Welcome to Day 1...

NIH Cloud Platforms Interoperability Fall 2021 Workshop

We'll be starting shortly!

Welcome

Stan Ahalt, Patrick Patton

Virtual Meeting Roles (Patton)

Role	Purpose	Assignee & Slack	
Maestro: Mute Master, Raised-Hand Monitor, & Security	Master of Zoom Ceremonies. Contact Amanda for questions about Zoom issues, breakout rooms, or other general questions or if you notice suspicious activity.	@Amanda Miller (amiller@renci.org)	
Screen Sharing	Will share screen and advance slides.	@Julie Hayes	
Slide Content	Will update slide content throughout the meeting.	@Sarah Davis	
Moderator	Moderator listed for each agenda item. Moderator will prompt slide transitions during presentations and foster productive conversation during discussions.	Becky Boyles (@rboyles)	Stan Ahalt (@stan)
Plenary Notetakers	All are encouraged to add comments to the Homepage and Meeting Notes	@Patrick Patton @Paul Kerr @Allie Gartland-Gray	@Joe Asare @Tom Madden @John Cheadle
Q&A Monitor	Monitor questions in #oct_workshop Slack channel as well as Zoom Chat. **Share Action Items, Decisions, and Outstanding Questions from Slack and Zoom** to the Homepage and Meeting Notes		
Time Watcher	Will try to keep us on time while still allowing room for important conversations.	@Sarah Davis	

Questions during the event? (Patton)

Verbal Questions: There will be time for questions throughout the meeting. If you want to verbally ask a question, use the Zoom feature to "raise your hand" and the host will enable your audio and then call on you to ask your question.

Zoom Chat: You can type questions via Zoom Chat throughout the meeting. Paul Kerr, Patrick Patton, Joe Asare, Allie Gartland-Gray, Tom Madden and John Cheadle will share questions from Slack and Zoom chat into the **Homepage and Meeting Notes**.

Slack: Questions can be asked throughout the meeting by using the #oct_workshop Slack channel. We encourage anyone to write questions, comments, answers, or discussion in Slack at any time. If you have not received an invitation to #oct_workshop, please email amiller@renci.org.

Zoom download…. (Patton)

The latest version

Want the ability to move independently between breakout sessions?

We updated the meeting settings to allow attendees to move freely between the breakout rooms. **This setting requires the latest version of Zoom.**

- **Follow these instructions or**
- **Watch this how-to video here: https://youtu.be/E7zERcVLUBM**

Registration (Patton)

BDC3 will reach out to attendees who have not yet registered to ensure they register via the form (**bit.ly/NCPI2021_Register**).

Note that future invitation lists are determined using past registration lists.

BDCatalyst Statement of Conduct (Ahalt)

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.

Community Rules of Engagement (Ahalt)

BDCatalyst “Santa Cruz Rules of Engagement”:

- Do not shy away from identifying problems & risks
- Be candid
- Be heard
  - Identify an ally or motivate via Slack
  - Reach out to a Contact for particular topic(s) - Slack or email bdc3@renci.org if you don’t know the Contact
- Be polite
  - If you are a “talker” remember to give others time/space to talk - if you are “quiet”, take advantage of any opening
  - Add your comments/ideas to notes if you don’t find space to talk!

Connecting Data, Enhancing Software…What Does a Data Ecosystem Look Like?

Susan Gregurick

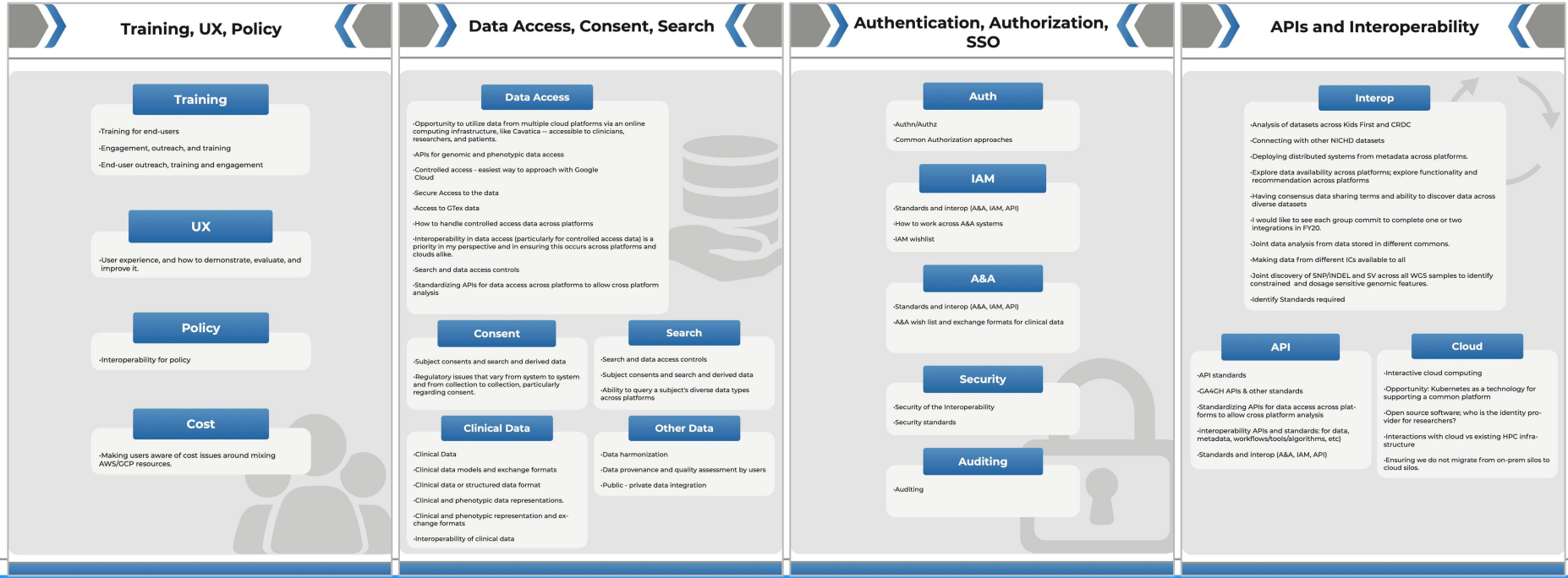
Goals Day 1:

Calibrate, Catalog, Identify Gaps/Challenges

Stan Ahalt

NCPI: Marking Progress (Ahalt)

- 2-years since first NCPI meeting in Chapel Hill, which focused on brainstorming the world of potential activities.



Since then... (Ahalt)

- 5 working groups moving forward on policy and development
- Multiple use cases driving progress
  - E.g. BDCatalyst used funds from ODSS and NHLBI to support development of interoperable AuthZN methods, search capabilities, semantic harmonization, and cross-platform compute on Kids First and AnVIL
  - More updates on driving use cases on Day 2

See all the good work accomplished to date in the Working Group Executive Summaries.

Addressing NIH/ODSS Goals (Ahalt)

What Does a Data Ecosystem Look Like?

Data, Software enhanced to support the FAIR and CARE Principles

Plan prospectively on how you will handle data;

Repository's ability to easily share data, metadata, and enhance findability across repositories

Software engineering and best practices enhanced for data science

Enhance an open community to work and communicate on software engineering

Cloud-enabled data analytics platforms can cross siloed boundaries, enable greater usability for researchers

- Participants in studies easily findable, data disambiguated
- Sustainability, sharing data, making available metadata and standards more compatible across systems

New colleagues trained and able to work across biomedical and data, information, computer science domains

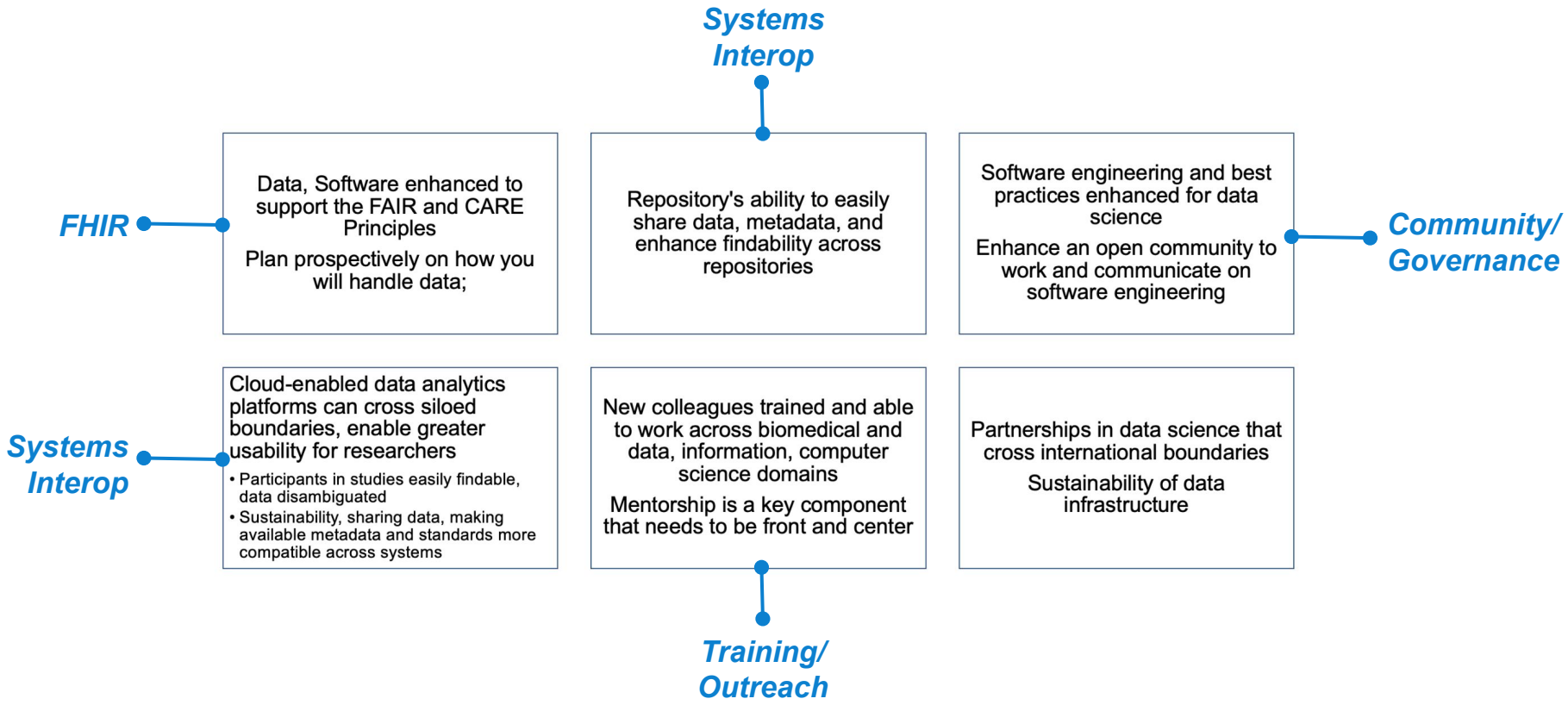
Mentorship is a key component that needs to be front and center

Partnerships in data science that cross international boundaries

Sustainability of data infrastructure

From Susan Gregurick, "Connecting Data, Enhancing Software…What Does a Data Ecosystem Look Like?", October 5, 2021.

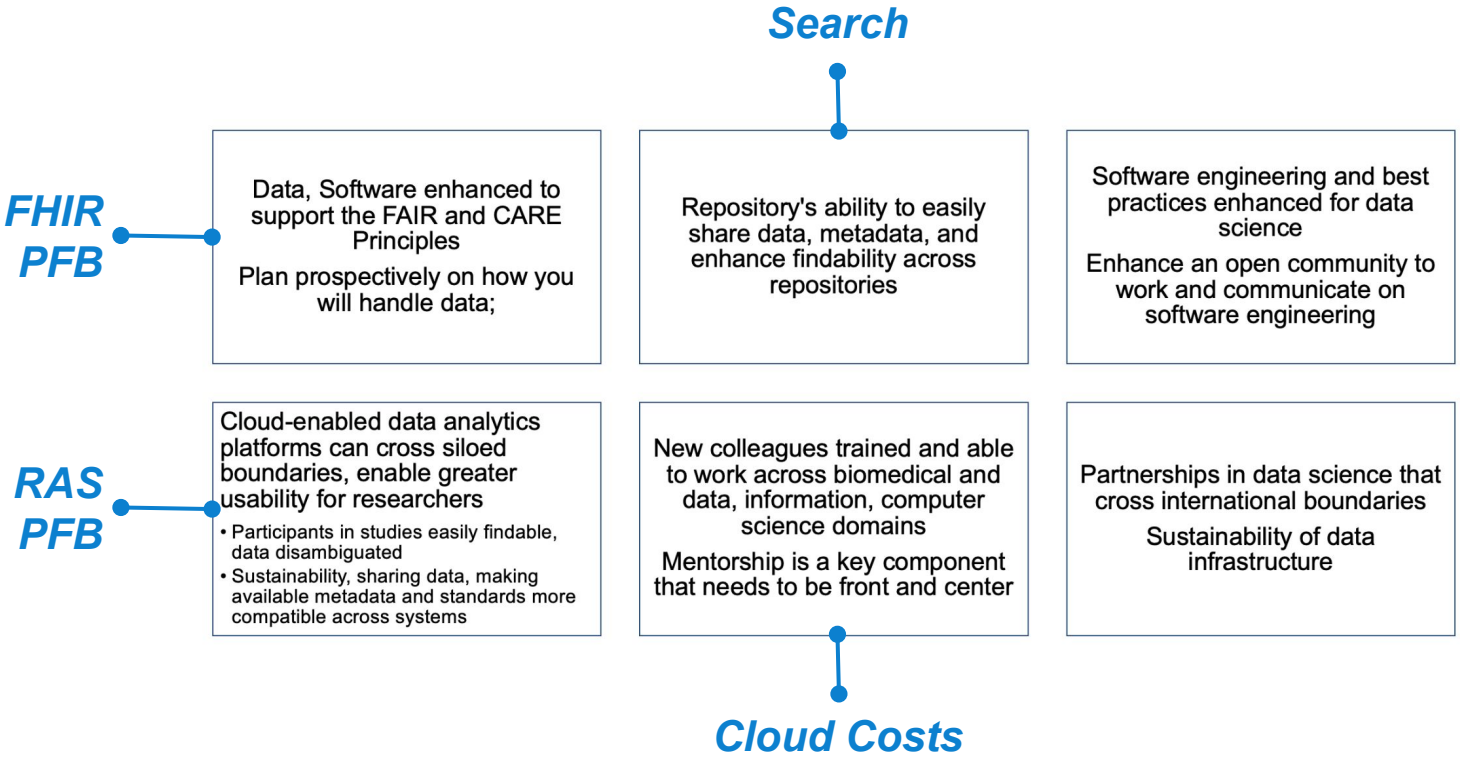
NCPI Working Groups (Ahalt)



Workshop Goals: Getting to “done” (Ahalt)

- Meeting Agenda is focused on actionable key topics to help reach the ODSS goals
  - RAS
  - PFB
  - FHIR
  - End User Cloud Costs
  - Search
- Catch-all Other Interoperability Efforts gathers other activities that we are working on and what’s coming next

Key Topics (Ahalt)



Workshop Goals: Getting to “done” (Ahalt)

- How can we **move the needle forward** on each key topic?
- What is the status of **use cases driving progress**?
- Where are the **gaps** for these topics that might need new use cases?
- What are **policy and development blockers** and how can we unblock them?
- What are the **next key pieces** that will help reach NIH goals?

Agenda: Day 1 **All times ET** (Ahalt)

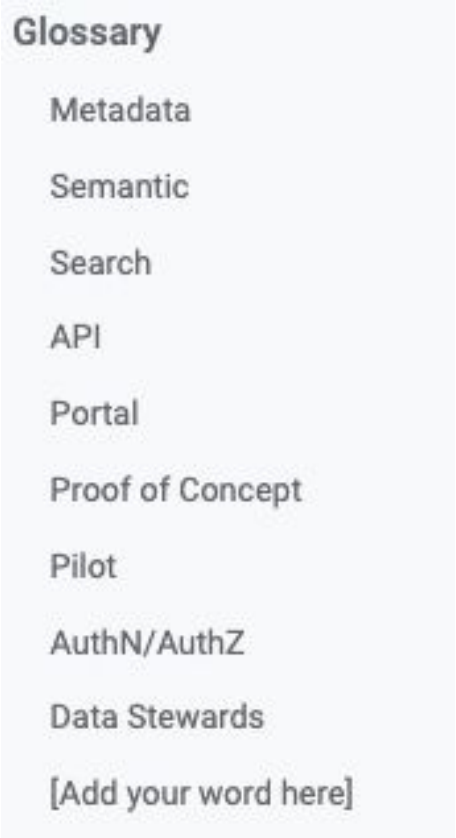
Time	Activity	Owner	Links
11:00-11:05am	Welcome	Stan Ahalt, Patrick Patton	Slides|Notes
11:05-11:40am	Connecting Data, Enhancing Software…What Does a Data Ecosystem Look Like?	Susan Gregurick	Slides|Notes
11:40-11:50am	Goals Day 1: Calibrate, Catalog, Identify Gaps/Challenges	Stan Ahalt	Slides|Notes
11:50 -12:15pm	Demo of Successful Federated Use Case (from search to FHIR to workspace)	Brian O'Connor, Jack DiGiovanna, Robert Carroll	Slides|Notes
12:15-1:00pm	Updates on Key Topics (Part 1) •PFB (10 min) (Grossman) •FHIR (15 min) (Carroll) •RAS (20 min) (O'Connor)	Moderator: Becky Boyles	Slides|Notes
1:00-1:45pm	Lunch Break		
1:15-1:45pm	Lunch Breakout 1: Discuss Gaps and Decide on Concrete Next Steps •RAS and data access (O'Connor)	Brian O'Connor	Slides|Notes
1:45-2:35pm	Updates on Key Topics (Part 2) •End-User Cloud Costs (20 min) (Schatz) •Search (20 min) (Rogers) •Other Interoperability Efforts (10 min) (Ahalt)	Moderator: Becky Boyles	Slides|Notes
2:35-3:05pm	Breakout Session 2: Discuss Gaps and Decide on Concrete Next Steps •PFB (VanTol) and FHIR (Carroll) •Other Interoperability Efforts (Ahalt)	Robert Carroll, Stan Ahalt	Slides|Notes
3:10-3:15pm	Break Plan for Day 2	Becky Boyles	Slides|Notes
3:10-4:00pm	Breakout Session 3: Discuss Gaps and Decide on Concrete Next Steps •End-user Cloud Costs (Schatz) •Search (Rogers) (EasyRetro)	Michael Schatz, David Rogers	Slides|Notes
Day 2: Wednesday, October 6			

Agenda: Day 2 **All times ET** (Ahalt)

Time	Activity	Owner	Links
11:00-11:10am	Welcome and Goals Day 2: Synthesize next steps, driving use cases, determine NIH/NCPI priorities	Stan Ahalt	Slides|Notes
11:10-12:40pm	Breakout Report Backs and Discussion •PFB (10 min) (Grossman) •FHIR (10 min) (Carroll) •RAS (20 min) (O'Connor) •End-User Cloud Costs (20 min) (Schatz) •Search (20 min) (Rogers) •Other Interoperability Efforts (10 min) (Ahalt)	Moderator: Becky Boyles	Slides|Notes
12:40-12:50pm	GA4GH Relationship	Brian O'Connor	Slides|Notes
12:50-2:00pm	Lunch Break		
1:30pm-2:00pm	NIH Breakout: NIH Coordination Working Group Discussion of Priority Next Steps	NIH Only (via separate invitation)	
2:00-2:15pm	Use Case Overview: The Journey of a NCPI Use Case	Asiyah Lin	Slides|Notes
2:15-3:20pm	Review of Current Scientific Use Cases	Moderator: Valentina Di Francesco	Slides|Notes
2:15-2:30pm	Leveraging Functionally Equivalent Pipelines for Long-Read Data on Different Systems	Owen Hirschi	Slides|Notes
2:30-2:50pm	Interoperability between Kids First & Undiagnosed Diseases Network (UDN) Data via dbGaP/SRA	Valerie Cotton, Allison Heath	Slides|Notes
2:50-3:05pm	Genetic Sex as a Biological Variable and X-inactivation	Melissa Wilson	Slides|Notes
3:05-3:20pm	Conducting reproducible science in PIC-SURE interoperating with Seven Bridges/Terra	Simran Makwana	Slides|Notes
3:20-4:00pm	Synthesize Goals and Next Steps for the next 6 Months, with focus on driving use cases	Stan Ahalt, Jon Kaltman	Slides|Notes

Meeting Deliverable: NCPI Glossary (Ahalt)

- While we often use the same words, we sometimes use them to mean different things.
- We hope this Glossary will be a concrete deliverable at the end of the meeting to help us coalesce around common definitions and/or highlight differences.
- Please review and add your definitions to listed words or add new words



Demo of Successful Federated Use Case:

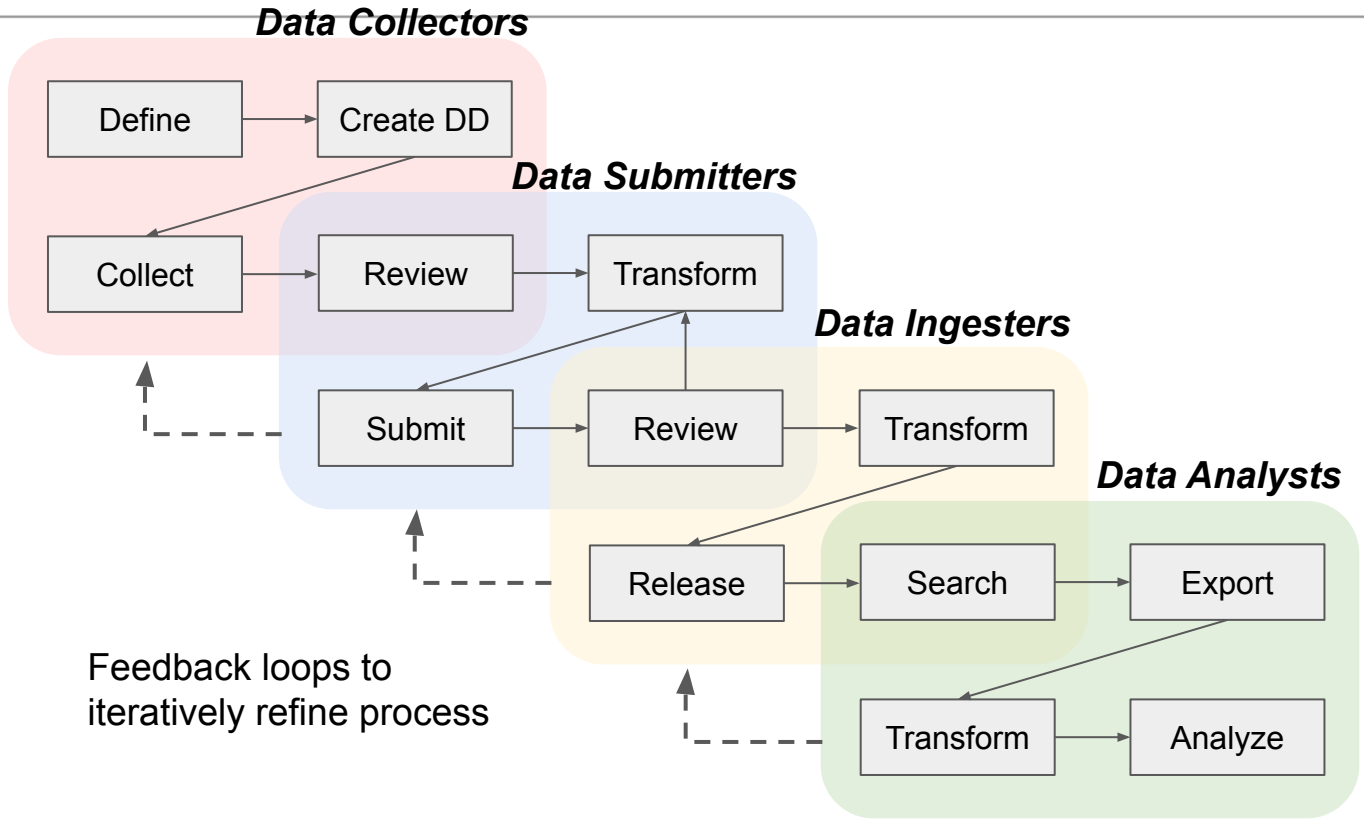
From Search to FHIR to Workspace

Brian O'Connor, Jack DiGiovanna, Robert Carroll

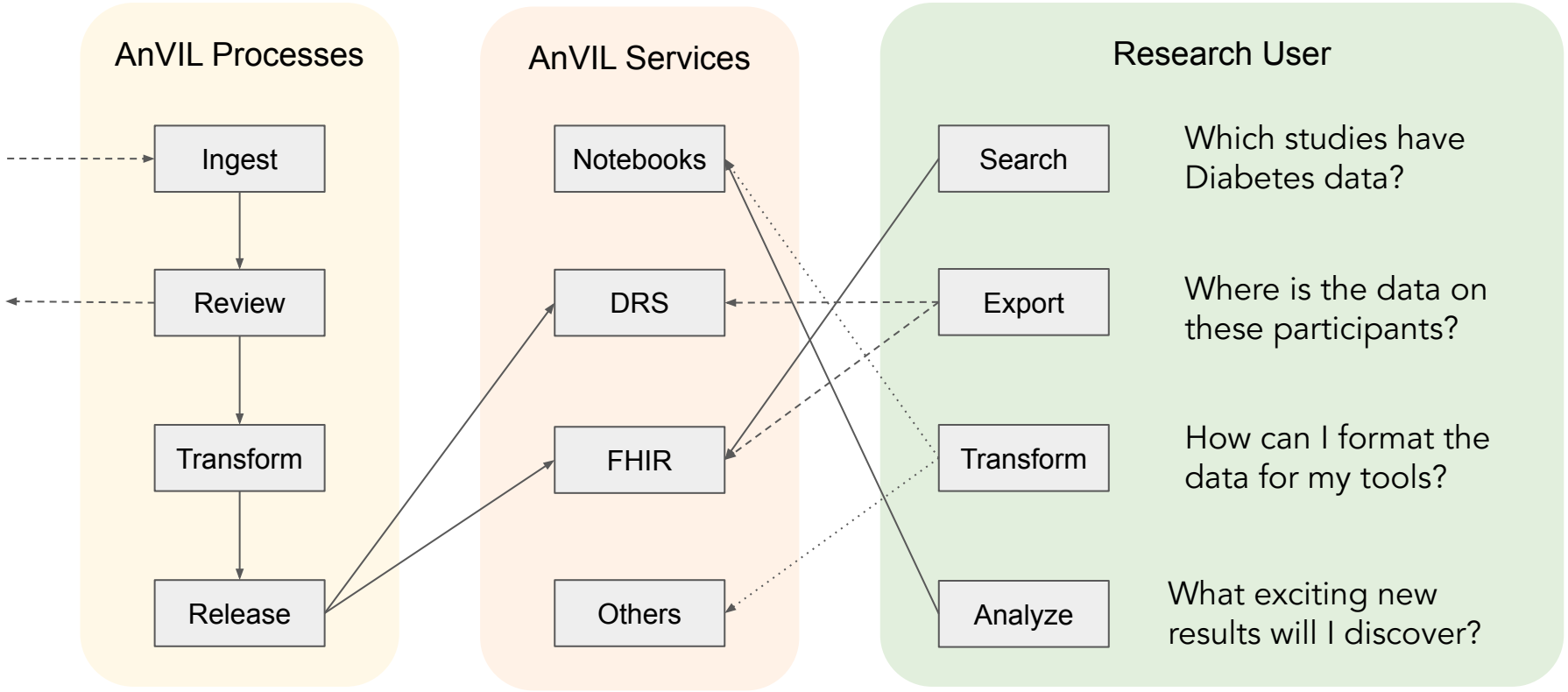
FHIR & Search

Data Access & Compute

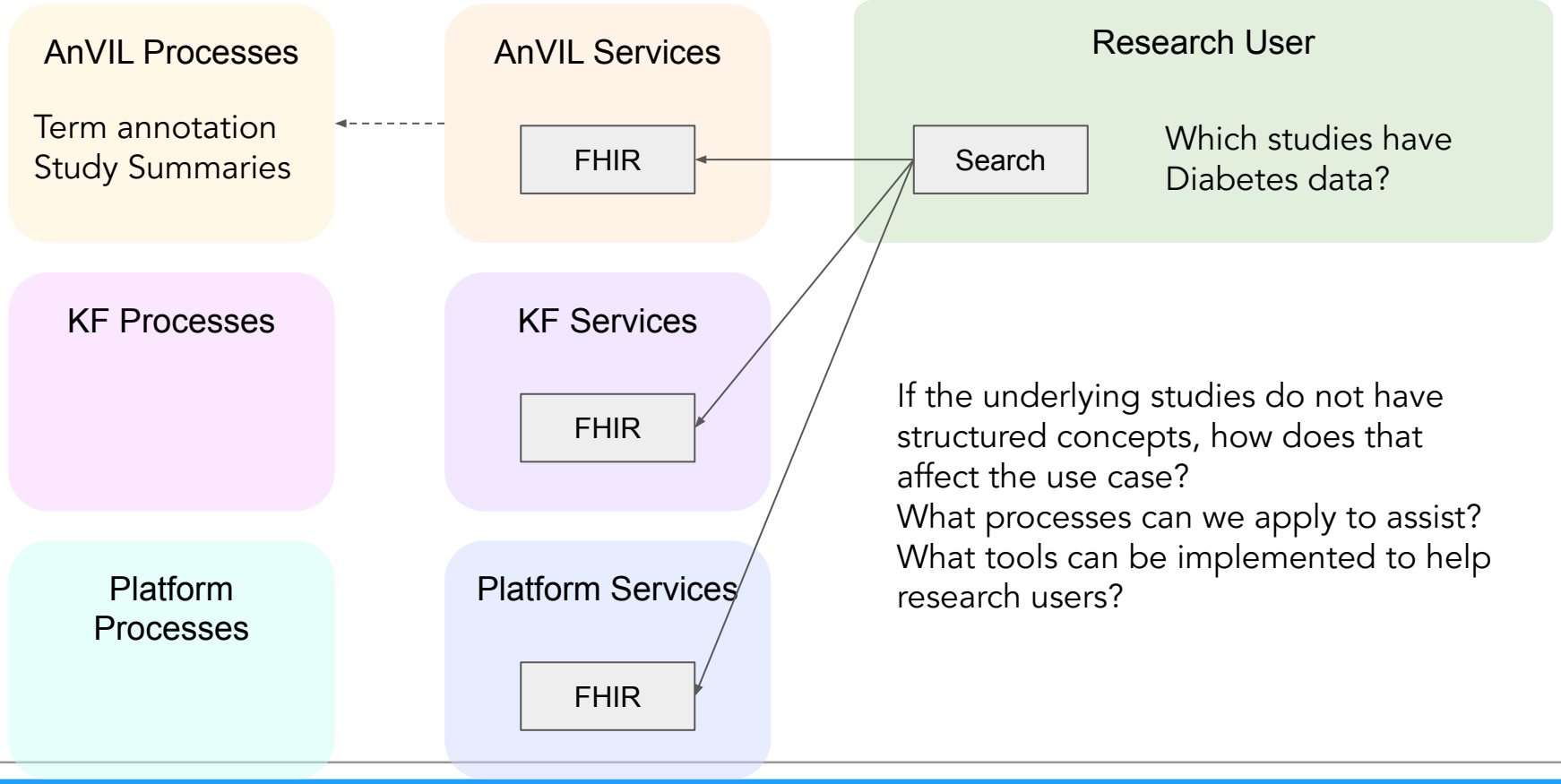
Data Life Cycle



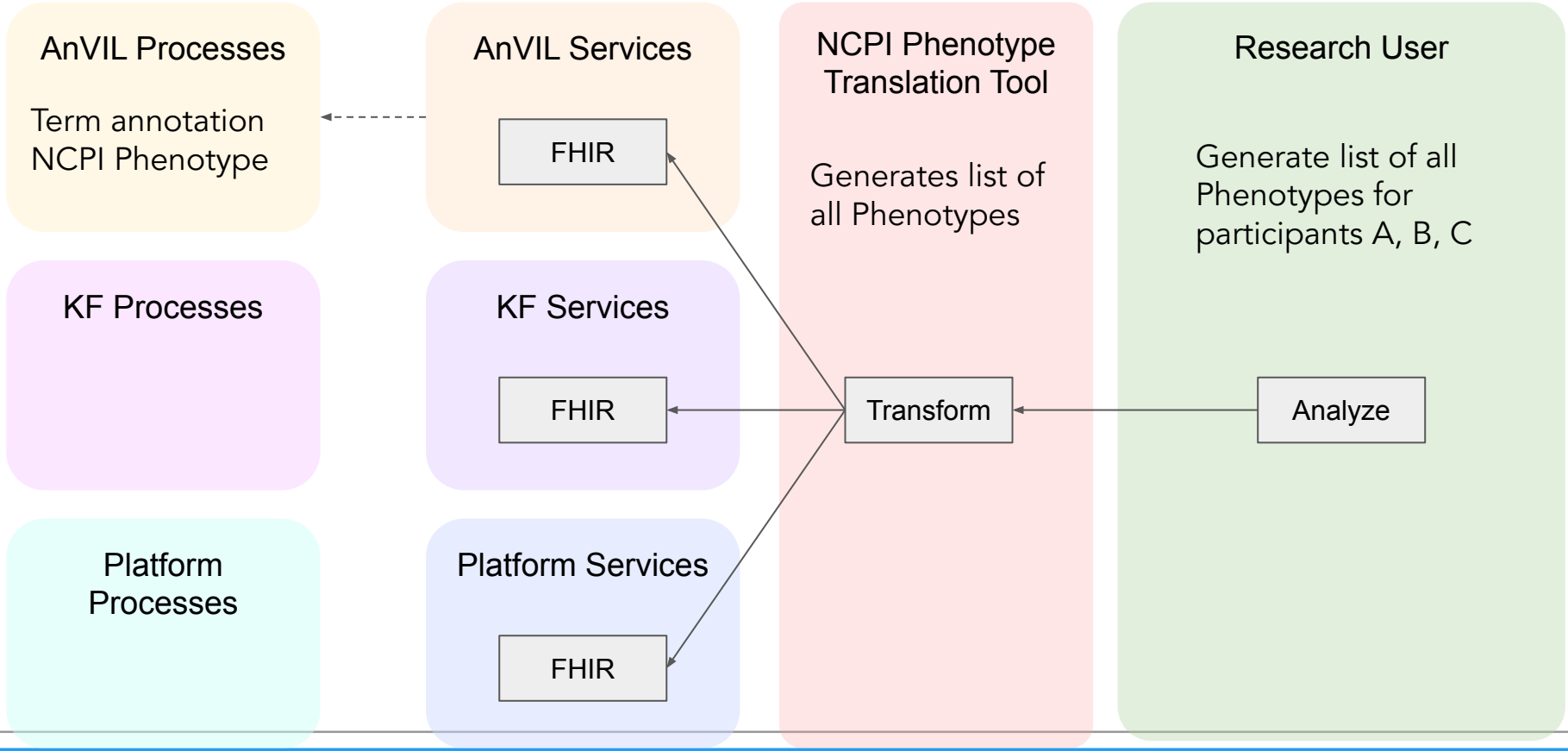
Platforms and Research Users



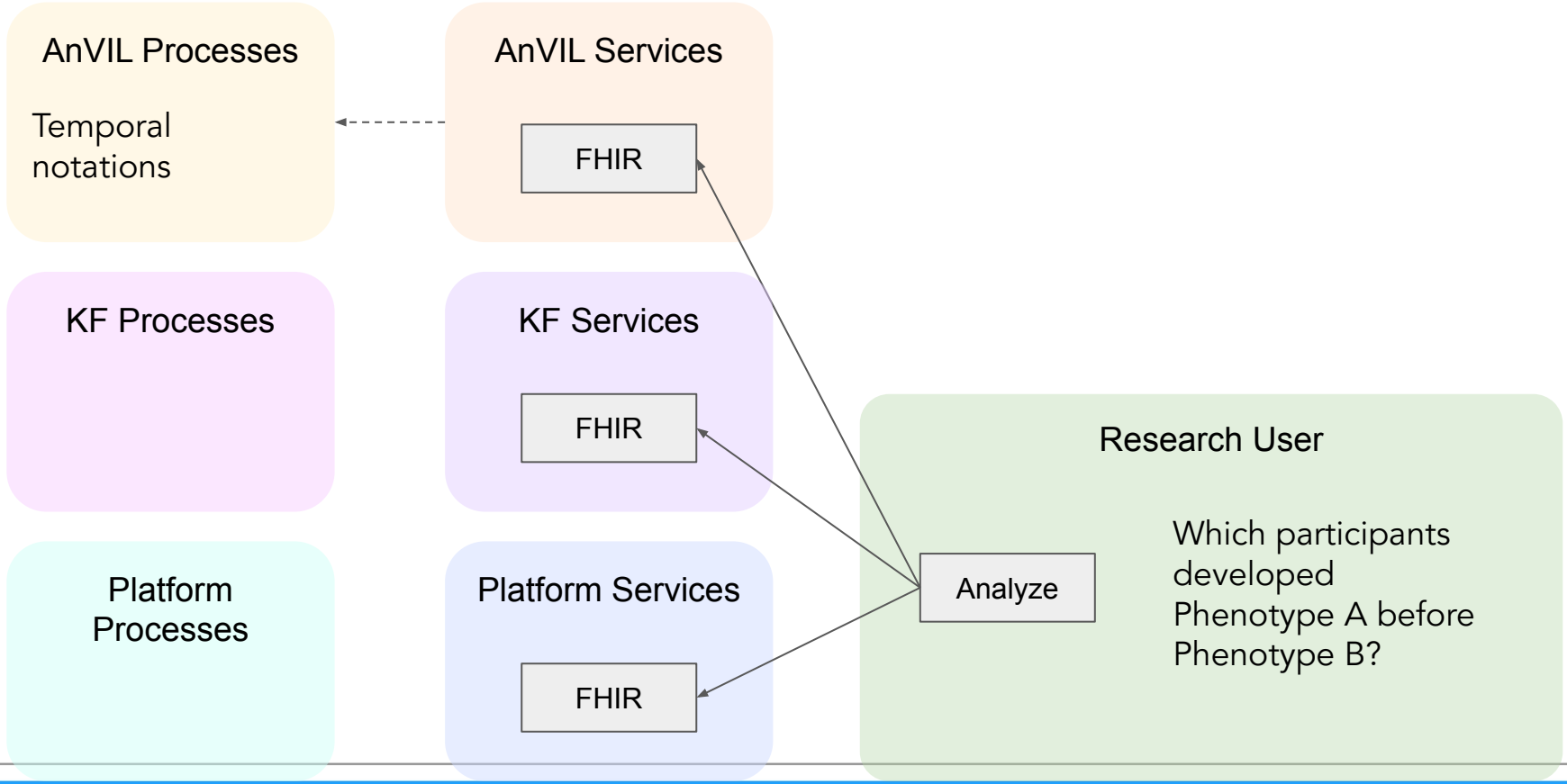
Study summary use cases



Additional tool / service layers



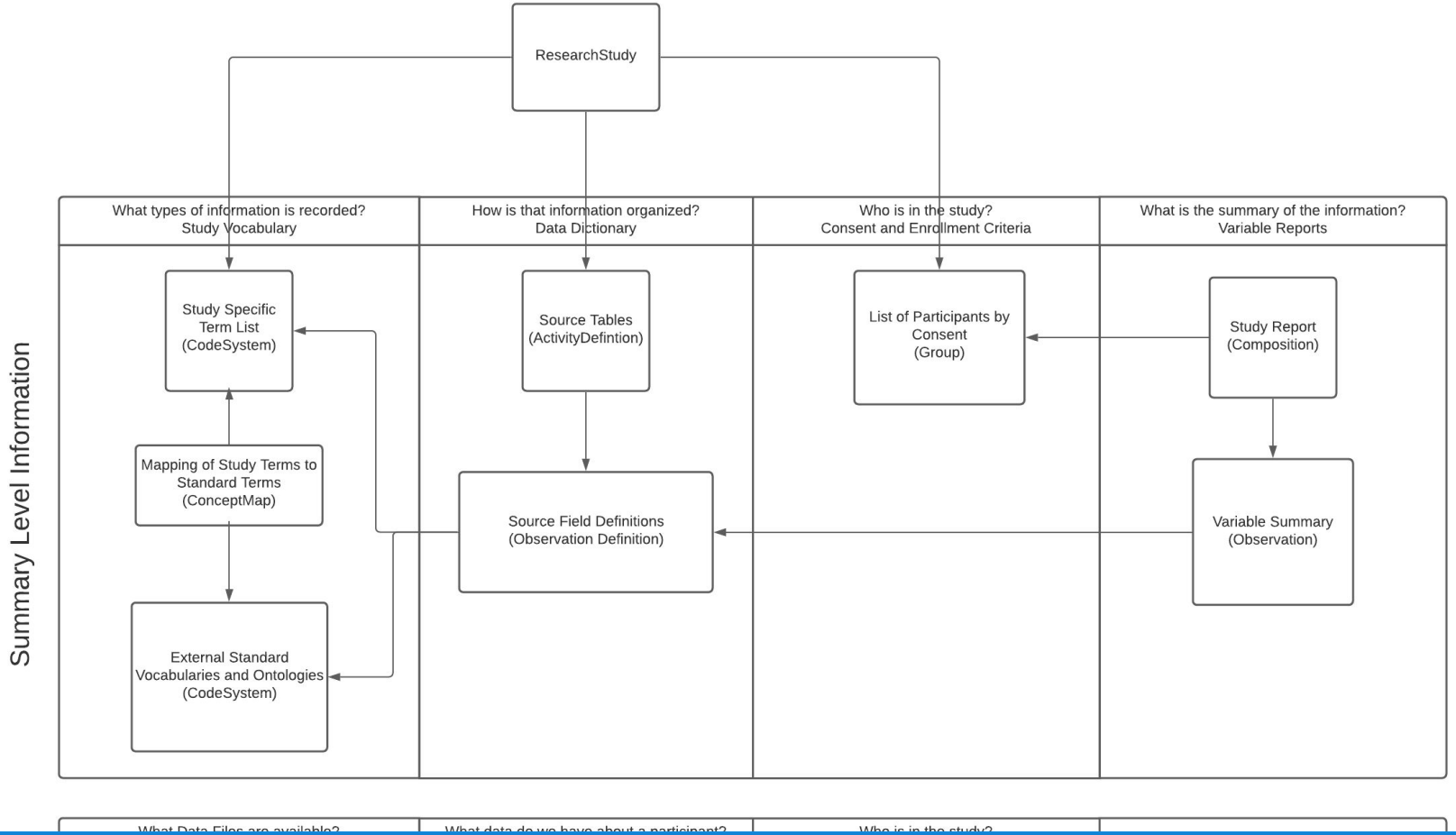
Platforms and Research Users



Representing Study Data

- Providing detailed study metadata is very important to understanding the data that's presented.
- dbGaP has set the standard for information that's available, and they are working towards "modernizing" the representation.
- This is currently organized in FHIR, but it's using a custom extension approach.
- We have built a proposal using more FHIR native approaches that should enable easy lift-over for existing data, programmatic definitions for new data, and structured links within the metadata.
- This model is developed in the context of researchers accessing existing research data.

Big picture


ResearchStudy
What types of information is recorded?
Study Vocabulary
How is that information organized?
Data Dictionary
Who is in the study?
Consent and Enrollment Criteria
What is the summary of the information?
Variable Reports
Summary Level Information
Study Specific Term List (CodeSystem)
Source Tables (ActivityDefintion)
List of Participants by Consent (Group)
Study Report (Composition)
Mapping of Study Terms to Standard Terms (ConceptMap)
Source Field Definitions (Observation Definition)
Variable Summary (Observation)
External Standard Vocabularies and Ontologies (CodeSystem)

Products

- FHIR Example: https://github.com/anvilproject/DD-On-FHIR
- Study Summary Tool: https://github.com/NIH-NCPI/ncpi-study-summary-generation-tool
- Study Browser Tool: https://github.com/NIH-NCPI/ncpi-fhir-study-summary-browser

Demo!

- Using the NCPI FHIR Implementation guide, we have several studies loaded into FHIR servers.
  - AnVIL internal test server on Google Healthcare API
  - KF development server running Smiles CDR on AWS
- Eric Torstenson developed and ran a ResearchStudy summary tool, which generated summary objects that could be made available publicly.
- I've written a quick Shiny app that looks at those summaries to generate some interactive content.
- Live demo if possible

NCPI Study Summary FHIR Browser

Select a study:

Show 10 entries Search:

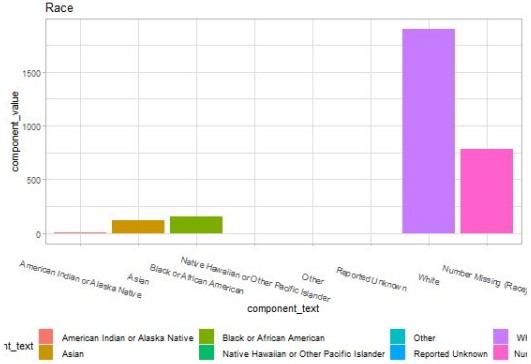
Study Title	Participants
Baylor Hopkins Center for Mendelian Genomics (BH CMG)	1621
National Heart, Lung, and Blood Institute (NHLBI) Bench to Bassinet Program: The Gabriella Miller Kids First Pediatric Research Program of the Pediatric Cardiac Genetics Consortium (PCGC)	2966
University of Washington Center for Mendelian Genomics (UW-CMG)	2802
Yale Center for Mendelian Genomics (Y CMG)	6979

Showing 1 to 4 of 4 entries Previous 1 Next

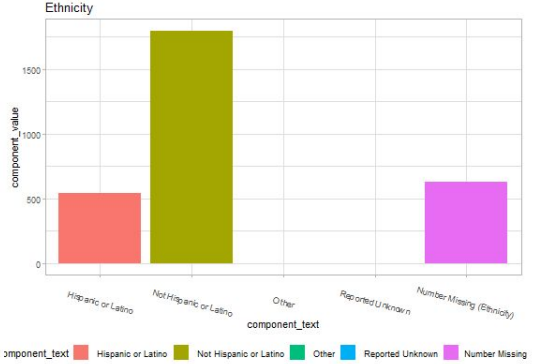
National Heart, Lung, and Blood Institute (NHLBI) Bench to Bassinet Program: The Gabriella Miller Kids First Pediatric Research Program of the Pediatric Cardiac Genetics Consortium (PCGC)

Study ID: f8fe498c-faa4-46eb-81b1-10231dac52d7

Race



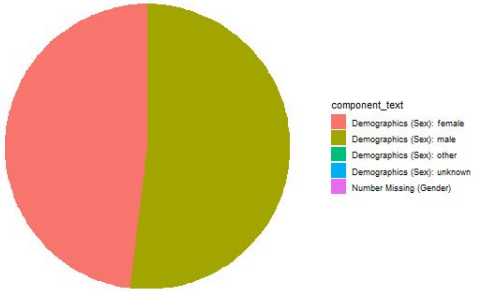
Ethnicity



Category	Participants
Demographics (Race): American Indian or Alaska Native	5
Demographics (Race): Asian	119
Demographics (Race): Black or African American	159
Demographics (Race): Native Hawaiian or Other Pacific Islander	2
Demographics (Race): White	1897
Demographics (Race): Other	0
Demographics (Race): Reported Unknown	0
Number Missing (Race)	784

Category	Participant
Demographics (Sex): male	154
Demographics (Sex): female	142
Demographics (Sex): other	
Demographics (Sex): unknown	
Number Missing (Gender)	

Gender



NCPI Study Summary FHIR Browser

National Heart, Lung, and Blood Institute (NHLBI) Bench to Bassinet Program: The Gabriella Miller Kids First Pediatric Research Program of the Pediatric Cardiac Genetics Consortium (PCGC)

Study ID: f8fe498c-faa4-46eb-81b1-10231dac52d7

List of groups in this study:

Show 10 entries Search:

Group Name	Participants
SD_PREASA7S-complete	2966

Showing 1 to 1 of 1 entries Previous 1 Next

Phenotype Summary

Show 10 entries Search:

Phenotype	Phenotype Present	Phenotype Absent	Number Missing Phenotype	Phenotype Reported Unknown
Conotruncal Left-sided Lesion	545	191	2230	0
Abnormal Ventricular Septum	419	309	2230	8
Abnormal Ventriculo-arterial Connection	337	396	2230	3
Abnormal Pulmonary Valve	301	421	2230	14
Abnormal Atrial Septum	291	365	2230	80
Abnormal Aorta	285	387	2230	64
Abnormal Right Ventricle	243	491	2230	2
Abnormal Aortic Valve	227	394	2230	115
Abnormal Mitral Valve	194	539	2230	3
Left Ventricular Outflow Tract Obstruction	158	578	2230	0

Showing 1 to 10 of 367 entries Previous 1 2 3 4 5 … 37 Next

NCPI Study Summary FHIR Brow

localhost:1221/?state=jylxsl5FO1Zmu2ysbNvw&code=4/0AX4XfWghybWRaPjm9Ip_kMI08OurSvxqAhGqwuHD_pf0WCoegHwoef_20p2r-F3XEltq7Q&scope=https://www.googleapis.com/auth/cloud-platform

NCPI Study Summary FHIR Browser

=> | ResearchStudy Browser | Study Phenotypes Browser | Configuration

Select a study:

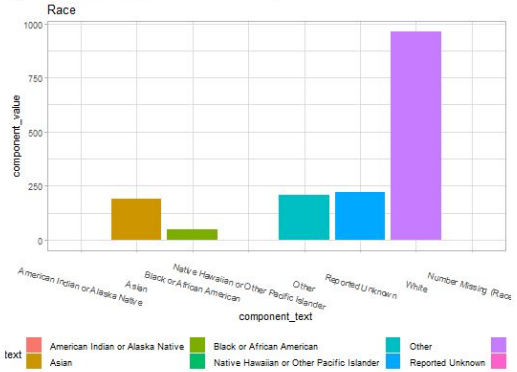
Show 10 entries Search:

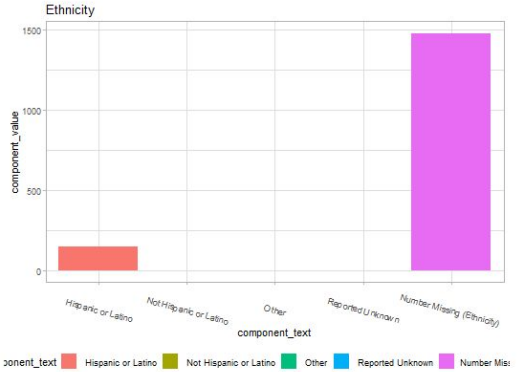
Study Title	Participants
Baylor Hopkins Center for Mendelian Genomics (BH CMG)	1621
National Heart, Lung, and Blood Institute (NHLBI) Bench to Bassinet Program: The Gabriella Miller Kids First Pediatric Research Program of the Pediatric Cardiac Genetics Consortium (PCGC)	2966
University of Washington Center for Mendelian Genomics (UW-CMG)	2802
Yale Center for Mendelian Genomics (Y CMG)	6979

Showing 1 to 4 of 4 entries Previous 1 Next

Baylor Hopkins Center for Mendelian Genomics (BH CMG)

Study ID: cdbdac2b-ff44-4c30-bdaf-e0b806851713

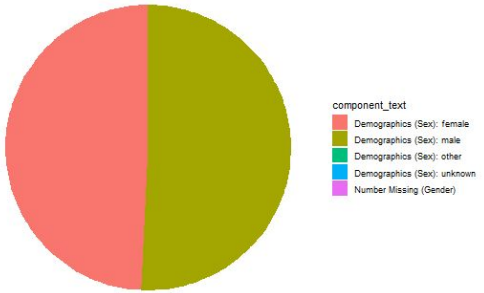




Category	Participants
Demographics (Race): American Indian or Alaska Native	0
Demographics (Race): Asian	189
Demographics (Race): Black or African American	46
Demographics (Race): Native Hawaiian or Other Pacific Islander	0
Demographics (Race): White	961
Demographics (Race): Other	207
Demographics (Race): Reported Unknown	218
Number Missing (Race)	0

Category	Participa
Demographics (Sex): male	
Demographics (Sex): female	
Demographics (Sex): other	
Demographics (Sex): unknown	
Number Missing (Gender)	

Gender



NCPI Study Summary FHIR Browser

Baylor Hopkins Center for Mendelian Genomics (BH CMG)

Study ID: cdbdac2b-ff44-4c30-bdaf-e0b806851713

List of groups in this study:

Show 10 entries Search:

Group Name	Participants
HMB-IRB-NPU	804
HMB-NPU	817
BH_CMG-complete	1621

Showing 1 to 3 of 3 entries Previous 1 Next

Phenotype Summary

Show 10 entries Search:

Phenotype	Phenotype Present	Phenotype Absent	Number Missing Phenotype
Global developmental delay	79	0	1542
Scoliosis	76	0	1545
Joint laxity	61	0	1560
Microcephaly	56	0	1565
Hypotonia	51	0	1570
Seizure	46	0	1575
Expressive language delay	45	0	1576
Decreased body weight	43	0	1578
Proportionate short stature	40	0	1581
High palate	39	1	1581

Showing 1 to 10 of 1,312 entries Previous 1 2 3 4 5 … 132 Next

NCPI Study Summary FHIR Browser

Baylor Hopkins Center for Mendelian Genomics (BH CMG)

Study ID: cdbdac2b-ff44-4c30-bdaf-e0b806851713

List of groups in this study:

Show 10 entries Search:

Group Name	Participants
HMB-IRB-NPU	804
HMB-NPU	817
BH_CMG-complete	1621

Showing 1 to 3 of 3 entries Previous 1 Next

Phenotype Summary

Show 10 entries Search:

Phenotype	Phenotype Present	Phenotype Absent	Number Missing Phenotype
Global developmental delay	42	0	775
Scoliosis	41	0	776
Microcephaly	34	0	783
Seizure	28	0	789
Intellectual disability	22	0	795
Expressive language delay	22	0	795
Peripheral neuropathy	22	0	795
Recurrent infections	21	0	796
Intellectual disability, moderate	15	0	802
Abnormality of the face	15	0	802

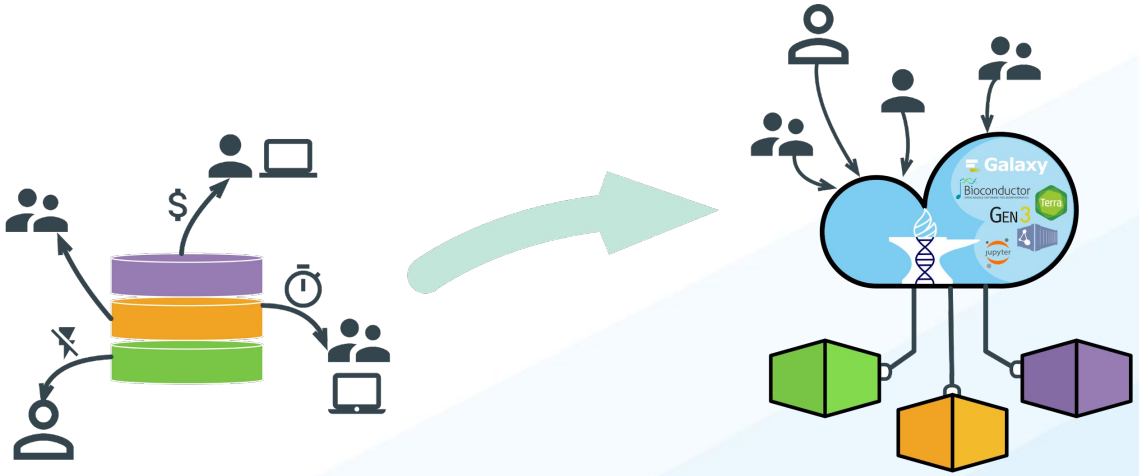
Showing 1 to 10 of 727 entries Previous 1 2 3 4 5 … 73 Next

FHIR & Search

Data Access & Compute

Inverting the Model of Genomic Data Sharing

AnVIL, BioData Catalyst, CRDC, and GMKF have
~11PB of data accessible on the cloud for ~831K participants



Traditional: Bring data to the researcher

Goal: Bring researcher to the data

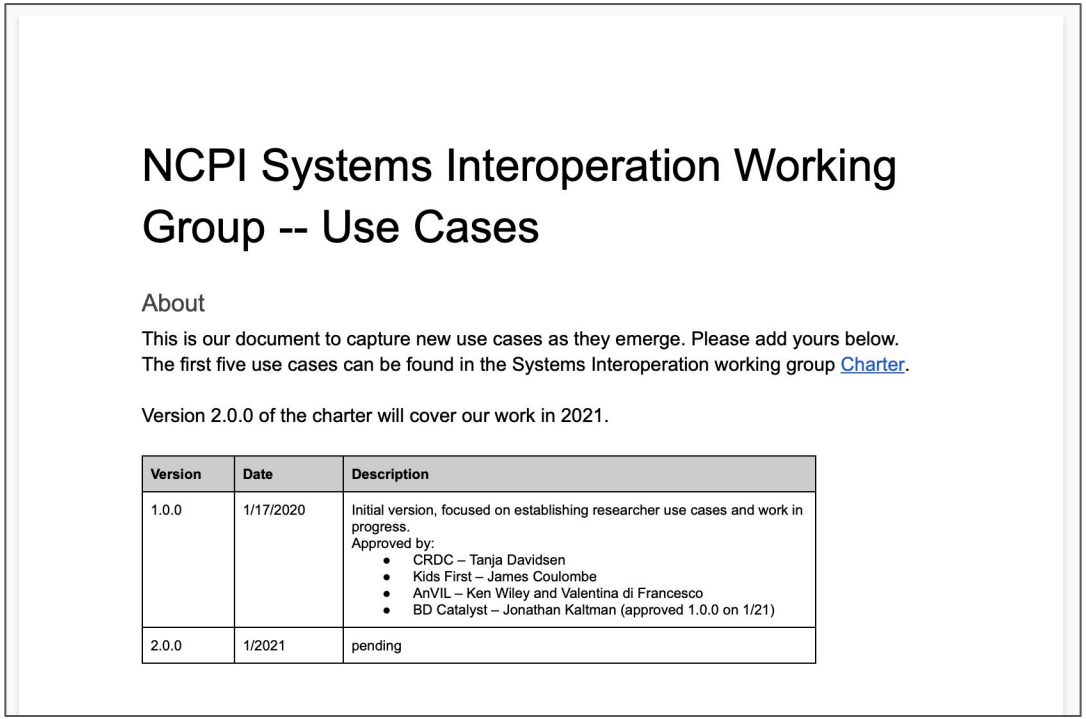
Schatz, Philippakis et al. (2021) *bioRxiv*. doi: https://doi.org/10.1101/2021.04.22.436044

NCPI Systems Interoperation WG

*The **NCPI Systems Interoperation Working Group** spearheads technical improvements to the NCPI participating cloud-based platforms that enable improved interoperability.*

https://anvilproject.org/ncpi

Researcher Use Cases Driving Work



NCPI Systems Interoperation Working Group -- Use Cases

About

This is our document to capture new use cases as they emerge. Please add yours below.
The first five use cases can be found in the Systems Interoperation working group Charter.

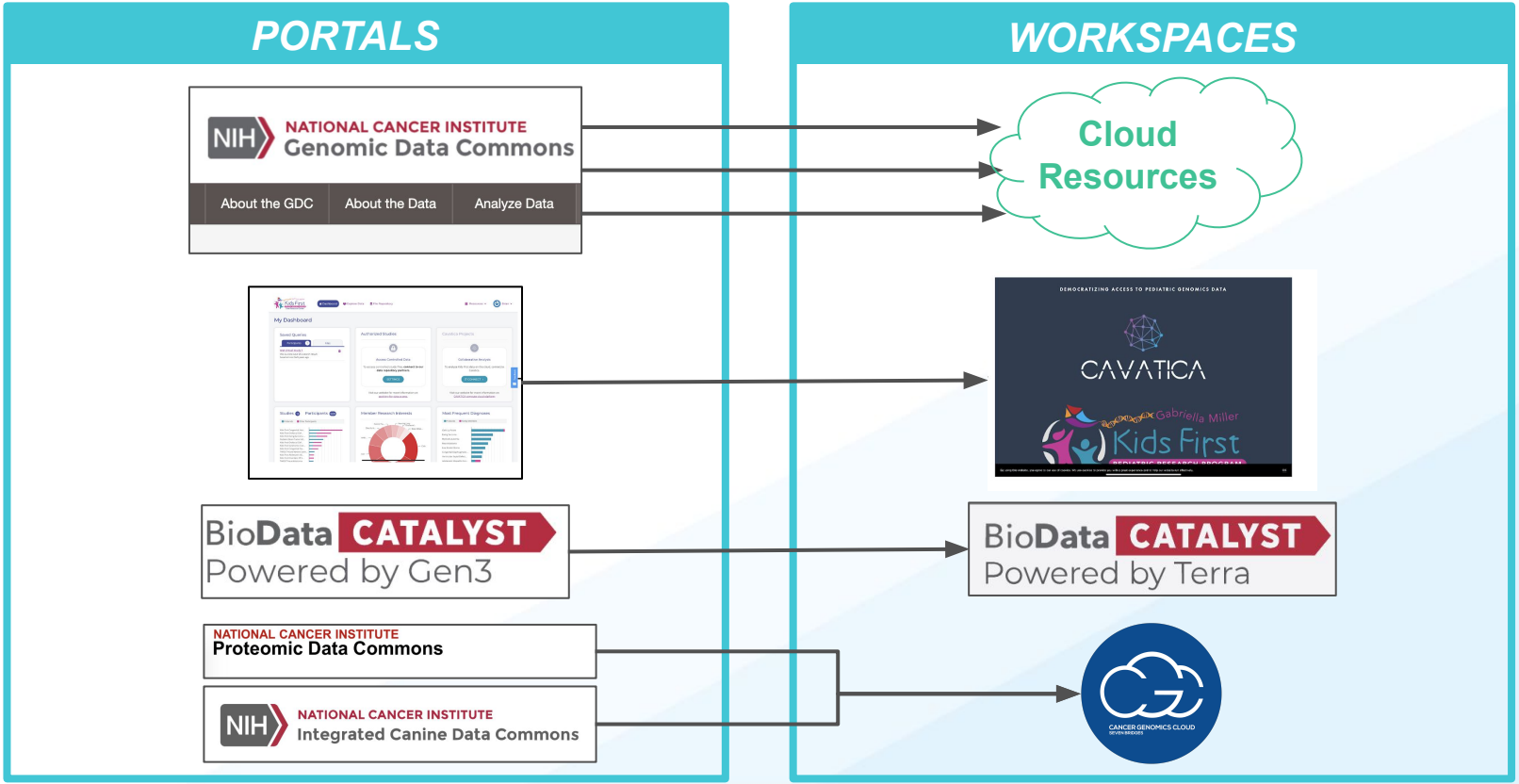
Version 2.0.0 of the charter will cover our work in 2021.

Version	Date	Description
1.0.0	1/17/2020	Initial version, focused on establishing researcher use cases and work in progress. Approved by: • CRDC – Tanja Davidsen • Kids First – James Coulombe • AnVIL – Ken Wiley and Valentina di Francesco • BD Catalyst – Jonathan Kaltman (approved 1.0.0 on 1/21)
2.0.0	1/2021	pending

*We worked with multiple researchers to define **11 driver use cases** for our work*

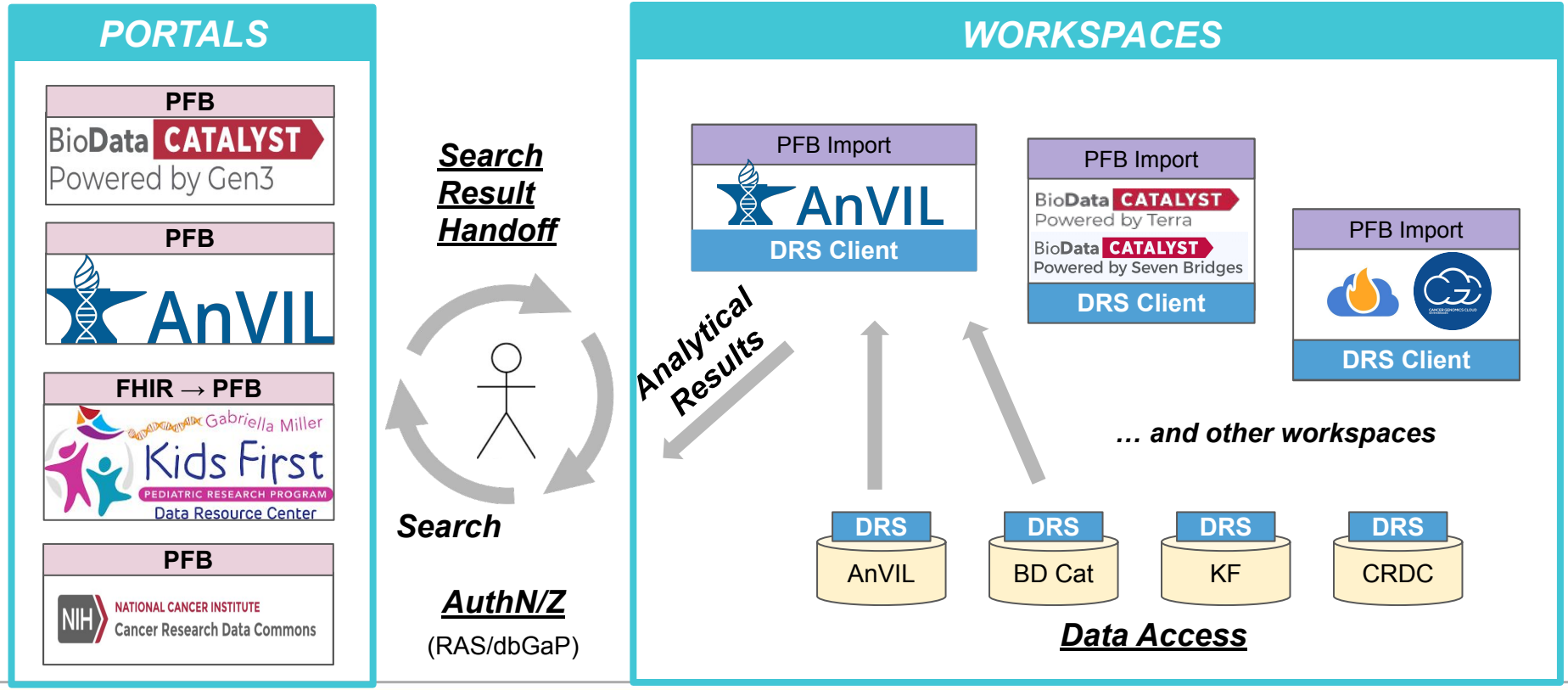
When NCPI Sys Interop Started (Jan 2020)

*Data **portals connect (intra-IC)** with **analysis systems (workspaces)***



Our Vision for Interoperability

*Data portals connect to any **workspaces (inter-IC)**, workspace access **data (inter-IC)***



3 Key Standards in NCPI Systems Interop

Search Result Handoff:
PFB (FHIR and Manifests)

Data Access: GA4GH DRS

Auth: RAS GA4GH Passports for AuthN/Z

NCPI Sys Interop's Progress

2020

- **MOUs/ISAs** for RAS and system interconnects
- **PFB for data handoff from portals to workspaces** (BDCat & AnVIL)
- **DRS for data access** to AnVIL, BDCat, Kids First, and CRDC
- Progress on **Researcher Use Cases**

2021

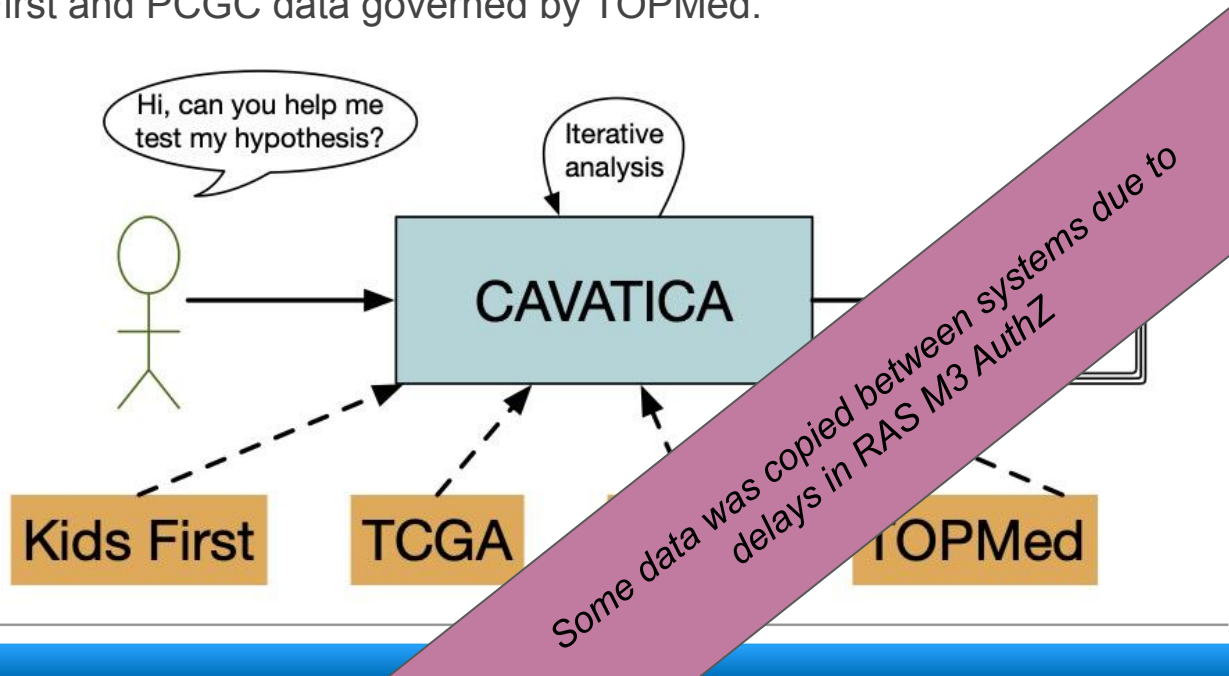
- NIH **RAS for authentication**
- **GA4GH standards** evolved (Passports, DRS, etc)
- More systems working on **PFB handoff** (PDC)
- Prototyping **FHIR → PFB bridge**
- More **workspaces supporting more DRS servers**
- **RAS Passports for Authorization designed**
- **Researcher Use Cases** finishing/expanding

Use Case Success Stories

Use Case #5: Wilson McKerrow et al. LINE1 analysis on the CGC spanned Proteomics Data Commons, TCGA, and GTEx

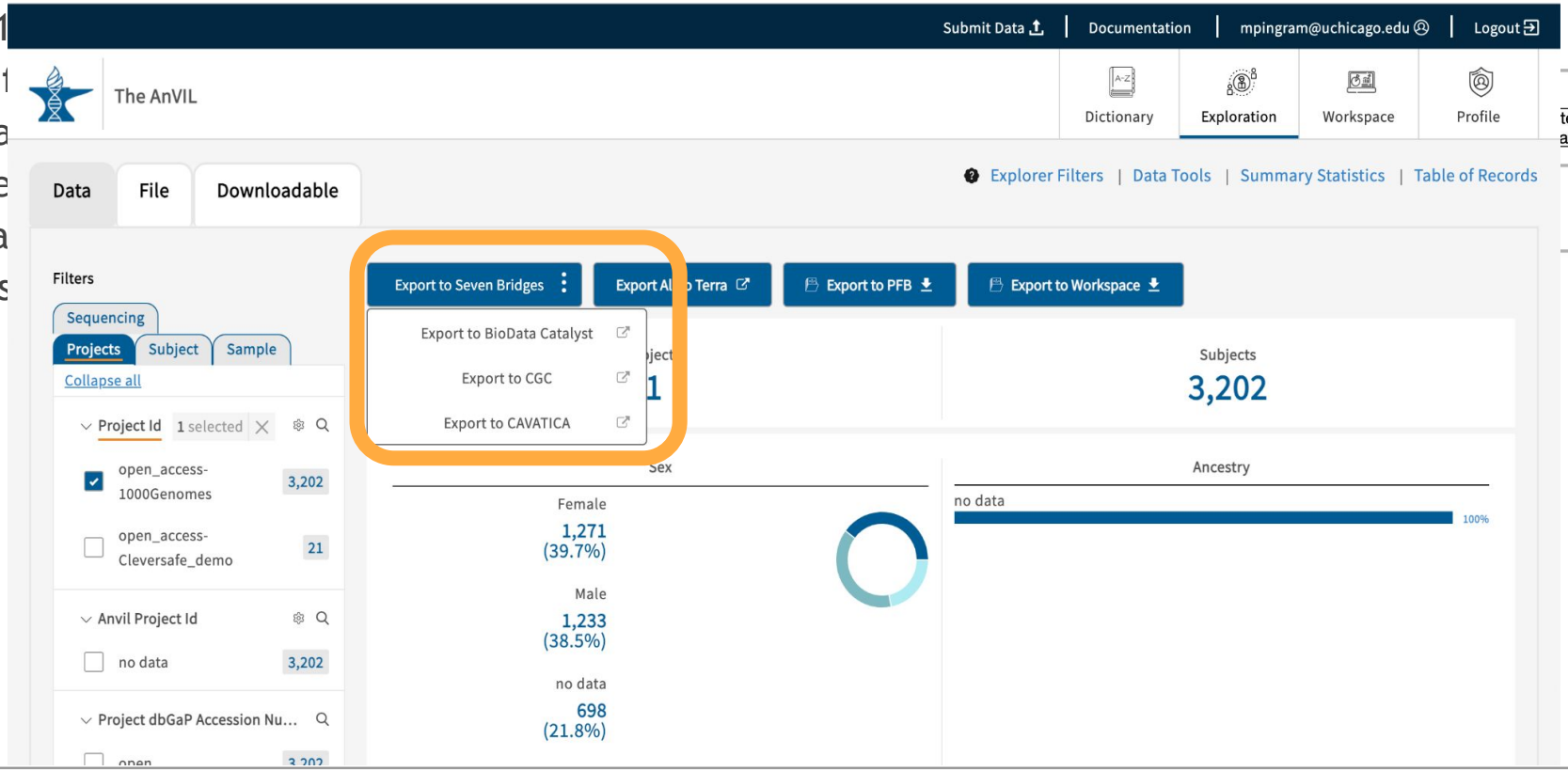
Use Case #1B: Deanne Taylor et al. PCGC analysis on CAVATICA and BDC powered by SB spanned the PCGC data governed by Kids First and PCGC data governed by TOPMed.

Proof of concept: KF, TCGA, GTEx, and TOPMed data in CAVATICA April 2021

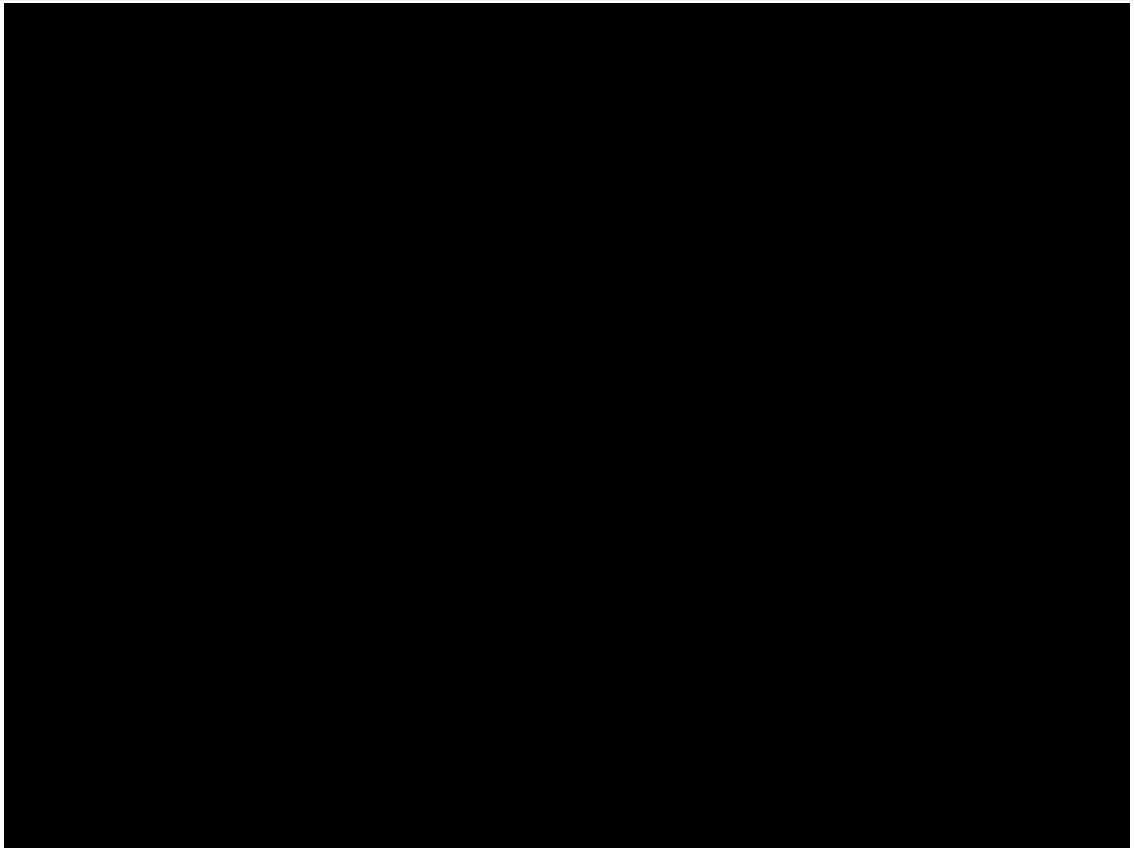


UX not yet optimal

Fast (<1
AuthN) f
but clea
improve
(RAS), a
user bas



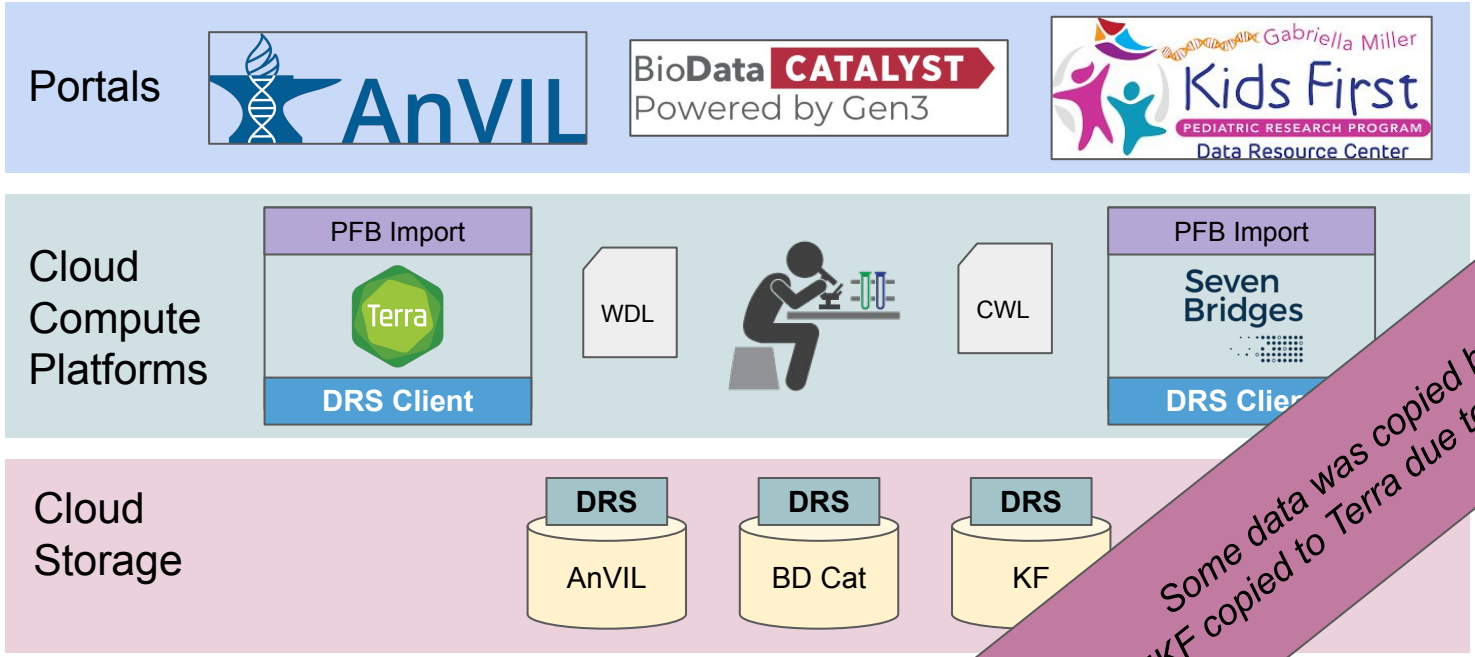
Improved UX



Use Case Success Stories

Use Case #7: Tim Majarian's cross dataset analysis for Congenital Heart Disease

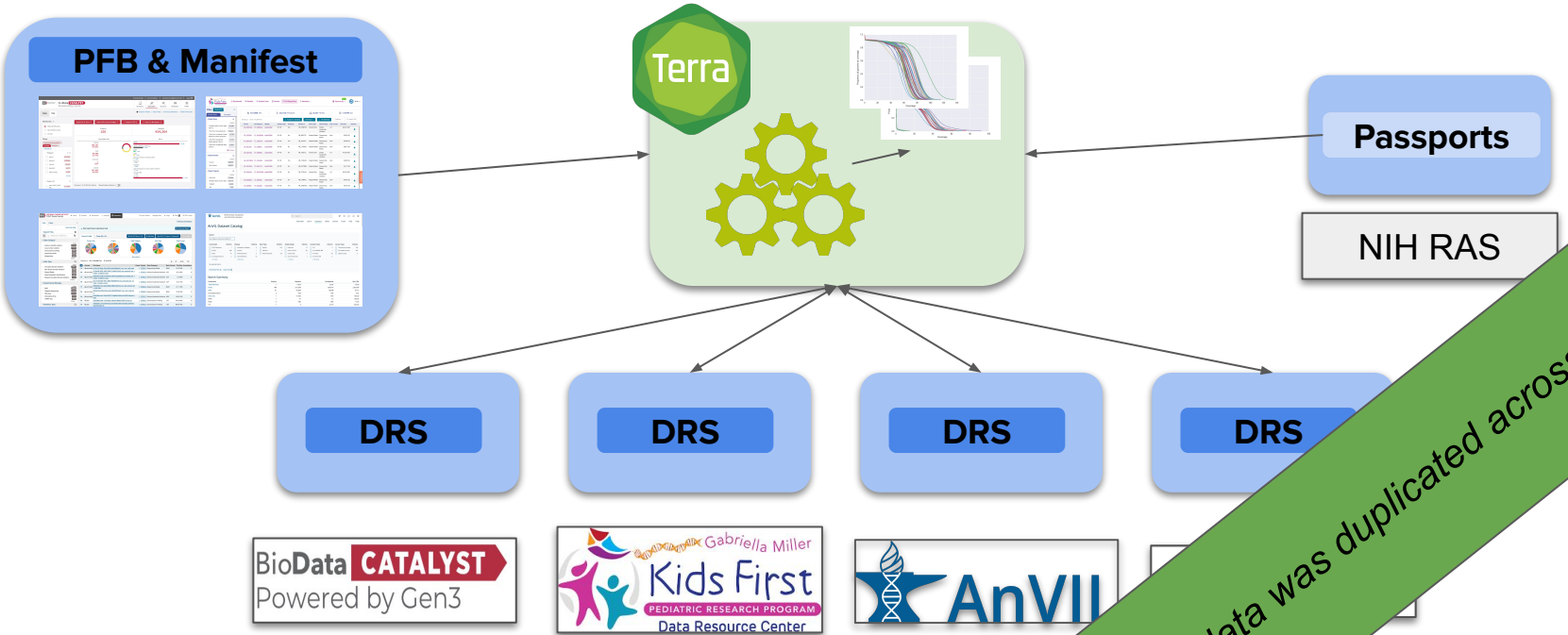
"We performed an association analysis, interrogating the effect of rare exonic variation on CHD risk at a fraction of the cost that would have otherwise been incurred without these interoperability tools."



Our Latest Use Case Success Story

Use Case #11: Melissa Wilson's use case examining Sex as a Biological Variable

Assessing the state of X and Y chromosome calling, we created a Terra workspace referencing AnVIL, BioData Catalyst, CRDC, and Kids First datasets. We used DRS to access data on demand without copying.




GDC Data Portal
Advanced Search
Add All Files to Cart
Workflow
Data Category
Data Type
Experimental Strategy
Workflow Type

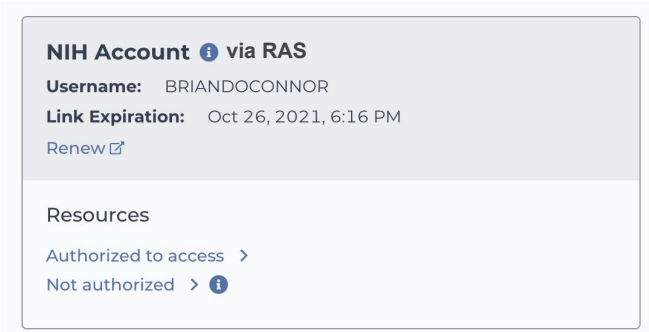
Priorities for 2022

Finish RAS Milestone 3

Multiple account links

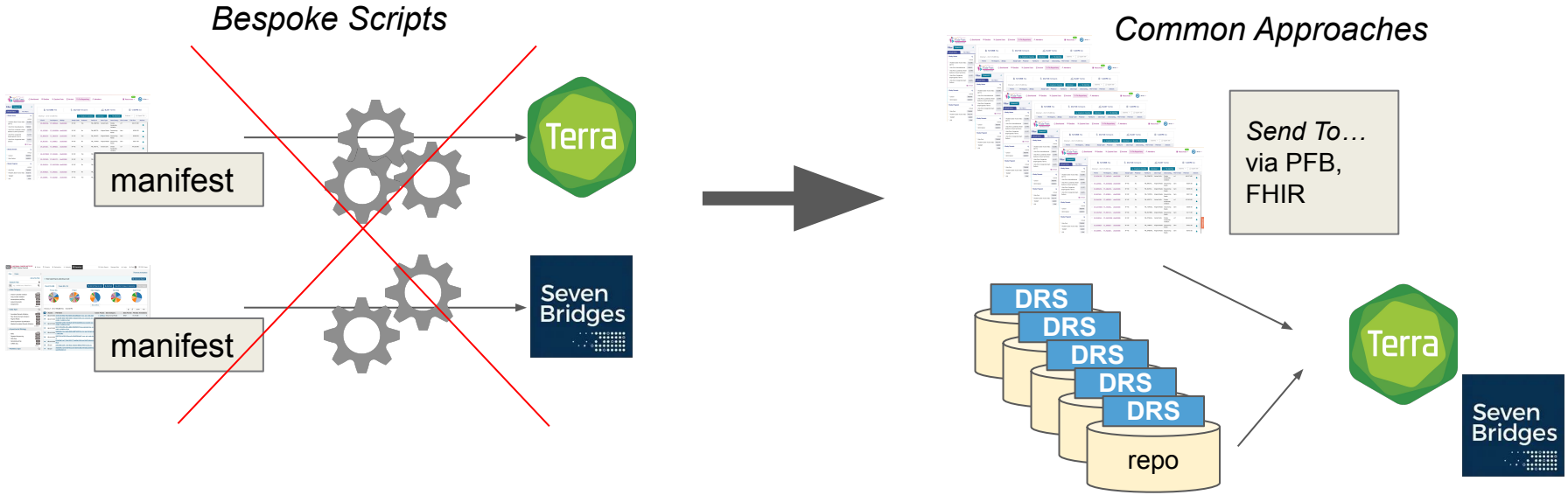
Simplify connecting data source through single, RAS identity/authorization

A single RAS-based account link



Priorities for 2022

Connect more portals + data repositories

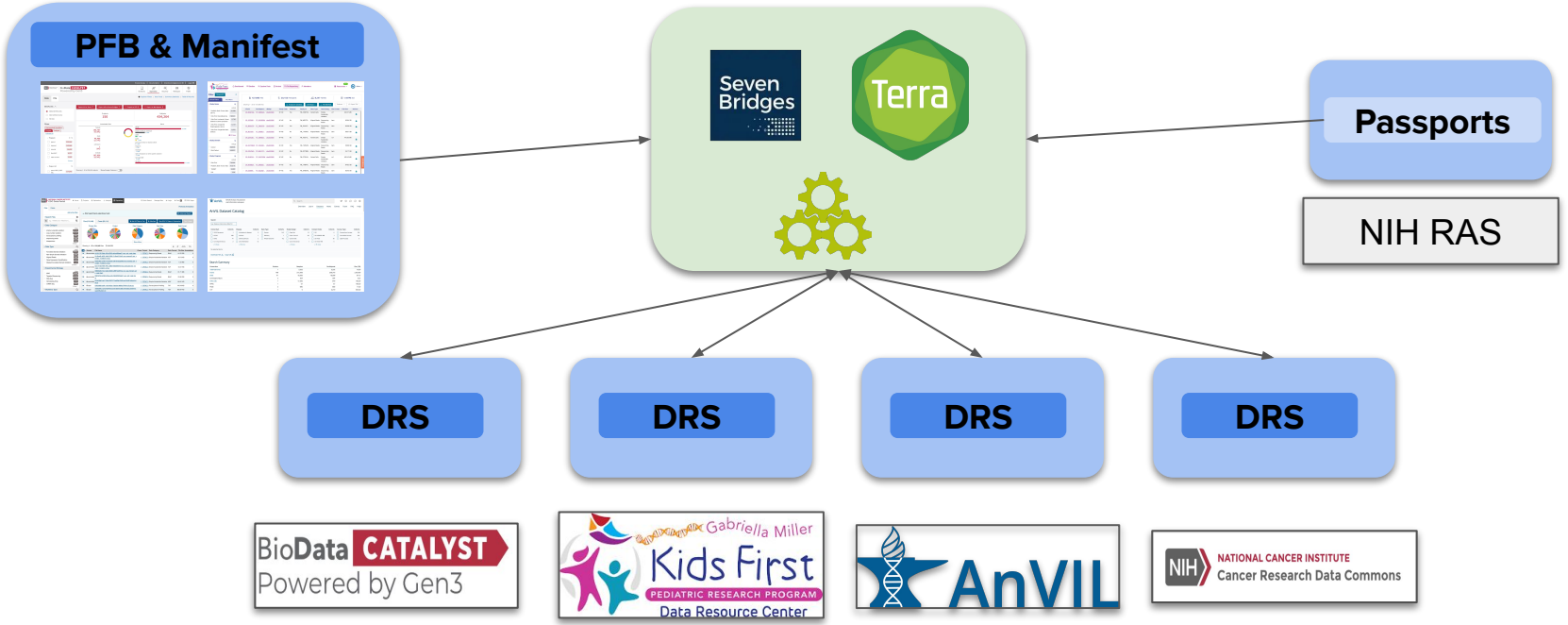


Users love the ability to "send to…" a workspace, add to more portals, make it easy to add new data repositories

Priorities for 2022

Tell Users!

Work with Outreach to let users know they can work with 11PB of data in these platforms today!



NCPI Systems Interoperation WG

Thank you *to everyone that has made NCPI Systems Interoperation possible!!*

Please consider joining our meetings, you can find more information at:

https://anvilproject.org/ncpi

Updates on Key Topics

Part 1

PFB, FHIR and RAS

Becky Boyles, Moderator

Reminder - What is PFB?

- The Portable Format for Biomedical Data (PFB) is a self-contained, self-describing, application independent **bulk format** for clinical, phenotype or other structured data.
- It is based upon **Avro**
- It encapsulates:
  - Data model / data dictionary
  - The bulk data itself
  - Pointers to third party controlled vocabularies for data elements
- It started based upon Gen3's graphical data model, but you can define PFB formats for **any data model**, relational, graph model, etc.
- For data versioning, support for multiple platforms and applications, long term support for data, it is helpful to have a self-contained bulk format

PFB - 1 (Grossman)

	Updates	Gaps/Next Steps
Gen3	• Gen3 is formulating a feature around functionality to allow external users to download study level PFBs via an API. The intent is to create a process to host PFB files that can be downloaded via DRS URIs externally existing data ingestion and submission would continue. An added ability to handoff PFBs would be available with this feature • Working closely with other groups to improve interop testing and Quality around PFB handoffs • Adding ability to export PFBs to 3 new destinations. These options will be provided in BioData Catalyst (export to export to CGC, CAVATICA, BDC powered by SBC) and in AnVIL (CGC and CAVATICA) supporting greater interoperability	• Complete review of feature document and design for providing ability to download study level PFBs. Plan for work to implement. • Continue commitment to Quality by supporting interop testing for various test cases around PFB handoffs.
Seven Bridges	Seven Bridges is developing interop solution to enable a user to send PFB from Gen3 systems outside of BioData Catalyst (like AnVIL) to BioData Catalyst Powered by Seven Bridges.	Pilot users to test the feature

PFB - 2 (Grossman)

	Updates	Gaps/Next Steps
Terra	• Terra currently supports PFB import from the AnVIL and BioData Catalyst data portals • Continued support of PFB as additional portals support the convention, currently working with the PDC portal	• Adding automation for ETL process of PFB in FHIR • Adding automated transfer of PFB onto Healthcare API (FHIR)
Kids First	FHAVRO: A generic Java library for converting FHIR resources into Avro and vice-versa. Avro schemas are obtained from project's FHIR implementation guide. • Enable developers to manage FHIR resources using the well-established Avro software ecosystem (e.g. Spark, Kafka) • Open source: https://github.com/Ferlab-Ste-Justine/fhavro Apache License 2.0 • Current status: in active development	• Generating Avro schema from a profile • Generating schema from NCPI implementation guide

PFB - 3 (Grossman)

	Updates	Gaps/Next Steps
NCBI	NCBI pioneered portable genomic data in 2011 with VDB, the foundation of SRA storage. It is a schema-driven columnar store with high compression, capable of representing any type of data, and organized into transportable units. The SRA uses these to model sequencing runs that was initially used across the INSDC.	**Gap:** Having provided APIs to access VDB, many tool vendors have not yet updated. **Next Step:** The VDB team is ready to guide tool vendors who are now willing to update in their adoption.
NCPI Outreach	Linking to documentation of PFB at https://anvilproject.org/ncpi/technologies	Keep PFB documentation up to date and expand as needed.

FHIR - 1 (Carroll)

	Updates	Gaps/Next Steps

AnVIL	FHIR Services
	• Proof of concept AnVIL FHIR server setup
	• Currently access only for AnVIL dev team as development continues
	FHIR Model
	• Implemented transform to NCPI Model for CMG data
	• Developed and implemented pilot release of a Study and Summary level model.
	• REDCap FHIR module has been updated to support export resources at minimum requirements.
		• Wider release of FHIR server will need to wait until Terra picks up managed service for FHIR
		  • Team engaged with Terra engineering
		  • Project plan in place to create process to configure service and ensure authorization
		• Continued onboarding of existing datasets
		• Testing and refinement of Study and Summary level model

BDCatalyst	• Loaded four test datasets using Bulk FHIR and built data ingestion pipelines to test Bulk FHIR standard in PIC-SURE • Prototyped FHIR server deployments in both the Google and Azure clouds • Prototyped conversion of synthetic HL7v2 and C-CDA documents into FHIR using Azure tools • Prototyped FHIR to PFB export	• Continue to load appropriate data from FHIR sources in PIC-SURE • Expand to use more real data sources and use cases (not test data) • Longer term, ensure appropriate data is accessible via FHIR as determined by the BDCatalyst project.

FHIR - 2 (Carroll)

	Updates	Gaps/Next Steps
Kids First	1. NIH NCPI FHIR Implementation Guide: https://nih-ncpi.github.io/ncpi-fhir-ig/index.html 2. Loaded five projects released on dbGap & KFDRC: ● Kids First: Enchondromatoses (SD_7NQ9151J): 285 Patients; 289 Specimens; 5,952 DocumentReferences ● Kids First: Congenital Heart Defects (SD_PREASA7S): 2,966 Patients; 2,987 Specimens; 16,506 DocumentReferences ● TARGET: Neuroblastoma (SD_YNSSAPHE): 277 Patients, 614 Specimens; 3,380 DocumentReferences ● Kids First: Familial Leukemia (SD_W0V965XZ): 620 Patients; 373 Specimens; 3,076 DocumentReferences ● Pediatric Brain Tumor Atlas - Children's Brain Tumor Tissue Consortium (SD_BHJXBDQK): 4,170 Patients; 48,240 Specimens; 43,004 DocumentReferences	● Replicating dbGaP's ResearchSubject model especially for curating various aggregate counts ● Developing a genomics module for sequencing and genomic workflow using Task and Observation ● Sustainable AuthN/AuthZ: The current AuthN/AuthZ flow requires an expiry ALB cookie and the acquisition of a cookie needs to be done manually. We therefore plan to implement OAuth2/OIDC setup supported via Keycloak. ● Exploring RAS-FHIR integration with Kurt R (UDN use case) ● Pedigree: Observation vs FamilyMemberHistory ● Phenotype: Condition vs Observation

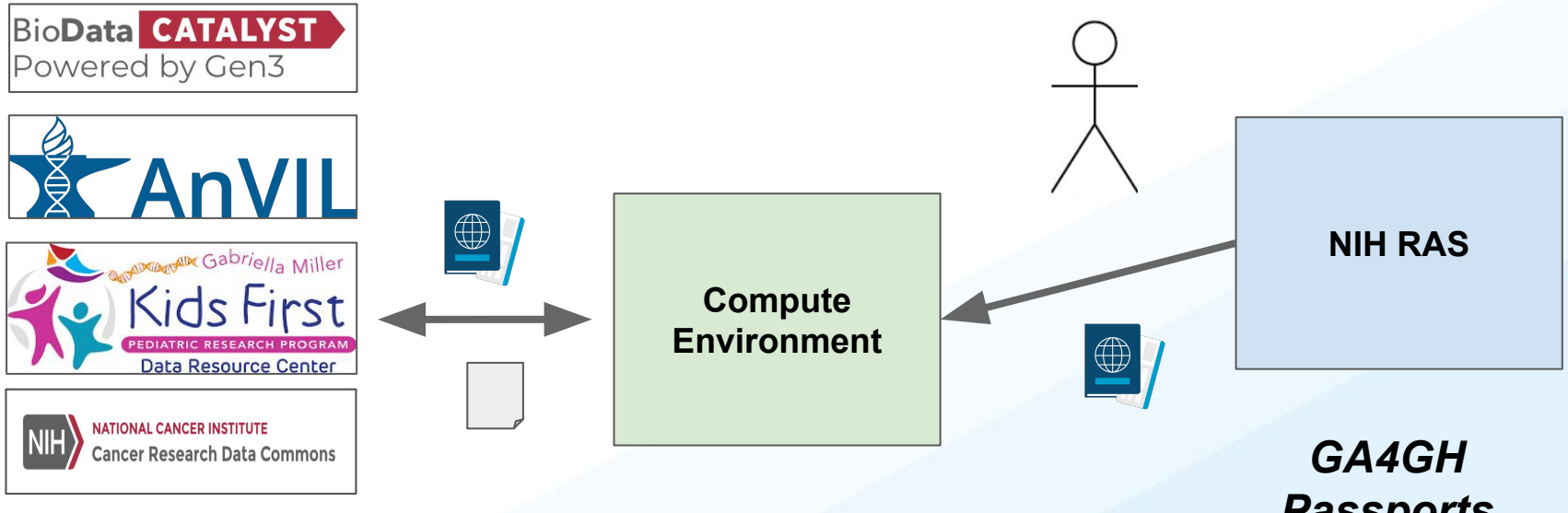
FHIR - 3 (Carroll)

	Updates	Gaps/Next Steps
NCBI dbGaP API	**Overview:** 1800 Studies comprising approx. 3 million subjects, 370,000 variables and 2.5 billion observations. **Study level meta-data:** The NCBI dbGaP FHIR API provide access to all of dbGaP studies meta data. Users can search using multiple criteria including study title, sponsor, type (prospective, longitudinal, cohort, case-control), keyword, condition, and many others. https://dbgap-api.ncbi.nlm.nih.gov/fhir/x1/ResearchStudy **Variable level data:** An initial FHIR research database populated with synthetic data from a few representative studies is in place for the development team and some limited beta testers to better understand the technology and how to best represent dbGaP data as a collaboration with NLM Research Data Finder. **RAS Integration:** A working prototype of RAS access mechanisms is expected to be completed in Q1 FY22 for testing with dbGaP control-access consent group and to allow authorized users to reach de-identified research.	• Current work is scaling up the servers and test loading more than 200 million observations.The ultimate goal is to provide seamless access to all of dbGaP metadata and phenotypic observations. • Continue development and testing to improve server performance.performance is a problem with big datasets such as dbGaP in native FHIR servers. • Integration of RAS will continue to be challenging due to constraints working with existing databases with different authorization systems • NCBI will continue to collaborate with LHC NLM to map and standardize the variable data. dbGaP variables have inconsistent and irregular labels that will require substantial effort to harmonize. • Continue with integration coordination with NLM Research Data Finder https://lhcforms.nlm.nih.gov/fhir/research-data-finder/ and NCPI dataset catalog NCPI Data | NCPI (anvilproject.org)

FHIR - 4 (Carroll)

	Updates	Gaps/Next Steps
NCPI Outreach	Linking to documentation of FHIR at https://anvilproject.org/ncpi/technologies	Keep FHIR documentation up to date and expand as needed.

RAS Update


BioData CATALYST
Powered by Gen3
AnVIL
Gabriella Miller
Kids First
PEDIATRIC RESEARCH PROGRAM
Data Resource Center
NIH
NATIONAL CANCER INSTITUTE
Cancer Research Data Commons
Compute Environment
NIH RAS
GA4GH DRS Data Servers (U. Chicago)
Workspaces (various)
GA4GH Passports (NIH)

RAS Key Docs & Milestones

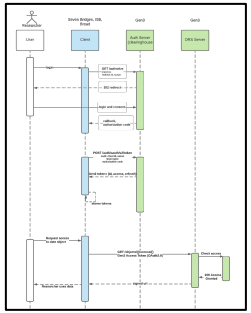
- **RAS design work** across a variety of teams and projects to date:
  - See: **RAS Authn/Authz "Milestone 3" Design with GA4GH Passports**

- Groups coordinated a 3 milestone plan:
  - **Milestone 1** : Login with RAS ✅
  - ~~**Milestone 2** : Gen3 uses RAS Visas as the authorization information instead of dbGaP telemetry files~~ Skipping this
  - **Milestone 3** : RAS Passport Visas can be used directly to access data resources, Central Fence is enabled by consistency across IC stacks

    → *designed in Q2-Q3 and now on an implementation timeline*

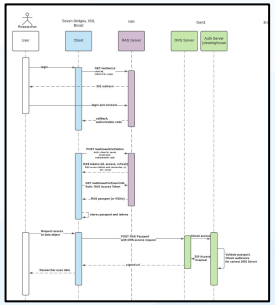
Summary of Milestone 3

- We've worked with Kids First, CRDC, AnVIL and BDCat to converge on a common approach for Milestone 3
- We've tried to help by putting together a summary of two preferred approaches and collaboratively address concerns… *goal is to add ability to access data with passports rather than taking away previous approach*



1: Current Gen3 Approach

&



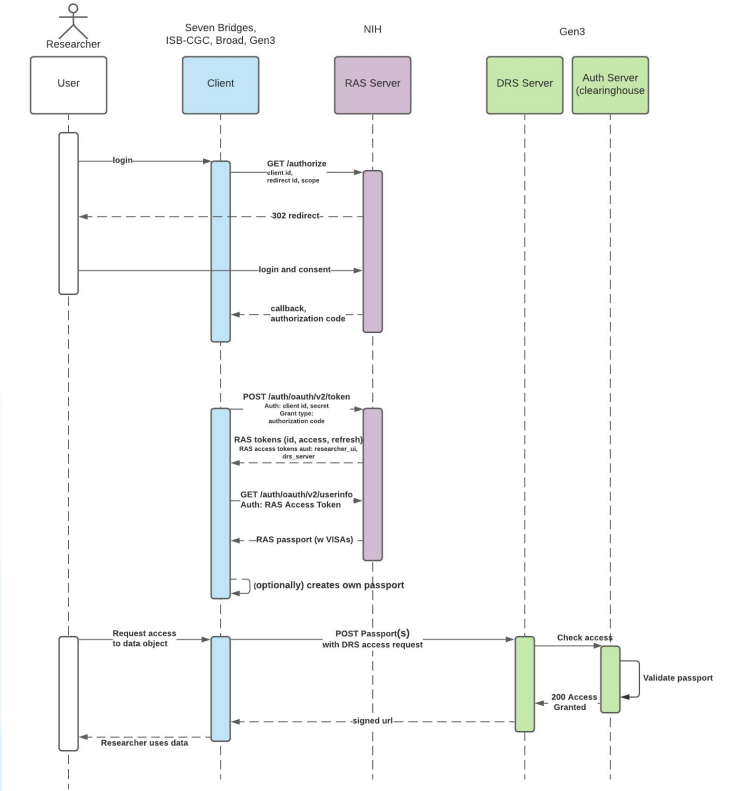
2: New Passport Approach

New Passports Approach

- Systems can interact with RAS directly, using RAS GA4GH Passports + Visas to access data from DRS data servers such as Gen3:
  - Client request RAS Passport directly from RAS
  - ~~Client repackages Passport while keeping RAS Visas intact~~
  - Passport is passed to DRS server in DRS data access request
  - DRS verifies and sends back a signed url

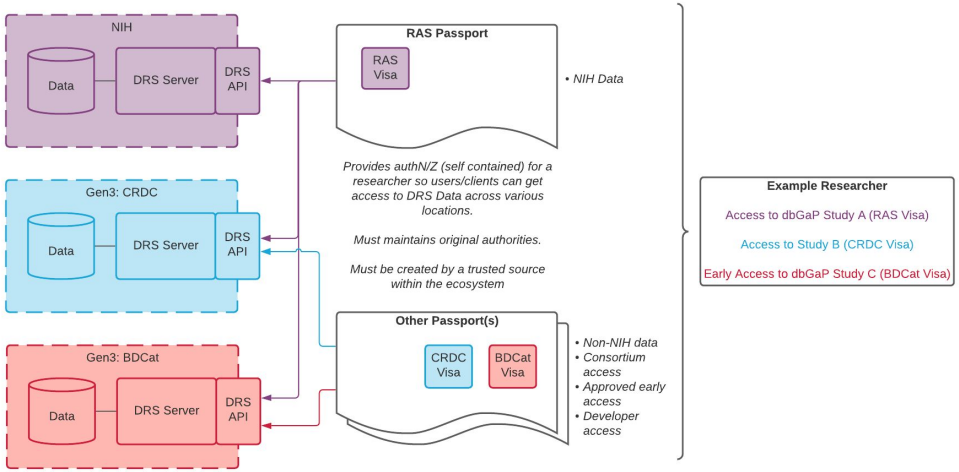
- Significantly improves the user experience for interoperability across NIH IC “stacks” by requiring just a single “account linking” with RAS (instead of multiple as is done today)

- RAS Passport + Visas then open up datasets that the users are authorized to use across systems like AnVIL, BDCat, CRDC, and GMKF as approved by researchers’ SO…
  this is transformative for interop!!



What We Learned Along the Way

- **Use the Passport from RAS unmodified (don't repackage)**
  - *Other passport brokers may use repackaged passports for developer/consortium access lists but don't mix with RAS visas*
  - ***DRS 1.2** now supports sending multiple, complete passports in a DRS data access request*



What We Learned Along the Way

- **Requirement for mutual TLS authentication for client verification**
  - *AnVIL, BioData Catalyst, CRDC, and GMKF indicated this is required*

Platform	Policy Requires Client Verification
AnVIL	Yes
BioData Catalyst	Yes
CRDC	Yes
GMKF	Yes

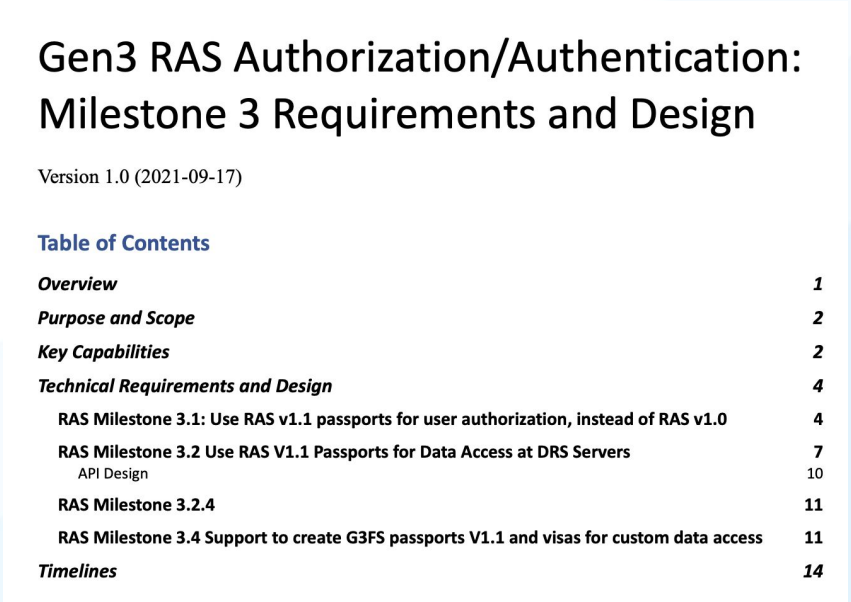
What We Learned Along the Way

- **Teams agreed to a timeline/plan**
  - *Implementation of staging/dev by U. Chicago before end of 2021*
  - *Workspace platforms implementing/testing by end of Q1 2022*
  - ***See signatures of platform architects, POs, and security team members***

3.2*	**Use RAS V1.1+ Passports for Data Access at DRS Servers** *This is a ***significant architecture change***, including • API Level Support for acceptance and validation of v1.1 passport(s) against DRS API as an alternate	• Clients can POST the full RAS v1.1 Passport to get controlled-access data from a DRS endpoint • Gen3 DRS Server uses GA4GH claims clearinghouse to validate unmodified RAS passports and visas for authorization decisions	See below for 3.2.1 and 3.2.3 target dates
	means of authentication and authorization • Parsing, validation, and interpretation of visas contained within v1.1 Passport(s) for means of realtime authorization upon data access requests • Caching support for scalability of average researcher workflows supporting thousands of data access requests in a short time frame • Final authorization decision by clearinghouse by aggregating information from parsing/interpretation of passport(s)/visa(s) and making a decision for controlled access data		
3.2.1	**Minimum Viable Product with support for RAS V1.1 Passports in Gen3 DRS endpoints** ***does not include** full integration tests nor performance support. *These are usually performed before*	• Stand up a development environment for clients to connect to, populated with mock control NIH data • An MVP deployment into respective development environment for clients of Gen3 to test respective flows (i.e. authorized users based on passport are returned a signed URL to data)	**Target Date:** 12/07/21
	rolling to environments		
3.2.2	**Load testing and profiling of 3.2.1 support**	• Validate performance is comparable to current support via OIDC and OAuth 2 tokens	Start date 12/1/21
3.2.3	**Performance improvements based on results of 3.2.2 and subsequent load testing**	• Performance improvements to ensure support is comparable to current support, as done via OIDC and OAuth 2 tokens	**Target Date** 02/18/22
3.3	**Mutual TLS Support as a mechanism for client authentication for controlled egress**	• Client authentication so that systems know which client is presenting the RAS Passport to their DRS endpoint *This is a system requirement for AnVIL and BDC *This support is not a RAS requirement but it is RAS recommended	**Target Date:** 12/17/21

What We Learned Along the Way

- **U. Chicago has shared a detailed technical plan with the RAS and other teams**
  - *Milestone 2 not needed*
  - *Gen3 plan needed for building clearinghouse function in G3FS*
  - *"The RAS team does not need to review another version of this technical planning document.*
  - *We are ready to provide support on the clearinghouse design as needed. "*



Gen3 RAS Authorization/Authentication: Milestone 3 Requirements and Design

Version 1.0 (2021-09-17)

Table of Contents

RAS - 1 (O'Connor)

	Updates	Gaps/Next Steps
Gen3	• Full implementation of RAS for AuthN • Continued discussion of milestones 3 to enable full AuthZ using RAS passports directly • Required SIA completed by U Chicago and signed off by CBIIT security	• Full consensus on plans for milestone 3 to enable use of RAS passports • Implementation of tasks established in the milestone 3 document
Seven Bridges	• Seven Bridges working to get RAS passports directly from RAS as part of UDN/NCBI/SRA use case. • Full implementation of RAS for AuthN • ISAs signed with all relevant systems • Approval of current RAS milestone 3 plans	• Upstream implementation of RAS for AuthZ and use of RAS passports in all systems • Rapid and robust SOP for support of end users

RAS - 2 (O'Connor)

	Updates	Gaps/Next Steps
Terra	Terra is code complete on a new service the External Credentials Manager (ECM) which can obtain a RAS-issued Passport from RAS using the RAS v1.1 passport specification and can monitor to identify expiring visas and request updated visas. This is currently on Terra's non-production environment and will not be deployed or utilized until Gen3's DRS work is complete and Terra has completed the work for sending a RAS Passport to a Gen3 DRS server for data access.	Terra is currently planning the development work for sending a RAS Passport to a Gen3 DRS server for data access.
PIC-SURE	Compatibility with Gen3 RAS based authentication, still using Gen3 based authorization.	Leveraging RAS Passports for Authorization once available.

RAS - 3 (O'Connor)

	Updates	Gaps/Next Steps
NCBI	NCBI has released a RAS Clearinghouse v1 service that processes RAS 1.1 passport tokens. This service is used by the dbGaP DRS v1.0+ service, with extended features for processing passports that have now been incorporated into DRS v1.2. Online dbGaP genomic data stored in the SRA can be reached by POSTing a RAS passport and DRS id to the DRS service.	• Minor adjustments to bring dbGaP DRS in line with v1.2 • Pilot passport v1.2 in support of FHIR • Provide externally accessible pre-release support for developers
NCPI Outreach	Linking to documentation of RAS at https://anvilproject.org/ncpi/technologies	Keep RAS documentation up to date and expand as needed.

Lunch Break 1:00-1:45pm ET

(and RAS Breakout 1:15-1:45pm ET)

Breakout 1 Instructions (Patton)

We will open the **RAS** breakout room in this same Zoom. This breakout will last from 1:15-1:45 p.m. ET

- If you have downloaded the latest version of Zoom (instructions and how-to video), you can move yourself into your preferred room.
- Otherwise, request that meeting host move you to your room.

We will have breakout rooms for other key topics this afternoon.

Breakout Report Backs will be first on the agenda for Day 2.

We will reconvene in the plenary session at 1:45 p.m. ET.

Updates on Key Topics

Part 2

End-User Cloud Costs, Search and Other Interoperability Efforts

Becky Boyles, Moderator

End-User Cloud Costs - 1 (Schatz)

	Updates	Gaps/Next Steps
CRDC	• Continued use of $300 pilot credits for new users • Continued use of benchmarking data for published tools • File archiving on AWS added to Seven Bridges for users to save on storage costs • FireCloud has expanded its tools for cloud cost estimation with in-app cost reporting. Users can now see cost incurred on a per submission and per workflow basis	• Continued documentation on AWS and Google costs, breakdown of costs for each analysis per NCI Cloud Resource • New tutorial coming soon describing how to estimate cloud costs
AnVIL	• AnVIL Cloud Credits (AC2) Program initiated • Offering $300 in cloud credits • Developed a cloud cost budget justification spreadsheet & template • Improved cloud cost calculations released on AnVIL/Terra • Started a project to empirically measure the costs of popular genomics tools: talk	• Continuation and possible expansion of the AnVIL Cloud Credits (AC2) Program • Develop a technical report on empirical cloud costs by Summer 2022

End-User Cloud Costs - 2 (Schatz)

	Updates	Gaps/Next Steps
BDCatalyst	• Offering $500 cloud credits through the NHLBI Cloud Credits program with an opportunity to request more funds for Heart, Lung & Blood research. • September Cloud Costs Community Hour: notes, slides, and Youtube videos available • Published documentation on managing team costs, estimating workflow costs, and setting up spend alerts on Terra and estimating total cloud costs on SBG • Launch of Project Per WorkSpace (PPWS) on Terra to support improved cost reporting functionality	• STRIDES enhancing Dashboard to provide users with more visibility and timeliness into cloud credits available and spent. • Seven Bridges investigating adding file archival options on Google Cloud. • Identify and evaluate solutions for BYO cloud credits
Kids First	• All new users receive $100 in pilot credits to explore CAVATICA (Seven Bridges). • Training materials related to costs are available through the Kids First DRC Help Center and Cavatica's Support Documents. • Wrote a report assessing the successes and lessons learned of a 2.5 year long pilot cloud credits program	• Finalize new guidelines and training materials to launch a public Cloud Credits program for Kids First users based on our own pilot's conclusions and recommendations of other NCPI platforms.

End-User Cloud Costs - 3 (Schatz)

	Updates	Gaps/Next Steps
NCBI	• No egress costs for SRA datasets stored in AWS Open Data (ODP) buckets and GCP Public Dataset Program • No egress costs to access SRA data on AWS or GCP from within the respective cloud compute environments, if running from the correct regions; compute in the cloud is at user expense • NCBI's Cloud Data Delivery Service provides free "thaw" from cold storage and delivery to users' buckets of SRA data in cold storage classes on AWS and GCP; per-user limits apply • Example user costs can be found here	• All public SRA data is in AWS ODP, but more controlled access (dbGaP) sequence data is coming

Search - 1 (Rogers)

	Updates	Gaps/Next Steps
CRDC	• Seven Bridges Cancer Genomics Cloud (CGC) complete UI for integration of Cancer Data Service (CDS) datasets now available • Ongoing work to harmonize metadata and identifier standards across the CRDC to better enable search • Cancer Data Aggregator (CDA) Release 1 launch to enable query of Genomic Data Commons (GDC) and Proteomic Data Commons (PDC) open access data • UAT testing of CDA Release 1 • Collaborated with Center for Cancer Data Harmonization (CCDH) team on end-to-end workflow demonstration of CRDC interoperability use cases	• Integration of Seven Bridges CGC and the CDA via Jupyter notebook • Continued efforts at harmonizing metadata and identifier standards across CRDC • CDA Release 2 launch scheduled to connect GDC, PDC and Imaging Data Commons (IDC) open access data • Obtain Authority to Operate (ATO) to publicly launch CDA API and enable controlled-access data query • Integrate with Cloud Resources (cloud analysis platforms) that provide CDA front end interface for UAT

Search - 2 (Rogers)

	Updates	Gaps/Next Steps
BDCatalyst	• Open Access search for phenotypic (PIC-SURE) and genomic data (Seven Bridges) prior to authorization • File and variable level search (phenotypic and genomic data) from User Interface on PIC-SURE • File and variable level search (phenotypic and genomic data) from PIC-SURE API on Terra and Seven Bridges • Semantic, public, full text, variable granularity search of TOPMed phenotypic concepts and dbGaP studies with explainable results, provenance in biomedical knowledge graphs, links to peer reviewed literature, and preliminary harmonization. (Dug) • Open Access Study-Level Search prior to Login through Gen3 Discovery Page • Subject, Study, and File Level Search (w/ secure limiting of results prior to authorization) through Gen3 Exploration Page • Exposed search API's for metadata, file object records, and genomic/phenotype data (GraphQL)	• Development of search use cases and personas • Development of integrated search strategy • Interoperability of handoff of search results to analysis workspaces across ecosystems - finding key use cases to drive development

Search - 3 (Rogers)

	Updates	Gaps/Next Steps
Kids First	• All clinical, phenotypic, demographic, file data searchable [both faceted and text] in registered-tier Portal that anyone can access (if they agree to click-through terms). Key non-Kids First datasets are also searchable in the Portal (interoperability with TARGET, CBTN etc). Filters applied to dynamic visualizations to build cohorts of multiple datasets and ability to identify children affected with multiple conditions (e.g., cancer & birth defects). • All source data (as provided by submitters) made searchable in addition to “harmonized” HPO, MONDO, and NCIt terms • Variant Database enables search of variants with annotations from ClinVar, TOPMed, Gnomad and the ability to identify which datasets include that variant and aggregated phenotypes. • All studies are searchable in dbGaP and grouped in an umbrella BioProject Study • (see slide about FHIR)	• Back-end API is migrating from custom data service to a FHIR-based data service for interoperability with Portal, CAVATICA, and other tools. • 6 out of 22 Kids First studies loaded into the Variant Database, more to be loaded • Improvement to Variant Database in the Portal and the launching of the Variant Workbench (controlled access tool) which using table/matrix formats to find participants of interest and run analyses on a SPARK cluster.

Search - 4 (Rogers)

	Updates	Gaps/Next Steps
AnVIL	• AnVILproject.org displays the dataset catalog • Dataset Catalog - Newly added detail page for each study populated via dbGaP FHIR API and Terra. • Dataset Catalog - Deep link from the study page to dbGap “Request Access” page preserving the study context. • Gen3 - Subject, Study, and File Level Search (w/ secure limiting of results prior to authorization) through Gen3 Exploration Page • Gen3 - Exposed search API’s for metadata, file object records, and genomic/phenotype data (GraphQL)	• UX Research / Dataset Catalog UI updates to make the study detail pages more informative and useful.

Search - 5 (Rogers)

	Updates	Gaps/Next Steps
NCBI	**Studies and Metadata:** Web / SOLR faceted search for: https://www.ncbi.nlm.nih.gov/gap/advanced_search/ FHIR Research Study resource: https://dbgap-api.ncbi.nlm.nih.gov/fhir/x1/ResearchStudy **Sequence Runs**: Sequence Read Archive: "controlled"[Access] - SRA - NCBI (nih.gov)	**Gap**: Coordinated Sequence catalog/API format **Next steps:** Linking access request system to search interfaces. Adding metadata and phenotypic data as RAS enabled FHIR API
NCPI-portal	• Held a “Mini Hackathon” for Dataset Catalog API integration AnVIL, BDC, CRDC db GaP studies all refreshed automatically via APIs. • Newly added study descriptions for each dataset with data from dbGaP • Deep link to dbGap “Request Access” from each study.	• Refresh KF dbGaP studies via API • Read non dbGaP studies via API. • UX research / Incremental UI updates. • Explore deeper integration with platform APIs and search engines. • POC of integration with Dug semantic search.

Other Interoperability Efforts - 1 (Ahalt)

	Updates	Gaps/Next Steps
CRDC	• DRS Client added to CGC (now has both Server and Client) • Push button connection between CGC, BDC, and Cavatica utilizing DRS endpoints • Ability to connect to any open DRS endpoint or add known DRS endpoints	• Broad FireCloud using DRS to integrate Proteomics Data Commons datasets
AnVIL	• Forward looking work at workflow interoperability • Forward looking work at utilizing generic "app" definitions for extending the AnVIL platform (e.g. expanding the Leo service that powers Galaxy integration)	• Continue to collaborate with BDCatalyst and other NCPI teams on the development of the "app" interface and extension of the Leo component to support it
BDCatalyst	• Imaging: POC Nifti ingestion workflow • New co-leads of Tools & Apps WG, proposed tiered approach for establishing criteria to support V3PAs • Established Tool Trust Tiger Team (T4) to address data protections and workflow credibility standards • Ongoing discussion between PIC-SURE and AnVIL	• Identification of use cases to drive next interop efforts • Test RAS using incoming SRA data (PCGC) • Continued exploration in imaging access and analysis, eg ability to support new image formats

Other Interoperability Efforts - 2 (Ahalt)

	Updates	Gaps/Next Steps
Kids First	Active cross-platform use cases include: ● *CFDE* (Kids First and HubMap: running common workflows, integrated knowledge graph; multiple DCCs: develop FHIR profiles for CFDE human datasets; exploring CAVATICA use) ● *INCLUDE* (Data Hub launching in March) - interop on genomic data of children with DS & leukemia and CHD ● *CARING for Children with COVID*: share pediatric COVID clinical data through FHIR API for ImmPort and BioData Catalyst to interoperate with. Additional use cases of interest: ● FaceBase - craniofacial birth defects data, human and model - also other model organism databases ● ABCD/HBCD (NDA) - pediatric genomics and imaging ● RDCRN - exploring CAVATICA use	New DATA Scholar starts 9/27, she will engage users, document use cases, propose, test and implement solutions, coordinate with NCPI and stakeholders

Other Interoperability Efforts - 3 (Ahalt)

	Updates	Gaps/Next Steps
NCBI	All dbGaP Approvals delivered to RAS dbGaP DRS in Public See: dbGaP DRS Documentation IDX service in Public See: SRA IDX Documentation FHIR Research Study API See: https://dbgap-api.ncbi.nlm.nih.gov/fhir/x1/ResearchStudy	dbGaP File Selector and SRA Run Selector configured for RAS Auth (in development)

Breakouts
PFB & FHIR,
Other Interoperability Efforts

Breakout 2 Instructions (Patton)

We will open the **PFB/FHIR and and Other Interoperability Efforts** breakout rooms in this same Zoom. These breakouts will last until **3:10 p.m. ET**

- If you have downloaded the latest version of Zoom (instructions and how-to video), you can move yourself into your preferred room.
- Otherwise, request that meeting host move you to your room.

We will then have a brief break and open the breakout rooms for End-user Cloud Costs, Search, and Other Interoperability Efforts.

Breakout Report Backs will be first on the agenda for Day 2.

Plan for Day 2

Becky Boyles

Agenda: Day 2 All times ET (Ahalt)

Time	Activity	Owner	Links
11:00-11:10am	Welcome and Goals Day 2: Synthesize next steps, driving use cases, determine NIH/NCPI priorities	Stan Ahalt	Slides|Notes
11:10-12:40pm	Breakout Report Backs and Discussion •PFB (10 min) (Grossman) •FHIR (10 min) (Carroll) •RAS (20 min) (O'Connor) •End-User Cloud Costs (20 min) (Schatz) •Search (20 min) (Rogers) •Other Interoperability Efforts (10 min) (Ahalt)	Moderator: Becky Boyles	Slides|Notes
12:40-12:50pm	GA4GH Relationship	Brian O'Connor	Slides|Notes
12:50-2:00pm	Lunch Break		
1:30pm-2:00pm	NIH Breakout: NIH Coordination Working Group Discussion of Priority Next Steps	NIH Only (via separate invitation)	
2:00-2:15pm	Use Case Overview: The Journey of a NCPI Use Case	Asiyah Lin	Slides|Notes
2:15-3:20pm	Review of Current Scientific Use Cases	Moderator: Valentina Di Francesco	Slides|Notes
2:15-2:30pm	Genetic Sex as a Biological Variable and X-inactivation	Melissa Wilson	Slides|Notes
2:30-2:50pm	Interoperability between Kids First & Undiagnosed Diseases Network (UDN) Data via dbGaP/SRA	Valerie Cotton, Allison Heath	Slides|Notes
2:50-3:05pm	Leveraging Functionally Equivalent Pipelines for Long-Read Data on Different Systems	Owen Hirschi	Slides|Notes
3:05-3:20pm	Conducting reproducible science in PIC-SURE interoperating with Seven Bridges/Terra	Simran Makwana	Slides|Notes
3:20-4:00pm	Synthesize Goals and Next Steps for the next 6 Months, with focus on driving use cases	Stan Ahalt, Jon Kaltman	Slides|Notes

Meeting Deliverable: NCPI Glossary

- Remember to keep populating the NCPI Glossary with new words or additional definitions
- We hope this Glossary will be a concrete deliverable at the end of the meeting to help us coalesce around common definitions and/or highlight differences.

Breakouts
End-user Cloud Costs, Search

Breakout 3 Instructions (Patton)

We will open the **End-user Cloud Costs and Search** breakout rooms in this same Zoom. These breakouts will last until **4:00pm ET**

- If you have downloaded the latest version of Zoom (instructions and how-to video), you can move yourself into your preferred room.
- Otherwise, request that meeting host move you to your room.

Breakout Report Backs are on the agenda at 11:10am ET on Day 2.