Data Sharing and Informatics Subcommittee
Monday, February 26, 2024

Present:

- NASA/JPL: Dan Crichton, Sean Kelly, Heather Kincaid
- Arizona State University: Ji Qiu
- Boston University: Absent
- EVMS: Julius Nyalwidhe
- DMCC: Greg Warnick, Ziding Feng
- Johns Hopkins: Zhen Zhang
- NCI: Guillermo Marquez, Christos Patriotis
- PNNL: Tao Liu
- University of North Carolina: Kristen Anton
- University of Texas MD Anderson Cancer Center: Absent

Action items from this call:

1. JPL to 1) populate the grid of Roles and Responsibilities for FAIR-based data presented on the call to the Public Portal, 2) document more, and 3) promote investigator trainings to ensure that NCI policies are followed
2. JPL to follow-up with lead PI of each project listed in LabCAS Holdings Nov 2023 spreadsheet rows 25-35 to determine if the projects can be considered public.

Agenda:

1. **Review roles and responsibilities for FAIR-based data**
   Discussion:
   JPL presents a slide of a grid mapping the data

1. Review roles and responsibilities for FAIR-based metadata

Responsibility	Investigators	JPL	NCI	DMCC	Future Needs
(F)indable	Submit data to JPL in a timely manner along with metadata; verify completeness	Capture, ingest, and post data and metadata so it is accessible; catalog metadata for search	Enforce that data is delivered per the data management plan	Data tracking for EDRN studies (EDRN DMCC Level 1,2&3 data)	Require **all data is delivered per a data management plan including all data** (completeness)
(A)ccessible	Identify sensitivity and sharing of data (e.g., public vs study)	Configure access to data based on sensitivity of data; ensure it can be accessed by the right groups	Enforce that data is made public per the NCI policy.	Notify JPL of data and metadata access (EDRN DMCC Level 1,2&3 data)	Require that **data be made public** per NCI policies
(I)nteroperable	Structure data so it can be integrated with other data; Validate data when posted to LabCAS	Provide API and tool access for data integration; support search; Link with other data via metadata	Enforce that meets specific structure requirements, where applicable	Support statistical analysis of validation study data, including structure of data	Enforce that **data standards are followed** including providing appropriate metadata and structuring data
(R)eusable	Structure data for reuse Validate data in LabCAS Need to deliver any required software Provide documentation	Provide ability to run central pipelines to support repeatable processing of derived data	Enforce sites meet usability process/requirements	Support analysis of validation study data	Need a process to ensure **usability of data** and metadata (e.g., peer review to validate quality, a reusability test); capture documentation and software

   - Based on 2016 Nature paper "The FAIR Guiding Principles for scientific data management and stewardship"https://www.nature.com/articles/sdata201618
   - JPL suggests populating this slide to the Public Portal, document more, and promote investigator trainings to ensure that NCI policies are followed. Ziding supports this direction. Dan asks for suggestions.

2. **Public access to EDRN datasets**
   Discussion:

Heather presents a catalogue of LabCAS data collections that are not public.
Christos suggests that a note be included to indicate that the project is continuing, or accumulating, its data collection. Other metadata fields should be considered; suggestions are appreciated.
JPL **TODO** will follow-up with lead PI of each project listed in rows 25-35 to determine if the projects can be considered public.

3. Linking data to external repositories

- LabCAS enables direct linkage to datasets deposited in external repositories, such as genomic data in dbGAP
- Data Collections can be created describing the data with links to the external data and relevant publication via the EDRN Portal Data Submission Form
- EDRN's definition of FAIR (Findable, Accessible, Interoperable, and Reusable) metadata underpins the integration of data across the knowledge system.
- By submitting FAIR metadata to JPL alongside external repositories, we support a federated system that enhances data discoverability and connectivity.
- This approach ensures that the research community can seamlessly explore EDRN biomarker research and access data hosted in the most appropriate repositories, including those designated for long-term stewardship and archiving.

Discussion:
Heather presents as an example the URS Ref Set sample submission, public portal instructions, etc. and will update with actual DB gap links provided by Dr. Leach.

4. Update on Cancer Biomarkers AI and Bioinformatics Workshop August 13 - 15, 2024 - Caltech, Pasadena

https://edrn.nci.nih.gov/news-and-events/upcoming-meetings/ai-bioinformatics-workshop/

Workshop Goals

- Define use cases and needs for cancer biomarker discovery and validation
- Discuss the state-of-the-art and current gaps in applying AI/ML and bioinformatics methods to cancer biomarker research
- Discuss challenges around reproducibility, interpretability, and explainability
- Discuss current informatics capabilities and infrastructure needs to support use cases
- Develop a set of goals and recommendations to address these needs in cancer biomarker research for the next few years

Hackathon

- Multiple organs
- Multiple modalities, but at least images
- With validation sets and ground truth
- Transfer learning, foundational possibilities, data fusion, and challenges from PIs/participants

Discussion:
Additional planning meeting to be held later today. JPL notes the following important dates for the planning of this event:

AI Workshop Planning Schedule

Item	Date
Announce Workshop / Call for Abstracts	March 1, 2024
Abstract Submissions Due	April 26, 2024
Registration Open	June 1, 2024
Authors Notified	June 15, 2024
Agenda Posted	July 15, 2024
Registration Closed	August 7, 2024
Workshop	August 13-15, 2024

5. **Other business (e.g., invite subcommittee members to share their experience/challenges/progress)**
   Discussion:

Next Call: Monday, Mar 18th, 2024, at 1 pm Eastern/10 am Pacific