

University of New Mexico

## UNM Digital Repository

---

Posters and Presentations

Research and Scholarship

---

2-22-2022

### Data and how to manage it

Lori Sloane

*Health Sciences Library and Informatics Center, University of New Mexico, lsloane@salud.unm.edu*

Follow this and additional works at: <https://digitalrepository.unm.edu/hslic-posters-presentations>

---

#### Recommended Citation

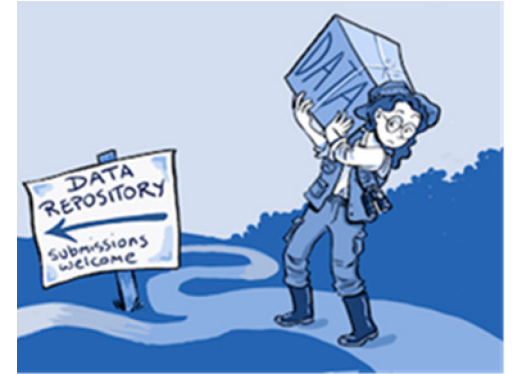
Sloane, Lori. "Data and how to manage it." (2022). <https://digitalrepository.unm.edu/hslic-posters-presentations/84>

This Presentation is brought to you for free and open access by the Research and Scholarship at UNM Digital Repository. It has been accepted for inclusion in Posters and Presentations by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).



## HEALTH SCIENCES LIBRARY & INFORMATICS CENTER

February 15, 2022



---

### MANAGING YOUR DATA

LORI SLOANE, HSLIC DATA MANAGER

Image: Ainsley Seago. "To deposit or not to deposit, that is the question." [doi:10.1371/journal.pbio.1001779](https://doi.org/10.1371/journal.pbio.1001779). Shared under [CC-BY License](#).

# NIH Data Management and Sharing Policy & Related Announcements

---

[NOT-OD-21-013](#) - Final NIH Policy for Data Management and Sharing

[NOT-OD-21-014](#) – Common Elements of an NIH Data Management and Sharing Plan

[NOT-OD-21-015](#) – Allowable Costs for Data Management and Sharing

[NOT-OD-21-016](#) – Selecting a Repository for Data Resulting from NIH-Supported Research

[NOT-OD-21-131](#) – Developing Consent Language for Future Use of Data and Biospecimens

# Logistics

---

1. Effective Date: January 25, 2023
2. Required submission of DMSP **at the time of application**
3. Compliance will be monitored
  - Extramural Awards: The Plan will become a Term and Condition of the Notice of Award.
  - After the end of the funding period, non-compliance with the NIH ICO-approved Plan may affect future funding decisions for the recipient institution
4. Scope: “The DMS Policy applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data.”



# New definition of Scientific Data

---

“The recorded factual material commonly accepted in the scientific community as **of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications.** “

*Scientific data* *do not include*

- *laboratory notebooks*
- *preliminary analyses*
- *completed case report forms*
- *drafts of scientific papers*
- *plans for future research*
- *peer reviews*
- *communications with colleagues*
- *physical objects (such as laboratory specimens)*

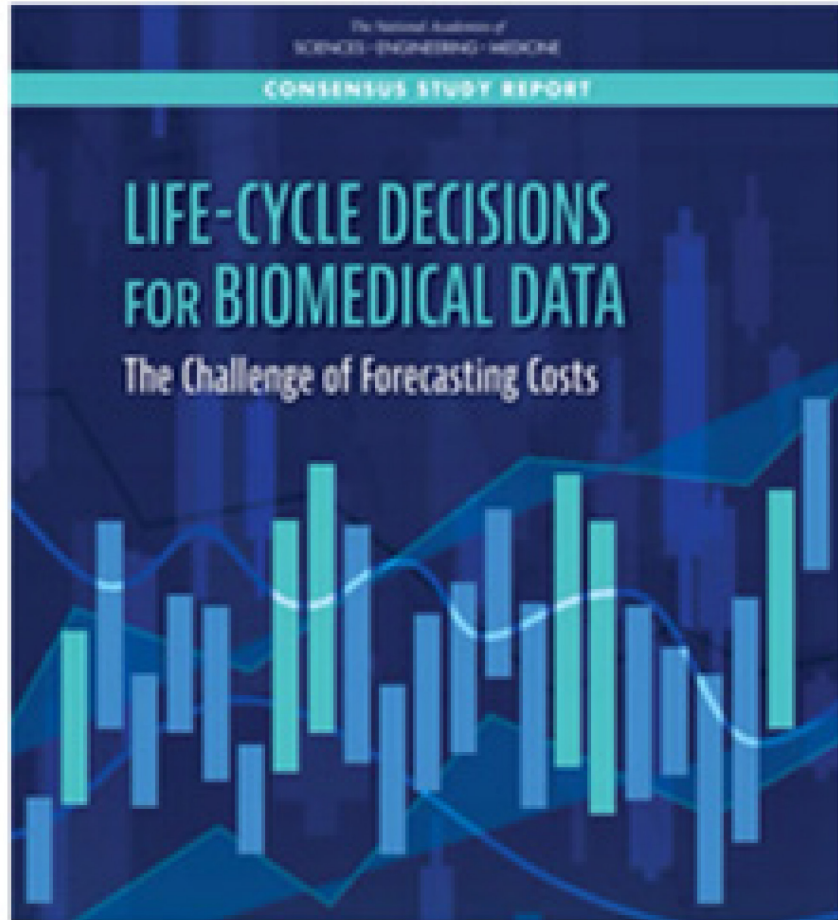
# Costs & Budget

---

“Personnel costs required to perform the types of data management and sharing activities described in the final Supplemental Information are allowable.”

“...funds for these activities must be spent during the performance period, even for scientific data and metadata preserved and shared beyond the award period. NIH funds cannot legally be spent after the award period.”

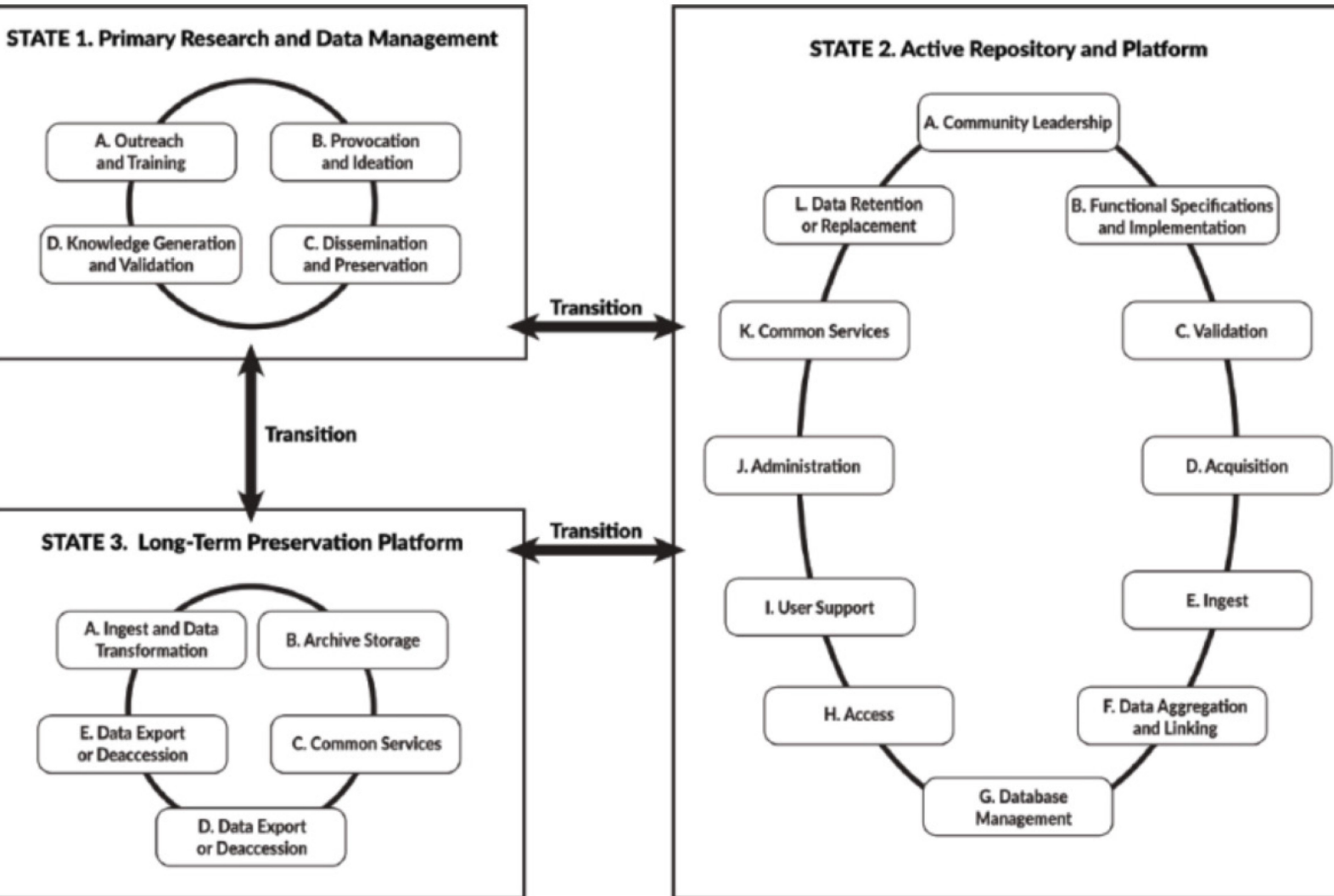
# Forecasting Cost for Preserving, Archiving, and Promoting Access to Biomedical Data



Funded by the National Library of Medicine, this study by the National Academies of Sciences, Engineering, and Medicine developed and demonstrated a **framework for forecasting long-term costs** for preserving, archiving, and accessing various types of biomedical data and estimating potential future benefits to research.

<https://vimeo.com/447601572>

<https://pubmed.ncbi.nlm.nih.gov/33026757/>



## Three data states

- **State 1:** Primary research/data management environment; data are captured and analyzed
- **State 2:** Active repository and platform; data may be acquired, curated, aggregated, accessed, and analyzed
- **State 3:** Long-term preservation platform

# Elements of a Data Management and Sharing Plan

---

## Data type

- Identifying data to be preserved and shared

## Related tools, software, code

- Tools and software needed to access and manipulate data

## Standards

- Standards to be applied to scientific data and metadata

## Data preservation, access, timelines

- Repository to be used, persistent unique identifier, and when/how long data will be available

## Access, distribution, reuse considerations

- Description of factors for data access, distribution, or reuse

## Oversight of data management

- How plan compliance will be monitored/ managed and by whom

# Choosing a data repository

---

“Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse.”

## Desirable characteristics (from [NOT-OD-21-016](#))

Unique persistent identifiers (DOI)	Clear use guidance
Long-term sustainability	Security and integrity
Metadata	Confidentiality
Curation & quality assurance	Common format
Free and easy access	Provenance
Broad and measure reuse	Retention policy

# Additional Considerations for Human Data (even if de-identified)

---

**Fidelity to Consent:** documented procedures to restrict dataset access and use to those that are consistent with participant consent

**Restricted Use Compliant:** documented procedures to communicate and enforce data use restrictions

**Privacy:** documentation to protect human subjects' data from inappropriate access

**Plan for Breach:** measures that include a response plan for detected data breaches

**Download Control:** Controls and audits access to and download of datasets (if download is permitted).

**Violations:** addressing violations of terms-of-use by users and data mismanagement by the repository

**Request Review:** makes use of an established and transparent process for reviewing data access requests



# Selecting a repository: the NIH data sharing landscape

NIH strongly encourages  
subject-specific, open access **Data Sharing Repositories**  
as a first choice.

[https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html)

Datasets up to **2 gigabytes**

## PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.



Datasets up to **20 gigabytes**

## Generalist Repositories

Datasets associated with publications or otherwise and links to PubMed.



High priority datasets, **petabyte-scale**

## Cloud Partners (STRIDES Program)

Store and manage large scale, high priority NIH datasets.



## BMIC Repository list

Trans-NIH  
BioMedical  
Informatics  
Coordinating Committee  
(BMIC)

To help researchers locate an appropriate resource for sharing their data, as well as to promote awareness of resources where datasets can be located for reuse, BMIC maintains lists of several types of data sharing resources:

- Open NIH-supported domain-specific repositories that house data of a specific type or related to a specific discipline;
- Other NIH-supported domain-specific resources, including repositories and knowledgebases, that have limitations on submitting and/or accessing data; and
- Generalist repositories that house data regardless of type, format, content, or subject matter.

[https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html)



re3data is a global registry of research data repositories. The registry covers research data repositories from different academic disciplines. re3data presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data aims to promote a culture of sharing, increased access and better visibility of research data.



1578 Standards	
Terminology Artifact	829
Model/Format	504
Reporting Guideline	226
Identifier Schema	19
<a href="#">VIEW ALL</a>	

1858 Databases	
Repositories	953
Knowledgebases	787
Knowledgebase/Repositories	118
<a href="#">VIEW ALL</a>	

158 Policies	
Journal	94
Funder	23
Project	14
Society	13
<a href="#">VIEW ALL</a>	

# Acceptable reasons to not share

---

- ❖ Informed consent limitations on sharing
- ❖ Existing consent (e.g. for previously collected biospecimens) prohibits sharing
- ❖ Privacy or safety of research participants would be compromised
- ❖ Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure

# Creating your DMP

DMPTool - <https://dmptool.org/>



## Sign in options

Option 1: If your institution is affiliated with DMPTool.

Your institution

- or -

Option 2: If your institution is not affiliated with DMPTool.

Email address

- or -

Option 3: If not affiliated and you need an account.

Create an account

[About](#) [Terms of use](#) [Privacy statement](#) [Accessibility](#) [Github](#) [Contact us](#)



DMPTool is a service of the University of California Curation Center of the California Digital Library  
Copyright 2010-2021 The Regents of the University of California

Version: v3.1.10;







## Descriptive research on COVID-19 materials provided to Medical Advisory Team

[Project Details](#)[Collaborators](#)[Write Plan](#)[Research outputs](#)[Download](#)[Finalize / Publish](#)

### \* Project title

☐ mock project for testing, practice, or educational purposes

### Project abstract

**B** *I*    

A team was assembled at the University of New Mexico Health Sciences Center to help advise the Medical Advisory Team to the Governor of New Mexico and Health Sciences Leadership in April of 2020. An overwhelming amount of information was coming out in the news and literature in regards to the COVID-19

### Research domain

### Project Start

### Project End

 ☐ Research outputs may have ethical concerns

## Select Guidance

To help you write your plan, DMPTool can show you guidance from a variety of organizations.

Select up to 6 organizations to see their guidance.

- ☒ DMPTool
- ☒ University of New Mexico (UNM)

Find guidance from additional organizations below

[See the full list](#)

[Save](#)

# Data collection best practices

---

- ❖ Have a plan developed beforehand and document it in detail
- ❖ Consider who, what, when, where, how you are going to collect your data from the outset
- ❖ Speak to a statistician and consider analysis throughout the process

# Data Collection and Data Management Documentation

---

Population: Teens with depression

Target: 100 subjects

Data collection information needed:

- Demographics, clinical measures, depression scale, MoCA

Resources:

- Data collection:
  - iPads (4)
  - REDCap software (<https://hsc.unm.edu/ctsc/services/informatics/redcap-login.html>)
  - Statistical software
  - Statistical Analysis Plan

Sites: After school programs in NM

- 25 YMCA, NM Teen Project, Camp Fire NM, etc.

Procedure: Meet with students during after school programs, collect data on iPads using REDCap

# Creating instruments/forms/CRFs

---

What is the best method for collecting your data?


- ❖ Single fields
- ❖ Multiple choice/select all
- ❖ Matrix of fields



Best practices:

- ❖ Keep values consistent
- ❖ Use validation whenever possible

# Collecting data: Example

Scenario: Collecting data on comorbidities

 Editing existing Record ID 1

<b>Record ID</b>	1
<b>List comorbidities:</b>	<div><input type="text"/></div>




<https://hsc-redcap-idp.health.unm.edu/my.policy>





# Collecting data: Example

Scenario: Collecting data on comorbidities

 Editing existing Record ID 1

**Record ID** 1

**List comorbidities:** 

**Select comorbidities** 

- ☐ Arthritis
- ☐ Sleep Apnea
- ☐ High Blood Pressure
- ☐ High Cholesterol
- ☐ Type 2 Diabetes
- ☐ Venous Stasis Disease
- ☐ Soft Tissue Infections
- ☐ Other

# Collecting data: Example

Scenario: Data validation

The screenshot shows a configuration interface for a data field. It includes a 'Field Type' dropdown set to 'Text Box (Short Text, Number, Date/Time, ...)', a 'Field Label' text box containing 'Visit Date' with a 'How to use Piping' link, a 'Variable Name' text box containing 'visit\_date' with a note 'ONLY letters, numbers, and underscores' and an unchecked checkbox for 'Enable auto naming of variable based upon its Field Label?', and a 'Validation?' dropdown set to 'Date (Y-M-D)' which is highlighted with a red box. Below the validation dropdown is a 'Minimum' label and an empty text box.

**Field Type:** Text Box (Short Text, Number, Date/Time, ...) ▼

**Field Label** [How to use Piping](#)

Visit Date

**Variable Name** (utilized during data export)

visit\_date














ONLY letters, numbers, and underscores

☐ Enable auto naming of variable based upon its Field Label?

**Validation?** (optional) Date (Y-M-D) ▼

Minimum:

# Documentation: Data Dictionary/Codebook

Instrument: <b>CDASH V 1.1 - Demographics</b> (cdash_v_11_demographics)																
	1	record_id	Record ID	text												
 	2	brthdat	Section Header: Demographics Information What is the subject's date of birth? <i>Record the date of birth using the DD-MM-YYYY format.</i>	text (date_dmy)												
 	3	age	What is the subject's age? <i>Record age of the subject in years.</i>	text												
 	4	dmdat	What is the date of collection? <i>Record the date the demographics data were collected using the DD-MM-YYYY format.</i>	text (date_dmy)												
 	5	sex	What is the sex of the subject? <i>Record the appropriate sex. Collect the sex or gender, as reported by the subject or caretaker. Select one.</i>	radio <table><tr><td>1</td><td>Female</td></tr><tr><td>2</td><td>Male</td></tr><tr><td>3</td><td>Undifferentiated</td></tr><tr><td>99</td><td>Unknown</td></tr></table>	1	Female	2	Male	3	Undifferentiated	99	Unknown				
1	Female															
2	Male															
3	Undifferentiated															
99	Unknown															
 	6	ethnic	What is the ethnicity of the subject? <i>Study participants should self-report ethnicity, with ethnicity being asked about before race. Select one.</i>	radio <table><tr><td>1</td><td>Hispanic or Latino</td></tr><tr><td>2</td><td>Not Hispanic or Latino</td></tr></table>	1	Hispanic or Latino	2	Not Hispanic or Latino								
1	Hispanic or Latino															
2	Not Hispanic or Latino															
 	7	race	What is the race of the subject? <i>Study participants should self-report race, with race being asked about after ethnicity. Check all that apply.</i>	checkbox <table><tr><td>1</td><td>race__1</td><td>Black or African American</td></tr><tr><td>2</td><td>race__2</td><td>American Indian or Alaska Native</td></tr><tr><td>3</td><td>race__3</td><td>Asian</td></tr><tr><td>4</td><td>race__4</td><td>Native Hawaiian or Other Pacific Islander</td></tr></table>	1	race__1	Black or African American	2	race__2	American Indian or Alaska Native	3	race__3	Asian	4	race__4	Native Hawaiian or Other Pacific Islander
1	race__1	Black or African American														
2	race__2	American Indian or Alaska Native														
3	race__3	Asian														
4	race__4	Native Hawaiian or Other Pacific Islander														



```
1 This DATSETNAMereadme.txt file was generated on YYYY-MM-DD by NAME
2 <help text is included in angle brackets, and can be deleted before saving>
3
4
5 GENERAL INFORMATION
6
7 1. Title of Dataset:
8
9 2. Author Information
10     A. Principal Investigator Contact Information
11         Name:
12         Institution:
13         Address:
14         Email:
15
16     B. Associate or Co-investigator Contact Information
17         Name:
18         Institution:
19         Address:
20         Email:
21
22     C. Alternate Contact Information
```

# Medical coding dictionaries

---

Helps categorize medical terms (e.g. diagnoses, symptoms, adverse events) so that they can be analyzed/reviewed

Standardizes data across studies

## **Dictionary/Coding Example:**

❖ ICD-10 codes

Record what medical coding dictionaries, validated instruments or standards will be used

Indicate version of dictionary/instrument/standard used

# Medical coding dictionaries

ICD-10 codes:

## Codes

- C00-C14 Malignant neoplasms of lip, oral cavity and pharynx
- C15-C26 Malignant neoplasms of digestive organs
- C30-C39 Malignant neoplasms of respiratory and intrathoracic organs
- C40-C41 Malignant neoplasms of bone and articular cartilage
- C43-C44 Melanoma and other malignant neoplasms of skin
- C45-C49 Malignant neoplasms of mesothelial and soft tissue
- C50-C50 Malignant neoplasms of breast
- C51-C58 Malignant neoplasms of female genital organs
- C60-C63 Malignant neoplasms of male genital organs
- C64-C68 Malignant neoplasms of urinary tract
- C69-C72 Malignant neoplasms of eye, brain and other parts of central nervous system
- C73-C75 Malignant neoplasms of thyroid and other endocrine glands
- C76-C80 Malignant neoplasms of ill-defined, other secondary and unspecified sites
- C7A-C7A Malignant neuroendocrine tumors
- C7B-C7B Secondary neuroendocrine tumors
- C81-C96 Malignant neoplasms of lymphoid, hematopoietic and related tissue
- D00-D09 In situ neoplasms
- D10-D36 Benign neoplasms, except benign neuroendocrine tumors
- D37-D48 Neoplasms of uncertain behavior, polycythemia vera and myelodysplastic syndromes
- D3A-D3A Benign neuroendocrine tumors
- D49-D49 Neoplasms of unspecified behavior



# Medical coding dictionaries

ICD-10 codes:

Adding new Record ID 5

Event Name: **Event 1 (Arm 1: Arm 1)**

Record ID 5

Last Name

Clinical Measures

Medication

BMI  View equation

Condition  Type to begin searching

econsent

Upload Consent

Form Status

Complete?

[K74] **Fibrosis** and cirrhosis of liver

[J94.1] **Fibro**thorax

[M72] **Fibro**blastic disorders

[N60.2] **Fibro**adenosis of breast

[M72.9] **Fibro**blastic disorder, unspecified

[M85.0] **Fibrous** dysplasia (monostotic)

[N60.3] **Fibros**clerosis of breast

[I42.4] Endocardial **fibro**elastosis

[M35.5] Multifocal **fibro**sclerosis

[M72.8] Other **fibro**blastic disorders

[J63.3] Graphite **fibro**sis (of lung)

[K71.7] Toxic liver disease with **fibro**sis and cirrhosis of liver

[I77.3] Arterial **fibro**muscular dysplasia

[K74.2] Hepatic **fibro**sis with hepatic sclerosis



# Medical coding dictionaries

Drug codes:

**Medication History**

Adding new Record ID 6

Record ID 6

Medication Name  Type to begin searching

Dose

Form Status

Complete?

- [1187316] Tylenol Severe Allergy Oral Product
- [1187331] Tylenol Sinus Congestion and Pain Daytime Pill
- [1296526] Tylenol Chewable Product
- [346199] Tylenol Severe Allergy
- [702405] Tylenol Sinus Congestion and Pain Daytime
- [1187330] Tylenol Sinus Congestion and Pain Daytime Oral Product
- [1187317] Tylenol Severe Allergy Pill
- [202433] Tylenol
- [1187091] Tylenol Cold Relief Nighttime Oral Liquid Product
- [1187315] Tylenol Pill
- [1187503] Tylenol with Codeine Pill
- [1187092] Tylenol Cold Relief Nighttime Oral Product



# REDCap Shared Library

[Return to REDCap](#)

Logged in as **Lori Sloane** (University of New Mexico)

**Keyword search:**

**Search options:**

Language:

Minimum downloads:

Recent additions:

Curated by REDLOC?

**Shared Library**

[Search](#)

**Library Metrics**

[My Activity](#)

[Institution Activity](#)

[Consortium Activity](#)

**REDLOC**

[REDLOC Login](#)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 >>

**Found 2527 results matching your search**


Didn't find what you were looking for? [Suggest a validated instrument for library inclusion](#)

Title	Downloads
▼ <b>FOR TESTING AND DEMONSTRATION PURPOSES ONLY - Nacho Craving Index Survey</b> ★	4490
<b>Details:</b>	
<b>Institution:</b> REDLOC	<a href="#">View as web page</a>
<b>Contact:</b> Brenda Minor	<a href="#">View as PDF</a>
<b>Contact email:</b> <a href="mailto:brenda.l.minor@vumc.org">brenda.l.minor@vumc.org</a>	<input type="button" value="Import into my REDCap project"/>
<b>Submitted by:</b> Brenda Minor	
<b>Description:</b> A fun survey for demonstration and testing only. THIS IS NOT A VALIDATED INSTRUMENT FOR REAL APPLICATIONS.	
<b>Acknowledgement:</b> REDLOC	
<b>Last updated:</b> August 6, 2012	
► Adolescent Outcomes Questionnaire Parent Reported ★	735
► Adolescent Outcomes Questionnaire Self Reported ★	407
► Agitated Behavior Scale ★	1034
► Alcohol Dependence Scale ★	322
► Alcohol Dependence Scale Scored ★	120
► Alcohol Use Disorders Identification Test Audit C Q ★	171
► Alcohol Use Disorders Identification Test Audit C Q Sc ★	59
► ALS Functional Rating Scale Revised ALSFRSR ★	29

# Data standards: CDASH



## Adverse Events:

 **CDASH V 1.1 - Adverse Events**

Editing existing record

**Adverse Events**

**Were any adverse events experienced?**

☐ No  
☐ Yes

reset value

Indicate if the subject experienced any adverse events.

**AE Identifier**

Record unique identifier for each adverse event for this subject.  
<br> <br> Number sequence for all following forms should not duplicate existing numbers for the subject.

**What is the adverse event term?**

Record only one diagnosis, sign or symptom per form (e.g., nausea and vomiting should not be recorded in the same entry, but as two separate entries). See eCRF completion instruction for more information.


**Does the subject have (specific adverse event)?**

☐ No  
☐ Yes

reset value

Please indicate if (specific adverse event) has occurred /is occurring by checking Yes or No.

**What is the date the adverse event started?**

 Today M-D-Y

Record the start date of the adverse event using the MM-DD-YYYY format.

# NIH CDE Repository

<https://cde.nlm.nih.gov/home>

NIH National Library of Medicine

NIH CDE Repository

Search: depression

596 data element results for **depression** Clear All (0.706 secs)

**Classification**

- > GRDR (1)
- > NCI (5)
- > NEI (1)
- > NHLBI (72)
- > NICHD (3)
- > NIDA (2)
- > NINDS (452)
- > NINR (18)
- > NLM (11)
- > ONC (4)
- > PhenX (27)
- > PROMIS / Neuro-QOL (46)
- > RADx-UP (1)
- > Women's CRN (1)

**Registration Status**

- ☐ Standard (46)
- ☐ Qualified (550)

**Data Types**

- ☐ Value List (478)
- ☐ Text (24)
- ☐ Number (89)
- ☐ Externally Defined (1)
- ☐ Date (4)

**Psychiatric depression indicator** Qualified

Indicator of history of **depression**

	Value	Code Name	Code
<b>Steward:</b>	NINDS	No	No
<b>Used By:</b>	NINDS, NHLBI	Yes	Yes
<b>Source:</b>	NINDS	Unknown	Unknown

**Feeling down, depressed, or hopeless in last 2 weeks [Reported.PHQ]** Standard

Score of how much the subject has been feeling down and **depressed** in the past two weeks, as part of the Patient Health Questionnaire (PHQ).

	Value	Code Name	Code
<b>Steward:</b>	NLM	Login to see the val...	Login t...
<b>Used By:</b>	NINDS, ONC, NLM	Login to see the val...	Login t...
<b>Source:</b>	LOINC	Login to see the val...	Login t...

**Depression** Qualified

Never Rarely Sometimes Often Always

<b>Steward:</b>	GRDR
<b>Used By:</b>	GRDR
<b>Source:</b>	GRDR



	A	B	C	D
1	Name	Question Texts	Value Type	Permissible Values
2	Psychiatric depression indicator		Value List	No; Yes; Unknown
3	Feeling down, depressed, or hopeless in last 2 weeks [Reported.PHQ]	Feeling down, depressed, or hopeless?; Over the last 2 weeks how often have you been bothered by feeling down depressed or hopeless?	Value List	0; 1; 2; 3
4	Depression		Externally Defined	
5	Depression	Depression	Value List	1; 0
6	PROMIS - Emotional Distress depression depressed scale		Value List	5; 4; 3; 2; 1
7	Dysphoria - depression		Value List	1; 2
8	Depressed BFI	I see myself as someone who is depressed, blue.	Value List	1; 2; 3; 4; 5
9	Depressed 30D K6	During the past 30 days, about how often did you feel...so depressed that nothing could cheer you up?	Value List	1; 2; 3; 4; 5
10	Patient Health Questionnaire Depression (PHQ) - Depressed score	Over the last 2 weeks how often have you been bothered by feeling down depressed or hopeless?	Value List	0; 1; 2; 3
11	Geriatric Depression Scale (GDS) Long Form - Depression scale		Value List	Mild depressive; Normal; Severe depressive
12	Center for Epidemiologic Studies-Depression Scale (CES-D) - depressed indicator		Value List	0; 1; 2; 3
13	Hamilton Depression Rating Scale (HDRS) - depressed mood indicator		Value List	0; 1; 2; 3; 4
14	Quality of Life - Depression assessment past week scale		Value List	5; 4; 3; 2; 1
15	Psychiatric depression onset date		Date	

# Data quality



Upload or download Data Quality Rules

Execute rules: All All except A&B Clear

Apply to: -- All records --

	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?	Total Discrepancies	Delete rule?
	A	Blank values*	-		Execute	
	B	Blank values* (required fields only)	-		Execute	
	C	Field validation errors (incorrect data type)	-		Execute	
	D	Field validation errors (out of range)	-		Execute	
	E	Outliers for numerical fields (numbers, integers, sliders, calc fields)**	-		Execute	
	F	Hidden fields that contain values***	-		Execute	
	G	Multiple choice fields with invalid values	-		Execute	
	H	Incorrect values for calculated fields	-		Execute	
	I	Fields containing "missing data codes"	-		Execute	
	Add	<div></div> <div>Enter descriptive name for new rule (e.g., Participants below age 18)</div>	<div></div> <div>Enter logic for new rule (e.g., [age] &lt; 18) <a href="#">How do I use special functions?</a></div>	<div></div> <div>Execute in real time on data entry forms ?</div>		

# Quality Assurance & Data Management Example


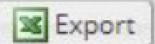

Data reconciliation responsibilities in REDCap:

Staff responsible:

- Research coordinator
- PI

If there is a discrepancy in data from the source documentation, follow this procedure:

1. Error flagged in REDCap
2. Review source documentation
3. Error (if it is not self-evident) to PI
  - a) [a] correct error, [b] irresolvable
4. Discrepancy is approved. Decision is documented below by the PI:

Data Resolution Dashboard		Filters: <div>All status types (1) ▼ All fields and rules ▼ User assigned (all users) or not assigned ▼</div>						
Click button to view data query	Record	Data Quality rule and/or Field	User Assigned	Days Open	First Update	Last Update		
 Export								
 2 comments	<a href="#">1 (#2)</a>	Field: medication (medication)	readk01	0	readk01 (04/30/2018 2:40pm): "Looks incorrect, is this a typo?"	readk01 (04/30/2018 2:40pm): "Fixed typo"		

# Where is the safest place to store my data?

---

UNM HSC Network Drive (H, O or N drives, Sharepoint)

REDCap (production mode)

❖ <https://hsc.unm.edu/ctsc/services/informatics/redcap-login.html>

❖ [UNM Main Campus High Performance Computing](#)

❖ [HSC-250 Systems and Network Security](#)



# Use open formats

---



# Data privacy: REDCap



**Edit Field**

You may add a new project field to this data collection instrument by completing the fields below and clicking the Save button at the bottom. When you add a new field, it will be added to the form on this page. For an overview of the different field types available, you may view the [Field Types video \(4 min\)](#).

Field Type: Text Box (Short Text, Number, Date/Time, ...)

Field Label: Email [How to use Piping](#)

Action Tags / Field Annotation (optional)

[Learn about Action Tags](#) or [using Field Annotation](#)

Variable Name (utilized during data export): email ☐ Enable auto naming of variable based upon its Field Label? (ONLY letters, numbers, and underscores)

Validation? (optional): Email

- OR -

Enable searching within a biomedical ontology? -- choose ontology to search --

Required?\* ☒ No ☐ Yes  
\* Prompt if field is blank

**Identifier?** ☐ No ☒ Yes  
Does the field contain identifying information (e.g., name, SSN, address)?

Custom Alignment: Right / Vertical (RV)

Field Note (optional):

Small reminder text displayed underneath field

Save Cancel

**De-identification options (optional)**

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

**Known Identifiers:**

- ☐ Remove all tagged Identifier fields (tagged in Data Dictionary)
- ☐ Hash the Record ID field (converts record name to an unrecognizable value)

**Free-form text:**

- ☐ Remove unvalidated Text fields (i.e. Text fields other than dates, numbers, etc.)
- ☐ Remove Notes/Essay box fields

**Date and datetime fields:**

- ☐ Remove all date and datetime fields

- OR -

- ☐ Shift all dates by value between 0 and 364 days (shifted amount determined by algorithm for each record) [What is date shifting?](#)

[Deselect all options](#)



# Data privacy: HIPAA Compliance

## Logging

[Download entire logging record to Microsoft Excel \(CSV\)](#)

This module lists all changes made to this project, including data exports, data changes, and the creation or deletion of users.

Filter by event: All event types (excluding page views) ▼

Filter by user name: All users ▼

Filter by record: All records ▼

Displaying events (by most recent): 1 - 100 ▼

Filter by time range from  to



Time / Date	Username	Action	List of Data Changes OR Fields Exported
10/30/2017 6:54pm	kdg294	Updated Record 18	item_name = 'None found', data_collection_instrument_worksheet_complete = '2'
10/30/2017 6:54pm	kdg294	Updated Record 18	dataset_information_for_data_collection_instrument_complete = '2'
10/30/2017 6:53pm	kdg294	Updated Record 17	data_collection_instrument_worksheet_complete = '2'
10/30/2017 6:53pm	kdg294	Updated Record 17	item_name = 'None found', data_collection_instrument_worksheet_complete = '0'
10/30/2017 6:53pm	kdg294	Updated Record 17	dataset_information_for_data_collection_instrument_complete = '2'
10/30/2017 6:52pm	kdg294	Updated Record 16	item_name = 'None found', data_collection_instrument_worksheet_complete = '2'
10/30/2017 6:52pm	kdg294	Updated Record 16	dataset_information_for_data_collection_instrument_complete = '2'
10/30/2017 6:52pm	kdg294	Updated Record 15	item_name = 'None found', data_collection_instrument_worksheet_complete = '2'
10/30/2017 6:51pm	kdg294	Updated Record 15	dataset_information_for_data_collection_instrument_complete = '2'

# Data Storage Documentation Example

## Data Storage Procedure and Workflow

### Data Storage During Collection:

- ❖ Data stored within REDCap
- ❖ Lab data uploaded to REDCap File Repository with file name **LabName\_DataType\_YYYYMMDD**

### Data Storage Before Analysis:

- ❖ REDCap data exported to H:/ drive **Project Folder** in CSV raw format with file name:
  - ❖ **ProjectName\_MasterDataset\_YYYYMMDD**
- ❖ **ProjectName\_MasterDataset\_YYYYMMDD** is cleaned and prepared for analysis
- ❖ Import into **SPSS v25** and saved to **Project Folder → Data Analysis Folder** with file name:
  - ❖ **ProjectName\_DataSubset\_YYYYMMDD**

### Data Storage During Analysis:

- ❖ All changes to analysis dataset including output logs and monitor files will be stored in **Project Folder → Data Analysis Folder** with the file name:
  - ❖ **DataSubset\_AnalysisType\_FileType\_YYYYMMDD**

# Data management and storage

---

Document where your data lives throughout your study

Who is responsible for the data at different stages (e.g., collection vs. analysis)?

How safe is your data in each location? How is that safety guaranteed?

Ensure that your file systems are planned out ahead of time and understandable

Back up data and save in open formats if possible

# Data transfer @ HSC

---

## Requesting HSC Central IT Transfer:

The default DUA language requires that transfers be handled by HSC Central IT. Investigators can request secure data transfer services by creating a self-service help ticket and attaching the request number with their documentation confirming that HSC Central IT security transfer has been requested.

### 1. File transfers

1. Submit a request through <http://help.health.unm.edu> and select -- "HSC Central IT secure transfer"

### 2. Data Entry into an External Web Application or Portal

1. Submit a request through <http://help.health.unm.edu> and select – "Request Security Review"

For questions concerning IT Security issues please refer to <http://hscsecurity.unm.edu/>.

<https://hsc.unm.edu/financialservices/preaward/ancillary-agreements/data-agreements/>

# Summary

---

Data management should be discussed **BEFORE** beginning a study and **any documentation** is better than no documentation



## HEALTH SCIENCES LIBRARY & INFORMATICS CENTER

February 15, 2022



---

### MANAGING YOUR DATA

LORI SLOANE, HSLIC DATA MANAGER

Image: Ainsley Seago. "To deposit or not to deposit, that is the question." [doi:10.1371/journal.pbio.1001779](https://doi.org/10.1371/journal.pbio.1001779). Shared under [CC-BY License](#).