

Training on FAIR Data Submission Guidance for EDRN

Heather Kincaid NASA JPL

7/2/2025

Data Sharing and FAIR *(Findable, Accessible, Interoperable and Reusable)* Data for EDRN NCI Requirements

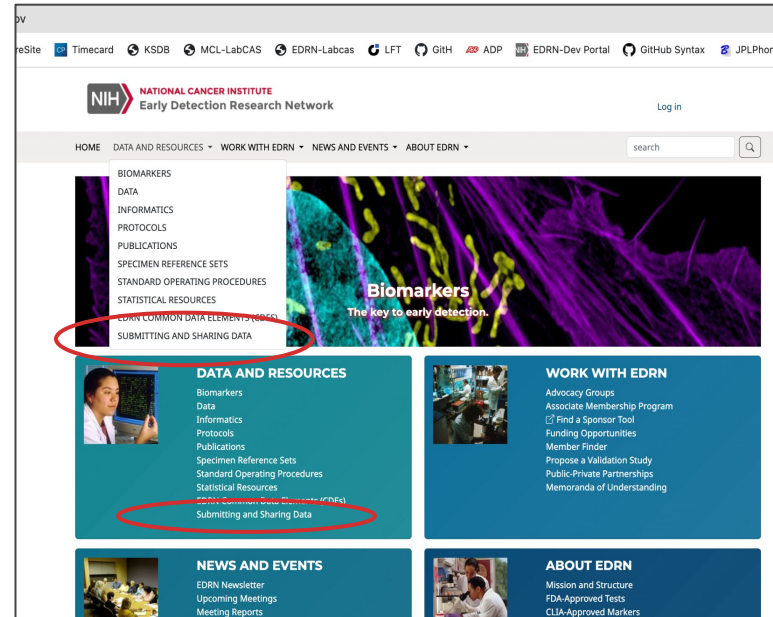
A primary goal of EDRN data collection is to ensure the reusability of the data by groups beyond those who originally collected it.

Per the [EDRN External Data Sharing and Reuse Policy Cycle V](#), data must be made available for public use.

To help EDRN sites transition to FAIR principles

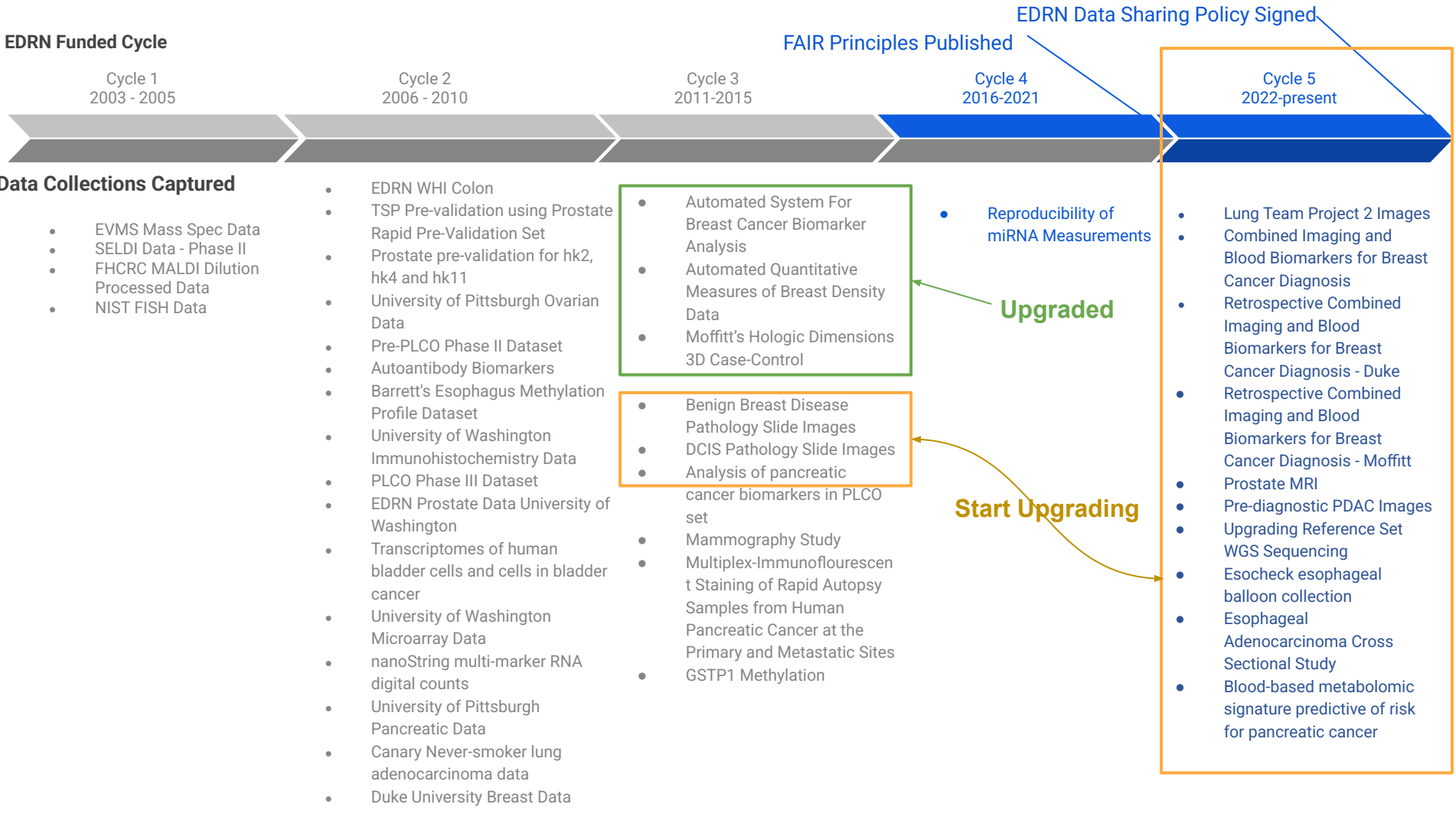
Drafted and Approved by the Data Sharing and Informatics Subcommittee, – now posted to EDRN Public Portal:

- [FAIR Data Submission Guidance for EDRN](#)
- [EDRN Roles for Making Data FAIR](#)
- [How to Submit New Data](#)
- [How to Update Existing Data](#)
- [Updated LabCAS Help Documentation](#)



<https://edrn.nci.nih.gov/>

EDRN Funded Cycle

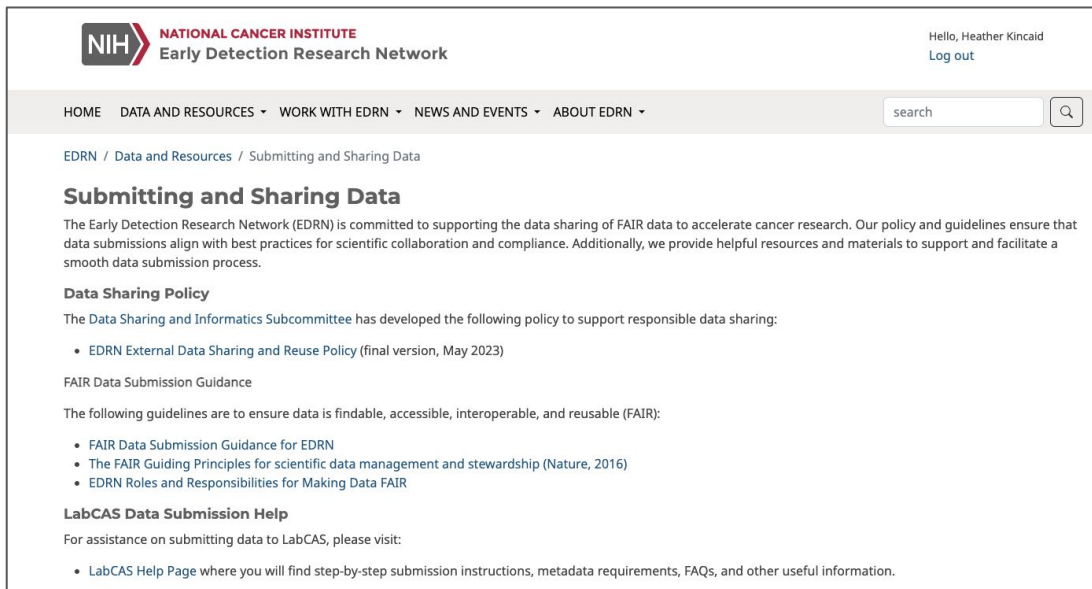


Essential Steps for Data Providers

1. Review EDRN Submitting and Sharing Data Page
2. Review the FAIR Data Submission Guidance for EDRN
3. Submit New Data Procedure

or

Update Existing Data Procedure



The screenshot displays the EDRN website interface. At the top, the NIH logo and 'NATIONAL CANCER INSTITUTE Early Detection Research Network' are visible. A user greeting 'Hello, Heather Kincaid' and a 'Log out' link are in the top right. A navigation bar includes links for HOME, DATA AND RESOURCES, WORK WITH EDRN, NEWS AND EVENTS, and ABOUT EDRN. A search bar is located on the right. The main content area shows the breadcrumb 'EDRN / Data and Resources / Submitting and Sharing Data'. The section title is 'Submitting and Sharing Data'. The text states: 'The Early Detection Research Network (EDRN) is committed to supporting the data sharing of FAIR data to accelerate cancer research. Our policy and guidelines ensure that data submissions align with best practices for scientific collaboration and compliance. Additionally, we provide helpful resources and materials to support and facilitate a smooth data submission process.' Below this is the 'Data Sharing Policy' section, which states: 'The Data Sharing and Informatics Subcommittee has developed the following policy to support responsible data sharing:' followed by a bullet point: '• EDRN External Data Sharing and Reuse Policy (final version, May 2023)'. The 'FAIR Data Submission Guidance' section follows, stating: 'The following guidelines are to ensure data is findable, accessible, interoperable, and reusable (FAIR):' followed by three bullet points: '• FAIR Data Submission Guidance for EDRN', '• The FAIR Guiding Principles for scientific data management and stewardship (Nature, 2016)', and '• EDRN Roles and Responsibilities for Making Data FAIR'. The 'LabCAS Data Submission Help' section states: 'For assistance on submitting data to LabCAS, please visit:' followed by a bullet point: '• LabCAS Help Page where you will find step-by-step submission instructions, metadata requirements, FAQs, and other useful information.'

Submitting and Sharing Data

<https://edrn.nci.nih.gov/>

The screenshot shows the NIH Early Detection Research Network homepage. The top navigation bar includes 'HOME', 'DATA AND RESOURCES', 'WORK WITH EDNRN', 'NEWS AND EVENTS', and 'ABOUT EDNRN'. The 'DATA AND RESOURCES' menu is open, showing options like 'BIOMARKERS', 'DATA', 'INFORMATICS', 'PROTOCOLS', 'PUBLICATIONS', 'SPECIMEN REFERENCE SETS', 'STANDARD OPERATING PROCEDURES', 'STATISTICAL RESOURCES', 'EDRN COMMON DATA ELEMENTS (CDEs)', and 'SUBMITTING AND SHARING DATA'. The 'SUBMITTING AND SHARING DATA' option is circled in red. Below the navigation bar, there is a large banner for the 'Early Detection Research Network' with the tagline 'Accelerative discovery, development, and validation of biomarkers for cancer risk and early detection.' Below the banner, there are four main sections: 'DATA AND RESOURCES', 'WORK WITH EDNRN', 'NEWS AND EVENTS', and 'ABOUT EDNRN'. The 'DATA AND RESOURCES' section is highlighted with a red circle around the 'SUBMITTING AND SHARING DATA' link. The 'WORK WITH EDNRN' section lists various programs and opportunities. The 'NEWS AND EVENTS' section lists newsletters, meetings, and reports. The 'ABOUT EDNRN' section lists mission, structure, and other information.

The screenshot shows the 'Submitting and Sharing Data' page on the NIH Early Detection Research Network website. The page title is 'Submitting and Sharing Data'. The main content area includes a paragraph about the EDNRN's commitment to supporting data sharing of FAIR data. Below this, there is a section titled 'Data Sharing Policy' which states that the Data Sharing and Informatics Subcommittee has developed a policy to support responsible data sharing. This section lists the 'EDRN External Data Sharing and Reuse Policy (final version, May 2023)' and 'FAIR Data Submission Guidance'. The 'FAIR Data Submission Guidance' section lists the following guidelines: 'The following guidelines are to ensure data is findable, accessible, interoperable, and reusable (FAIR):' followed by three bullet points: 'FAIR Data Submission Guidance for EDNRN', 'The FAIR Guiding Principles for scientific data management and stewardship (Nature, 2016)', and 'EDNRN Roles and Responsibilities for Making Data FAIR'. Below this, there is a section titled 'LabCAS Data Submission Help' which states 'For assistance on submitting data to LabCAS, please visit:' followed by a bullet point: 'LabCAS Help Page where you will find step-by-step submission instructions, metadata requirements, FAQs, and other useful information.' Red arrows point from the 'SUBMITTING AND SHARING DATA' link in the first screenshot to the 'Submitting and Sharing Data' page title, and from the 'LabCAS Help Page' link to the 'LabCAS Help Page' text.

Review the FAIR Data Submission Guidance for EDRN

The process is broken into five primary sections:

1. Core Metadata
2. De-identification
3. Data to Upload
4. Organization of Data
5. Review and Verification

Need help? Contact JPL at any time.

The screenshot shows the NIH National Cancer Institute Early Detection Research Network (EDRN) website. The header includes the NIH logo and the text "NATIONAL CANCER INSTITUTE Early Detection Research Network". A navigation bar contains links: HOME, DATA AND RESOURCES, WORK WITH EDRN, NEWS AND EVENTS, and ABOUT EDRN. A search bar is located on the right. The main content area is titled "FAIR Data Submission Guidance for EDRN" and includes the version (1.0.1) and date (2025-1-1). The text explains the purpose of the guidance and lists the five primary sections: 1.0 Core Metadata to Support Definition, Accessibility, and Structure of the Data; 1.1 Metadata for Collections, Datasets, and Files; 2.0 De-identification of Data; 3.0 Data to Upload; and 3.1 ReadMe File, Ancillary Data, Data Dictionaries and Other Information. The 3.2 Data Files section is partially visible at the bottom.

NIH NATIONAL CANCER INSTITUTE
Early Detection Research Network

Log in

HOME DATA AND RESOURCES WORK WITH EDRN NEWS AND EVENTS ABOUT EDRN

EDRN / Data and Resources / Submitting and Sharing Data / FAIR Data Submission Guidance for EDRN

FAIR Data Submission Guidance for EDRN

Version: 1.0.1
Date: 2025-1-1

To align with the FAIR principles outlined in the EDRN Data Sharing Policy, the EDRN has developed a set of minimal requirements for submitting data to LabCAS, EDRN's biomarker data commons repository.

A primary goal of EDRN data collection is to ensure the reusability of the data by groups beyond those who originally collected it. Per the EDRN Data Sharing Policy, data must be made available for public use. This includes providing sufficient metadata and documentation to help users understand the data's configuration and structure. Below is guidance on the critical metadata that should accompany your data submission. Additionally, supplemental documents and readme files can be included to support the enhanced use of the data.

1.0 Core Metadata to Support Definition, Accessibility, and Structure of the Data

Metadata is critical to support the discoverability, interpretability, and usability of the data. LabCAS organizes data into Collections, Datasets, and Files, each with its own set of minimal metadata requirements. Additional metadata is also defined for various assay types. Those metadata are coordinated as Common Data Elements by research groups and should be added to increase the usability of the data.

1.1 Metadata for Collections, Datasets, and Files

This Metadata Check List details the required and optional metadata for Collections, Datasets, and Files for LabCAS data submission. For more comprehensive information, please refer to the EDRN Data Model.

2.0 De-identification of Data

All data must be de-identified at your site before uploading to LabCAS. Additional guidance may be available in your study's SOP.

For Reference Sets (LTP2, PMRL, and BRSL), follow the Imaging Data Transfer SOP provided by the DMCC.

As described in the Metadata Check List, De-identification Method (Safe Harbor, or Expert Review) is required metadata for data submission. Please refer to Health and Human Services - Methods for De-identification of PHI for more information.

3.0 Data to Upload

This section details best practices for uploading data.

3.1 ReadMe File, Ancillary Data, Data Dictionaries and Other Information

Methodology details should be included as part of the metadata. You can also include supplemental information explaining the algorithms and computations applied to the raw data. Additional data, such as ancillary data or clinical records, may also be uploaded to LabCAS as supporting documentation. Examples of supporting files include:

- ReadMe file: A text file explaining your data for easier understanding and reuse.
- Standard Operating Procedures (SOPs): Any SOPs followed during your study.
- Ancillary Data: Clinical or other data captured during the study.
- Data Dictionaries: Definitions of any ancillary data.

3.2 Data Files

To support reproducibility and facilitate robust analyses by external researchers, each data submission should include the applicable core data components. Please note that summary or aggregated data alone is insufficient; raw data and relevant clinical data are essential for meaningful reanalysis.

- 3.2.1 Raw Data Files
Raw data files are the unprocessed outputs directly obtained from the instruments or data collection processes. Including raw data ensures that downstream analyses can be reproduced and verified. The raw data should be:
 - Unaltered and Complete: All data generated by the instruments or data collection processes should be included without modifications, filtering, or aggregation.
 - In Original Format: Whenever possible, submit raw data in the original file format generated by the instrument (e.g., .fastq for sequencing data, .dcm for imaging data, .csv for sensor data). This helps maintain fidelity and compatibility for future reprocessing.
 - Accompanied by Essential Parameters: Ensure any instrument parameters or settings (e.g., calibration details, machine specifications) are documented, either within the metadata or as a separate parameter file.

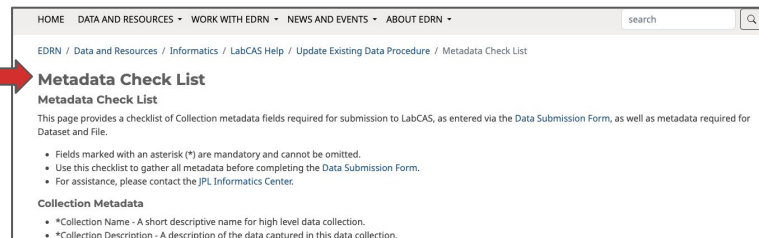
Core Metadata to Support Definition, Accessibility, and Structure of the Data

Steps:

- Gather Metadata
- Submit via the FAIR Submission Process

Key Required Fields to Support Reusability:

- Intuitive name for the data collection
- Description of the data collection
- Principal Investigator and Data Custodian
- EDRN Protocol (provided by DMCC)
- Instrument(s) used
- Method details
- Results and conclusions
- Description of data structure
- PubMed ID(s) (if applicable)
- Links to data and resource name (if data is linked, not deposited in LabCAS)



HOME DATA AND RESOURCES • WORK WITH EDRN • NEWS AND EVENTS • ABOUT EDRN

EDRN / Data and Resources / Informatics / LabCAS Help / Update Existing Data Procedure / Metadata Check List

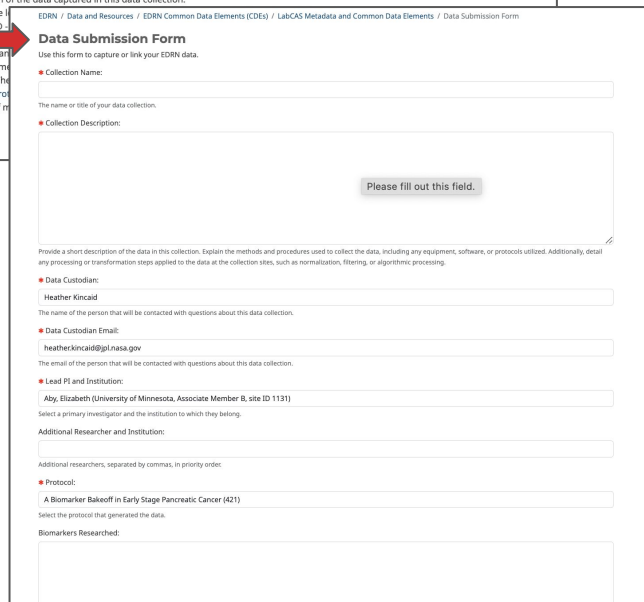
Metadata Check List

This page provides a checklist of Collection metadata fields required for submission to LabCAS, as entered via the Data Submission Form, as well as metadata required for Dataset and File.

- Fields marked with an asterisk (*) are mandatory and cannot be omitted.
- Use this checklist to gather all metadata before completing the Data Submission Form.
- For assistance, please contact the JPL Informatics Center.

Collection Metadata

- *Collection Name - A short descriptive name for high level data collection.
- *Collection Description - A description of the data captured in this data collection.
- *EDRN Lead PI Name/Lead PI ID - The name and ID of the lead principal investigator.
- *EDRN Institution Name/Institution ID - The name and ID of the institution.
- *Data Custodian - The name and email of the data custodian.
- *AdditionalInvolvedResearchers - Name and email of additional researchers.
- *EDRN Protocol Name/Protocol ID - The name and ID of the EDRN protocol.
- *Discipline - A domain or grouping of research.



EDRN / Data and Resources / EDRN Common Data Elements (CDEs) / LabCAS Metadata and Common Data Elements / Data Submission Form

Data Submission Form

Use this form to capture or link your EDRN data.

• Collection Name:

The name or title of your data collection.

• Collection Description:

Please fill out this field.

Provide a short description of the data in this collection. Explain the methods and procedures used to collect the data, including any equipment, software, or protocols utilized. Additionally, detail any processing or transformation steps applied to the data at the collection sites, such as normalization, filtering, or algorithmic processing.

• Data Custodian:

Heather Kincaid

The name of the person that will be contacted with questions about this data collection.

• Data Custodian Email:

heatherkincaid@jpl.nasa.gov

The email of the person that will be contacted with questions about this data collection.

• Lead PI and Institution:

Aby, Elizabeth (University of Minnesota, Associate Member II, site ID 1131)

Select a primary investigator and the institution to which they belong.

Additional Researcher and Institution:

Additional researchers, separated by commas, in priority order.

• Protocol:

A Biomarker Bakeoff in Early Stage Pancreatic Cancer (421)

Select the protocol that generated the data.

Biomarkers Researched:

Options for Sharing Data in Alignment with FAIR Principles

Deposit your data directly in LabCAS to ensure your dataset is:

- Citable
- Findable
- Accessible
- Reusable

if needed request a persistent identifier (DOI) to link to your publication

OR

If your data is already deposited or will be deposited in a public NIH/NCI-supported FAIR repository:

- Submit metadata only to LabCAS so it can be linked and made Findable and Accessible.

De-Identification of Data

All data must be de-identified at your site before uploading to LabCAS.

Data to Include in Upload

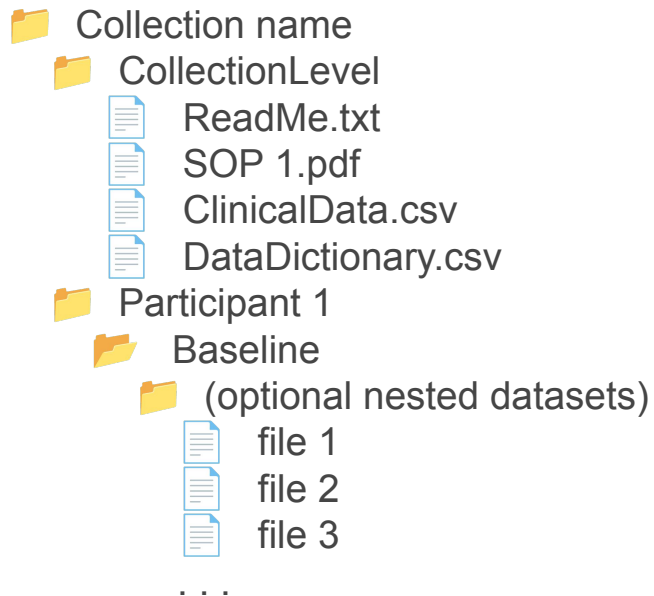
It's not *just* the data files—include supporting materials to make your data reusable:

- **ReadMe File:** A text file explaining your dataset, including supplemental details on algorithms or computations applied to the raw data—for easier understanding and reuse.
- **Standard Operating Procedures (SOPs):** Any SOPs followed during your study.
- **Ancillary Data:** Clinical or other relevant data captured during the study.
- **Data Dictionaries:** Definitions and descriptions of ancillary data elements.
- **Any Other Useful Documentation:** Additional materials that support interpretation or reuse of the data.

Data Organization

Match the logical organization of your data.

For example:





Review and Validation of Data in LabCAS

During upload to LabCAS required metadata will be validated.

Site and/or assigned domain expert must:

- Login to LabCAS
- Review and validate data in LabCAS.
- Ensure accuracy of metadata and data capture.
- Confirm that the data is usable and understandable by other researchers.

This step is ***critical*** to ensure the data can be shared and reused by the broader research community.

Example of FAIR Data Collection in LabCAS

<https://edrn-labcas.jpl.nasa.gov/labcas-ui/>

Automated System For Breast Cancer Biomarker Analysis - Collection 1

Collection Details

Collection Name:

Automated System For Breast Cancer Biomarker Analysis - Collection 1

Collection Description:

180 case-control pairs; 2D mammograms in both for presentati... [More](#)

Instrument:

General Electric (GE) Senographe 2000D full field digital mammography (FFDM), 2D units

Lead PI:

John Heine (3734)

Data Custodian:

Erin Fowler

Data Custodian Email:

Erin.Fowler@moffitt.org

Institution:

H. Lee Moffitt Cancer Center and Research Institute Inc. (816)

Protocol Name:

[An Automated System for Breast Cancer Biomarker Analysis](#)

Discipline:

Radiology

Data Category:

Mammography

Organ:

Breast (1)

Collaborative Group:

Breast and Gynecologic

Method Details:

Approach: all collections were developed to make breast dens... [More](#)

Results And Conclusion Summary:

Select findings from studies using these collections are bri... [More](#)

Delivery Status:

Complete

Data Structure Description:

R01CA166269 breast density and breast cancer study (2013-2017). Hologic Selenia 2D FFDM and Hologic Dimensions 2D/3D digital breast tomosynthesis (DBT) units operating in 2D mode. 319 case-control pairs of women 18 and over. 2D mammograms in for presentation and for processing representations. Demographical, clinical and pathological data approx 50 CDE. This dataset includes study images with a matched design, featuring case and control images for analysis. Case images represent the unaffected breast from the case group, while control images correspond to the ipsilateral breast of matched controls for the case's unaffected breast. The images include the craniocaudal (CC) view and are available in various types, including raw data for processing, processed data for presentation, C-View images, and DBT volumetric images. Key attributes are captured in the image file names, such as the case or control status (indicated as "C" for case and "N" for control) and match pair information, where the combination of case/control status and pair number creates a unique Study ID. Additionally, the file names provide two-field image information, detailing the mammography technology used, data format, image type, image view, laterality, breast orientation, and flags for extra views or special images. For example, in Case 0683, file names include: "C0683_MG_DAT_LCC" for raw data, "C0683_MG_PRO_LCC" for processed data, "C0683_BT_SYN_LCC" for C-View images, and "C0683_BT_VOL_LCC" for DBT volumetric images. Similarly, for Control 0683, file names include: "N0683_MG_DAT_LCC" for raw data, "N0683_MG_PRO_LCC" for processed data, "N0683_BT_SYN_LCC" for C-View images, and "N0683_BT_VOL_LCC" for DBT volumetric images. This structure ensures comprehensive metadata for accurate and detailed analysis. See pdf slides under the collection level files for more details.

Data Capture Start Date:

2013

Data Capture End Date:

2017

FAIRAlignment Status:

Data Ready

Pub Med ID:

[18373863,20095250,21080916,21371912,21474058,21971260,23289950,24320473,24207013,25018067,25652480,30523909,30545682,30525207,33304615,37907569](#)

Species:

Homo sapiens

Data Disclaimer:

Data and information released from the National Cancer Insti... [More](#)

Reference URLLink:

Reference URLDescription:

id:

Automated_System_For_Breast_Cancer_Biomarker_Analysis

Collection Level Files

Explore or Download

An Automated System for Breast Cancer Biomarker Analysis Log 2024.xls

Clinical Validation Center at Moffitt Cancer Center- Mammography Case-Control CollectionsOverview.pdf

GE Data Dictionary 447.xlsx

README.pdf

screening_protocol.pdf

Datasets in this Collection

Explore or Download

Add Folder Grouping Heirarchy:

Default

+

C0001

★

+

C0002

★

+

C0007

★

+

C0008

★

+

C0009

★

+

C0010

★

+

C0011

★

+

C0012

★

+

C0014

★

+

C0016

★

+

C0019

★

+

C0020

★

+

C0021

★

+

C0022

★

+

C0023

★

+

C0024

★

+

C0025

★

+

C0026

★

+

C0027

★

+

C0029

★

+

C0030

★

+

C0031

★

+

C0032

★

+

C0033

★

+

C0035

★

+

C0036

★

+

C0037

★

Step by step process and specific site training

LabCAS Help

LabCAS is EDRN's Cancer Biomarker Data Commons, supporting data storage, management, and sharing for biomarker discovery. LabCAS ensures data is reusable, publicly accessible, and aligned with the [EDRN Data Sharing Policy](#) and [EDRN FAIR Submission Guidance](#). Learn more on the [About LabCAS](#) page.



SUPPORT RESOURCES

The following support resources are comprehensive and helpful:

- Access LabCAS
- Submit New Data
- Update Existing Data
- Cite Data and Obtain DOI
- FAIR Data Submission Guidance for EDRN
- Data Model and FAIR Alignment
- EDRN Data Sharing Policy
- Quick Reference Guides



TRAINING

The following materials can enhance your LabCAS expertise:

- Training Materials

Need Help?

Contact the JPL Informatics Center via email.

[EDRN](#) / [Data and Resources](#) / [Informatics](#) / [LabCAS Cancer Biomarker Data Co](#)

LabCAS Quick Reference Guides

- [Requesting a LabCAS Account](#)
- [Requesting an Aspera Account](#)
- [Uploading Data](#)
- [Downloading Data](#)
- [LabCAS Users Guide](#)
- [LabCAS DICOM Viewer Users Guide](#)
- [Metadata Checklist](#)
- [Collection Metadata Submission Form](#)
- [Verify Completeness and FAIR Alignment](#)

Questions?

Contact Info: Heather Kincaid - heather.kincaid@jpl.nasa.gov