

Starting Soon: Ecosystem & Research Highlights

Reflect on 2023 and look forward to 2024 with BDC program leaders and community members



Regina Bures, PhD

NHLBI BDC Scientific
Program Director



Sweta Ladwa, MPH

NHLBI BDC Data Management Core
Scientific Program Director



Brandon Lê, PhD

BDC Fellows Program
Alumnus and Researcher

Statement of Conduct

The BioData Catalyst Consortium is dedicated to **providing a harassment-free experience for everyone**, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, or religion (or lack thereof). We do not tolerate harassment of community members in any form. Sexual language and imagery is generally not appropriate for any venue, including meetings, presentations, or discussions.

Resource: [Statement of Conduct](#)

Session Materials and Housekeeping

- Slides and recording will be available on the [community forum](#)
- We encourage you to submit unanswered questions, no matter how big or small, to our [help desk](#)
- Join the ecosystem: <https://biodatacatalyst.nhlbi.nih.gov/contact/ecosystem>

Interact with today's slides: <https://bit.ly/BDC-Dec-Slides>

Support and Guests



Amber Voght

User Engagement Specialist,
BDC Coordinating Center



Nathalie Volkheimer

User Engagement Specialist,
BDC Coordinating Center



Regina Bures, PhD

NHLBI BDC
Scientific Program Director



Sweta Ladwa, MPH

NHLBI BDC Data Management Core
Scientific Program Director



Brandon Lê, PhD

BDC Fellows Program
Alumnus and Researcher

Support and Guests



Amber Voght

User Engagement Specialist,
BDC Coordinating Center



Nathalie Volkheimer

User Engagement Specialist,
BDC Coordinating Center



Regina Bures, PhD

NHLBI BDC
Scientific Program Director



Sweta Ladwa, MPH

NHLBI BDC Data Management Core
Scientific Program Director



Brandon Lê, PhD

BDC Fellows Program
Alumnus and Researcher

Regina Bures, PhD

NHLBI BDC Scientific Program Director



National Heart, Lung,
and Blood Institute

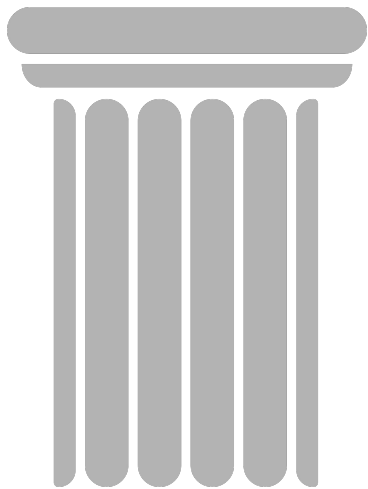
BioData

CATALYST

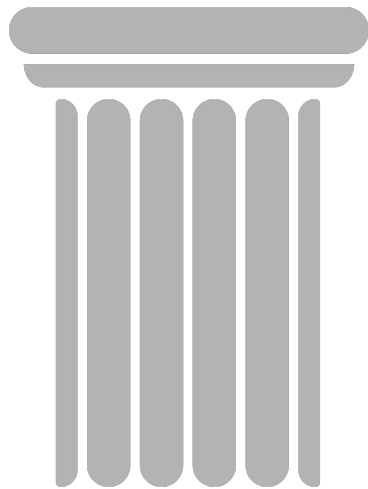
®

NHLBI BioData Catalyst® (BDC)

Mission



Vision



The *mission* is to develop and integrate advanced cyberinfrastructure, leading edge tools, and FAIR data to support the NHLBI research community.

The *vision* is to be a community-driven ecosystem implementing data science solutions to democratize data and computational access to advance Heart, Lung, Blood, and Sleep science.



2023 GA4GH **DRIVER PROJECTS**

Sickle cell disease (SCD) is the most common inherited blood disorder in the United States.



www.nhlbi.nih.gov/blood



National Heart, Lung,
and Blood Institute

**Blood Diseases
& Disorders
Education Program**

Sickle Cell Disease: Six Clinical Datasets Added

- Cooperative Study of Sickle Cell Disease ([CSSCD](#))
- Hematopoietic Cell Transplant for Sickle Cell Disease ([HCT for SCD](#))
- Multicenter Study of Hydroxyurea ([MSH](#))
- Optimizing Primary Stroke Prevention in Children with Sickle Cell Anemia ([STOP II](#))
- Treatment of Pulmonary Hypertension and Sickle Cell Disease with Sildenafil Therapy ([Walk-PHaSST](#))
- Hydroxyurea to Prevent Organ Damage in Children with Sickle Cell Anemia ([Baby Hug](#))

Additional Resources

→ Read:

BDC supports sickle cell research by hosting key datasets

→ Watch:

BDC: Sickle Cell Disease Datasets and Research

Be part of the NHLBI's Celebration and Strategic Planning for the next five years.



**National Heart, Lung,
and Blood Institute**

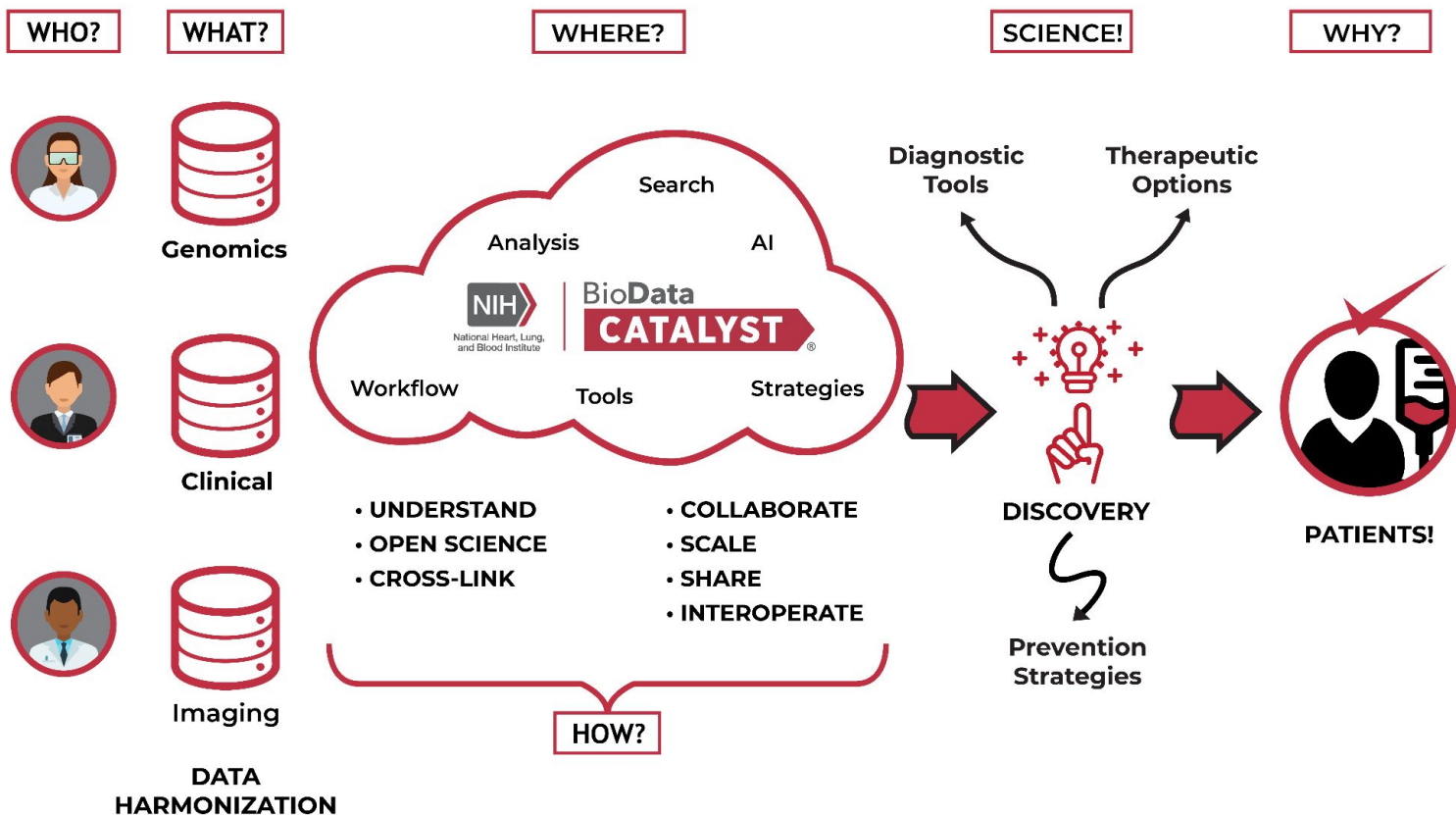


Request for Information on Research Needs

This [RFI](#) has a particular interest in seeking input on novel research needs and approaches in the following focus areas by **December 15th**:

- Harnessing **data science and new technologies** to drive scientific discovery and precision health.
- Using novel approaches for **addressing health disparities** and tackling their biological underpinnings for heart, lung, blood, and sleep diseases and conditions.
- Leveraging the power of **community and patient engagement**.
- Furthering the science on the **importance of lifestyle** behaviors.
- Supporting **women's health** through the lifespan.
- Addressing and reducing the **impact of “place”** (geography, climate, rural/urban, neighborhood) on heart, lung, blood, and sleep health.

Deep appreciation for the “Who” and “Why”



Democratizing Data and Facilitating Data Sharing

Sweta Ladwa, MPH

NHLBI BDC Data Management Core
Scientific Program Director



National Heart, Lung,
and Blood Institute

BioData

CATALYST

®

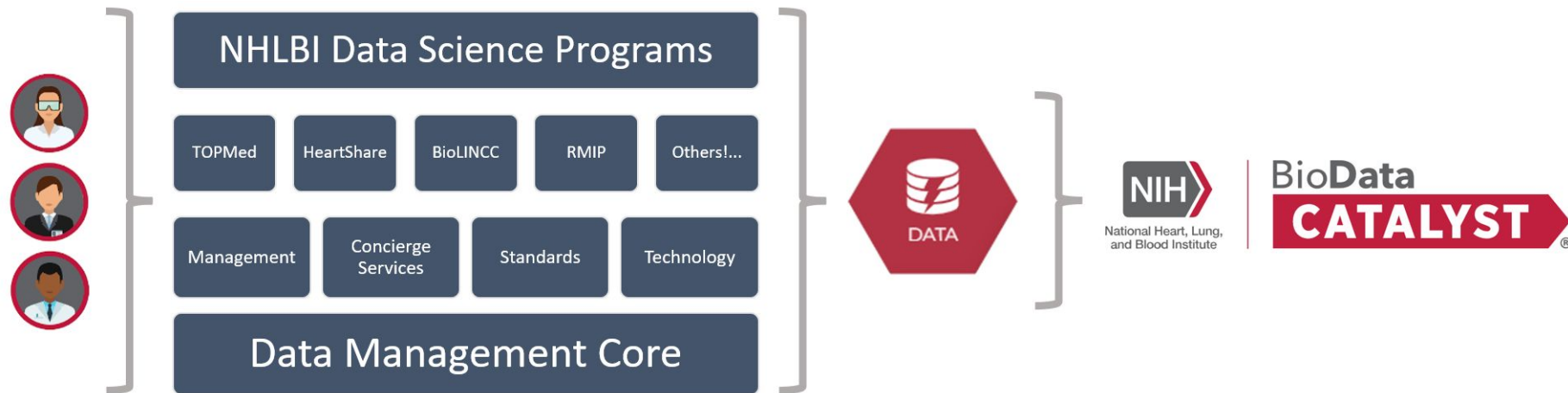
BDC Data Management Core

Enable data generators to submit their datasets into the BDC ecosystem such that high value data is available for research

- Identify and prioritize data for onboarding into BDC
- Standardize data and streamline the data onboarding process
- Help ensure the efficient and effective management of data before and during ingest into BDC

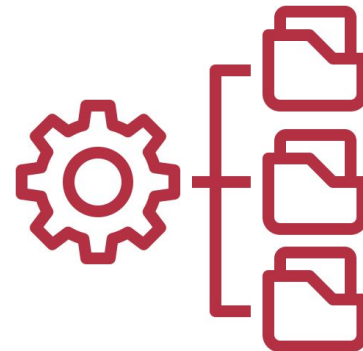
BDC Data Management Core

Enable data generators to submit their datasets into the BDC ecosystem such that high value data is available for research



Data Management In Support of BDC

- Goal of increasing quality and availability of data
- Individual study data submission
- Consortium/batch study data submission
- Support for released studies
- Developing methods for prioritizing data ingest



Data Ingestion over the past year (since 10/2022)

- Accelerating COVID-19 Therapeutic Interventions and Vaccines 4 ACUTE (ACTIV4a) v1.0, v1.1
- Clinical-trial of COVID-19 Convalescent Plasma in Outpatients (C3PO)
- Complement Inhibition Using Eculizumab to Overcome Platelet Transfusion Refractoriness in Patients with Severe Thrombocytopenia (DIR-Eculizumab)
- COVID-19 Post-hospital Thrombosis Prevention Study (ACTIV-4C)
- COVID-19 Post-hospital Thrombosis Prevention Study (ACTIV4c)
- Hydroxyurea to Prevent Organ Damage in Children with Sickle Cell Anemia (BABYHUG)
- Molecular Atlas of Lung Development (LungMAP)
- Multi-Ethnic Study of Atherosclerosis (BioLINCC)
- Multicenter Study of Hydroxyurea (MSH)
- NHLBI TOPMed - NHGRI CCDG: The Vanderbilt AF Ablation Registry
- NHLBI TOPMed: Boston Early-Onset COPD Study (EOCOPD)
- NHLBI TOPMed: Cleveland Clinic Atrial Fibrillation (CCAF) Study
- NHLBI TOPMed: Diabetes Heart Study (DHS) African American Coronary Artery Calcification (AACAC)
- NHLBI TOPMed: Genomic Activities such as Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study (FHS)
- NHLBI TOPMed: Heart and Vascular Health Study (HVH)
- NHLBI TOPMed: MESA and MESA Family AA-CAC (MESA)
- NHLBI TOPMed: The Cleveland Family Study (CFS)
- NHLBI TOPMed: The Genetics and Epidemiology of Asthma in Barbados
- NHLBI TOPMed: The Jackson Heart Study (JHS)
- NHLBI TOPMed: Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Project: Cardiovascular Health Study (CHS)
- Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b)
- The Cleveland Family Study (NSRR-CFS)
- The Genetic Epidemiology of Asthma in Costa Rica (CRA)
- The Pediatric Cardiac Genomics Consortium (PCGC)

Present and Future State

- Centralized access to data
- Readily available tools for conducting primary and secondary analysis with development of program-specific tools
- Guidance for preparing data for maximal reuse including data standards and metadata
- Collaborate with program investigators and stakeholders to propose data standards, harmonization strategies and metadata requirements for various data types.



- Facilitate efficient and secure data sharing with collaborators
- Host tools and workflows for conducting primary and secondary analysis
- Enable BioData Catalyst to host new data types
- Encourage collaboration to advance scientific innovation

NHLBI Program Integration

- Primary focus on **three areas** to best support NHLBI programs transitioning to BDC
 - **Metadata** - to find, access, and effectively reuse/reproduce the data and results
 - **Tools** - to effectively reuse and reproduce the data and results
 - **Linking** - for individuals in programs generating multi-modal data in a privacy preserving way
- Support FAIR Data for NHLBI data generating programs
 - Including support for the NIH Final Policy for Data Management and Sharing

2023 Retrospective

Brandon Lê, PhD



National Heart, Lung,
and Blood Institute

BioData

CATALYST

®

Overview

- Genetic and omic modifiers of Sickle Cell Disease (SCD)
- BDC Fellow - Cohort II
- Dissertation defended!

Activity on BDC

- Performed multiple analyses through the ecosystem – association studies of renal function across the TOPMed SCD studies
- Developed workflows and imported tools on Seven Bridges
- Participated in the BDC community: community hours, quarterly meetings, working groups

Datasets and toolkits

NHLBI TOPMed

- Whole-genome sequencing program, part of the Precision Medicine Initiative in the United States
- Composed of many clinical studies containing genetic, omic, and phenotypic data
- Four SCD patient cohorts utilized
- Renal outcomes obtained from the study investigators



NHLBI BioData Catalyst (BDC)

- A community of NHLBI-affiliated researchers working on TOPMed datasets
- Hosts several cloud-based ecosystems intended to facilitate genetic and omic analyses, connect researchers, and improve reproducibility in computational pipelines
- Supported through computational credits, data storage, and technical support



Tool and Workflow Development

- Imported and modified apps
- Developed workflows
- Tested bleeding-edge tools (e.g. Annotation Explorer)
- Participated in the BDC community: community hours, quarterly meetings, working groups

Tool and Workflow Development

Base fileset: TOPMed WGS Freeze 8, minDP10

Filter SNPs by:

- Filter: only pass
- Include SNPs missing in < 15% of individuals
- Minor allele frequency (MAF) > 0.05
- Exclude indels or multiple alternate alleles
- Hardy-Weinberg equilibrium (HWE) test p-value > 1e-6



Filtered VCFs containing relevant SNPs

Control for population substructure using EIGENSTRAT:

Above filtering, plus:

- Minor allele frequency (MAF) > 0.20
- Prune for SNPs under linkage disequilibrium (LD prune)
- Analyze cryptic relatedness



Principal components generated from subset of filtered SNPs

GWAS performed using PLINK

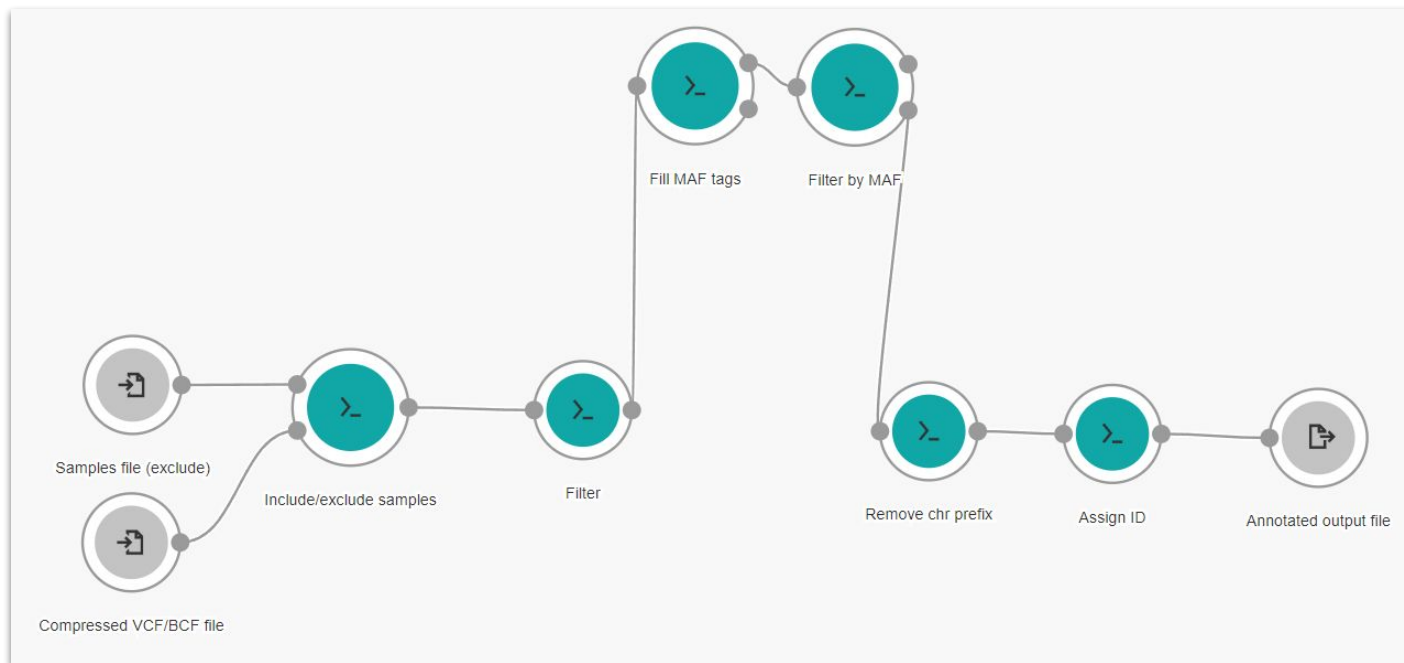
Covariates (eGFR, eGFR < 90, proteinuria):

- Top 2 principal components
- Hemoglobin diagnosis (SS / SB⁰ or SB⁺ / SC)

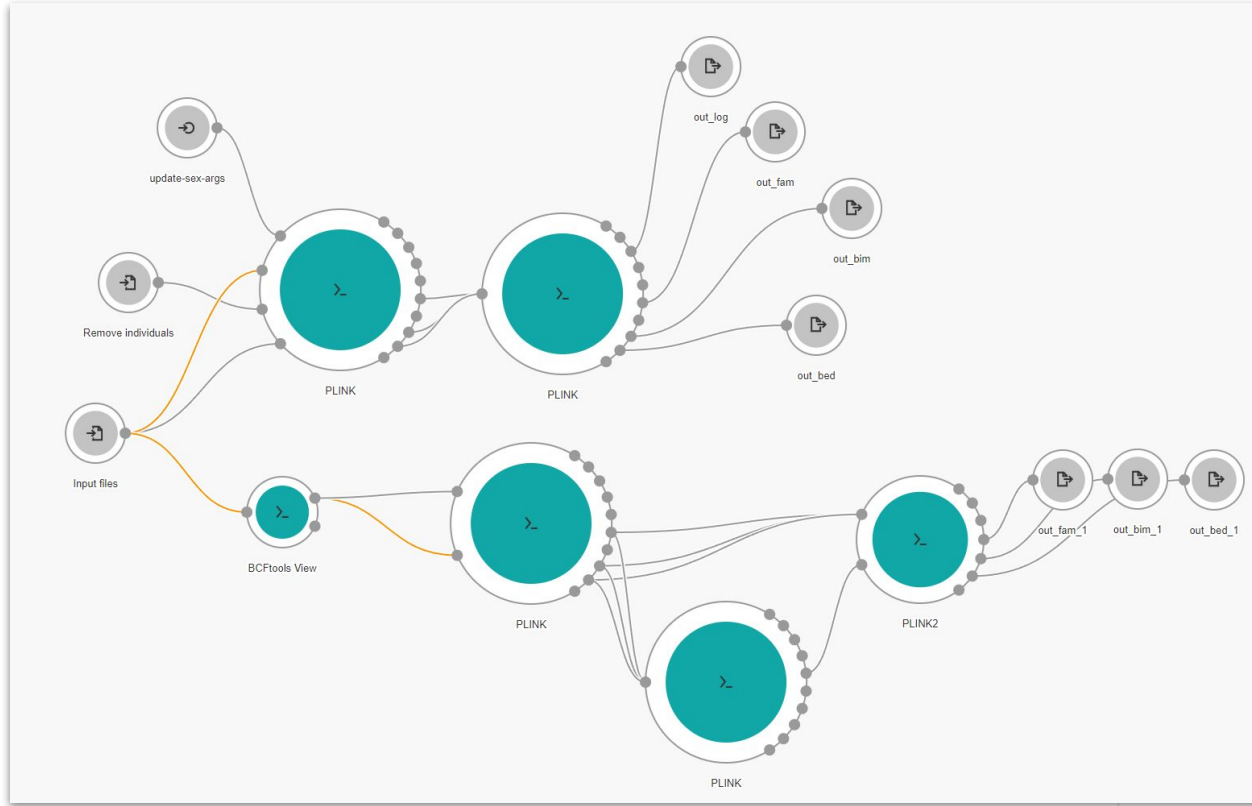
Covariates (proteinuria) – also include:

- Duffy antigen
- Age at enrollment

Tool and Workflow Development



Tool and Workflow Development



Retrospective

Highlights

- Robust support system
- Accessibility of HPC to broader range of interests / skillsets
- Scaling, parallelization, reproducibility

Continued growth

- Inclusion of additional tools
- Facilitating data access
- Onboarding of additional users



Before we leave...

Thank you to the BDC Consortium

Thank you to our BDC ecosystem reps

... and thank YOU !!!

We had a wonderful year showcasing the functionality and research happening on BDC - and it wouldn't be possible without you!

Summary and Discussion

Submit unanswered questions to our [help desk](#):

<https://biodatacatalyst.nhlbi.nih.gov/contact/>

Your feedback matters.

Please take the short anonymous survey
at the end of this meeting.

Join the ecosystem

Join the BDC Community

<https://biodatacatalyst.nhlbi.nih.gov/contact/ecosystem>

Thank you!

The next Community Hours is
Wednesday, January 17th at 1 PM ET

Topic: Art of the Start Panel

Register now: <https://bit.ly/BDC-Jan>