	33
Other	0	0	0

2 Combined software/data umbrella (p. 7)



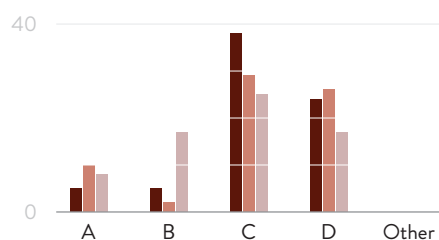
	Proposal Funded	Under Review	Proposal Rejected
2a	14	14	25
2b	20	2	8
2c	17	19	8
2d	14	5	8
Other	0	0	0

3 Presentation of domain-specific interests (p. 8)



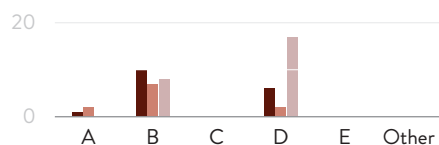
	Proposal Funded	Under Review	Proposal Rejected
3a	11	12	0
3b	20	21	25
3c	12	29	25
Other	0	0	0

4 Solicitation-specific criteria (p. 9)



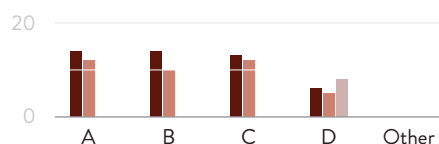
	Proposal Funded	Under Review	Proposal Rejected
4a	5	10	8
4b	5	2	17
4c	38	29	25
4d	24	26	17
Other	0	0	0

5 Solicitation logistics (p. 10)



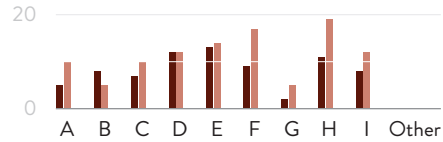
	Proposal Funded	Under Review	Proposal Rejected
5a	1	2	0
5b	10	7	8
5c	0	0	0
5d	6	2	17
5e	0	0	0
Other	0	0	0

6 Guidance included in the solicitation (p. 11)



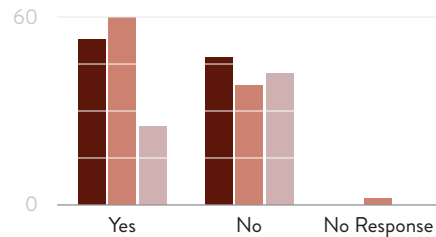
	Proposal Funded	Under Review	Proposal Rejected
6a	14	12	0
6b	14	10	0
6c	13	12	0
6d	6	5	8
Other	0	0	0

7 Support provided beyond the written solicitation (p. 12)



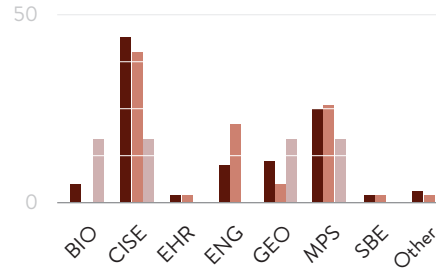
	Proposal Funded	Under Review	Proposal Rejected
7a	5	10	0
7b	8	5	0
7c	7	10	0
7d	12	12	0
7e	13	14	0
7f	9	17	0
7g	2	5	0
7h	11	19	0
7i	8	12	0
Other	0	0	0

D2 Served as Reviewer (p. 3)



Yes	53	60	25
No	47	38	42
N/R	0	2	0

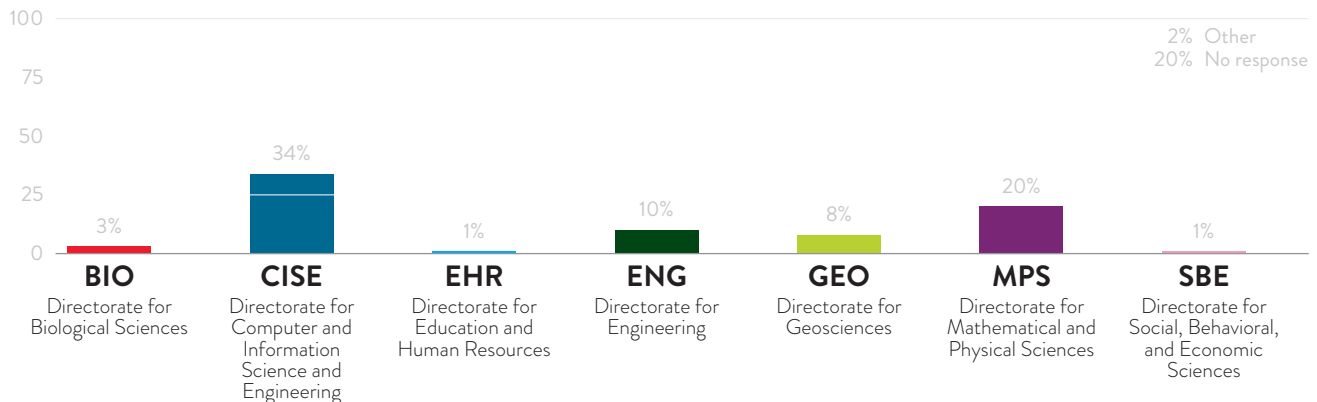
D5 Directorate (p. 3)



BIO	5	0	17
CISE	44	40	17
EHR	2	2	0
ENG	10	21	0
GEO	11	5	17
MPS	25	26	17
SBE	2	2	0
Other	3	2	0

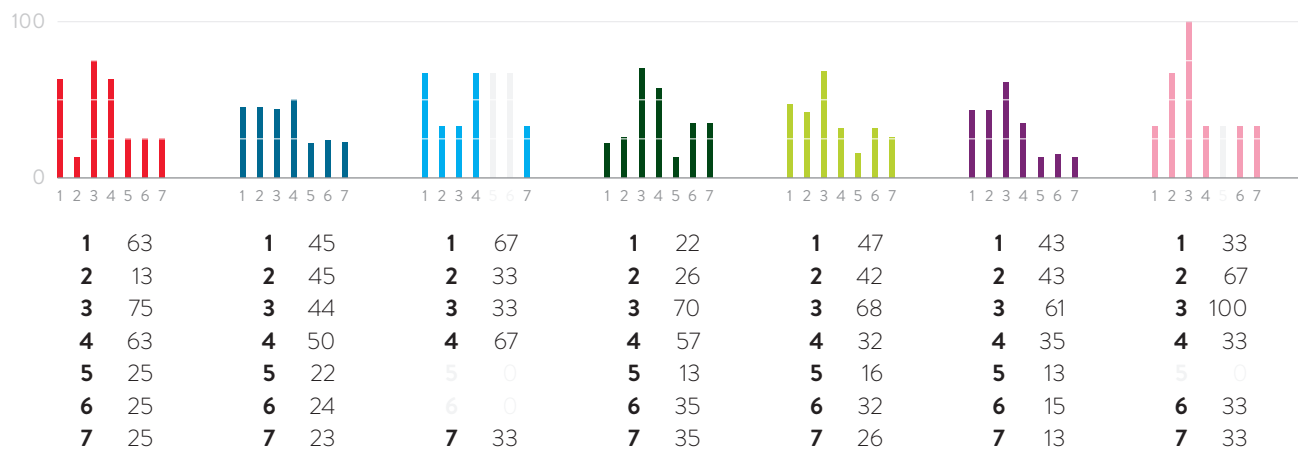
C Answers to the survey's solicitation and selected demographic questions, sorted by groups determined by respondent directorates (demographic question 5).

Respondents were asked to indicate which NSF Directorate is most related to their field:

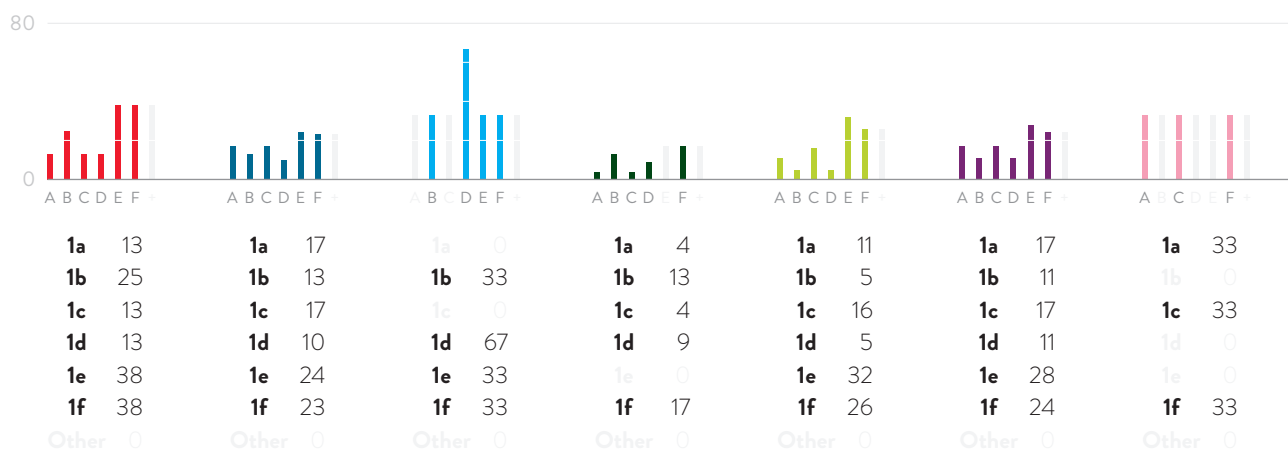


BIO**CISE****EHR****ENG****GEO****MPS****SBE**

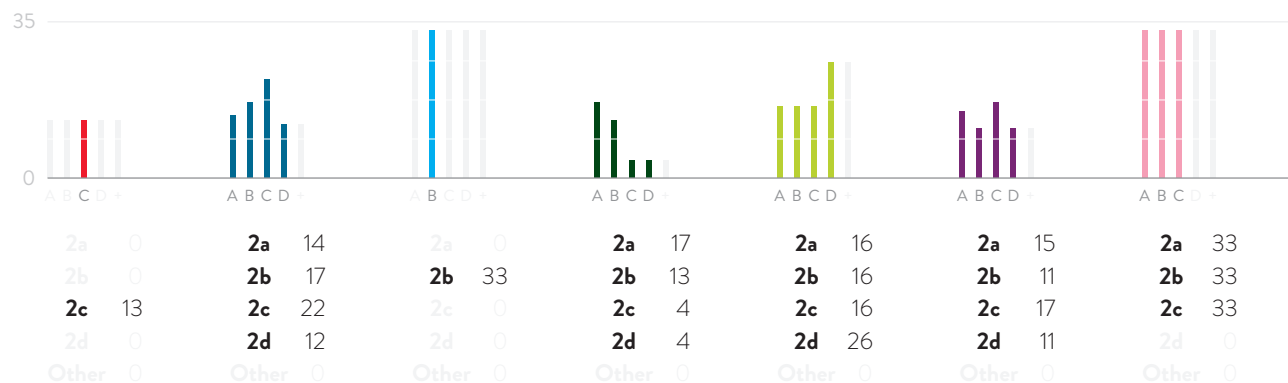
Participants were asked to **select three of seven** indicated aspects of the current solicitation where improvement is most needed or could make the most difference.

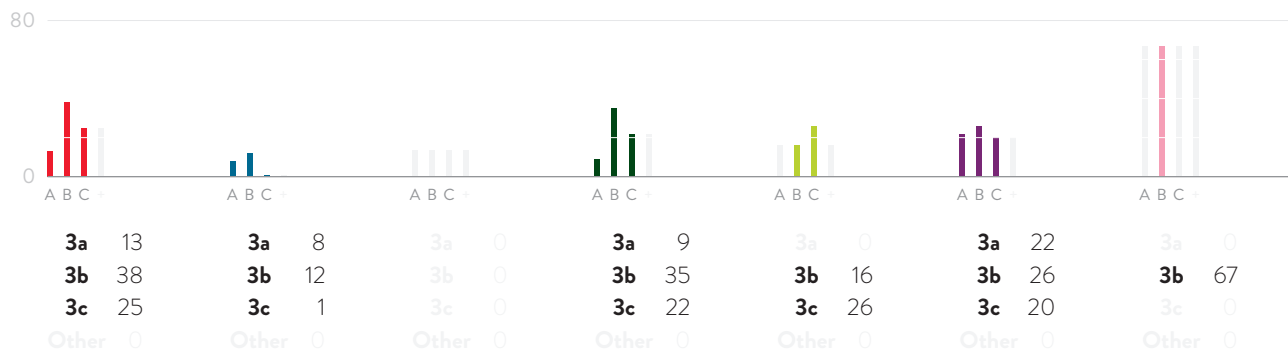
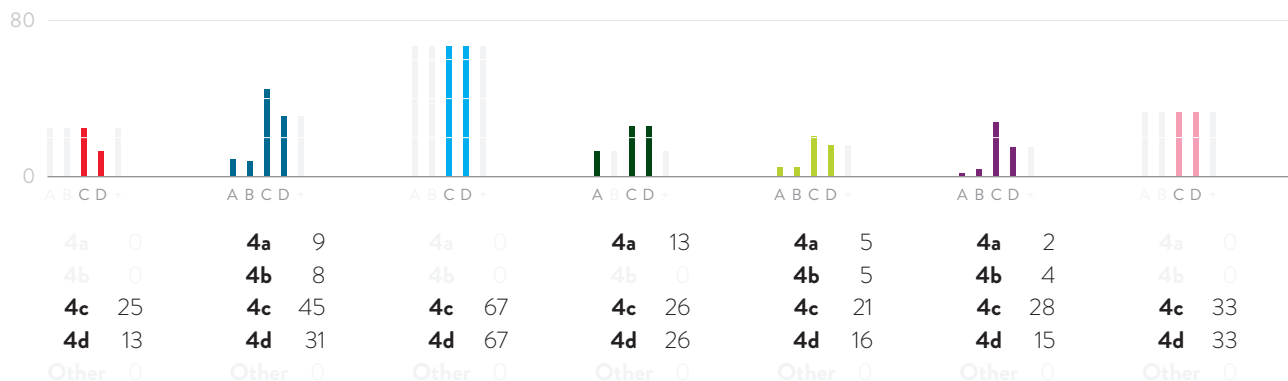
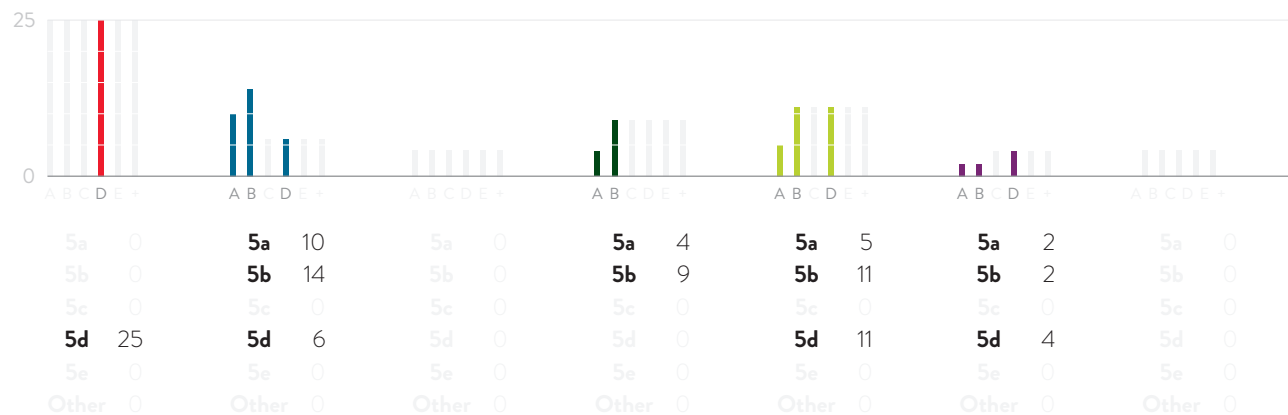


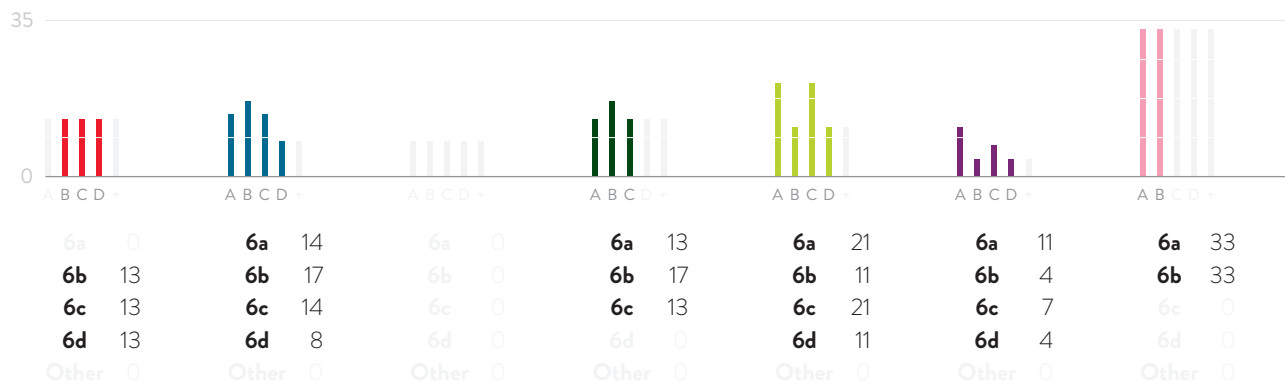
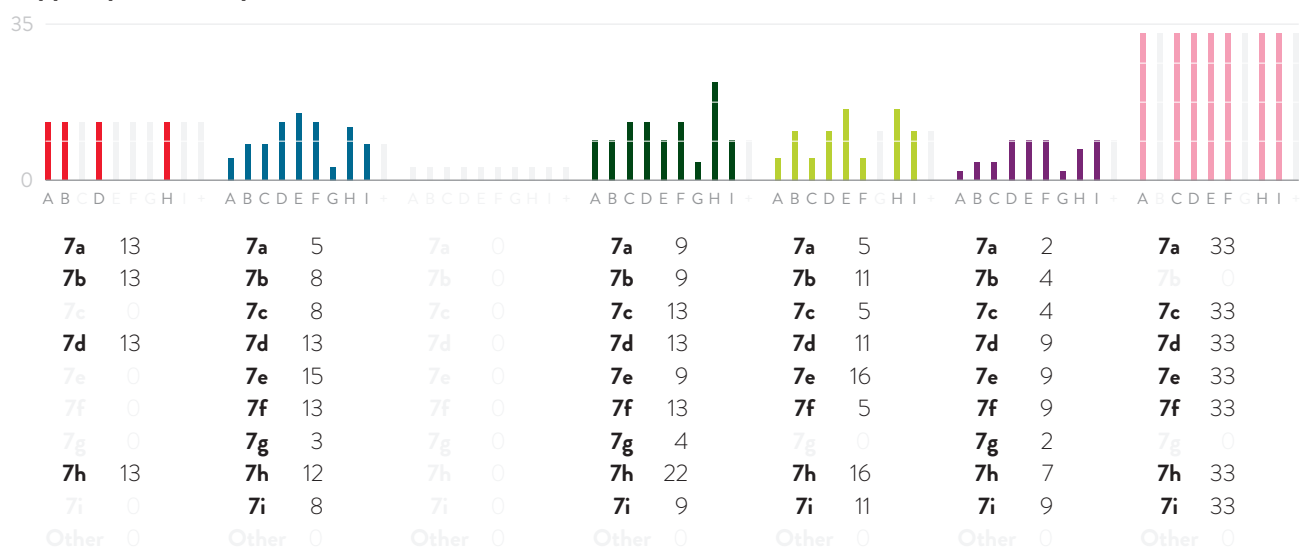
1 Types of awards (p. 6)

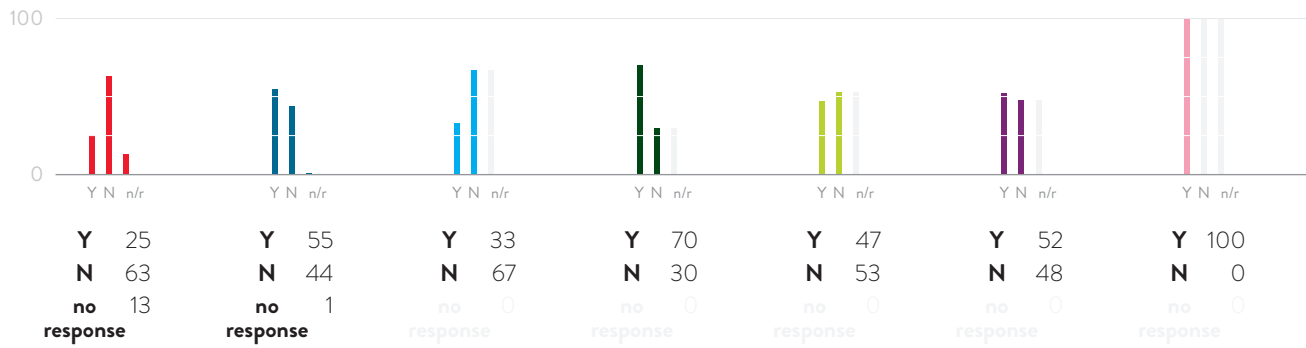


2 Combined software/data umbrella (p. 7)



BIO**CISE****EHR****ENG****GEO****MPS****SBE****3 Presentation of domain-specific interests (p. 8)****4 Solicitation-specific criteria (p. 9)****5 Solicitation logistics (p. 10)**

BIO**CISE****EHR****ENG****GEO****MPS****SBE****6 Guidance included in the solicitation (p. 11)****7 Support provided beyond the written solicitation (p. 12)**

BIO**CISE****EHR****ENG****GEO****MPS****SBE****D2 Served as Reviewer** (p. 3)**D4b Status of Proposal** (p. 3)