

The Journey of a NCPI Use Case

From a seed to a forest

Asiyah Y. Lin

The 5th NCPI Workshop

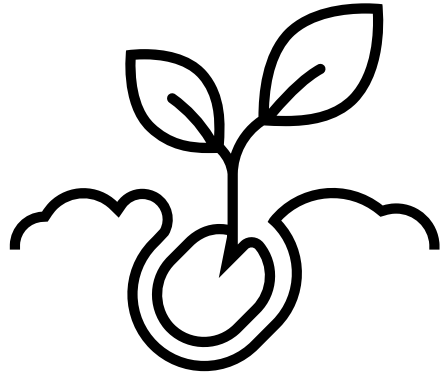
Oct 6, 2021

What is considered as a NCPI Use Case?

- To answer a scientific question, access and analyze data from NCPI platforms ($n \geq 2$) becomes necessary
- Demonstrate the NCPI interoperability
 - Create cohorts from at least two NCPI platforms
 - Access data from at least two NCPI platforms for analysis in one workspace
 - Develop portable software and tools across at least two NCPI platforms
 - More examples ...
- Drive the development of NCPI interoperability technology specification

*NCPI platforms: platforms support AnVIL, BDCat, CRDC, Kids First, NCBI.

Proposal Phase



- NIH staff or a researcher identifies a potential scientific use case.
- In collaboration with NCPI WG Leads and platform PIs:
 - Identify scientific lead
 - Identify platform lead (a.k.a. interoperability tech lead)
 - Develop the interoperability plan and challenges
 - Identify funding resources
- Develop one pager.
- NIH Coordination WG keeps the one pager for documentation and management purposes.

Interoperability between Kids First/CAVATICA and SRA's copy of the Undiagnosed Diseases

NCPI Use Case Details

Status: NCBI actively moving all files (BAMs) to hot AWS/SRA storage. Files become immediate DRS as they are moved into S3. Next steps: Seven Bridges development work to obtain RAS present them to NCBI/SRA DRS server to access files in CAVATICA workspaces.

Platform contact for genomic interop: Michele Mattioni and Kurt Rodarmer

Platform contact for FHIR structuring: TBD (one from dbGaP, one from Kids First)

Researcher contact: ~~TBD – assigned to Adam Resnick to resolve~~ Lisa Bastarache

Dataset: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001233
and [insert Kids First datasets once determined, listed here:

<https://commonfund.nih.gov/kidsfirst/x01projects>]

NCPI use case link for genomic data interop: https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/18

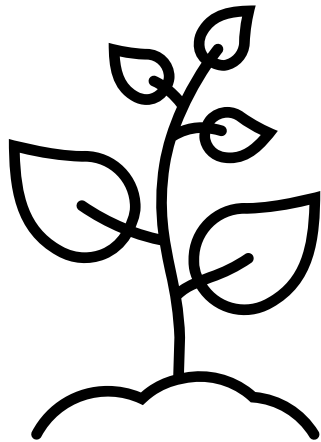
NCPI use case link for FHIR structuring of phenotypes: https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/18

Summary:

- **Goal:** enable co-analysis of Kids First data (BAMs/CRAMs + phenotype data) with U (phenotype data) in Seven Bridges CAVATICA. This requires 1) search/finding the di

One pager
Example

Implementation Phase:



- Scientific and platform leads coordinate with the System Interoperability WG to carry out the technical plan.
- Scientific and platform leads are responsible for reporting implementation progress.
- Demos at the bi-annual workshop.
- Provide updates on the GitHub Use Case Tracker.
- May become inactive use cases if no progress is made.

Dissemination Phase:



- A NCPI use case is completed with a demo of the implemented interoperability technical plan (**Note:** completion of the research plan is not necessary).
- Work with Outreach WG (Dave Rogers) to develop training materials:
 1. Training videos
 2. Necessary documentations
 3. Any publications (if relevant)
- Reach out and educate users to implement and grow the user community!

Training
video
example

Demo of Search Result Hand-off



<https://anvilproject.org/ncpi#demo-of-search-result-hand-off>

- ☐

🕒


FHIR UC1: ResearchStudies representation in rare disease (CMGs & Kids First)
#16 opened 22 days ago by cottonva

📄 Needs One Pager
- ☐

🕒

UC 13: Leverage functionally equivalent pipelines for long-reads data on different systems

one pager done

 2
- ☐

🕒

UC 12 - (Xihong) Whole Genome Sequencing Association Analysis pipeline

one pager done

3
- ☐

🕒

UC 11. (Wilson) Sex as a Biological Variable


one pager done

3
- ☐

🕒

UC 10. SRA & Kids First DRC for Kids First & UDN co-analysis

one pager done

 2
- ☐

🕒

UC 9. Whole slide images

need one pager

2
- ☐

🕒

UC 8. PIC-SURE API search of clinical and genomic data available from Seven Bridges Platform

need one pager

2
- ☐

🕒

7. NHGRI AnVIL + Kids First DRC + NHLBI BioData Catalyst

need training material

4
- ☐

🕒

UC 1a. NHLBI BioData Catalyst + Kids First DRC

inactive

2

Questions and Suggestions are welcomed!

Team: Asiyah Lin, Dave Rogers, Jack DiGiovanna, Ken Wiley, Valerie Cotton,
Valentina Di Francesco