Data Asset Inventory Report:
Summary of Responses

NIH HEAL Initiative and Helping to End Addiction Long-term are service marks of the U.S. Department of Health and Human Services.

Prepared for: **The NIH HEAL Initiative®**
Prepared by: **The HEAL Data Stewardship Group**

Table of Contents

Executive Summary

The Helping to End Addiction Long-term (HEAL) Data Stewardship Group (HEAL Stewards), in collaboration with the National Institutes of Health (NIH) HEAL Initiative® Office, conducted the HEAL Data Asset Inventory (DAI) to understand the current plans and future needs for data processing and management in HEAL studies. The findings, summarized in this report, identify opportunities and challenges facing HEAL principal investigators (PIs), HEAL Stewards or other stakeholders. Overall, the HEAL DAI catalyzed considerable engagement between the NIH HEAL Stewards and investigators and their teams, with approximately 18% of principal investigators (representing 104 out of 564 awards) submitting responses. Additionally, the distribution of the various scientific Research Focus Areas among DAI respondents roughly mirrored that of all HEAL awardees.

The first round of DAI responses offers a valuable snapshot of data management plans across the HEAL Initiative®. The key findings that emerged from the analysis of their responses are summarized in the bulleted list below, and are explained in more detail in the section titled "Key Findings." A detailed characterization of the teams that responded to the DAI can be found in the section titled "Data Asset Inventory Response Summary."

- Most studies have not begun planning for repository use.

- The majority of the respondents (57%) reported that they are not collecting metadata to describe their HEAL data, although many studies do collect metadata.

- Most studies rely on workstation-based analysis and graphing software (e.g., SAS, R) for data analysis. The majority of respondents reported that they are not generating common biomedical sources of "big data" such as fMRI, microscopy images, and EEG/EKG.

- Most studies rely on local small data files (e.g., spreadsheets, flat files) for storing their data.

- Nearly 80% of DAI respondents have begun generating data, yet less than a quarter (24.4%) have made plans to share their data with the scientific community. The anticipated date for data sharing varies from July 2021 to August 2026.

- The majority of respondents (74%) do not have concerns with their ability to comply with the NIH HEAL data sharing policy, and less than half of respondents expressed interest in assistance in this area. However, responses in other areas of the DAI suggest that additional guidance is needed to align studies' public access and data sharing plans with the HEAL Public Access and Data Sharing Policy.

This report endeavored to capture the vast, complex landscape of HEAL Initiative® awards by characterizing DAI responses and identifying relevant opportunities and challenges. This exercise informs a robust strategy for HEAL Steward engagement with HEAL awardees moving forward. Broadly, the HEAL Stewards plan to provide additional outreach and education surrounding metadata. Additional assistance with HEAL data sharing and public access plans as well as data sharing concerns will also be broadly addressed. Ongoing assistance for repository selection and use will be important for the success of the HEAL Initiative®. Finally, the HEAL Stewards will focus outreach on segments of the HEAL awardee population with lower response rates to the DAI, particularly those with smaller awards who may have less access to support. Access resources here ⤴, to learn more about how the HEAL Stewards are working across the HEAL data community to provide solutions for managing and coordinating the diverse data across the NIH HEAL Initiative®.

Background

The HEAL research portfolio represents a unique collection of scientific investigations aimed at improving our understanding of addiction and pain, with the ultimate goal of improving patient lives. The HEAL Data Asset Inventory (henceforth referred to as the Inventory) was created to inform data management efforts by collecting information on HEAL research and the data it will produce. This effort will build upon and expand the pilot data management landscape analyses conducted over the last several years by our BioTeam colleagues (Osborne et al., 2021).

The Inventory was shared with HEAL projects awarded during the fiscal years 2018, 2019, and 2020. Throughout this period a total of 564 projects were awarded. Projects were distributed across six HEAL-defined Research Focus Areas aimed at enhancing pain management and improving treatments for opioid misuse and addiction. Types of projects included Research Projects (R), Training and Career Development Programs (T), Small Business Programs (SBIR, STTR), Supplements, Multiproject (P, U), Cooperative Agreements (U), and Other Transactions (OT). The duration of the projects varied from one to five years, with some expected to go longer after transitioning to the next stage of research within the HEAL Initiative® (e.g., from pilot/exploratory to clinical trial). Total funding allocated to the projects ranged from the low-five figures to the mid-eight figures.

The results of the Inventory are summarized in three main sections below: Data Asset Inventory Response Summary, Key Findings, and Conclusion. A more comprehensive treatment of the results can be found in subsequent appendices.

Data Asset Inventory Response Summary

The characterization of respondents took the following variables into account: Research Focus Area, Award Mechanism, Data Coordination Center, Study Size, and Administering Institutes and Centers (ICs). Two metrics were used to characterize inventory responses compared to the HEAL awards population: (1) response rate, as the percentage between inventory responses and HEAL awards, and (2) representativeness, as the count of

respondents belonging to a given variable group compared to the count of the variable population. The HEAL awards population data used for characterization was obtained from two sources: The Helping to End Addiction Long-term® Initiative Funded Projects website and the NIH RePORTER website.

The inventory was structured in two parts, each targeting a predetermined respondent or group of respondents best fit to answer the questions comprehensively. Part 1 gathered general project information, to be answered by either PIs, project managers (PM), or any other personnel; Part 2 included specific data related questions to be answered by data managers. So, a complete Inventory response for a given award constituted successful submission of both parts of the assessment for that specific award. The Inventory was sent to 564 HEAL investigators and resulted in 104 responses in Part 1 and 46 responses in part 2, with corresponding response rates of 18% and 8%, respectively. A subset of research programs in HEAL have access to expert data management infrastructure and additional guidance through a Data Coordinating Center (DCC).[1] The lower response rate to part 2 of the inventory occurred partly because individual DCCs acted as the data manager by submitting one global part 2 response on behalf of the multiple HEAL studies they support. Thus, each targeted award (out of 564 total awards) could submit a response to part 1 (564 total possible individual submissions); and, since many awards receive support through a common DCC, only one part 2 response was required for all studies under a common DCC.

Research Focus Areas

The inventory asked respondents to characterize their study for HEAL by identifying their research focus areas. Figure 1 illustrates the inventory response rate across the six HEAL research focus areas. Response rates differed across various research focus areas, for reasons that are unclear. Further scrutiny of the Inventory data could help explain these response rate differences and delineate strategies to increase outreach.

---

[1] The term Data Coordinating Center (DCC) is used in reference to any groups doing data harmonization and management for a set of HEAL awards, despite their official titles or center names.
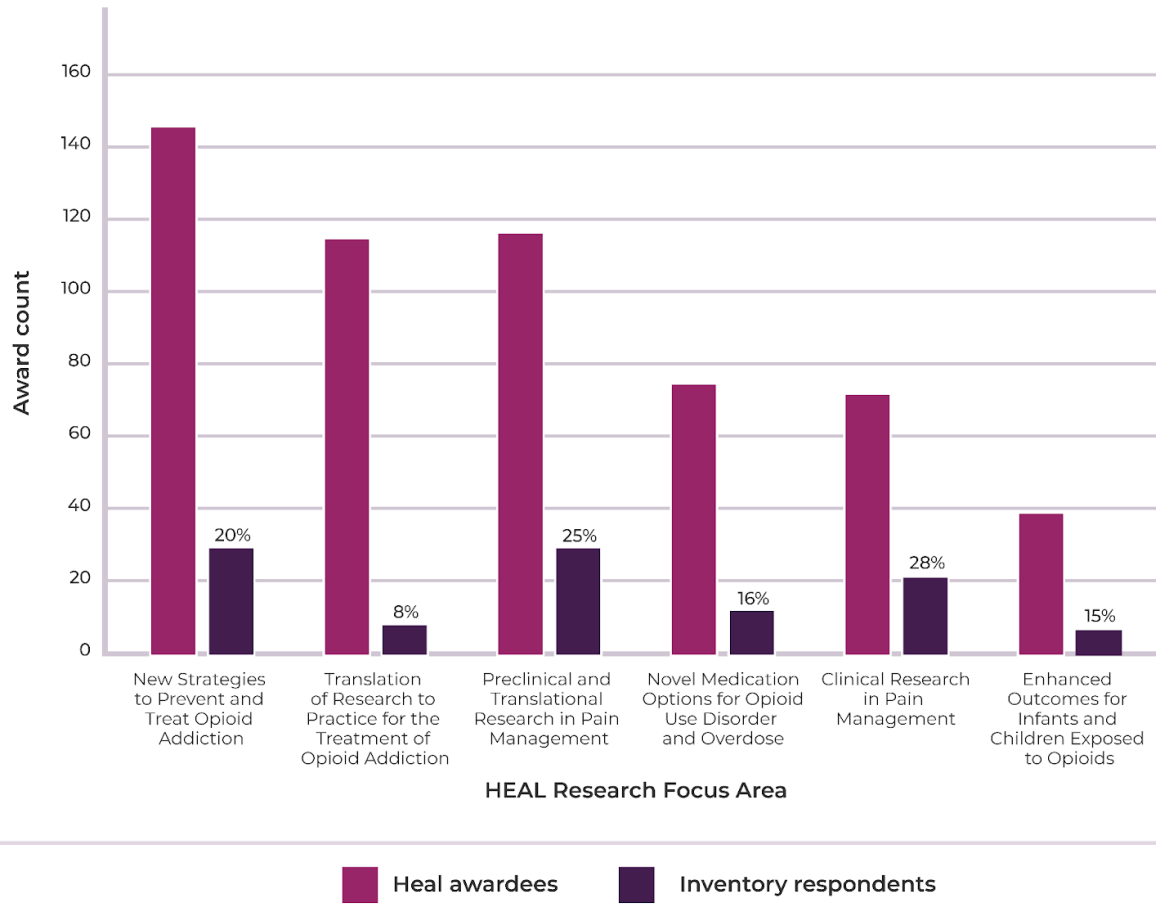
Figure 1. Responses by HEAL Research Focus Area. The Y axis represents the award count, the "%" on top of the Inventory respondents represents the percentage between Inventory responses and HEAL awards by research focus area.
Note: A small proportion of respondents submitted inaccurate responses for certain descriptive fields (e.g., research focus area). Figures in this section represent corrected responses for research focus areas submitted by respondents according to data in the HEAL Initiative® Funded Projects website.

We compared the proportion of studies in each research focus area to all HEAL studies vs. a comparison of the proportion of respondents within each research focus area to total respondents to the Inventory. For each dataset, values were represented as percentages of the number of individuals with studies from a given research focus area relative to the total number of studies. Inventory respondent data was generated from the most recent export of the database (7/26/21). HEAL Ecosystem data was generated from the HEAL Funded Projects page.

The research focus area "Translation of Research to Practice for the Treatment of Opioid Addiction" was underrepresented in the inventory, whereas researchers in "Preclinical and Translational Research in Pain Management" and "Clinical Research in Pain Management" were slightly overrepresented, as shown in Figure 2.
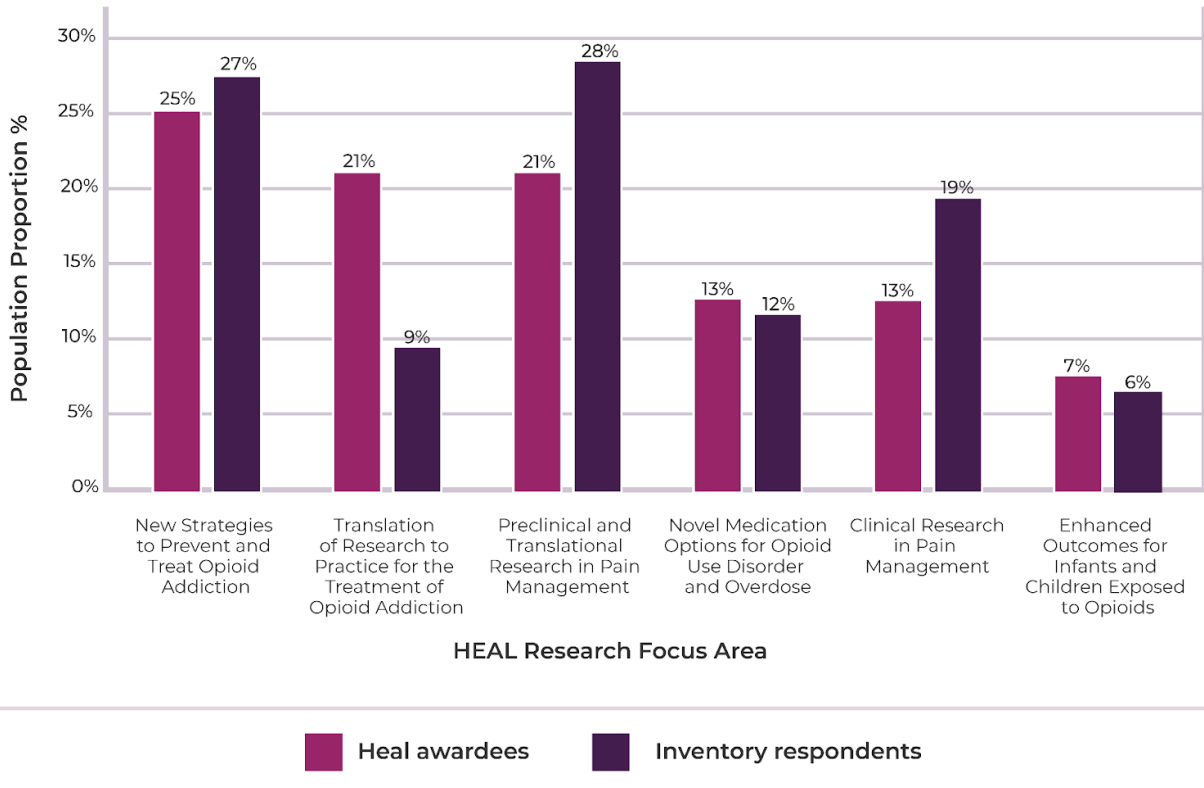

Figure 2. Overall proportion of HEAL awards and of inventory respondents by HEAL Research Focus Area (total HEAL awardees vs. total inventory respondents belonging to a given Research Focus Area).

Type of Grant Program - Activity Code

Following the contextualization of responses by Research Focus Area, trends by activity code were identified and measured in response rates. NIH utilizes a broad selection of award types for HEAL, mainly distributed among R and U grant programs, Table 1 describes the activity codes frequently utilized by HEAL.

R awards: research grant programs	
R01	NIH's main independent research project grants
R21	Exploratory and developmental research projects
R34	Clinical trial planning grants

R awards: research grant programs	
R41/R43 /R44	Small business research
R61	Exploratory and/or developmental research projects with larger budgets than R21
U awards: cooperative agreements for high-priority research areas that require substantial involvement from NIH program or scientific staff	
UG1	Clinical evaluation of various methods of therapy and/or prevention (in specific disease areas)
UG3	Bi-phasic approach to funding exploratory and/or developmental research
U01	Discrete, specified, circumscribed project to be performed by the named investigator(s) in an area representing his or her specific interest and competencies
U24	Improvement of the capability of resources to serve biomedical research
U19	Multiple projects directed toward a specific major objective
Other activity codes: P50 and PL1 awards for multi project studies, K01 and K23 awards for research career development, and S06 awards to strengthen the biomedical research and research training capability of ethnic minority institutions.	

Table 1. Selection of NIH activity codes largely utilized by HEAL, definitions obtained from Activity Codes Search Results (nih.gov).

A breakdown by activity code displays a disproportionate response rate across different activity codes. To facilitate the graphical visualization of the different activity codes in HEAL, codes were divided in two types, R codes plus all others with less than five awards and U codes. For instance, 234 out of 564 HEAL projects are R type grants (Figure 3a). While this group was somewhat responsive to the Inventory (16%), codes like R61, R21, and R33 had a response rate equal or above 20%, whereas codes such as R01 and R34 had a response rate below 10%. Correspondingly, 295 out of 564 HEAL awards are U type grants (Figure 3b). The overall response rate of this group was 21%, with codes like UH3, U44, and UG3 reaching a response rate above 30%, while codes like U01, U19, and UG1 attained a response rate of 10% or less. The higher response rate among U awardees might be the product of their high involvement with the NIH programs or scientific staff. Activity codes identified as "all others" in Figure 3a show a combined response rate of 25%, some of these grant types are P, K, and S.
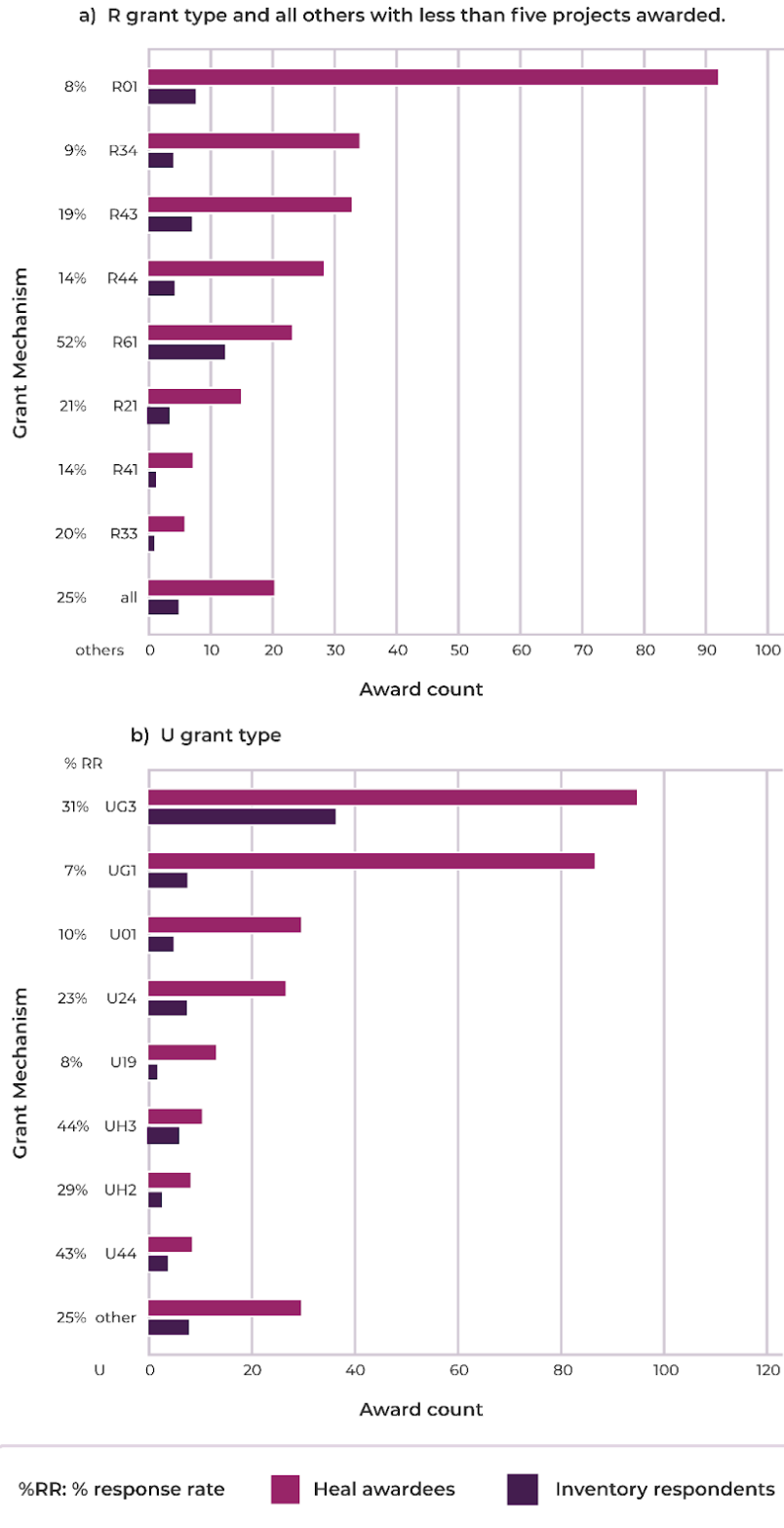
Figure 3. Inventory responses by activity code, breakdown by type: (a) this section accounts for R types and all others with less than five projects awarded, and (b) this section includes the U grant type.

Size of Award

Awards of a smaller size (total funding) likely do not have access to resources such as a DCC, have smaller budgets, and fewer years to complete the project relative to the other studies in the HEAL portfolio, and therefore may need additional data management support. Figure 4 illustrates inventory response by total funding. In general, awards under $2 million, responded at a lower rate to the Inventory.
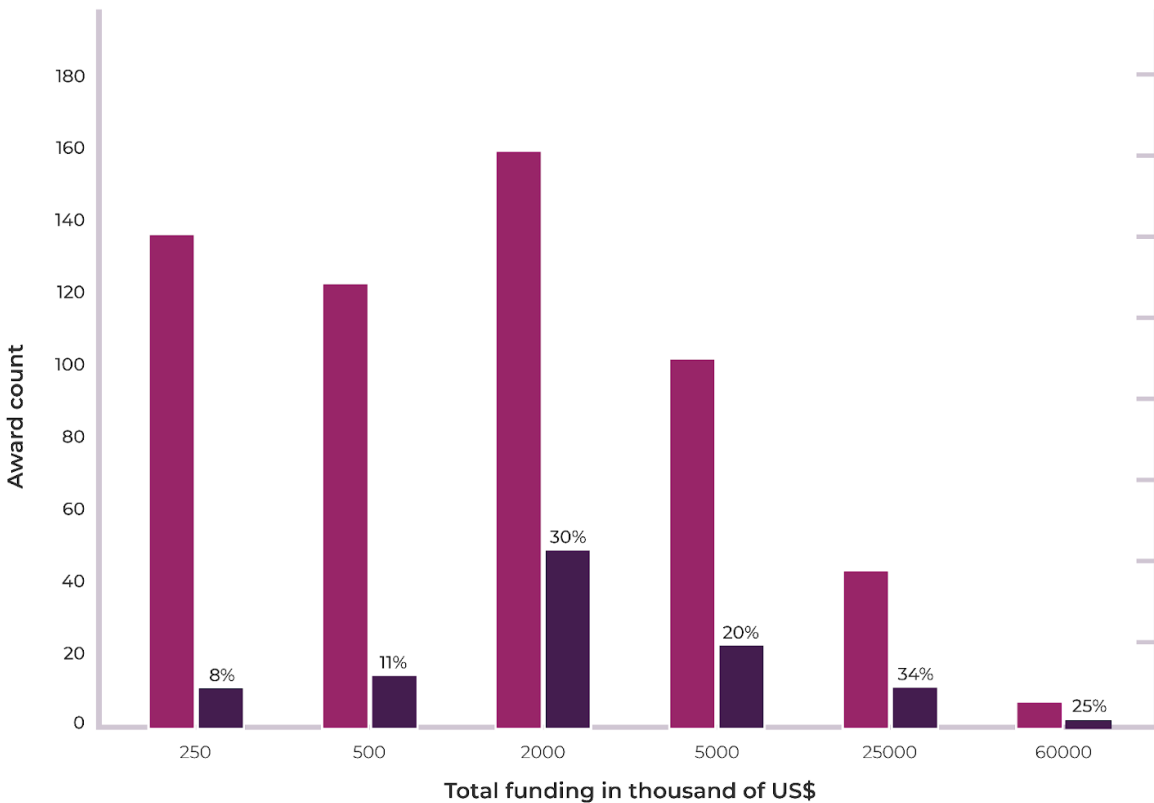

Figure 4. Inventory responses by total award funding in US$. The left Y axis represents the award count, the right Y axis represents the percentage between Inventory responses and HEAL awards.

Data Coordination Center (DCC)

The HEAL DCCs provide data management for multisite studies and collaborative research networks. They are responsible for coordinating data management efforts for

the different studies, collaboratives, or clinical centers under their umbrella. Table 2 shows that awardees associated with a DCCs completed the general and data sections of the Inventory at a higher response rate than the ones without DCC.

There was a high (40%) response rate from HEAL DCCs on both the general and data questionnaires (Part 1 and Part 2 of Inventory); given that roughly 13% of awardees are

DCC Status	General Q	Data Q
No DCC	16%	6%
DCC assoc.	28%	14%

Table 2. Response rate by association to a DCC

associated with a DCC, their responses provide insight about a broader number of awardees. It is important to mention that HEAL Stewards requested DCCs to answer Part 2 of the Inventory on behalf of the studies they manage data for.

Other Considerations

Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) response rates consisted of roughly 12% of all Inventory responses (12 respondents out of 104 were small business awards) and represented 17% of HEAL small business awards. It is important to note that data sharing expectations for SBIRs are not defined in the NIH HEAL Public Access and Data Sharing Policy.

Key Findings

Study Data Types

In order to guide repository selection and long-term data storage plans, respondents were asked about which large data files they plan to create (see Figure 5) such as fMRI or EKG. Most respondents indicated that they had no plans to create large data files as part of their HEAL work. However, a subset (29 respondents or 27.9%) indicated that they had such plans. It will be worthwhile for the HEAL data stewards to identify repositories that accommodate needs for the varied types of "big data" listed in Table 3.
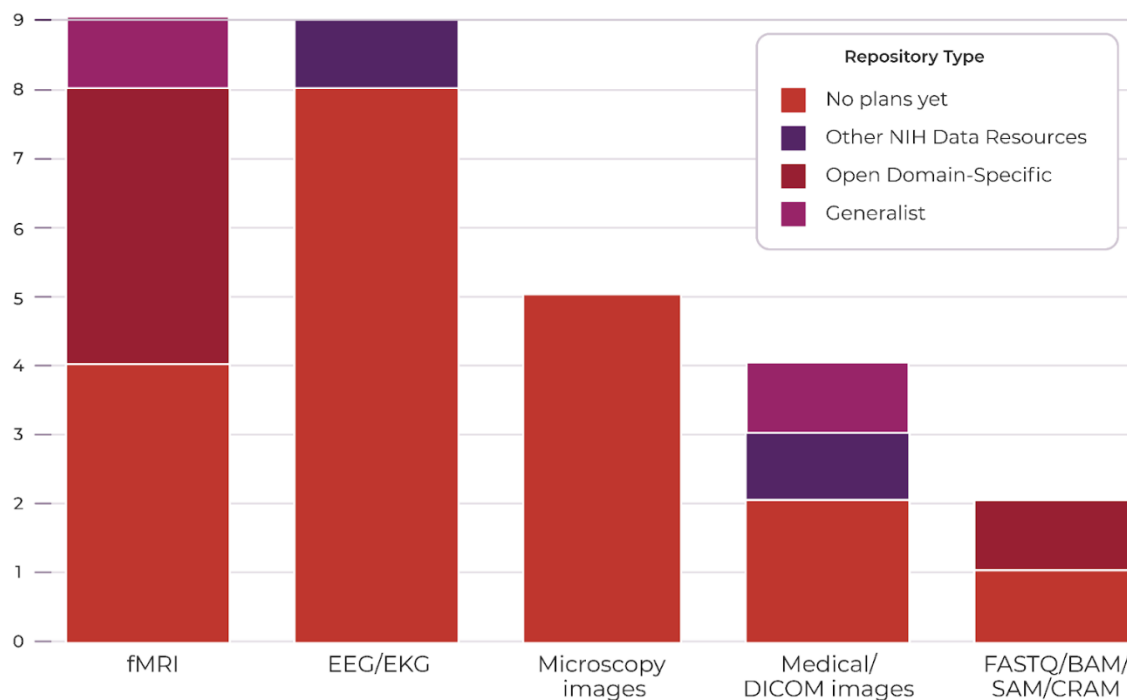
Figure 5. Repository use and HEAL big data

Research Focus Area	EEG/EKG	FASTQ/ BAM/SA M/CRAM	fMRI	Medical/ DICOM images	Microscopy images	Grand Total
Clinical Research in Pain Management	-	-	2	2	-	4
Enhances Outcomes for Infants and Children Exposed to Opioids	1	-	3	-	-	4
New Strategies to Prevent and Treat Opioid Addiction	2	1	-	-	-	3
Novel Medication Options for Opioid Use Disorder and Overdose	3	-	1	-	1	5
Preclinical and Translational Research in Pain Management	3	1	3	2	4	13
Grand Total	9	2	9	4	5	29

Table 3. Breakdown of Inventory respondents creating large data files, divided by Research Focus Area and by data type.
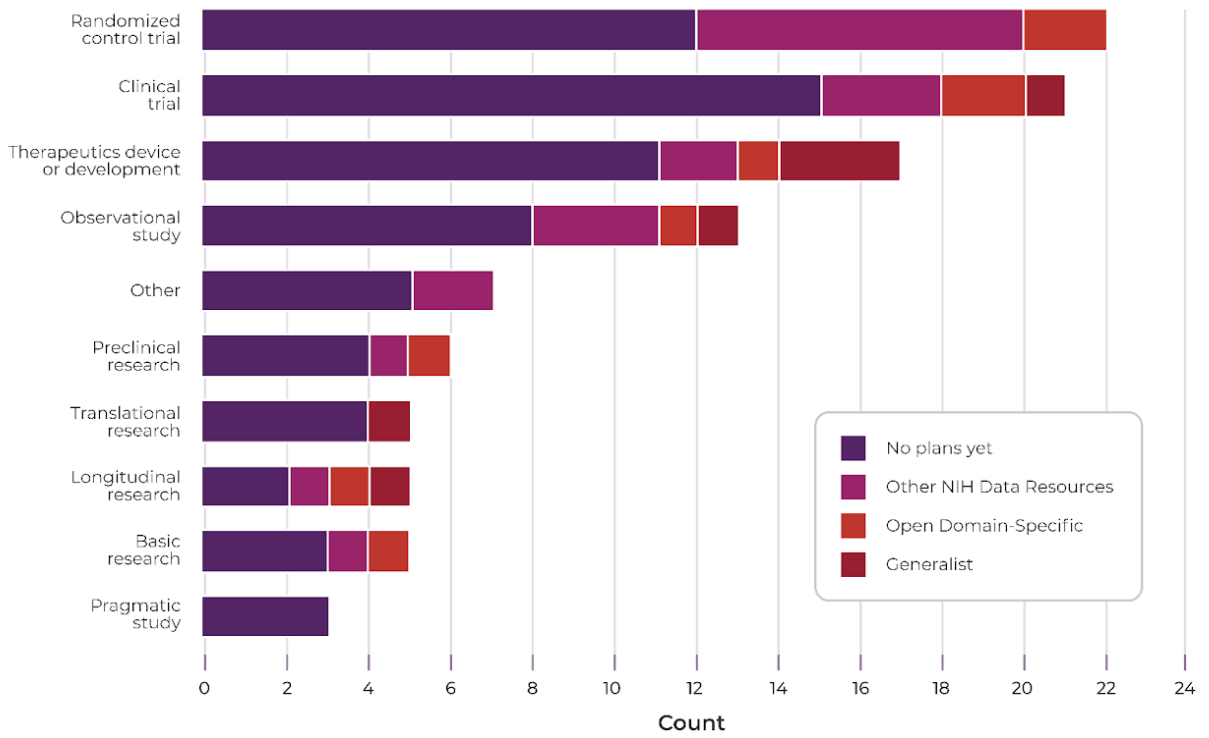
Repository Plans

Depositing research data in a repository promotes longevity and renders the data

(re)usable to future researchers (*File Formats for Long-Term Access*, n.d.). To determine

whether HEAL researchers were prepared to promote the longevity of their HEAL data, respondents were asked questions about research database and repository use. Only half of respondents (27/53 respondents) confirmed that their HEAL studies were currently registered on clinicaltrials.gov, a database of privately and publicly funded clinical studies. In addition, only one study in clinical research in pain management (1/53) indicated that their clinicaltrials.gov registration page referenced support from the HEAL Initiative®. The denominator (53 studies) included studies from various programs and research areas and descriptions including clinical and observational studies, as well as therapeutics or device development.

Additionally, the majority of respondents (64.4%) reported not having any plans for repository use, despite the fact that most studies who submitted a response have started generating data as part of their HEAL work (see Figure 6). If this subset of respondents reflects the entire HEAL population, then helping programs to plan for repository use will be a top priority effort.

a) Repository use plans by study types

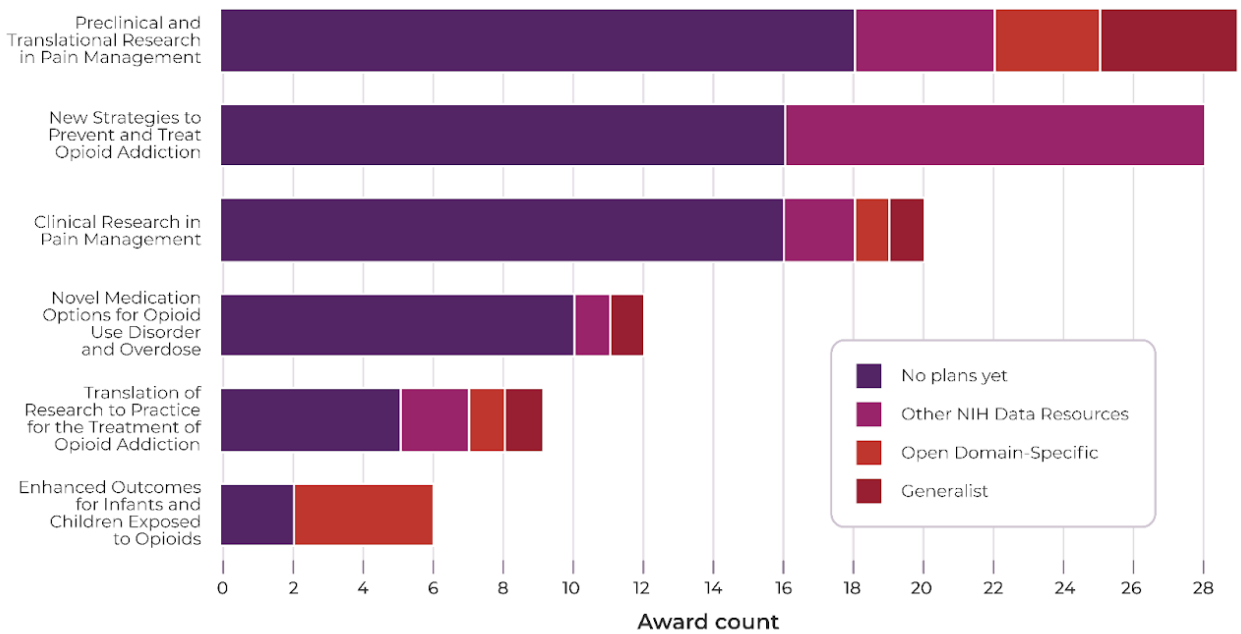b) Repository use plans by HEAL research focus area

Figure 6. Repository use plans in HEAL study types and research focus areas, categorized by repository type indicated, **if applicable**).

Further analysis of resource sharing plans submitted by respondents revealed specific repositories that respondents have mentioned and plan to use for data deposition. A list of 12 repositories was used to look for matches in the documents and count the frequency each name appeared. See summary in Table 4.

Mentioned Repositories	DASH	XNAT	OpenNeuro	Figshare
Mentions	19	6	1	1

Table 4. Specific repositories mentioned by DAI respondents as data deposition sites

Big Data and Repository Plans

Out of the 29 respondents who said they plan to produce large data files, only 9 indicated having concrete plans for repository use (Figure 5). The other 20 respondents indicated that they have no current plans for repository use. This finding is cause for concern given that some large data files may require specific storage needs and be more difficult for some repositories to accommodate (e.g., image data).

Public Access and Data Sharing

Data Sharing

The NIH HEAL Initiative® is committed to building infrastructure that supports research needs and goals through dissemination of new findings, and sharing collective data and knowledge about opioid misuse and pain. To determine whether HEAL researchers were prepared to fulfill their data sharing requirements, DAI respondents were asked questions about data generation, data sharing, and public access plans. Nearly 80% (82 out of 100) of HEAL studies that submitted a DAI response have begun generating study data as part of their HEAL work. Yet, less than a quarter (24.4%) have made plans to share their data with the scientific community. 28 respondents out of 104 (34.1%) have no plans for data sharing and 34 out of 104 respondents (41.5%) indicated uncertainty around data sharing plans. Seven respondents anticipate being ready to share data between July 2021 and December 2021.

Public Access Plans

When asked if they currently had a public access and data sharing plan, 68 out 104 (65.4%) responded positively ("Yes"). The rest (36 out of 104) said "No" or indicated uncertainty. 77 respondents (74%) expressed no concerns regarding compliance with the NIH HEAL Public Access and Data Sharing Policy. Five out of 104 DAI respondents (4.8%) indicated concerns about their ability to comply, and 22 (21.2%) were unsure whether they had concerns. 57 out of 104 respondents (54.8%) did not want assistance developing public access and data sharing plans.

The respondents who confirmed having plans were prompted to upload their public access and data sharing plans to a database for further analysis. 16 out of the 57 respondents (28%) that did not want assistance with developing public access and data sharing plans did not upload existing plans for further analysis.

Overall, analysis of the public access and data sharing plans submitted by Inventory respondents revealed a lack of consistency in structure and content of the plans. This suggests that some studies may have an unclear understanding of data sharing and public access expectations, and would benefit from additional guidance and assistance, to be more closely aligned with the HEAL Public Access and Data Sharing Policy. Other plans that were submitted offered a more detailed approach with clearly outlined steps describing the journey from data generation to data sharing (or what data sharing means for their HEAL work).

It is important to note that slightly over half of respondents (54.8%) did **not** express interest in assistance in the development of a plan. Twenty-seven of 104 respondents (26%) said they would like assistance developing a data management and sharing plan, while 20 (19%) were not sure.

Metadata Standards

The term metadata is used to designate any data that describes other data. Inventory respondents were asked about metadata to better understand what metadata elements are being used across HEAL. The majority of respondents (57.5%) responded No or

indicated uncertainty ("Not sure") to collecting metadata to describe their HEAL data. Approximately 43% of respondents said they used metadata to describe their HEAL data.

Of the forty-seven respondents who completed Part 2 of the Inventory, six reported that they would like assistance with metadata standards, 9 were not sure, and 32 would not like assistance. Of the 15 who would like assistance or are not sure, 4 reported that they do not collect metadata and 3 were unsure if they collect metadata. Of the 32 respondents that would **not** like assistance with metadata standards, 16 reported that they do not collect metadata and 4 are unsure if they collect metadata. Finally, 12 belong to a data coordinating center (DCC), perhaps explaining why they do not feel the need for outside assistance.

These statistics likely reflect a lack of consistent definition for metadata. Figure 7 illustrates a sequence of responses from an anonymous respondent that would **not** like assistance with metadata standards and reported **not** collecting metadata for their study; however it is clear from their responses to other questions that they do in fact collect metadata. This discrepancy highlights a need for targeted outreach, education, and workshops. Many respondents will still need assistance with understanding metadata-related topics (e.g., definition, standards), despite indicating otherwise.
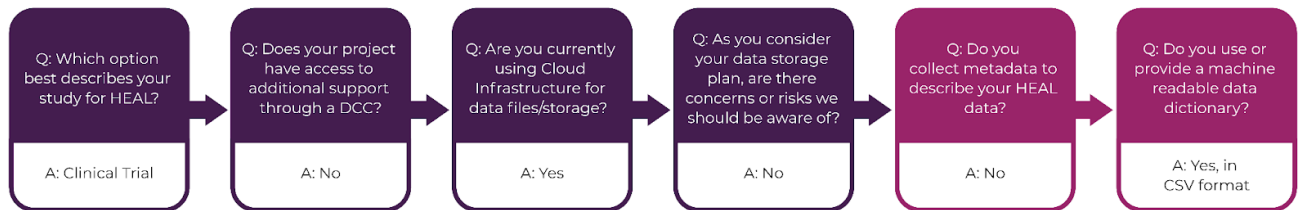

Figure 7. Contradictory DAI responses highlighted in magenta, indicating unfamiliarity with metadata collection.

Conclusion

The HEAL Initiative® is collecting a wide and diverse array of data, and the data described in this report represent a snapshot of HEAL studies during the eight-week response collection period. The insights gleaned from this inventory and subsequent analysis provide a robust foundation for HEAL Stewards' priorities and activities going forward. Four main takeaways emerged from this process:

Version 2.0

- HEAL researchers would benefit from outreach and education surrounding metadata standards.

- Respondents will likely need assistance with repository selection and use, despite lack of stated interest.

- Many respondents will need assistance with HEAL data sharing and public access plans, as well as addressing data sharing concerns.

- Targeted outreach within the HEAL investigator cohort, via research focus areas, research programs, and small business grants with low response rates may be necessary in the future.

For more information about the inventory or standard terminology definitions, please see attached Appendix A and Appendix B.

19

Appendix A: Definitions

- **HEAL Research Focus Areas** constitute:

  - Translation of Research to Practice for the Treatment of Opioid Addiction

  - Enhanced Outcomes for Infants and Children Exposed to Opioids

  - Novel Medication Options for Opioid Use Disorder and Overdose

  - New Strategies to Prevent and Treat Opioid Addiction

  - Preclinical and Translational Research in Pain Management

  - Clinical Research in Pain Management

- **HEAL Research Programs** are led by 12 NIH Institutes and Centers within HEAL research focus areas to find scientific solutions to the opioid crisis. There were 25 HEAL research programs as of the year 2019 (*Research*, 2019).

- **Basic research** is curiosity-driven and describes research meant to enrich scientific knowledge base and general understanding of phenomena or behavior, but without seeking to solve or treat these problems (*Basic vs. Applied Research*, n.d.).

- **Clinical trials** are research studies that evaluate medical, surgical, or behavioral intervention, and are the primary way that researchers assess the safety and effectiveness of new treatments (e.g., a new drug, or medical device) in people (*What Are Clinical Trials and Studies?*, n.d.).

- **Observational study** involves the study of participants in a setting where the independent variable is not under the control of the researcher (Gilmartin-Thomas et al., 2018).

- **Therapeutics, or device development** is the process of creating a concept or an idea for a new device to address an unmet medical need, or improve existing treatment options in a given context. The process may require scientists to invent, refine, and test the devices

- **Preclinical research** can be in vitro or in vivo research, and involves the evaluation of potential therapeutic interventions in cells and animals before testing on human subjects. The goal is to determine the safety of a therapeutic intervention being evaluated and assess the potential to cause serious harm (*Preclinical Research*, 2021).

- **Randomized control trial** is a study design that randomly assigns participants to experimental groups or a control group (*Randomized Controlled Trial*, n.d.).

- **Pragmatic research** is research focused on translation of new data into the clinic and health decision making (Holtrop & Glasgow, 2020).

- **Longitudinal research** is a type of correlational research study that uses continuous or repeated measures to follow particular individuals (participants) over prolonged periods of time, such as weeks, months, or several years (Caruana et al., 2015).

- **Translational research** takes scientific discoveries made in the preceding phases of the research lifecycle (e.g., preclinical or clinical research) and transforms them into new treatments and approaches to safe, impactful, and effective medical care (Henderson, n.d.).

Appendix B: Additional Materials

1. [How to reference the HEAL Initiative® in clinicaltrials.gov entries](#) ⬈

2. [GitHub Documentation](#) ⬈

3. [HEAL Data Asset Inventory Questions](#) ⬈

References

Basic vs. Applied Research. (n.d.). Retrieved August 2, 2021, from

   https://www.utep.edu/couri/dept/about-ug-research/basic-vs-applied-

   research.html ⬈

Buchanan, E. M., Crain, S. E., Cunningham, A. L., Johnson, H. R., Stash, H., Papadatou-

   Pastou, M., Isager, P. M., Carlsson, R., & Aczel, B. (2021). Getting Started Creating Data

   Dictionaries: How to Create a Shareable Data Set. *Advances in Methods and*

   Practices in Psychological Science, 4(1), 2515245920928007.

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal Studies.

   Journal of Thoracic Disease, 7(11), E537–E540.

File Formats for Long-Term Access. (n.d.). Retrieved September 16, 2021, from

   https://libraries.mit.edu/data-management/store/formats/ ⬈

Gilmartin-Thomas, J. F., Liew, D., & Hopper, I. (2018). Observational Studies and Their

   Utility for Practice. *Australian Prescriber, 41*(3), 82–85.

Hayslett, M. (2012). *LibGuides: Metadata for Data Management: A Tutorial: Definition.*

   https://guides.lib.unc.edu/metadata/definition ⬈

Henderson, C. (n.d.). *Clinical & Translational Research. Retrieved August 3, 2021, from*

   https://tracs.unc.edu/index.php/clinical-translational-research ⬈

Holtrop, J. S., & Glasgow, R. E. (2020). Pragmatic Research: An Introduction for Clinical

   Practitioners. *Family Practice, 37*(3), 424–428.

Office of the Commissioner. (2018, March 11). *Step 1: Device Discovery and Concept.*

https://www.fda.gov/patients/device-development-process/step-1-device-discovery-

and-concept

Osborne, B., Cleveland, K., & Bayly, M. (2021). *HEAL Data Call Report.*

Preclinical Research. (2021, July 31). https://www.nature.com/subjects/pre-clinical-studies

Randomized Controlled Trial. (n.d.). Retrieved August 2, 2021, from

https://himmelfarb.gwu.edu/tutorials/studydesign101/rcts.cfm ⬈

Research. (2019). https://heal.nih.gov/research

What Are Clinical Trials and Studies? (n.d.). Retrieved August 3, 2021, from

https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies