BioData Catalyst Implementation Plan

V2.0 - 20200403

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

BioData Catalyst Implementation Plan

V2 - 20200403

Document Status

Version

V2.0

Approvals

Signatures presented below denote review and approval of the BioData Catalyst Implementation Plan. These approvals are given based on the understanding that the Implementation Plan, and the information herein, will be revised at regular periods over the course of the program. It is the responsibility of the Principal Investigator (PI) of each funded team and select NHLBI program staff to add their name(s) in the indicated space below.

Approved Date

4/03/2020

PI Approvals:

PI	Team	Approval Date
Robert L. Grossman (University of Chicago)	Calcium	3/31/2020
Anthony Philippakis (Broad Institute)		3/30/20
Benedict Paten (UCSC)		3/30/2020
Paul Avillach	Carbon	03/31/20
Ashok Krishnamurthy	Helium	3/31/2020
Brandi Davis-Dusenbery	Xenon	3/31/2020

NIH Approvals:

Responsible Person	NIH NHLBI BioData Catalyst Role	Approval Date
Jonathan Kaltman	Program Manager	4/3/2020
Alastair Thomson, NHLBI CIO	Information Security	4/3/2020

Next Review Date

4/03/2021

Document Owner

BDC3

Revision History

Date (YYYYMMDD)	Versio n Numbe r	Revision Reviewed/ Approved By	Brief Description of Change
20190110	V0	N/A	Draft document created
20190206	V.0.1	Stan Ahalt	Content update: all sections
20190214	V0.2	Paul Avillach (Carbon team)	Added i2b2/tranSMART platform and PIC-SURE metaAPI as the “gold master” for clinical data in DataSTAGE
20190226	V0.3	Rebecca Boyles	Changes accepted and comments addressed
20190314	V0.3	Marcie Rathbun	At the end of section 3.2, added link to Operationalization document: NHLBI DataSTAGE 60 Day o16n Plan v1-2
20190426	V1.0	NHLBI	V1.0 reviewed and approved by NHLBI Links, graphics, & editing updates [Marcie]
20200403	V2.0		Changes based on annual consortium review: - re-branded as BioData Catalyst & updated relevant graphics - replaced the Ambassadors Program section with a section on the Fellows Program - Small edits to the Beta-user training section to make more general and add the Help Desk initiative - Updated the org. chart & WG/TT list

TABLE OF CONTENTS

1 INTRODUCTION

1.1 PURPOSE OF THE IMPLEMENTATION PLAN

The BioData Catalyst Implementation Plan describes the process by which the BioData Catalyst Consortium will incrementally progress towards the vision of the program described in the BioData Catalyst Strategic Framework. The Implementation Plan will enable the teams to decompose the strategic vision into concrete steps and define measures of completion for each step. Additionally, the Implementation Plan and Strategic Framework will focus the Consortium on the steps necessary to execute on the BioData Catalyst strategy.

1.2 BACKGROUND

This document outlines how the various elements from the planning phase of the BioData Catalyst project will come together to form a concrete, operationalized BioData Catalyst platform. The platform will offer the ability to perform novel science and access an unprecedented array of data to a diverse set of users. The platform will advance groundbreaking research and significant advances in medicine.

The Implementation Plan, coupled with the Project Management Plan, establishes priorities and accountabilities for resource use. The BioData Catalyst Project Management Plan describes how BioData Catalyst will execute, monitor, and control work towards deliverables within the program. To create project priorities and transparency, the Implementation Plan uses the following guiding principles:

- Maximizing availability of resources;
- Positive impact on the NHLBI mission;
- Responsible stewardship of funds;
- Utilization of technologies that maximize data security and integrity;
- Implementation of cost-effective solutions; and
- Consistency with the BioData Catalyst Consortium Charter and values.

2 OVERVIEW

The BioData Catalyst Consortium (BDCC) is a collection of teams and stakeholders working to deliver on the common goals of integrated and advanced cyberinfrastructure, leading-edge data management and analysis tools, FAIR data, and HLBS researcher engagement. The BDCC takes an Agile development approach towards implementation to be flexible and responsive to user needs and react to user feedback. Accordingly, work within the Consortium is described at various levels of detail with a focus on collaboration with the user community.

The organization of the program will be coordinated according to objectives that are captured in User Narratives, which are further decomposed into Features, Epics, and User Stories.

Coordination along this framework will support the BioData Catalyst teams working in a coordinated manner towards common goals. User Narratives are an orthogonal construct to Work Streams, and are critical in the integration of user participation into the development cycle.

Project milestones are defined and tracked via the delivery of User Narratives, Features, and Epics.

2.1 KEY TERMS AND CONCEPTS

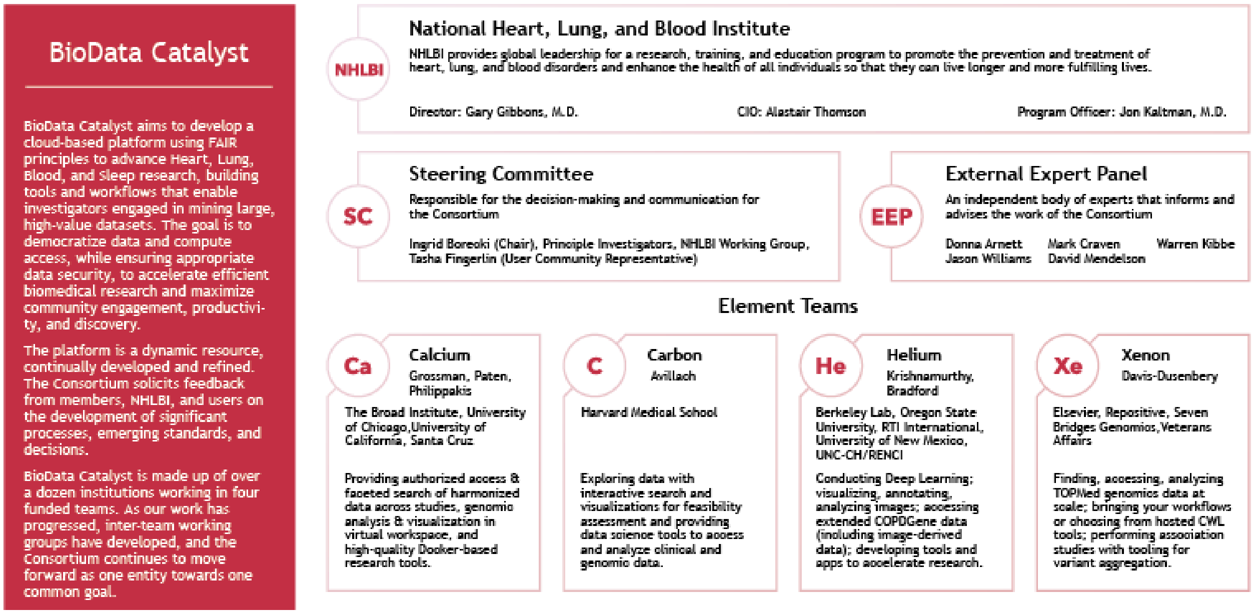
In the BioData Catalyst program, the following key terms are used:

- **User Narrative**: A description of a user interaction experience within the system from the perspective of a particular persona. Example: An experienced bioinformatician wants to search TOPMed studies for a qualitative trait to be used in a GWAS study.
- **Feature**: A functionality at the system level that fulfills a meaningful stakeholder need. Example: Search TOPMed datasets using the PIC-SURE platform.
- **Epic:** A (very) large User Story described at the program level that can be broken into executable stories. Example: PIC-SURE is accessible on BioData Catalyst.
- **User Story**: An item that describes a requirement or functionality for a user. Example: A user can access PIC-SURE through an icon on BioData Catalyst to initiate a search.
- **Work Stream**: A collection of related features; orthogonal to a User Narrative. Example: Work Streams impacted by the above User Narrative include production system, data analysis, data access, and data management.

2.2 PROGRAM ORGANIZATION

Consortium Groups

The BioData Catalyst program is composed of several groups who each bring various resources towards executing the vision of BioData Catalyst. The organizational chart below displays the teams and their responsibilities.



BioData Catalyst

BioData Catalyst aims to develop a cloud-based platform using FAIR principles to advance Heart, Lung, Blood, and Sleep research, building tools and workflows that enable investigators engaged in mining large, high-value datasets. The goal is to democratize data and compute access, while ensuring appropriate data security, to accelerate efficient biomedical research and maximize community engagement, productivity, and discovery.

The platform is a dynamic resource, continually developed and refined. The Consortium solicits feedback from members, NHLBI, and users on the development of significant processes, emerging standards, and decisions.

BioData Catalyst is made up of over a dozen institutions working in four funded teams. As our work has progressed, inter-team working groups have developed, and the Consortium continues to move forward as one entity towards one common goal.

NHLBI — National Heart, Lung, and Blood Institute

NHLBI provides global leadership for a research, training, and education program to promote the prevention and treatment of heart, lung, and blood disorders and enhance the health of all individuals so that they can live longer and more fulfilling lives.

Director: Gary Gibbons, M.D. CIO: Alastair Thomson Program Officer: Jon Kaltman, M.D.

SC — Steering Committee

Responsible for the decision-making and communication for the Consortium

Ingrid Borecki (Chair), Principle Investigators, NHLBI Working Group, Tasha Fingerlin (User Community Representative)

EEP — External Expert Panel

An independent body of experts that informs and advises the work of the Consortium

Donna Arnett, Mark Craven, Warren Kibbe, Jason Williams, David Mendelson

Element Teams

Ca — Calcium

Grossman, Paten, Philippakis

The Broad Institute, University of Chicago,University of California, Santa Cruz

Providing authorized access & faceted search of harmonized data across studies, genomic analysis & visualization in virtual workspace, and high-quality Docker-based research tools.

C — Carbon

Avillach

Harvard Medical School

Exploring data with interactive search and visualizations for feasibility assessment and providing data science tools to access and analyze clinical and genomic data.

He — Helium

Krishnamurthy, Bradford

Berkeley Lab, Oregon State University, RTI International, University of New Mexico, UNC-CH/RENCI

Conducting Deep Learning; visualizing, annotating, analyzing images; accessing extended COPDGene data (including image-derived data); developing tools and apps to accelerate research.

Xe — Xenon

Davis-Dusenbery

Elsevier, Repositive, Seven Bridges Genomics,Veterans Affairs

Finding, accessing, analyzing TOPMed genomics data at scale; bringing your workflows or choosing from hosted CWL tools; performing association studies with tooling for variant aggregation.

The BioData Catalyst Coordinating Center (BDC3), in collaboration with NHLBI, develops and maintains the Strategic Framework, Implementation, and Project Management Plans. All members of the Consortium are periodically invited to provide feedback on these plans to the BDC3, with a particular focus on integrating feedback from the Data Stewards (TOPMed) and users. The draft documents and any significant changes are reviewed by the Steering Committee as well as the External Expert Panel. The teams are responsible for collaborating to deliver Features, Epics, and User Stories and advance the BioData Catalyst ecosystem. Additional details on the membership, roles, and responsibilities of each group can be found in the Project Management Plan.

Additionally, the BioData Catalyst has created multiple boards, working groups, and tiger teams that develop standards, protocols, and best practices to ensure ecosystem development. Boards are decision-making entities; working groups make recommendations for standards and protocols; and tiger teams are working groups that operate for a short period of time. Current groups include

Collaboration Groups

Group	**Chair**	**Description**
Change Control Board	Schwartz	Evaluating requests for changes that impact project risk, cost, scope, or schedule of the NHLBI-approved workplan and requests for intra-team changes.
Data Release Management	Ladwa, Culotti	Making recommendations around prioritization and organization of data, metadata schemas, and data ingestion and release.
Integration Testing Tiger Team	Osborn	Designing and building an integration testing framework with tests to regularly verify the basic functionality of the ecosystem.
Data Access	Bradford, Lyons	Identifying, outlining, and developing policies and procedures to guide accessing data on the platform.
Data Harmonization	Carroll, Heavner	Supporting phenotype harmonization across the platform to reduce duplication and maximize expertise/efficiency; defining requirements for search and analytics across TOPMed.
Tools & Applications	Cox, O’Connor	Maintaining a list of tools, workflows, or “apps” feasible for inclusion in the environment.
User Engagement	Krishnamurthy, DiGiovanna,	Coordinating the platform recruitment of TOPMed investigators; documenting steps to

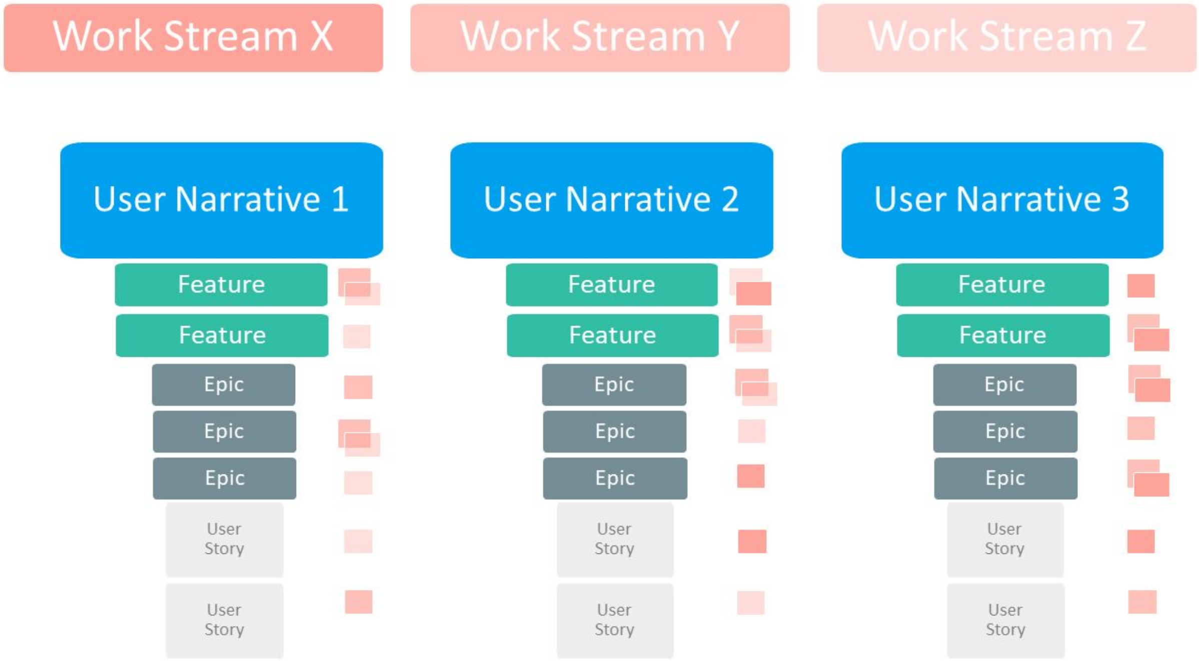
	Bis	operationalize data/tools/ computational workflows needed; supporting training.

Coordination of Activities

Initially executed by four teams, meeting the goals of BioData Catalyst requires intense and ongoing collaboration to create cyberinfrastructure, tools, processes, and a community of practice. The software development teams within the BioData Catalyst Consortium will be largely self-organizing around Epics, which the BDC3 will coordinate to ensure synchronization across shared Features. Multiple Features will commonly compose a User Narrative. Successful completion of work will be measured against the ability for a user to complete the work outlined in a User Narrative.

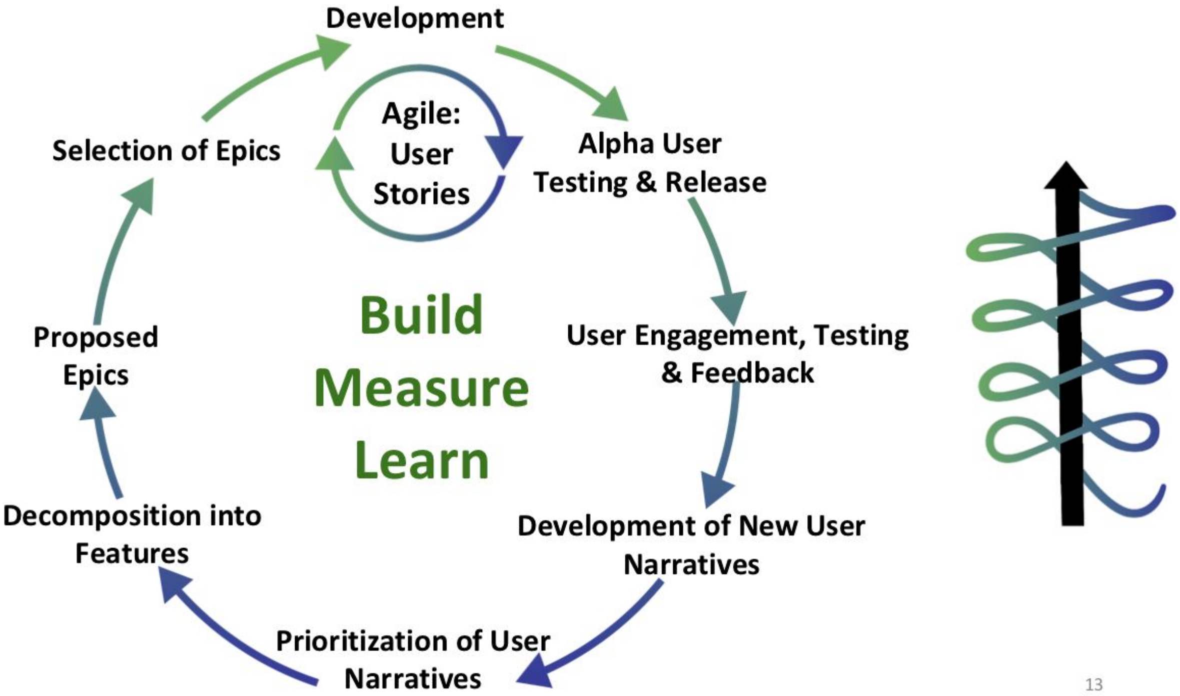
The ability for a user to complete a User Narrative on the BioData Catalyst system will indicate meaningful progress towards completion.

Independently, Features, Epics, and the more granular User Stories can be mapped to Work Streams, which are useful for reporting on aggregations of specific types of work, as shown in the figure below.



BioData Catalyst maintains a Consortium glossary of terms that is regularly updated in the BDCatalyst-RFC-2_BioData Catalyst_Strategic_Planning_Nomenclature.

A cyclical evaluation and revision of the BioData Catalyst User Narratives will be critical to the execution of the BioData Catalyst Agile program. Regular collection of user feedback and needs will feed into the development process and be represented in new or revised User Narratives that will be prioritized in coordination with NHLBI. This continued and organized refinement of priorities for Consortium development work will support close coordination and ground the BioData Catalyst program in the needs of the user community.


Development
Agile: User Stories
Alpha User Testing & Release
User Engagement, Testing & Feedback
Development of New User Narratives
Prioritization of User Narratives
Decomposition into Features
Proposed Epics
Selection of Epics
Build
Measure
Learn
13

3 BIODATA CATALYST PLATFORM OVERVIEW

3.1 SYSTEM DESCRIPTION

Inherent in the approach to the system design is the recognition that the current state of data and computational resources places onerous limits on the HLBS research community. Examples of limitations include an inability to execute arbitrary code, inability to access and work on very large data (e.g., TOPMed CRAMs) due to technical constraints, inability to search on one platform and execute on another, difficulties for groups of investigators to share controlled-access data and work together in a common workspace, as well as a laborious, several month process for a researcher gaining access to data.

The BioData Catalyst architecture provides an early cyberinfrastructure to researchers as quickly and responsibly as possible with an eye towards addressing the above limitations. BioData Catalyst will balance early delivery with ambitious goals by extending functionality through phased rollouts.

To accomplish this goal, the Consortium will abide by the below design principles:

- Meet user needs and incorporate feedback
- Leverage existing tools and infrastructure, when feasible

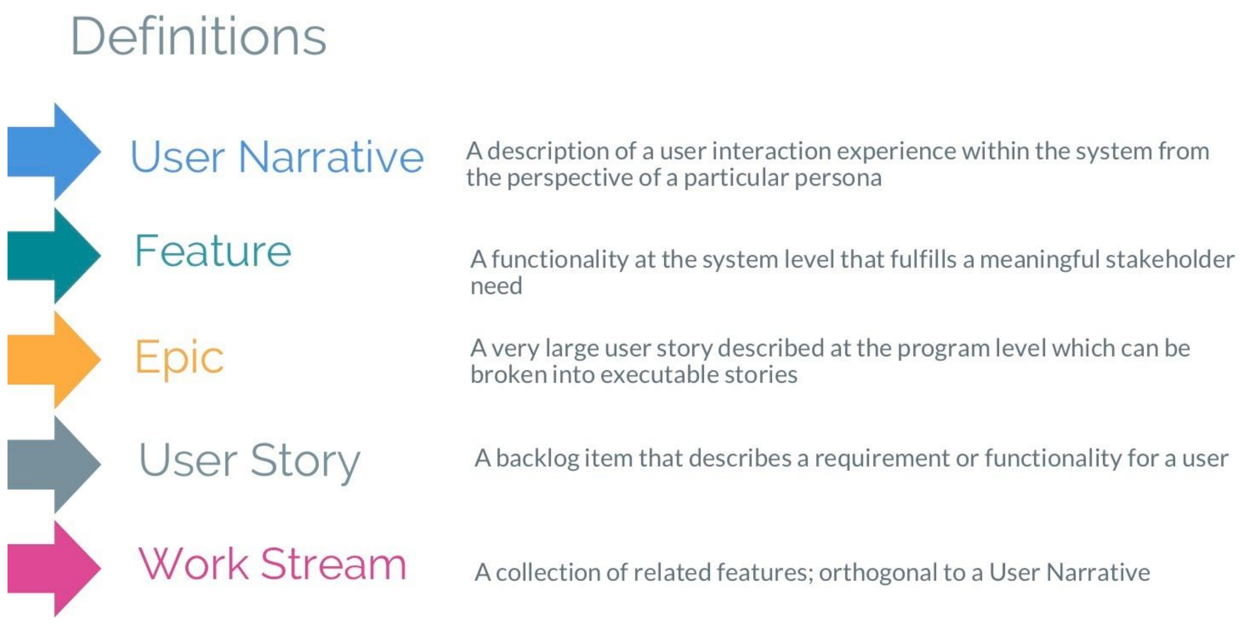
- Duplicate functionality when intentional and reasonable
- Architect interoperability with relevant systems
- Encounter a seamless experience, regardless of underlying components
- Leverage cost-advantageous cloud resources
- Support scalability and extension of functionality
- Have an early impact on computational-driven HLBS science
- Enable easy access to applications and tools for users across BioData Catalyst
- Provide systems security for hosting identifiable data
- Implement rigorous testing and Quality Assurance measures for components and data

Applying these design principles, our initial architecture of the BioData Catalyst platform is pictured in the figure below. The teams will leverage the Data Commons Framework Services (DCFS) of Gen3 to provide critical infrastructure, common security, data access services, and the genomic data gold master. The DCFS is a set of software services designed specifically to support this kind of Data Commons platform. The DCFS is powered by the Gen3 platform and were initially developed to support the National Cancer Institute's (NCI) Genomic Data Commons (Grossman, 2018). The PIC-SURE platform will be the clinical data gold master database leveraging its metaAPI. These data services will make use of the NIH STRIDES partnerships that offer NIH investigators cloud services and storage at discount pricing to support research (NIH, 2018).

The DCFS will include authentication and authorization services and digital object globally unique IDs for indexing. The current Terra, Seven Bridges, and PIC-SURE platforms will establish appropriate memos for interoperation with the DCFS. These memos are means through which groups will formalize cooperation with one another to develop interoperability solutions that meet functional, technical, and security/compliance requirements.

BioData Catalyst will be extended through the integration of third-party applications. There are a number of possible models in which a third-party application can operate within the BioData Catalyst platform. The terms of operation for these applications are being developed collaboratively between the Tools and Applications Working Group and the Operationalization Tiger Team.

3.2 SYSTEM DEVELOPMENT



Definitions

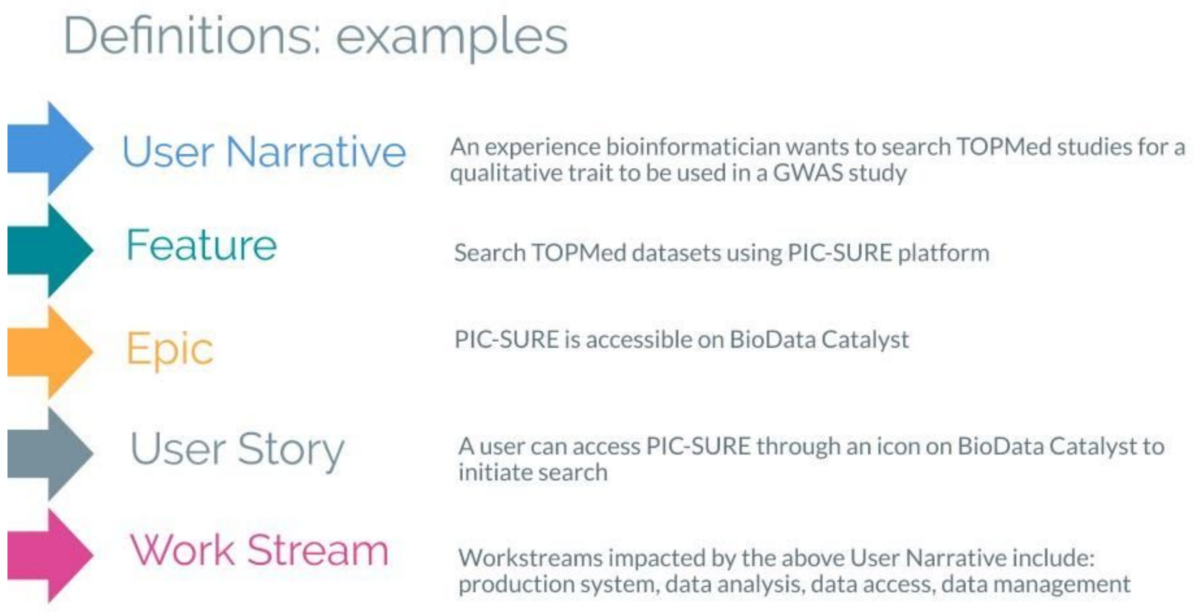
User Narrative
A description of a user interaction experience within the system from the perspective of a particular persona

Feature
A functionality at the system level that fulfills a meaningful stakeholder need

Epic
A very large user story described at the program level which can be broken into executable stories

User Story
A backlog item that describes a requirement or functionality for a user

Work Stream
A collection of related features; orthogonal to a User Narrative



Definitions: examples

User Narrative
An experience bioinformatician wants to search TOPMed studies for a qualitative trait to be used in a GWAS study

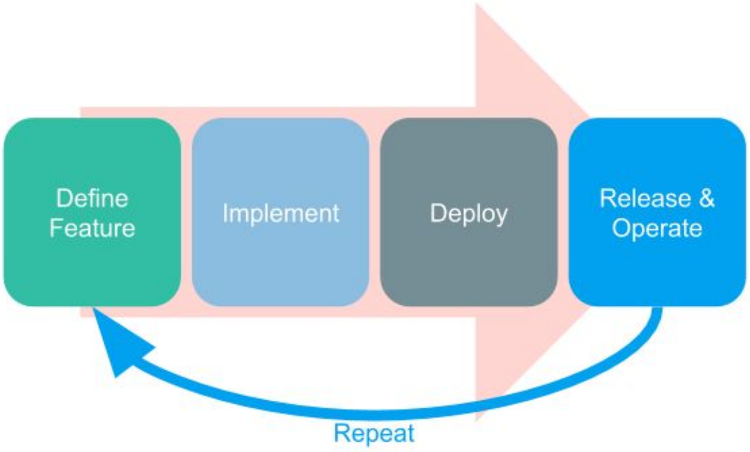
Feature
Search TOPMed datasets using PIC-SURE platform

Epic
PIC-SURE is accessible on BioData Catalyst

User Story
A user can access PIC-SURE through an icon on BioData Catalyst to initiate search

Work Stream
Workstreams impacted by the above User Narrative include: production system, data analysis, data access, data management

User Narratives 2019-2021



The phased development of the BioData Catalyst platform will be orchestrated through the collection, prioritization, and execution of User Narratives. User testing will be performed against these narratives to ensure appropriate completion. User Narratives offer an opportunity to engage potential users in the development process. As these users begin to work within BioData Catalyst, they will identify additional User Narratives that are needed to advance their research. These User Narrative updates will be reflected in regular revisions to the Strategic Framework and Implementation Plan to be reflected in future development efforts.

Features, Epics, and User Stories

The relationship between Features, Epics, and User Stories is hierarchical. A Feature is a service that fulfills a stakeholder need. Releases are managed at the level of the Feature and will be coordinated by the BDC3 in conjunction with the development teams and stakeholders. To support prioritization and acceptance testing, Feature descriptions include their benefits and criteria for acceptance.

Features conceptually map to Work Streams, which are groupings of technologically-related work. In BioData Catalyst, these Work Streams map to BioData Catalyst Working Groups and Tiger Teams who will help to gather additional information and inform the creation of Features. These small Working Groups and Tiger Teams will center on collaborative opportunities within the Work Stream areas. vbThe BDC3 will work with the teams and appropriate Working Groups and/or Tiger Teams to coordinate activities, deliverables, and releases. The initial Working Groups and Tiger Teams are listed and reflect the current focus of BioData Catalyst efforts. By design, these groups will evolve over the course of the project.

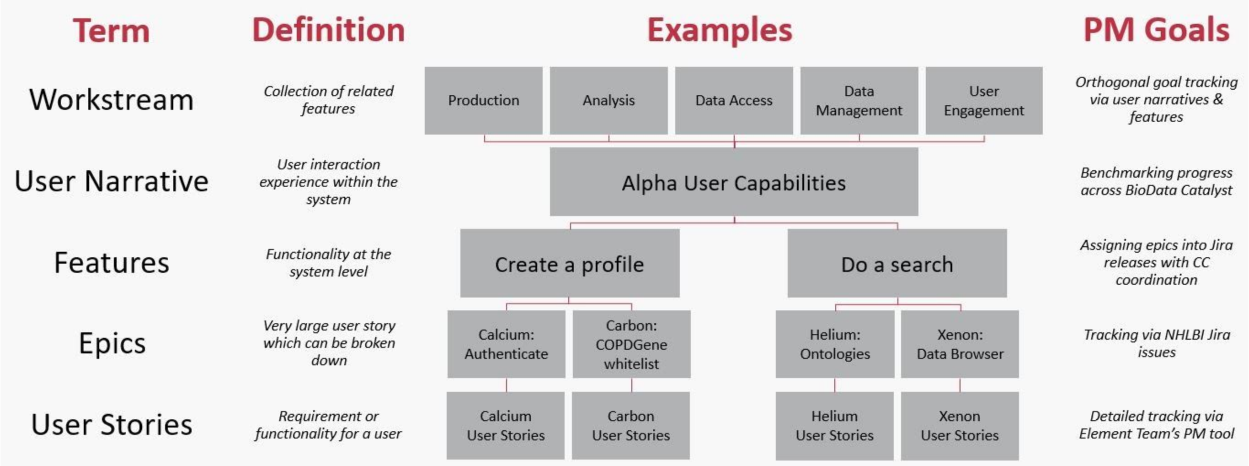
Initial Groups
Operationalization Tiger Team
Tools and Apps Working Group
Data Harmonization Work Group
Data Access / UX-UI Working Group

Cross-Team Development Coordination

Coordinating across the Consortium towards creating a cutting-edge cyberinfrastructure requires a framework to support individual development as well as collaboration. The BioData Catalyst program will work within the framework below to communicate with teams and track outcomes and dependencies.

A User Narrative is a description of how a particular user will interact with the system, often crossing Work Streams and including many different types of Features.

As briefly described in the previous section, the BDC3 will coordinate across teams, and for the purpose of software releases, at the Feature level of development. Individual teams will manage development work according to Epics; Epics are best described as very large User Stories. High-level requirements are gathered at this level and are further refined by teams into User Stories. BioData Catalystteams will report the cost to NHLBI at the Epic level. The individual teams will track resources and schedules at the User Story level. Additional details on reporting can be found in the Project Management Plan.



Term	Definition	Examples	PM Goals
Workstream	*Collection of related features*	Production Analysis Data Access Data Management User Engagement	*Orthogonal goal tracking via user narratives & features*
User Narrative	*User interaction experience within the system*	Alpha User Capabilities	*Benchmarking progress across BioData Catalyst*
Features	*Functionality at the system level*	Create a profile Do a search	*Assigning epics into Jira releases with CC coordination*
Epics	*Very large user story which can be broken down*	Calcium: Authenticate Carbon: COPDGene whitelist Helium: Ontologies Xenon: Data Browser	*Tracking via NHLBI Jira issues*
User Stories	*Requirement or functionality for a user*	Calcium User Stories Carbon User Stories Helium User Stories Xenon User Stories	*Detailed tracking via Element Team's PM tool*

Important elements of building a successful platform are user experience and feedback. BDC3 intends to develop the training, user engagement, and assessment strategy to ensure efficient onboarding of users, training, and feedback leading to streaming software and solution improvement. Refer to section "Training, User Engagement, and Assessment" below for more details.

4 TRAINING, USER ENGAGEMENT, AND ASSESSMENT

The training, engagement, and assessment strategy within BioData Catalyst will be a phased approach intentionally developed towards the needs of particular user communities as defined by the User Narratives. By approaching training in phases, we anticipate the establishment of an early BioData Catalyst Fellows Program which will recruit early adopters who will be invested in the development of the BioData Catalyst platform and become active contributors of feedback to the development teams.

4.1 TRAINING

As BioData Catalyst evolves as a program and a system, our approach to training will also evolve. The assessment of programs will lead to modifications as we take an Agile approach to training.

Fellows Program

The BioData Catalyst Fellows Program will seek to attract eager, risk-tolerant users to BDCatalyst to provide real world user feedback, especially scientific guidance, to the BDCatalyst developers. The program will be open to academic disciplines conducting biomedical research or related research in heart, lung, blood, or sleep domains. BioData Catalyst Fellows will be adept in the types of technology to be incorporated in BioData Catalyst, e.g., command-line facile and distributed computing.

Fellows will be awarded a stipend to cover salary, travel, training, publication, and conference fees. During their period of performance Fellows will work closely with mentors from the BDCatalyst development teams and with each other. Fellows should be willing and enthusiastic representatives for BioData Catalyst and comfortable with learning from success and failure.

Fellows will be expected to participate in a platform orientation at the beginning of their period of performance. This will be the first of monthly meetings between the BDC3 and Fellows and liaisons from BDCatalyst Working Groups. The Fellows will be mentored by liaisons from within the BDCatalyst user engagement teams and development teams for additional information and onboarding specific to their research projects.

Beta-User Training

Initial training will be focused on support for Fellows leveraging the BioData Catalyst platform as part of a User Narrative. As additional users are onboarded, the training will broaden. Once the platform is available to a broader audience, we will support freely-accessible online training for Beta-Users at any time, as well as Carpentries-inspired workshops led by trained instructors, Fellows Trainers, and BioData Catalyst trainers. These instructors will also serve the role of providing feedback, either by communications or formal surveys, to the BioData Catalyst developers on common issues encountered during training that can be included in the development pipeline. Including a user feedback component within the training effort will further increase the touchpoints with potential users and serve to ultimately ground the platform development in the broader community needs. Training resources will evolve to meet the needs of the BDCatalyst community.

4.2 USER ENGAGEMENT

Along with plans to engage users through training, we recognize that BioData Catalyst development teams will need to interact with specific users through the development of the Features that support a particular User Narrative. These users will be carefully identified in concert with the Steering Committee, NHLBI, and Data Stewards to ensure a handful of knowledgeable and appropriate users are virtually embedded within the development teams. This time commitment will necessitate compensation for time as consultants to the project.

The users will be provided with specific training and support documentation by the BioData Catalyst teams so they can contribute to and test a User Narrative beginning at the earliest point of development. BDC3 will organize a centralized library of training and support resources.

Preliminary advertising of BioData Catalyst to the general user base will be through publications announcing the platform and its capabilities.

In addition, the BioData Catalyst Help Desk will function as a connection point with the broader community of users and surface bugs and features that will both inform platform and services developers and training and support resources.

Community Outreach

The end goal of BioData Catalyst community outreach is a sustainable and engaged community of scientific researchers who both use and invest in the BioData Catalyst platform because of the utility it presents to their research. The BioData Catalyst Consortium will seek to help BioData Catalyst partners grow their own training programs that can expand upon existing training materials. The BDC3 will provide support to these efforts through assistance in materials development, video conferencing, and/or meeting hosting in an effort to establish users as training leaders, as well as facilitating integration of these materials into the training library.

4.3 ASSESSMENT

Successful operationalization of the Implementation Plan will require ongoing, but targeted, assessment. Assessment is directed towards gaining insight into alignment between the BioData Catalyst Strategic Vision, the particular Work Streams driving implementation, and feedback from users. In-progress assessment related to the development of Epics and Work Plan activities is built into the Agile development process. Project management assessment and operational metrics are addressed in the Project Management Plan.

Training assessment includes activities such as:

- Online Pre-, Post-, and Follow-up Short Surveys

Fellows and Beta-User assessment includes activities such as:

- Focus Group/Ethnographic analysis
- Online Pre-, Post-, and Follow-up Short Surveys

Community Outreach assessment includes activities such as:

- Tracking Consortium membership recruitment
- Usage patterns of the platform
- Identification of new collaborative projects

Training workshops will utilize Poll Everywhere or a similar tool to gather feedback on the training experience when the event closes. Additional surveys to gather later impressions will be circulated; the results synthesized and incorporated in the next round of training.

5 SUMMARY

Over the course of this Implementation Plan, we anticipate that there will be some evolution of the User Stories and Features as science and technology evolve. However, by 2021, we anticipate that the BioData Catalyst ecosystem will serve as a novel, fully-functioning resource in which users from a variety of disciplines and levels can perform complex operations and access newly-available scientific data to make significant strides in research and beyond.

6 REFERENCE DOCUMENTS

- Strategic Framework Plan
- Project Management Plan
- NHLBI DataSTAGE 60 Day o16n Plan v1-2 (drafted by the Operationalization Tiger Team)
- DataSTAGE User Narratives, Features, and Epics
- STAGE-RFC-2_DataSTAGE_Strategic_Planning_Nomenclature

REFERENCES

"A Strategic Framework for BioData Catalyst, v.2.0", BioData Catalyst Coordination Center, March 2020. Strategic Framework

"Data Storage, Toolspace, Access, and Analytics for biG-data Empowerment (DataSTAGE) Project Management Plan, v.1.0", DataSTAGE Coordination Center, February 2019. Project Management Plan

"STAGE-RFC-2 DataSTAGE Strategic Planning Nomenclature", DataSTAGE Coordination Center, February 2019. STAGE-RFC-2_DataSTAGE_Strategic_Planning_Nomenclature

"DataSTAGE", Jonathan Kaltman, presented to Dr. Gibbons, NHLBI, Jan 22, 2019. Jon's presentation on Jan 22 to NHLBI/Dr. Gibbons

"DataSTAGE User Narratives, Features, and Epics", DataSTAGE Consortium Members, February 2019. DataSTAGE User Narratives, Features, and Epics

"Training and User Engagement Plan by Copper Team 2018", 5M.5. Product, DCPPC Copper Team (Titus Brown et al), 2018, Training and User Engagement Plan

"Progress Toward Cancer Data Ecosystems". Grossman, Robert L., PhD. The Cancer Journal: May/June 2018 - Volume 24 - Issue 3 - p 126–130 doi: 10.1097/PPO.0000000000000318

"STRIDES Initiative – NIH Common Fund." National Institutes of Health, U.S. Department of Health and Human Services, commonfund.nih.gov/strides.