

# ARDaC: Common Data Model Facilitates Data Dissemination and Enables Data Commons for Modern Clinical Studies

Nanxin Jin<sup>a,b</sup>, Zuotian Li<sup>a,c</sup>, Carla Kettler<sup>b</sup>, Baijian Yang<sup>b</sup>, Wanzhu Tu<sup>a</sup>, Jing Su<sup>a</sup>

<sup>a</sup> Biostatistics and Health Data Science, Indiana University School of Medicine, Indiana, USA

<sup>b</sup> Computer and Information Technology, Purdue University, Indiana, USA

<sup>c</sup> Computer Graphics Technology, Purdue University, Indiana, USA





# Motivations

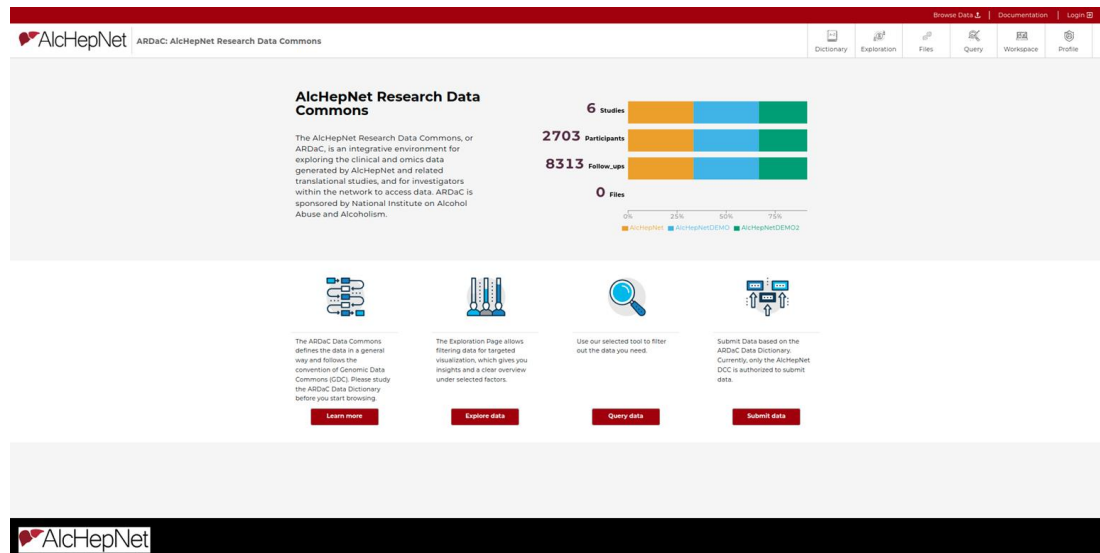
---

- Recording Longitudinal clinical information about participants.
- Following up visits for clinical events or responses to interventions.
- Collecting and biobank biospecimens.



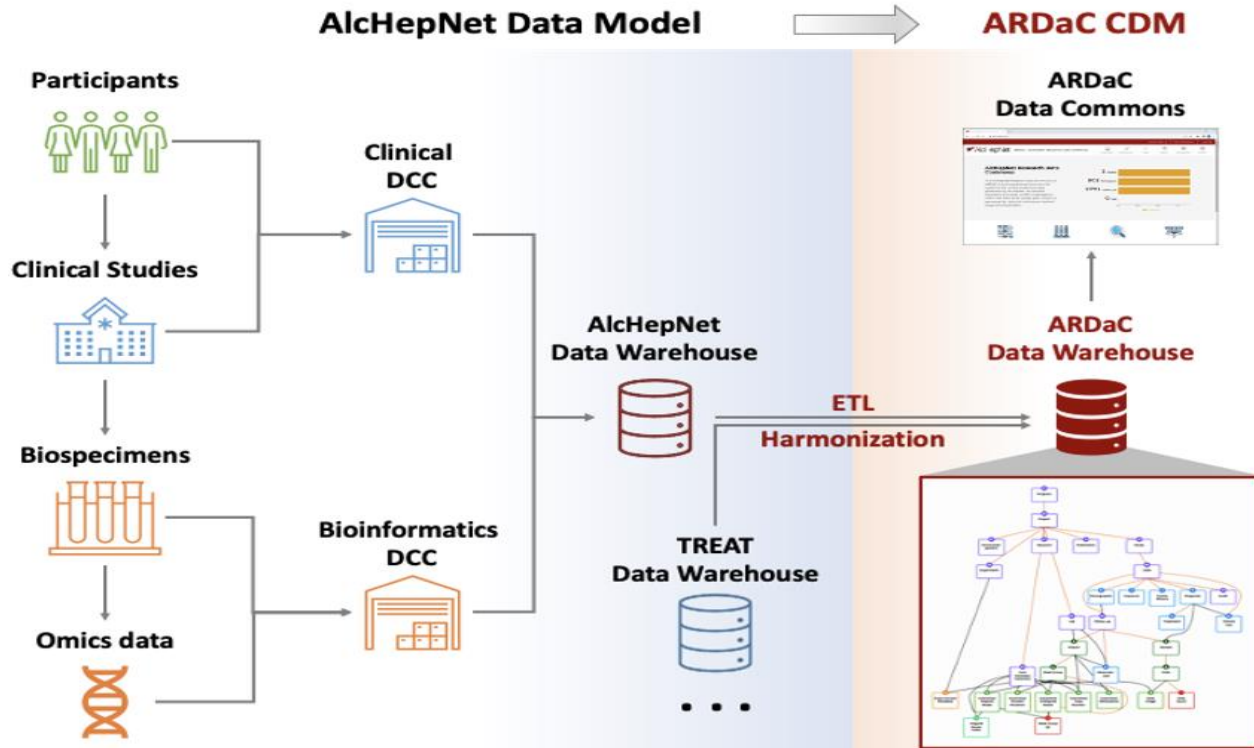
# Alcoholic Hepatitis Network Research Data Commons (ARDaC)

- Following FAIR principles.
- Three data domains:
  - Clinical data
  - Biospecimen metadata
  - Molecular and multiomics data
- Extend GDC to ARDaC
- Features:
  - Compatible with GDC
  - Gen3 framework
  - ETL toolkit



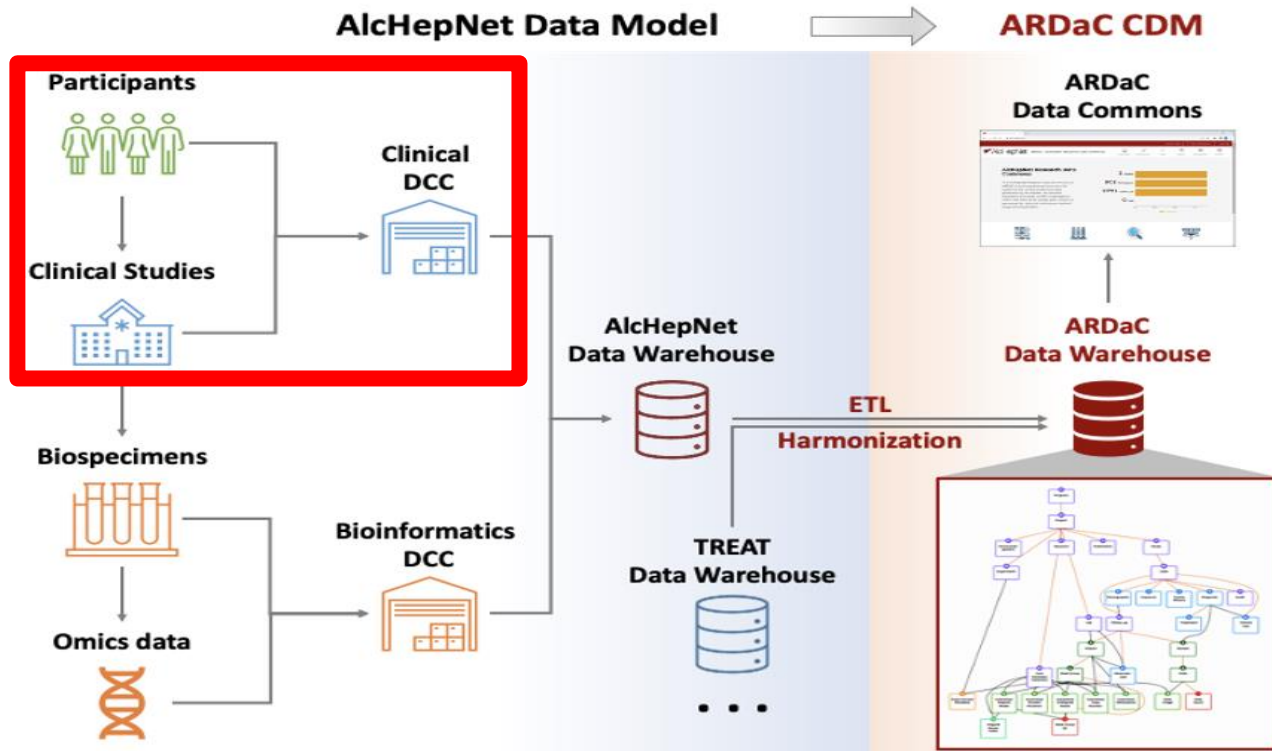


## Data Flow



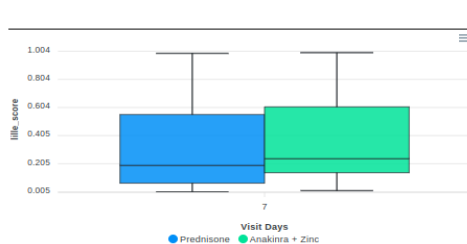
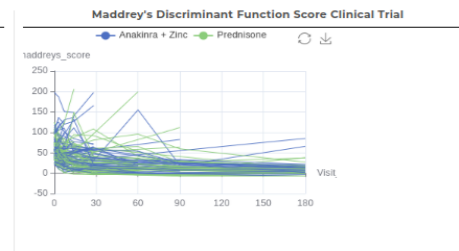
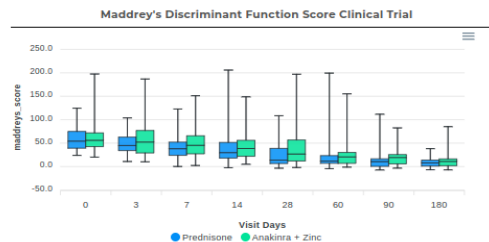
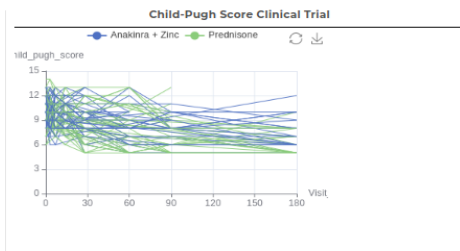
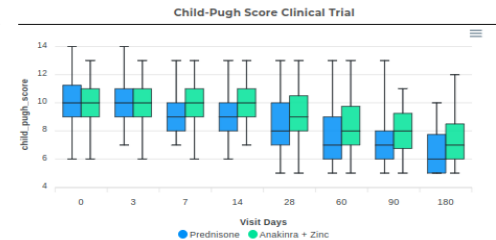
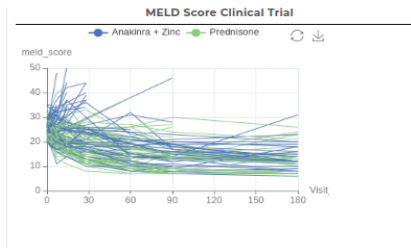
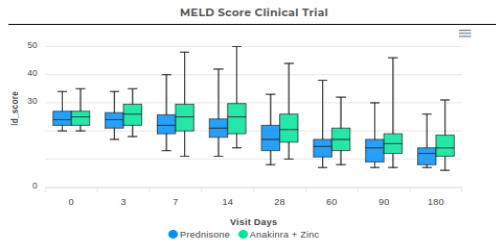


## Data Flow



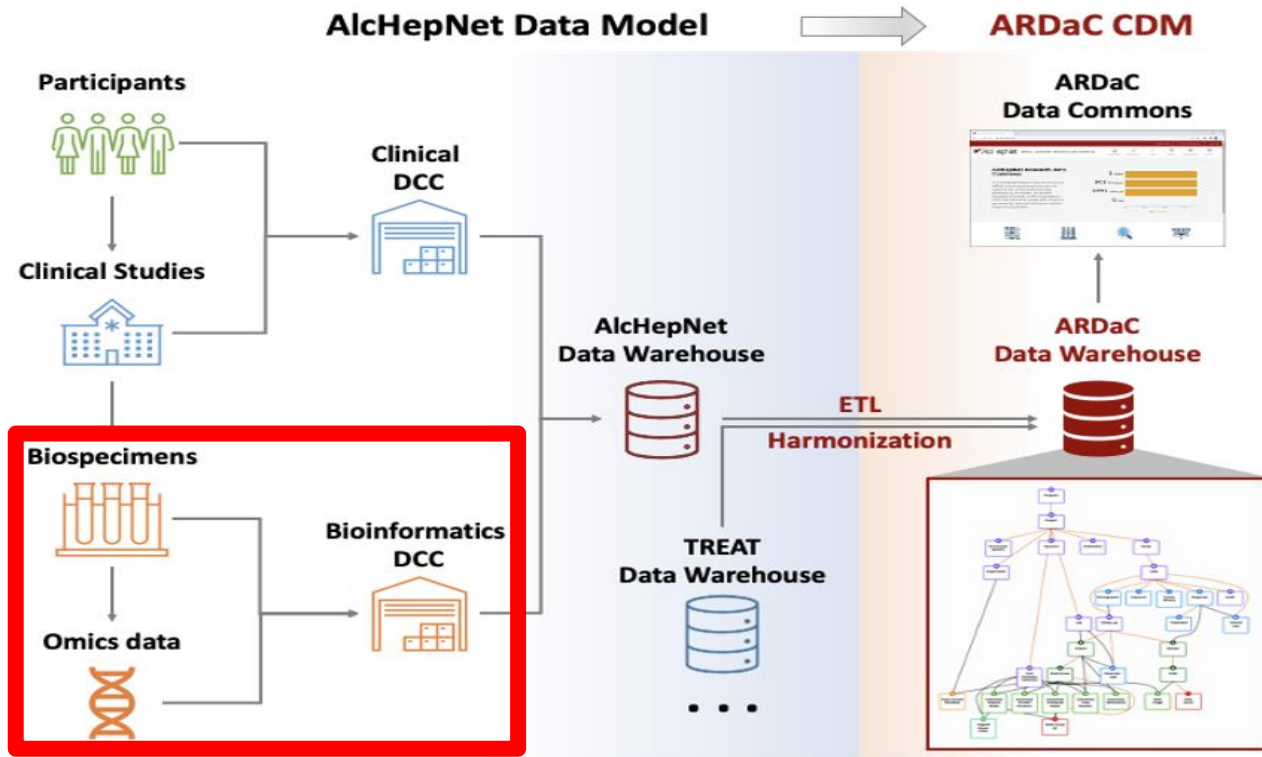
# MEDINFO 23

8 - 12 JULY 2023 | SYDNEY, AUSTRALIA



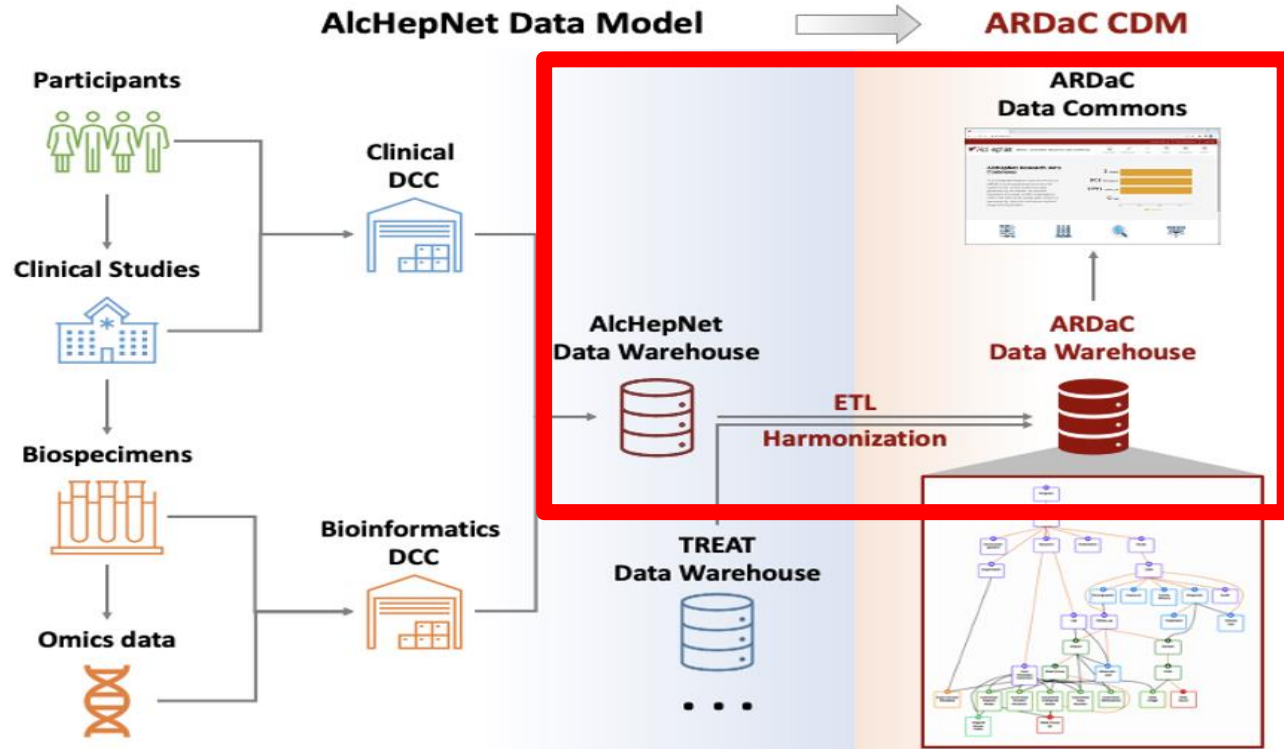


## Data Flow





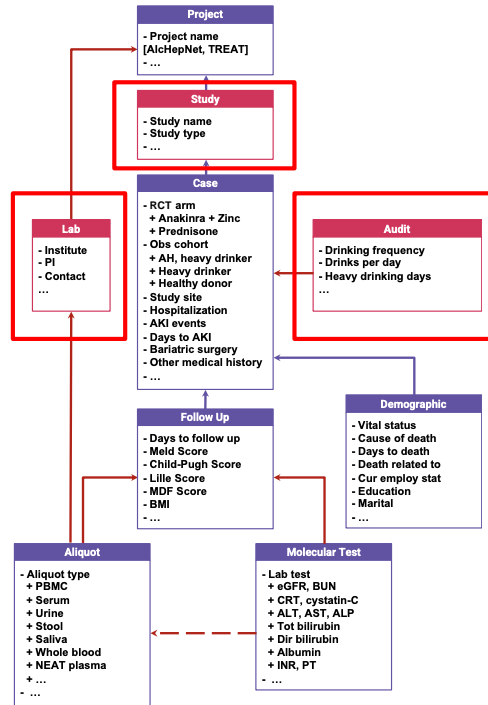
## Data Flow





## Data Modelling

- New nodes



# MEDINFO 23

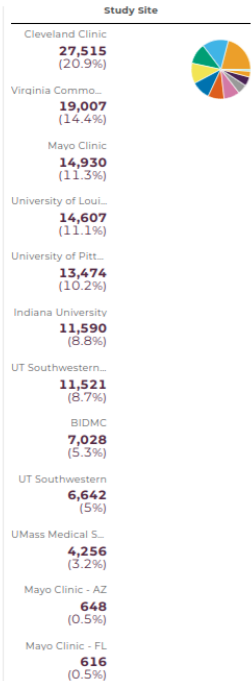
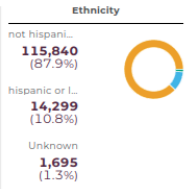
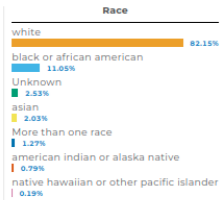
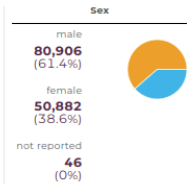
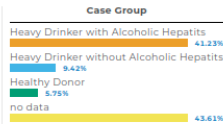
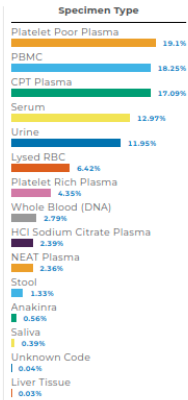
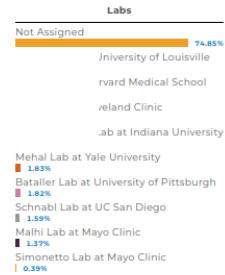
8 - 12 JULY 2023 | SYDNEY, AUSTRALIA



Biospecimens  
**131,834**

Follow Ups  
**1,378**

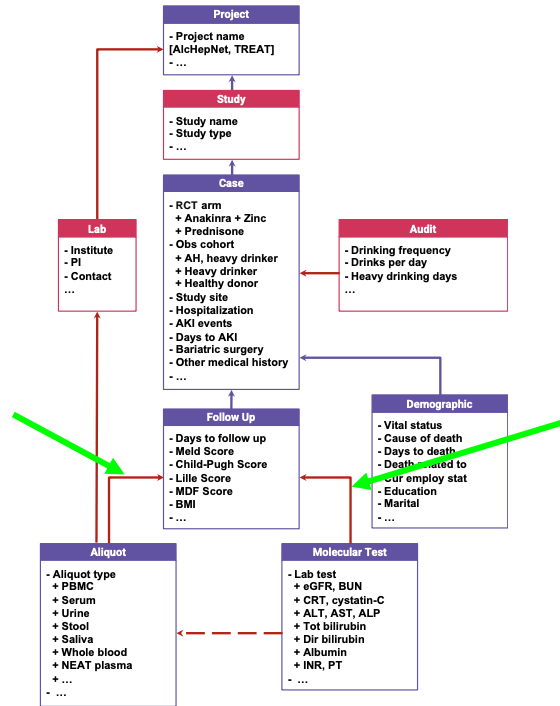
Labs  
**10**





## Data Modelling

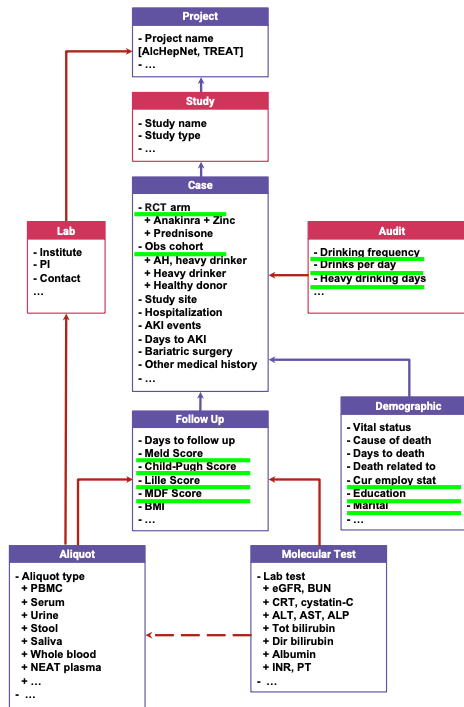
- Remodelling





## Data Modelling

- New properties





## Future Works

- Generating multiomics
- AI/ML scripts
- TREAT data

A screenshot of the AlcHepNet ARDaC (AlcHepNet Research Data Commons) web interface. The interface has a red header bar with navigation links: "Submit/Browse Data", "Documentation", "jinn@iu.edu", and "Logout". Below the header is a navigation menu with icons for "Dictionary", "Exploration", "Files", "Query", "Workspace", and "Profile". The main content area displays three Jupyter notebook launch cards. Each card features the Jupyter logo and a "Launch" button. The first card is titled "Bioinfo - Python/R" with "0.5CPU, 256Mi memory". The second card is also titled "Bioinfo - Python/R" with "0.8CPU, 1024Mi memory". The third card is titled "Brain - Python/R/Freesurfer" with "0.8CPU, 1.5Gi memory".



# Conclusion

---

- Met the urgent needs of the data dissemination in modern **clinical studies**.
- Proposed new data model for **clinical studies**.
- Enabled development of research data commons for **clinical studies**.
- <https://dev.ardac.org/>



# Acknowledgement

---

- National Library of Medicine of the National Institutes of Health (award number U24AA02696 to W.T..)
- National Library of Medicine of the National Institutes of Health (award number R01LM013771 to J.S.)
- Indiana University Precision Health Initiative (to J.S.)



# Thank you!

---