AACR American Association for Cancer Research®

ANNUAL MEETING
2024 • SAN DIEGO

APRIL 5-10
#AACR24
AACR.ORG/AACR24

Year of Open Science: Impact of the Cancer Research Data Commons

Erika Kim, PhD
Ina Felau, MS
Esmeralda Casas-Silva, PhD
Anthony Kerlavage, PhD
Center for Biomedical Informatics and Information Technology
National Cancer Institute, NIH, US

CANCER RESEARCH

The Foundational Cancer Journal Driving Transformative Science

AACRJournals.org
@CR_AACR

AACR American Association for Cancer Research®

2302007Nn

AACR American Association for Cancer Research®
ANNUAL MEETING
2024 • SAN DIEGO
APRIL 5-10
#AACR24
AACR.ORG/AACR24

Overview: NCI Cancer Research Data Commons (CRDC)

Ina Felau, MS
Health Science Administrator
Informatics and Data Science Program
Center for Biomedical Informatics and Information Technology
National Cancer Institute, NIH, US

NIH NATIONAL CANCER INSTITUTE

Disclosure Information

Ina Felau, MS

I am a full-time paid employee of the NIH/NCI.

I have no financial relationships to disclose.

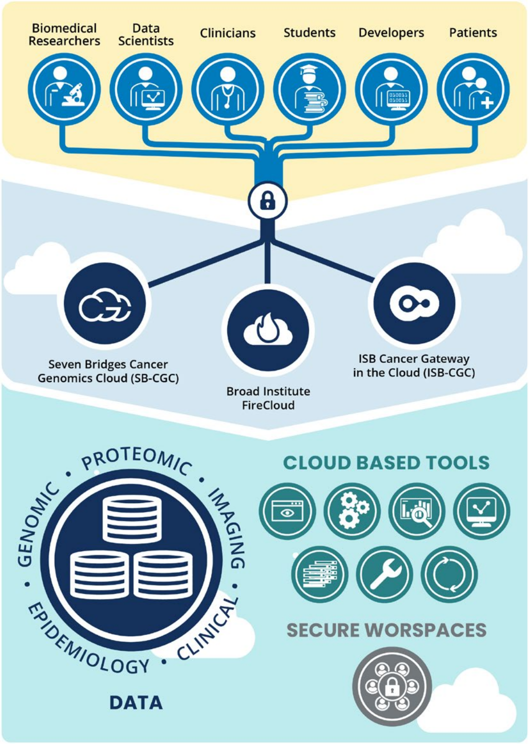
Agenda

- **Overview: CRDC Vision**
- CRDC Ecosystem
  - Data Commons
  - Cloud Resources
- Interoperability Initiatives
  - Initiatives within CRDC
  - Trans-NIH initiative (NCPI)
- Resources for Researchers

NIH NATIONAL CANCER INSTITUTE
Cancer Research Data Commons

datacommons.cancer.gov

Vision for the CRDC


Biomedical Researchers
Data Scientists
Clinicians
Students
Developers
Patients
Seven Bridges Cancer Genomics Cloud (SB-CGC)
Broad Institute FireCloud
ISB Cancer Gateway in the Cloud (ISB-CGC)
PROTEOMIC • IMAGING • CLINICAL • EPIDEMIOLOGY • GENOMIC
DATA
CLOUD BASED TOOLS
SECURE WORSPACES

🔒 = Secure data storage and access

Mission

- Empower researchers by providing a secure, accessible cancer data ecosystem
- Provide state-of-the-art visualization, analysis, and interoperability tools in a flexible, cloud-based computational environment

Lower Barriers

- Data submission
- FAIR data access, search, retrieval
- Integration of data for cross-domain analysis
- Analysis platforms, tools, and workflows

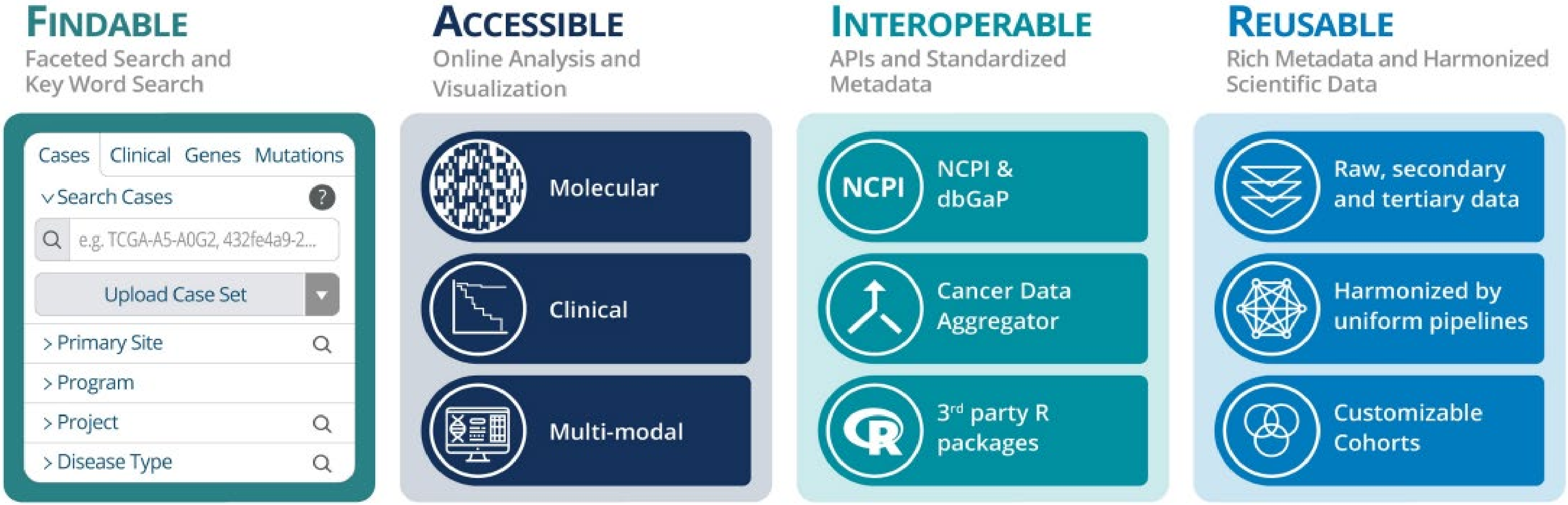
Infrastructure & Sustainability

- Security and appropriate access for sensitive data
- Sustainable, reusable, and uniform architecture
- Comprehensive plan for long term data storage and accessibility to tools

Stakeholder Focus

- Include all scientists and clinicians (of all technical abilities) using the data

FAIR Principles



FINDABLE
Faceted Search and Key Word Search

ACCESSIBLE
Online Analysis and Visualization

- Molecular
- Clinical
- Multi-modal

INTEROPERABLE
APIs and Standardized Metadata

- NCPI & dbGaP
- Cancer Data Aggregator
- 3rd party R packages

REUSABLE
Rich Metadata and Harmonized Scientific Data

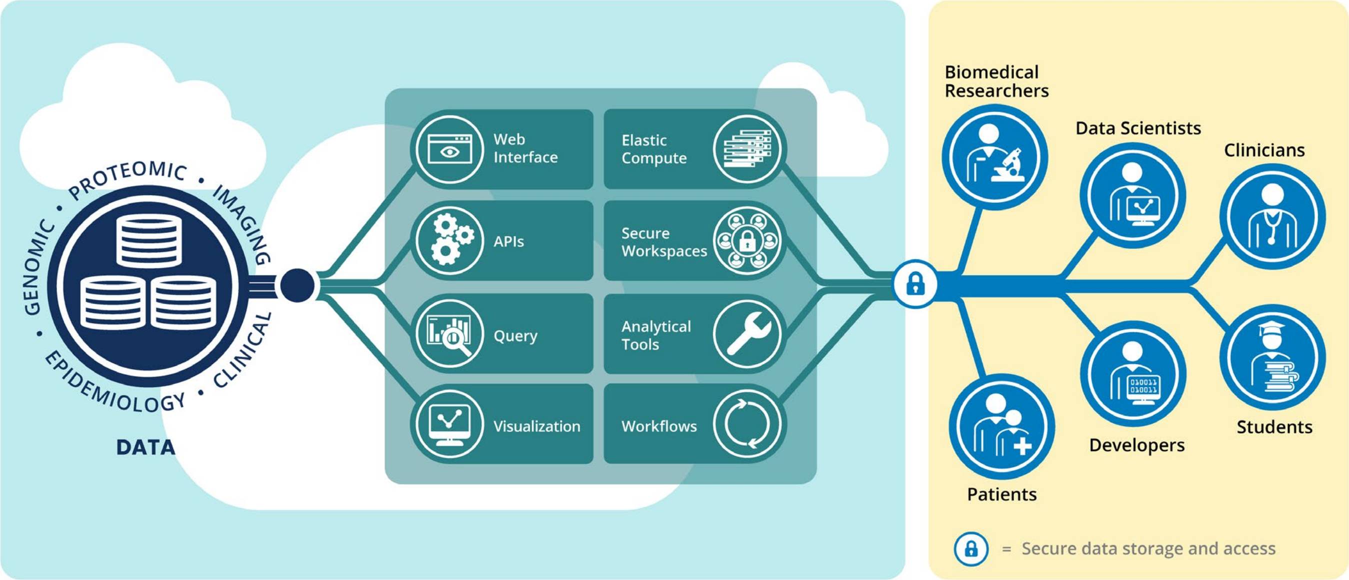
- Raw, secondary and tertiary data
- Harmonized by uniform pipelines
- Customizable Cohorts

At the bottom are six specialized data commons (GDC, PDC, ICDC, CDS, IDC and CTDC). Selected CRDC features are used to demonstrate the implementation of FAIR principles.

Agenda

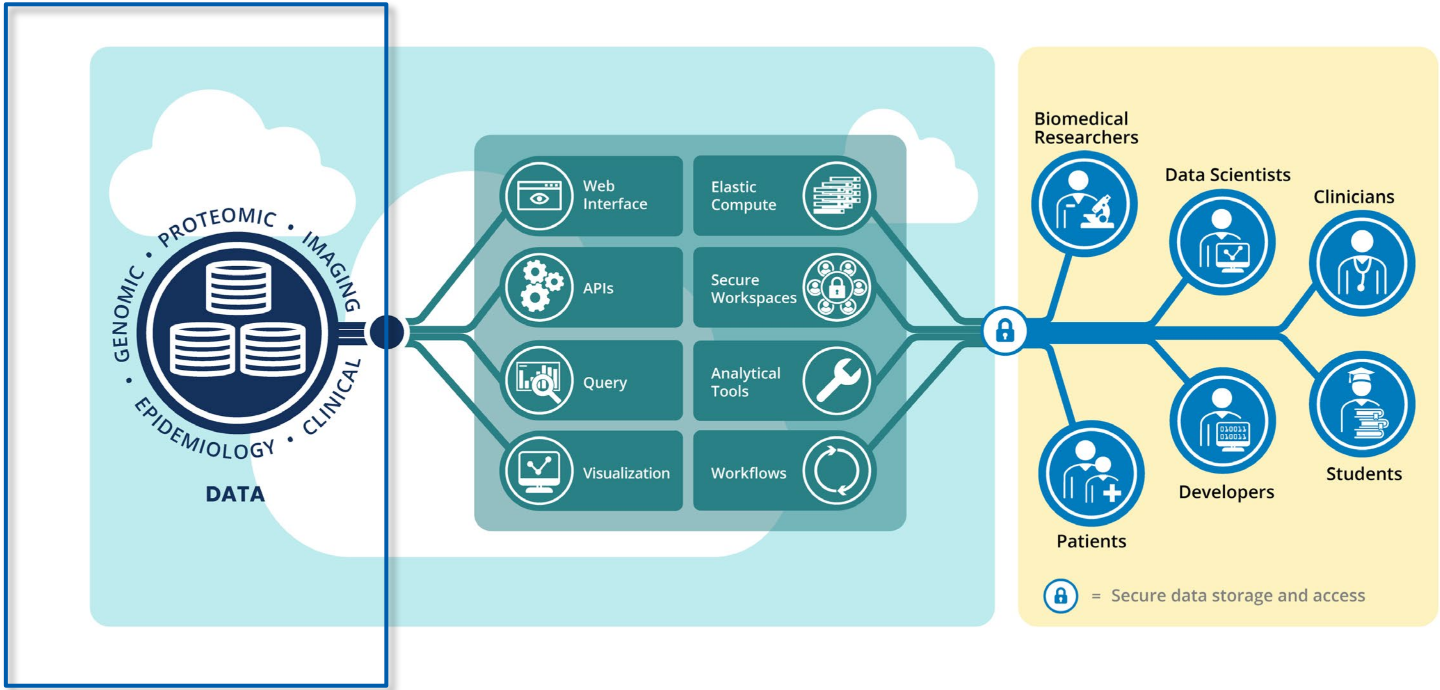
- Overview: CRDC Vision
- **CRDC Ecosystem**
  - **Data Commons**
  - **Cloud Resources**
- Interoperability Initiatives
  - Initiatives within CRDC
  - Trans-NIH initiative (NCPI)
- Resources for Researchers

CRDC Ecosystem



https://datacommons.cancer.gov/

CRDC Ecosystem



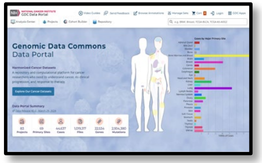
Data Commons

https://datacommons.cancer.gov/

CRDC Data Commons

PROTEOMIC • IMAGING
GENOMIC •
• CLINICAL
EPIDEMIOLOGY •
DATA

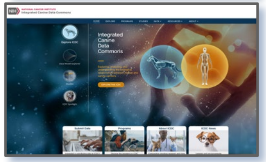
Data Commons



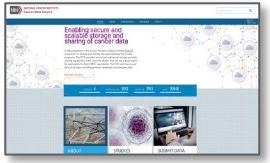
Genomic Data Commons

Proteomic Data Commons

Imaging Data Commons

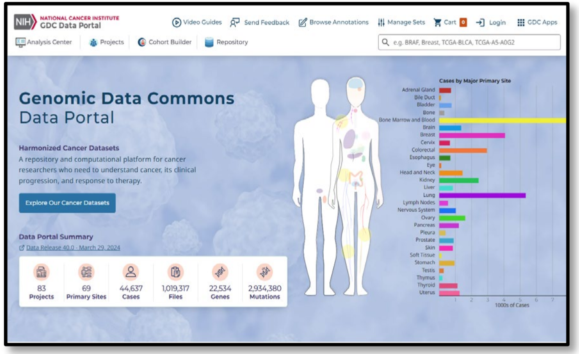


Integrated Canine Data Commons


Enabling secure and scalable storage and sharing of cancer data

Cancer Data Service

CRDC Data Commons



Genomic Data Commons
Data Portal

Harmonized Cancer Datasets

A repository and computational platform for cancer researchers who need to understand cancer, its clinical progression, and response to therapy.

Data Portal Summary

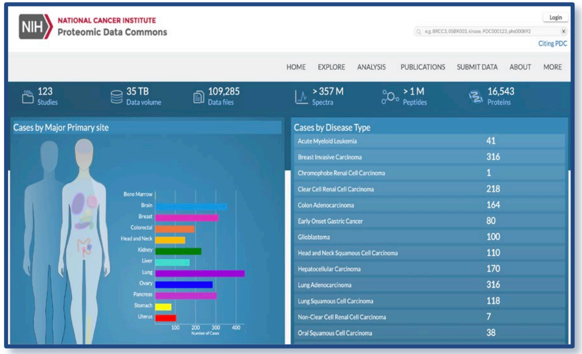
Data Release 40.0 - March 29, 2024

83	69	44,637	1,019,317	22,534	2,934,380
Projects	Primary Sites	Cases	Files	Genes	Mutations

Genomic Data Commons

- Share, analyze, and visualize genomic data
- Harmonized to the same genome standard and variant calling pipeline

https://portal.gdc.cancer.gov/



123	35 TB	109,285	> 357 M	> 1 M	16,543
Studies	Data volume	Data files	Spectra	Peptides	Proteins

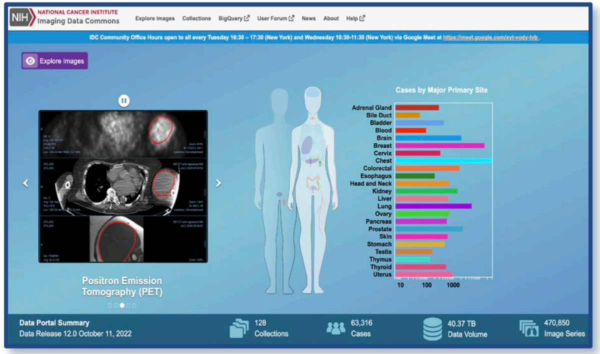
Cases by Disease Type

Acute Myeloid Leukemia	41
Breast Invasive Carcinoma	316
Chromophobe Renal Cell Carcinoma	1
Clear Cell Renal Cell Carcinoma	218
Colon Adenocarcinoma	164
Early Onset Gastric Cancer	80
Glioblastoma	100
Head and Neck Squamous Cell Carcinoma	110
Hepatocellular Carcinoma	170
Lung Adenocarcinoma	316
Lung Squamous Cell Carcinoma	118
Non-Clear Cell Renal Cell Carcinoma	7
Oral Squamous Cell Carcinoma	38

Proteomic Data Commons

- Filter, query, search, visualize and download proteomic data and metadata
- Data harmonization pipeline to uniformly analyze all PDC data

https://pdc.cancer.gov/



IDC Community Office Hours open to all every Tuesday 16:30 – 17:30 (New York) and Wednesday 10:30-11:30 (New York) via Google Meet at https://meet.google.com/xyt-vody-tvb .

Data Portal Summary
Data Release 12.0 October 11, 2022

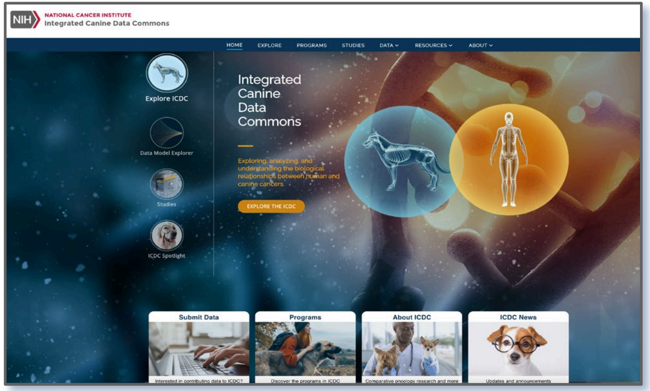
128	63,316	40.37 TB	470,850
Collections	Cases	Data Volume	Image Series

Imaging Data Commons

- Share, analyze, and visualize de-identified multi-modal imaging data, as medical images (MRI, PET, CT)
- Uses DICOM standard

https://imaging.datacommons.cancer.gov/

CRDC Data Commons



Integrated Canine Data Commons

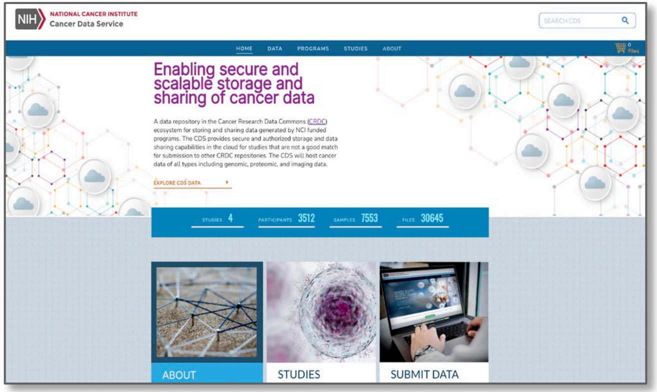
Exploring, analyzing, and understanding the biological relationships between human and canine cancers.

Interested in contributing data to ICDC?

Discover the programs in ICDC

Comparative oncology research and more

Updates and announcements



Enabling secure and scalable storage and sharing of cancer data

A data repository in the Cancer Research Data Commons (CRDC) ecosystem for storing and sharing data generated by NCI funded programs. The CDS provides secure and authorized storage and data sharing capabilities in the cloud for studies that are not a good match for submission to other CRDC repositories. The CDS will host cancer data of all types including genomic, proteomic, and imaging data.

STUDIES 4 PARTICIPANTS 3512 SAMPLES 7553 FILES 30645

Integrated Canine Data Commons

- Share data from canine clinical trials
- All data (including raw sequence data) are open access.

https://caninecommons.cancer.gov/

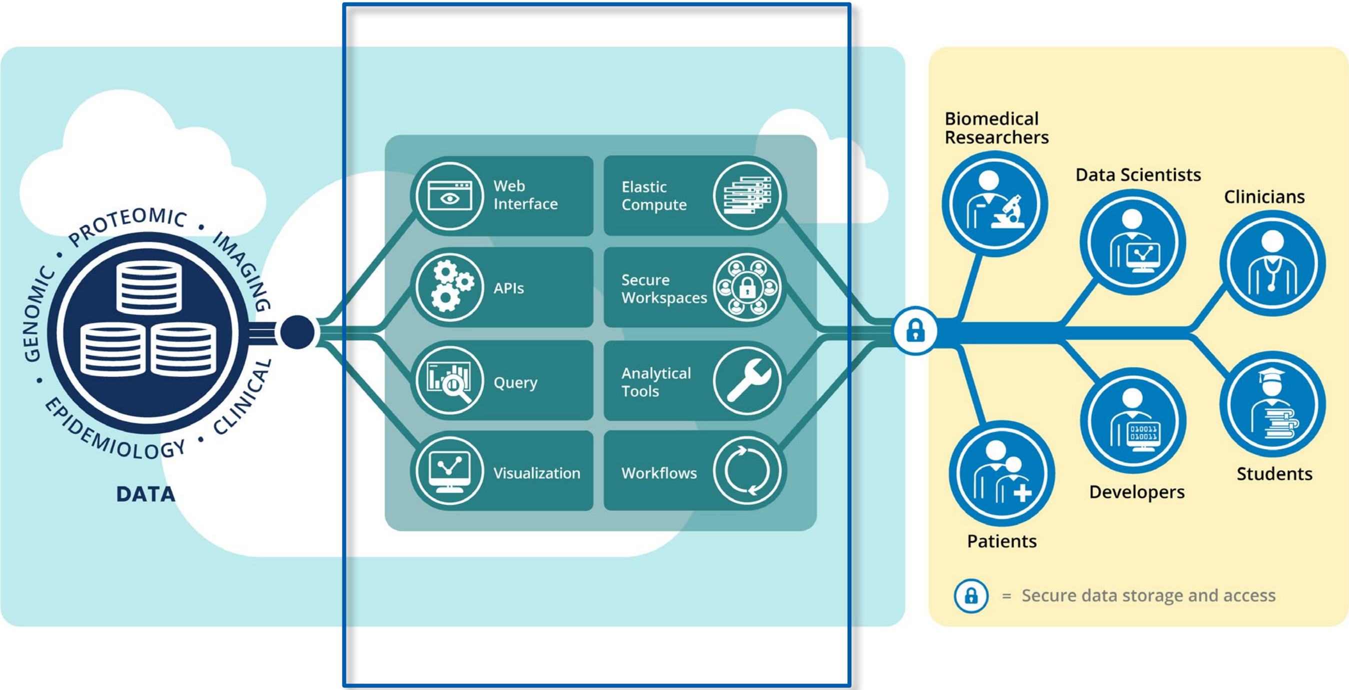
Cancer Data Service

- Access NCI-funded data currently not hosted by other CRDC data commons
- All datatypes accepted

https://dataservice.datacommons.cancer.gov/

Data types available across data commons: WGS, WXS, RNAseq, miRNA-seq, scRNAseq, ATAC-seq, DNA methylation, mass spectrometry-based proteomic data, DICOM.

CRDC Ecosystem



Cloud Resources

https://datacommons.cancer.gov/

CRDC Cloud Resources

Broad FireCloud (FC), powered by Terra

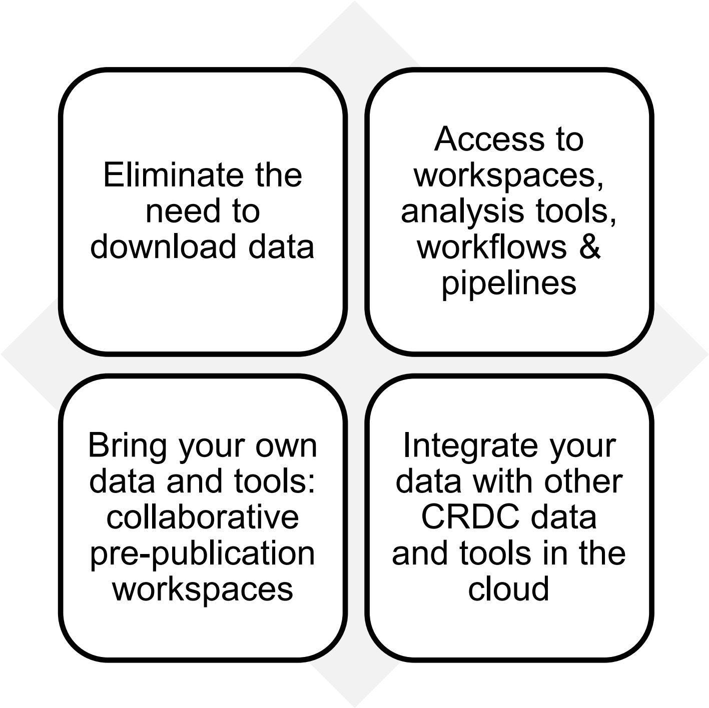
- Based on the Google Cloud Platform (GCP)
- Offers extensive repositories of pre-built tools and workflows in the Workflow Definition Language (WDL).

The ISB Cancer Gateway in the Cloud (ISB-CGC)

- Offers Google Cloud Platform (GCP) native tools and Google BigQuery for big data analytics and Google Compute Engine for complex workflow execution.
- Designed for users looking to use derived data.

The Seven Bridges Cancer Genomics Cloud (SB-CGC), powered by Velsera

- Based on the Amazon Web Services (AWS) platform
- Offers a curated library of over 850 tools and workflows optimized for the cloud using the Common Workflow Language (CWL).



Eliminate the need to download data

Access to workspaces, analysis tools, workflows & pipelines

Bring your own data and tools: collaborative pre-publication workspaces

Integrate your data with other CRDC data and tools in the cloud

Agenda

- Overview: CRDC Vision
- CRDC Ecosystem
  - Data Commons
  - Cloud Resources
- **Interoperability Initiatives**
  - **Initiatives within CRDC**
  - **Trans-NIH initiative (NCPI)**
- Resources for Researchers

CRDC: Interoperability Needs for Cancer Data

Challenge: Access comprehensive datasets like TCGA and CPTAC from multiple repositories for integrative analysis

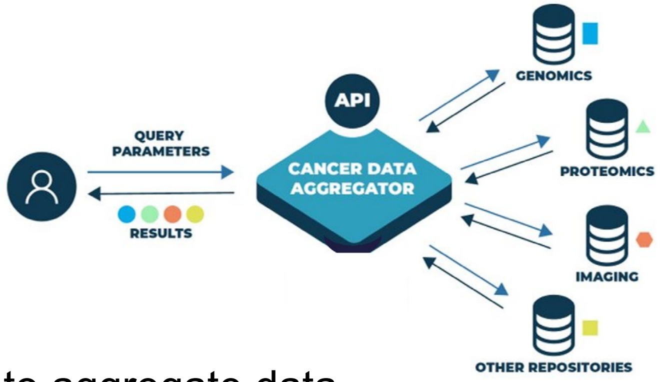
- Discover relevant datasets *across multiple resources using common standards*

- *Aggregate and analyze data* housed in separate data repositories using latest analytical tools

CRDC: Internal Interoperability Projects

CRDC Data Standards Services (DSS)

- Semantic harmonization across CRDC datasets
- Shared data models for submission & search
- Leverage existing standards, eg NCIt

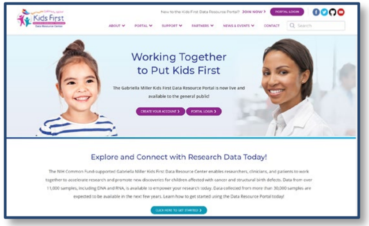


CRDC Cancer Data Aggregator (CDA)

- Search by harmonized, common language terms to aggregate data distributed across CRDC repositories
- Get information about subjects, files or specimens in a standard tsv format that can be opened in Excel, integrated into a pipeline or uploaded to a cloud resource
- cdapython available via interactive browser, notebooks or local install

NCPI: NIH Cloud Platform Interoperability

Connecting with a Greater Data Ecosystem

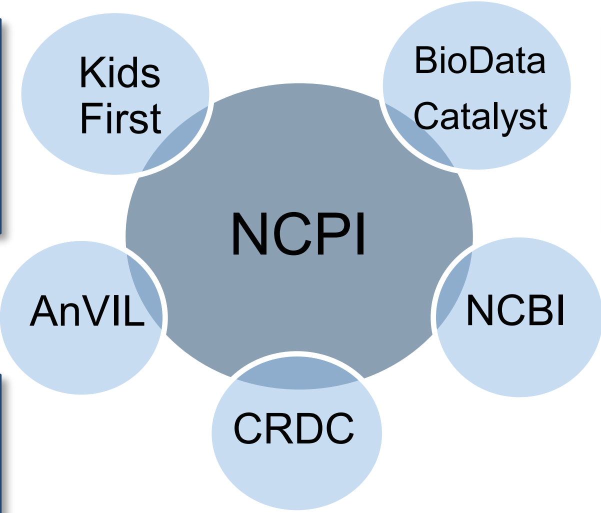

Gabriella Miller Kids First Data Resource Center
New to the Kids First Data Resource Portal? JOIN NOW ›
PORTAL LOGIN
ABOUT
PORTAL
SUPPORT
PARTNERS
NEWS & EVENTS
CONTACT
Search
CREATE YOUR ACCOUNT ›
PORTAL LOGIN ›
CLICK HERE TO GET STARTED! ›

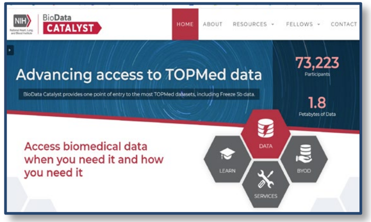
Working Together to Put Kids First

The Gabriella Miller Kids First Data Resource Portal is now live and available to the general public!

Explore and Connect with Research Data Today!

The NIH Common Fund-supported Gabriella Miller Kids First Data Resource Center enables researchers, clinicians, and patients to work together to accelerate research and promote new discoveries for children affected with cancer and structural birth defects. Data from over 11,000 samples, including DNA and RNA, is available to empower your research today. Data collected from more than 30,000 samples are expected to be available in the next few years. Learn how to get started using the Data Resource Portal today!

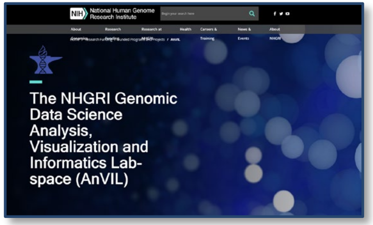




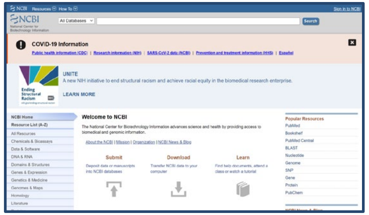
Advancing access to TOPMed data

BioData Catalyst provides one point of entry to the most TOPMed datasets, including Freeze 5b data.

Access biomedical data when you need it and how you need it



The NHGRI Genomic Data Science Analysis, Visualization and Informatics Lab-space (AnVIL)



COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

UNITE
A new NIH initiative to end structural racism and achieve racial equity in the biomedical research enterprise.

Welcome to NCBI

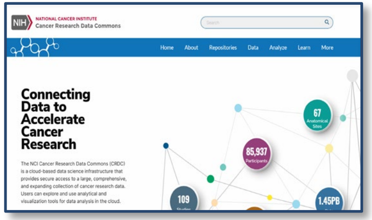
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit: Deposit data or manuscripts into NCBI databases

Download: Transfer NCBI data to your computer

Learn: Find help documents, attend a class or watch a tutorial



Connecting Data to Accelerate Cancer Research

The NCI Cancer Research Data Commons (CRDC) is a cloud-based data science infrastructure that provides secure access to a large, comprehensive, and expanding collection of cancer research data. Users can explore and use analytical and visualization tools for data analysis in the cloud.

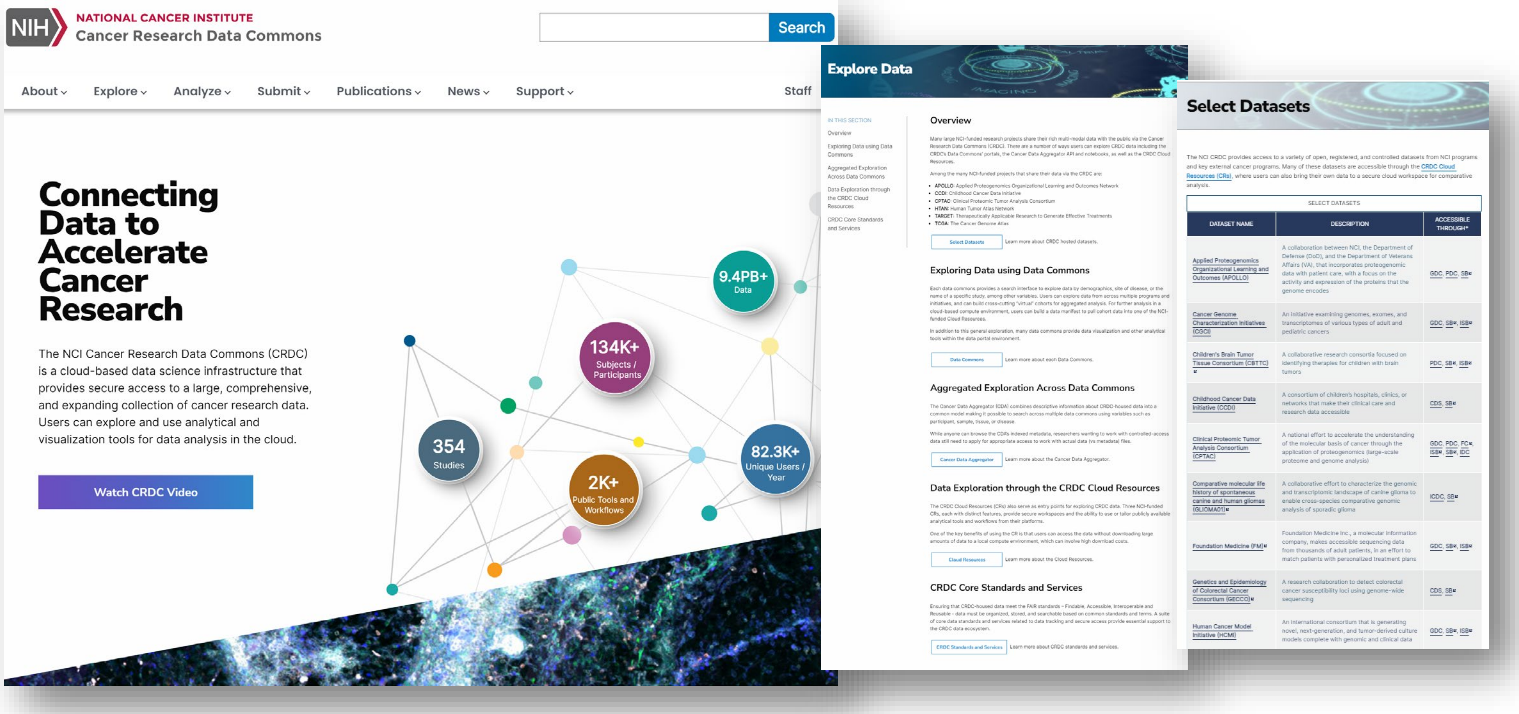
https://www.ncpi-acc.org/

Agenda

- Overview: CRDC Vision
- CRDC Ecosystem
  - Data Commons
  - Cloud Resources
- Interoperability Initiatives
  - Initiatives within CRDC
  - Trans-NIH initiative (NCPI)
- **Resources for Researchers**

CRDC Website

https://datacommons.cancer.gov/


NIH
NATIONAL CANCER INSTITUTE
Cancer Research Data Commons
Search
About
Explore
Analyze
Submit
Publications
News
Support
Staff
9.4PB+
Data
134K+
Subjects / Participants
354
Studies
2K+
Public Tools and Workflows
82.3K+
Unique Users / Year
Watch CRDC Video

Connecting Data to Accelerate Cancer Research

The NCI Cancer Research Data Commons (CRDC) is a cloud-based data science infrastructure that provides secure access to a large, comprehensive, and expanding collection of cancer research data. Users can explore and use analytical and visualization tools for data analysis in the cloud.

Explore Data

IN THIS SECTION
Overview
Exploring Data using Data Commons
Aggregated Exploration Across Data Commons
Data Exploration through the CRDC Cloud Resources
CRDC Core Standards and Services

Overview

Many large NCI-funded research projects share their rich multi-modal data with the public via the Cancer Research Data Commons (CRDC). There are a number of ways users can explore CRDC data including the CRDC's Data Commons' portals, the Cancer Data Aggregator API and notebooks, as well as the CRDC Cloud Resources.

Among the many NCI-funded projects that share their data via the CRDC are:

- **APOLLO**: Applied Proteogenomics Organizational Learning and Outcomes Network
- **CCDI**: Childhood Cancer Data Initiative
- **CPTAC**: Clinical Proteomic Tumor Analysis Consortium
- **HTAN**: Human Tumor Atlas Network
- **TARGET**: Therapeutically Applicable Research to Generate Effective Treatments
- **TCGA**: The Cancer Genome Atlas

Select Datasets Learn more about CRDC hosted datasets.

Exploring Data using Data Commons

Each data commons provides a search interface to explore data by demographics, site of disease, or the name of a specific study, among other variables. Users can explore data from across multiple programs and initiatives, and can build cross-cutting "virtual" cohorts for aggregated analysis. For further analysis in a cloud-based compute environment, users can build a data manifest to pull cohort data into one of the NCI-funded Cloud Resources.

In addition to this general exploration, many data commons provide data visualization and other analytical tools within the data portal environment.

Data Commons Learn more about each Data Commons.

Aggregated Exploration Across Data Commons

The Cancer Data Aggregator (CDA) combines descriptive information about CRDC-housed data into a common model making it possible to search across multiple data commons using variables such as participant, sample, tissue, or disease.

While anyone can browse the CDA's indexed metadata, researchers wanting to work with controlled-access data still need to apply for appropriate access to work with actual data (vs metadata) files.

Cancer Data Aggregator Learn more about the Cancer Data Aggregator.

Data Exploration through the CRDC Cloud Resources

The CRDC Cloud Resources (CRs) also serve as entry points for exploring CRDC data. Three NCI-funded CRs, each with distinct features, provide secure workspaces and the ability to use or tailor publicly available analytical tools and workflows from their platforms.

One of the key benefits of using the CR is that users can access the data without downloading large amounts of data to a local compute environment, which can involve high download costs.

Cloud Resources Learn more about the Cloud Resources.

CRDC Core Standards and Services

Ensuring that CRDC-housed data meet the FAIR standards – Findable, Accessible, Interoperable and Reusable - data must be organized, stored, and searchable based on common standards and terms. A suite of core data standards and services related to data tracking and secure access provide essential support to the CRDC data ecosystem.

CRDC Standards and Services Learn more about CRDC standards and services.

Select Datasets

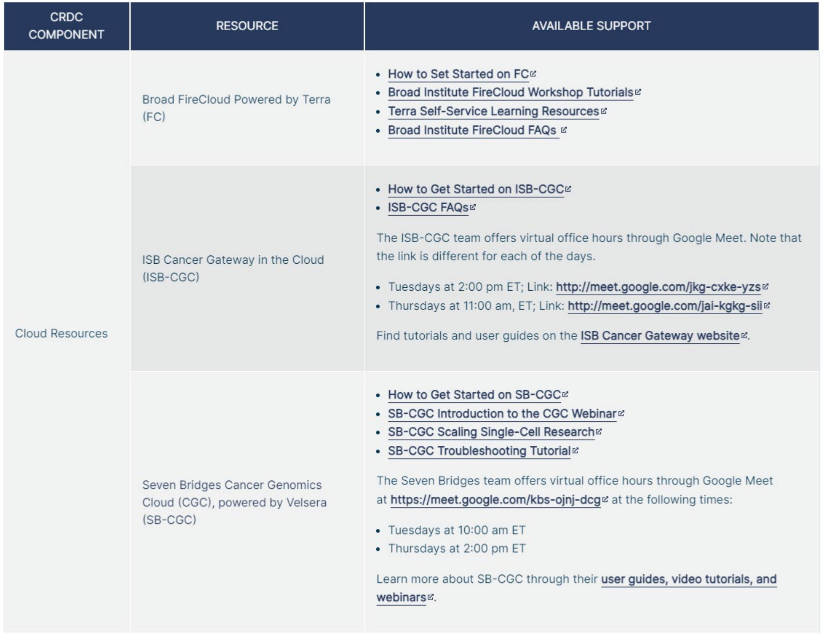
The NCI CRDC provides access to a variety of open, registered, and controlled datasets from NCI programs and key external cancer programs. Many of these datasets are accessible through the CRDC Cloud Resources (CRs), where users can also bring their own data to a secure cloud workspace for comparative analysis.

SELECT DATASETS

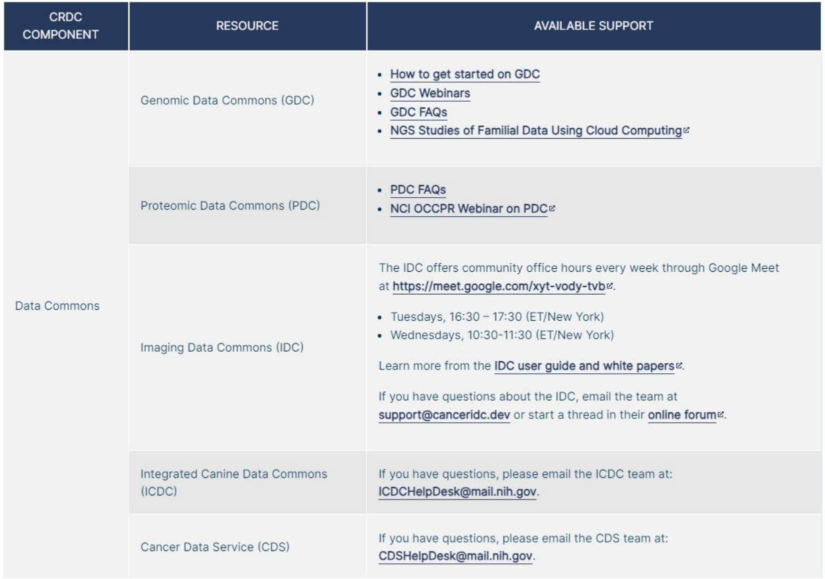
DATASET NAME	DESCRIPTION	ACCESSIBLE THROUGH*
Applied Proteogenomics Organizational Learning and Outcomes (APOLLO)	A collaboration between NCI, the Department of Defense (DoD), and the Department of Veterans Affairs (VA), that incorporates proteogenomic data with patient care, with a focus on the activity and expression of the proteins that the genome encodes	GDC, PDC, SB
Cancer Genome Characterization Initiatives (CGCI)	An initiative examining genomes, exomes, and transcriptomes of various types of adult and pediatric cancers	GDC, SB, ISB
Children's Brain Tumor Tissue Consortium (CBTTC)	A collaborative research consortia focused on identifying therapies for children with brain tumors	PDC, SB, ISB
Childhood Cancer Data Initiative (CCDI)	A consortium of children's hospitals, clinics, or networks that make their clinical care and research data accessible	CDS, SB
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	A national effort to accelerate the understanding of the molecular basis of cancer through the application of proteogenomics (large-scale proteome and genome analysis)	GDC, PDC, FC, ISB, SB, IDC
Comparative molecular life history of spontaneous canine and human gliomas (GLIOMA01)	A collaborative effort to characterize the genomic and transcriptomic landscape of canine glioma to enable cross-species comparative genomic analysis of sporadic glioma	ICDC, SB
Foundation Medicine (FM)	Foundation Medicine Inc., a molecular information company, makes accessible sequencing data from thousands of adult patients, in an effort to match patients with personalized treatment plans	GDC, SB, ISB
Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)	A research collaboration to detect colorectal cancer susceptibility loci using genome-wide sequencing	CDS, SB
Human Cancer Model Initiative (HCMI)	An international consortium that is generating novel, next-generation, and tumor-derived culture models complete with genomic and clinical data	GDC, SB, ISB

Support for Researchers

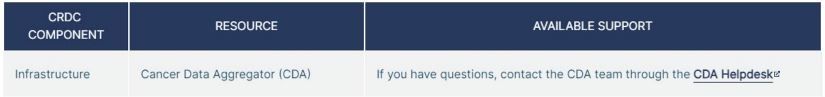
https://datacommons.cancer.gov/support-for-researchers



CRDC COMPONENT	RESOURCE	AVAILABLE SUPPORT
Cloud Resources	Broad FireCloud Powered by Terra (FC)	• How to Set Started on FC • Broad Institute FireCloud Workshop Tutorials • Terra Self-Service Learning Resources • Broad Institute FireCloud FAQs
	ISB Cancer Gateway in the Cloud (ISB-CGC)	• How to Get Started on ISB-CGC • ISB-CGC FAQs The ISB-CGC team offers virtual office hours through Google Meet. Note that the link is different for each of the days. • Tuesdays at 2:00 pm ET; Link: http://meet.google.com/jkg-cxke-yzs • Thursdays at 11:00 am, ET; Link: http://meet.google.com/jai-kgkg-sii Find tutorials and user guides on the ISB Cancer Gateway website.
	Seven Bridges Cancer Genomics Cloud (CGC), powered by Velsera (SB-CGC)	• How to Get Started on SB-CGC • SB-CGC Introduction to the CGC Webinar • SB-CGC Scaling Single-Cell Research • SB-CGC Troubleshooting Tutorial The Seven Bridges team offers virtual office hours through Google Meet at https://meet.google.com/kbs-ojnj-dcg at the following times: • Tuesdays at 10:00 am ET • Thursdays at 2:00 pm ET Learn more about SB-CGC through their user guides, video tutorials, and webinars.



CRDC COMPONENT	RESOURCE	AVAILABLE SUPPORT
Data Commons	Genomic Data Commons (GDC)	• How to get started on GDC • GDC Webinars • GDC FAQs • NGS Studies of Familial Data Using Cloud Computing
	Proteomic Data Commons (PDC)	• PDC FAQs • NCI OCCPR Webinar on PDC
	Imaging Data Commons (IDC)	The IDC offers community office hours every week through Google Meet at https://meet.google.com/xyt-vody-tvb. • Tuesdays, 16:30 – 17:30 (ET/New York) • Wednesdays, 10:30-11:30 (ET/New York) Learn more from the IDC user guide and white papers. If you have questions about the IDC, email the team at support@canceridc.dev or start a thread in their online forum.
	Integrated Canine Data Commons (ICDC)	If you have questions, please email the ICDC team at: ICDCHelpDesk@mail.nih.gov.
	Cancer Data Service (CDS)	If you have questions, please email the CDS team at: CDSHelpDesk@mail.nih.gov.



CRDC COMPONENT	RESOURCE	AVAILABLE SUPPORT
Infrastructure	Cancer Data Aggregator (CDA)	If you have questions, contact the CDA team through the CDA Helpdesk

CRDC Insights: External Newsletter (quarterly)

https://datacommons.cancer.gov/crdc-insights

NIH NATIONAL CANCER INSTITUTE
Cancer Research Data Commons

CRDC Insights
Quarterly: March 2023

NIH Data Management and Sharing Policy: CRDC's Role

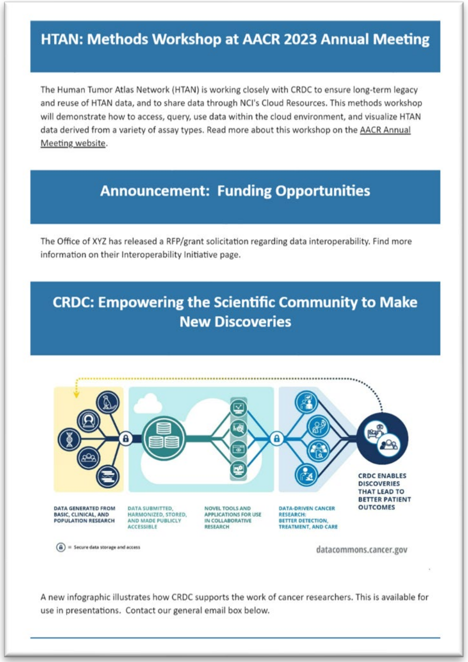
The new policy is in effect, and applies to new grant applications, competitive renewals, or competitive revisions. In brief:

- Data sharing now pertains to all researchers with no budget minimum.
- Applications, renewals, or revisions must include a data management and sharing plan.
- Data must be shared at time of publication or by the end of the performance period, whichever is sooner.

The Cancer Research Data Commons (CRDC) is home to a collection of data commons and cloud resources that host datasets from NCI-funded research, and make those datasets accessible to the research community. Learn more about submitting and accessing data, and using CRDC tools for your research. (link to data landing page)

CRDC Resources in the Classroom

Faculty members and data scientists with Purdue University and Velsara (Seven Bridges) teamed up to produce a four-part online workshop that introduces the Cancer Genomics Cloud. The series also provides hands-on lessons in bulk- and single-cell RNA-seq analysis using datasets provided by Purdue researchers. *Read the full story.*



HTAN: Methods Workshop at AACR 2023 Annual Meeting

The Human Tumor Atlas Network (HTAN) is working closely with CRDC to ensure long-term legacy and reuse of HTAN data, and to share data through NCI's Cloud Resources. This methods workshop will demonstrate how to access, query, use data within the cloud environment, and visualize HTAN data derived from a variety of assay types. Read more about this workshop on the AACR Annual Meeting website.

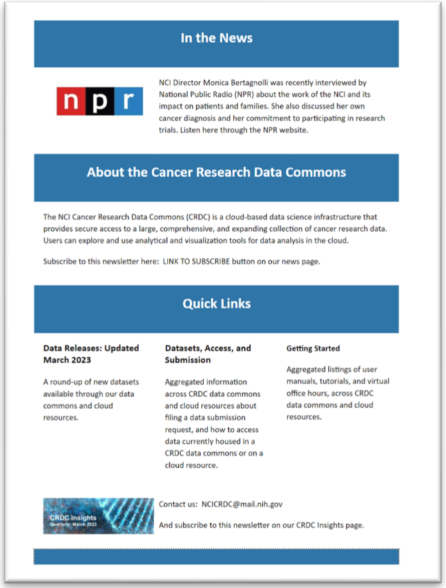
Announcement: Funding Opportunities

The Office of XYZ has released a RFP/grant solicitation regarding data interoperability. Find more information on their Interoperability Initiative page.

CRDC: Empowering the Scientific Community to Make New Discoveries

DATA GENERATED FROM BASIC, CLINICAL, AND POPULATION RESEARCH
DATA SUBMITTED, HARMONIZED, STORED, AND MADE PUBLICLY ACCESSIBLE
NOVEL TOOLS AND APPLICATIONS FOR USE IN COLLABORATIVE RESEARCH
DATA-DRIVEN CANCER RESEARCH: BETTER DETECTION, TREATMENT, AND CARE
CRDC ENABLES DISCOVERIES THAT LEAD TO BETTER PATIENT OUTCOMES
= Secure data storage and access
datacommons.cancer.gov

A new infographic illustrates how CRDC supports the work of cancer researchers. This is available for use in presentations. Contact our general email box below.



In the News

npr

NCI Director Monica Bertagnolli was recently interviewed by National Public Radio (NPR) about the work of the NCI and its impact on patients and families. She also discussed her own cancer diagnosis and her commitment to participating in research trials. Listen here through the NPR website.

About the Cancer Research Data Commons

The NCI Cancer Research Data Commons (CRDC) is a cloud-based data science infrastructure that provides secure access to a large, comprehensive, and expanding collection of cancer research data. Users can explore and use analytical and visualization tools for data analysis in the cloud.

Subscribe to this newsletter here: LINK TO SUBSCRIBE button on our news page.

Quick Links

Data Releases: Updated March 2023

A round-up of new datasets available through our data commons and cloud resources.

Datasets, Access, and Submission

Aggregated information across CRDC data commons and cloud resources about filing a data submission request, and how to access data currently housed in a CRDC data commons or on a cloud resource.

Getting Started

Aggregated listings of user manuals, tutorials, and virtual office hours, across CRDC data commons and cloud resources.

CRDC Insights
Quarterly: March 2023

Contact us: NCICRDC@mail.nih.gov

And subscribe to this newsletter on our CRDC Insights page.

2024 CRDC Fall Symposium: October 16-17, 2024

Save the Date

A one-and-a-half day event highlighting the 10th anniversary of the CRDC as well as plans for the future.

NIH MASUR AUDITORIUM, BETHESDA MD (10/16)
NCI CAMPUS, ROCKVILLE MD (10/17)

PRE-REGISTRATION REQUIRED

Register & More Information at:
DATACOMMONS.CANCER.GOV

CRDC and ODS Collaboration Session
Wednesday, October 16 @ 1:30 PM ET

- Data Sharing & Access within CRDC
- CRDC Symposium Kick-Off

Immediately following NCI Office of Data Sharing Symposium (separate event registration)

CRDC Session
Thursday, October 17 @ 9:00 AM ET

- CRDC History & Current State
- Success Stories & Impactful Programs
- Future Spotlight
- Fireside Chat

Overview:
NCI Cancer Research Data Commons (CRDC)

NIH NATIONAL CANCER INSTITUTE
Cancer Research Data Commons

bitly

datacommons.cancer.gov

AACR American Association for Cancer Research®
ANNUAL MEETING
2024 • SAN DIEGO

APRIL 5-10
#AACR24
AACR.ORG/AACR24

CRDC Impact and Success Stories

Esmeralda Casas-Silva, Ph.D.
Health Science Administrator
Center for Biomedical Informatics and Information Technology
National Cancer Institute, NIH, US

NIH NATIONAL CANCER INSTITUTE

Disclosure Information

Esmeralda Casas-Silva, Ph.D.

I am a full-time paid employee of the NIH/NCI.

I have no financial relationships to disclose.

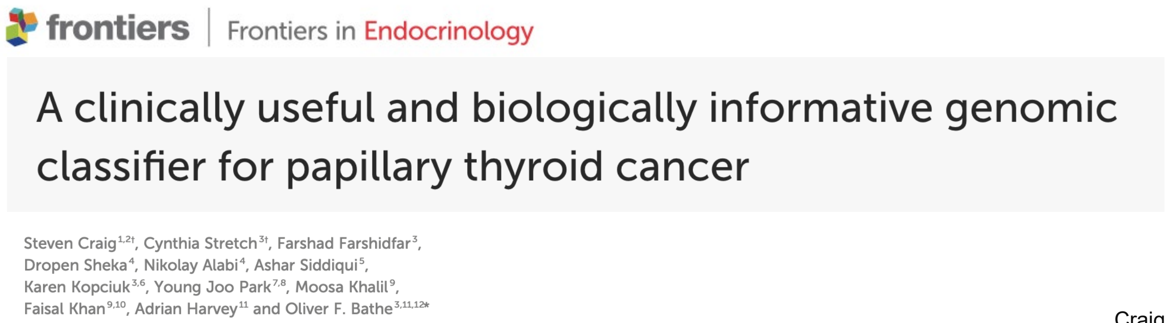
CRDC Success Story 1: Using GDC to Advance Papillary Thyroid Cancer (PTC) Care Through Genomic Classifiers

Problem:

- Higher recurrence and drug resistance in PTC subset
- No reliable way to predict which PTCs will progress

Goal:

- Identify molecular risk factors for papillary thyroid cancer progression
- Develop PTC genomic classifiers, stratify patients based on recurrence risk
  - Improve prognosis predictions



frontiers | Frontiers in Endocrinology

A clinically useful and biologically informative genomic classifier for papillary thyroid cancer

Steven Craig[1,2†], Cynthia Stretch[3†], Farshad Farshidfar[3], Dropen Sheka[4], Nikolay Alabi[4], Ashar Siddiqui[5], Karen Kopciuk[3,6], Young Joo Park[7,8], Moosa Khalil[9], Faisal Khan[9,10], Adrian Harvey[11] and Oliver F. Bathe[3,11,12*]

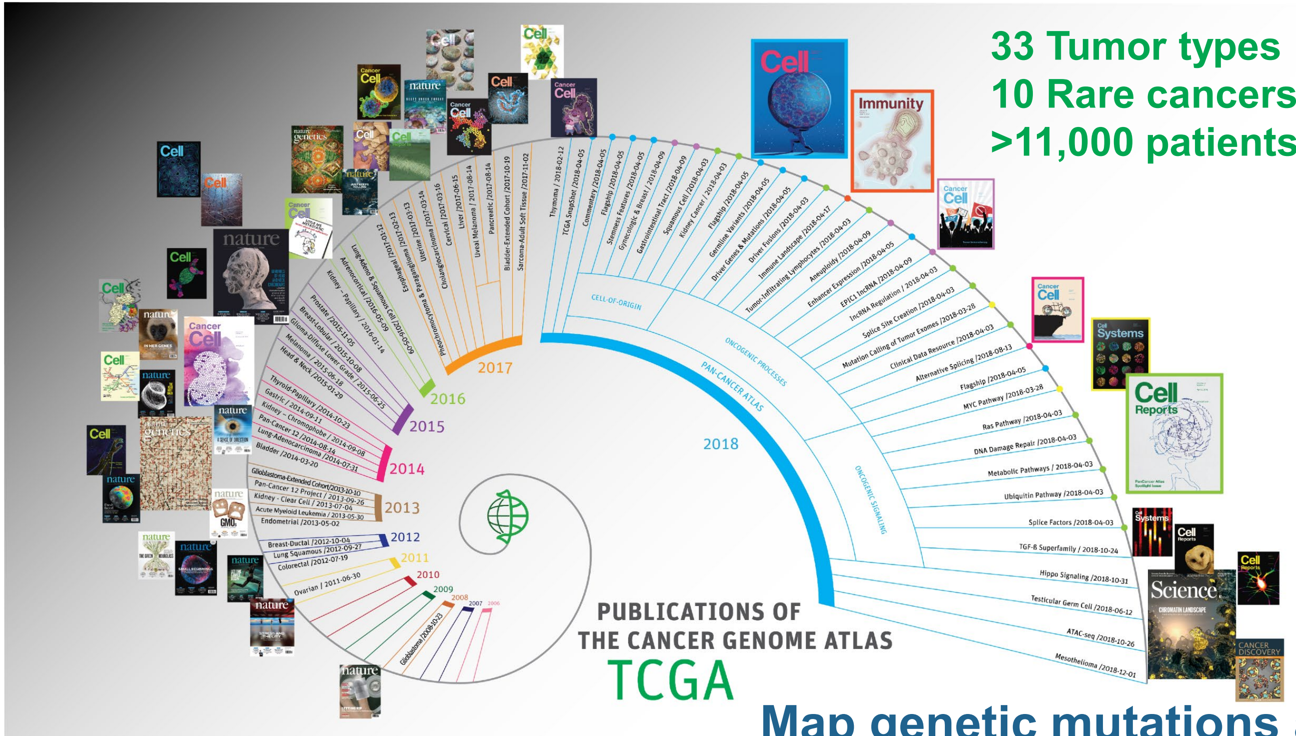
Craig et al., 2023, *Front. Endocrinol.*, 14:1220617

AACR American Association for Cancer Research®
ANNUAL MEETING
2024 • SAN DIEGO
APRIL 5-10 • AACR.ORG/AACR24 • #AACR24

The Cancer Genome Atlas (TCGA)

NIH NATIONAL CANCER INSTITUTE

NIH National Human Genome Research Institute


PUBLICATIONS OF
THE CANCER GENOME ATLAS
TCGA
2006
2007
2008
Glioblastoma /2008-10-23
2009
2010
2011
Ovarian / 2011-06-30
2012
Breast-Ductal /2012-10-04
Lung Squamous /2012-09-27
Colorectal /2012-07-19
2013
Glioblastoma-Extended Cohort/2013-10-10
Pan-Cancer 12 Project / 2013-09-26
Kidney - Clear Cell / 2013-07-04
Acute Myeloid Leukemia / 2013-05-30
Endometrial /2013-05-02
2014
Thyroid-Papillary /2014-10-23
Gastric / 2014-09-11
Kidney – Chromophobe / 2014-09-08
Pan-Cancer 12 /2014-08-14
Lung-Adenocarcinoma /2014-07-31
Bladder /2014-03-20
2015
Prostate /2015-11-05
Breast-Lobular / 2015-10-08
Glioma-Diffuse Lower Grade / 2015-06-25
Melanoma / 2015-06-18
Head & Neck /2015-01-29
2016
Lung-Adeno & Squamous Cell /2016-05-09
Adrenocortical /2016-05-09
Kidney – Papillary / 2016-01-14
2017
Esophageal /2017-01-12
Pheochromocytoma & Paraganglioma /2017-02-13
Uterine /2017-03-13
Cholangiocarcinoma /2017-03-14
Cervical /2017-03-16
Liver /2017-06-15
Uveal Melanoma / 2017-08-14
Pancreatic /2017-08-14
Bladder-Extended Cohort /2017-10-19
Sarcoma-Adult Soft Tissue /2017-11-02
2018
PAN-CANCER ATLAS
CELL-OF-ORIGIN
ONCOGENIC PROCESSES
ONCOGENIC SIGNALING
Thymoma / 2018-02-12
TCGA SnapShot /2018-04-05
Commentary /2018-04-05
Flagship /2018-04-05
Stemness Features /2018-04-05
Gynecologic & Breast / 2018-04-09
Gastrointestinal Tract /2018-04-09
Squamous Cell /2018-04-03
Kidney Cancer / 2018-04-03
Flagship /2018-04-05
Germline Variants /2018-04-05
Driver Genes & Mutations /2018-04-05
Driver Fusions /2018-04-03
Immune Landscape /2018-04-17
Tumor-Infiltrating Lymphocytes /2018-04-03
Aneuploidy /2018-04-09
Enhancer Expression /2018-04-05
EPIC1 lncRNA /2018-04-09
lncRNA Regulation / 2018-04-03
Splice Site Creation /2018-04-03
Mutation Calling of Tumor Exomes /2018-03-28
Clinical Data Resource /2018-04-03
Alternative Splicing /2018-08-13
Flagship /2018-04-05
MYC Pathway /2018-03-28
Ras Pathway /2018-04-03
DNA Damage Repair /2018-04-03
Metabolic Pathways / 2018-04-03
Ubiquitin Pathway /2018-04-03
Splice Factors /2018-04-03
TGF-ß Superfamily / 2018-10-24
Hippo Signaling /2018-10-31
Testicular Germ Cell /2018-06-12
ATAC-seq /2018-10-26
Mesothelioma /2018-12-01
Cell
Cancer Cell
nature
nature genetics
Immunity
Cell Systems
Cell Reports
Science
CANCER DISCOVERY

33 Tumor types
10 Rare cancers
>11,000 patients

2.5 Petabytes of data

Genomic Data Commons
Proteomic Data Commons
Imaging Data Commons

- ✓ Genomic
- ✓ Transcriptomic
- ✓ Epigenomic
- ✓ Imaging
- ✓ Clinical
- ✓ More…

Map genetic mutations across cancers

CRDC Success Story 1: Using GDC to Advance Papillary Thyroid Cancer (PTC) Care Through Genomic Classifiers

Approach:

- Leverage GDC to access cases from landmark TCGA study detailing genomic profile of PTC
  - Transcriptional data
  - Copy Number Variation
  - Methylation status
- Apply machine learning to 500+ cases
  - Stratify into molecular subtypes based on recurrence risk
- Used TCGA methylation data in GDC to explore epigenetic differences between cancer subtypes

CRDC Success Story 1: Using GDC to Advance Papillary Thyroid Cancer (PTC) Care Through Genomic Classifiers

Key findings:

- 3 unique molecular subtypes
  - Subtype 1: Lowest recurrence rate
    - Lower BRAFV600E, higher RAS mutations
  - Subtype 2: Moderate recurrence
    - Higher BRAFV600E, inflammatory, EMT pathways
  - Subtype 3: High recurrence rate
    - immunosuppressive microenvironment, high EZH2-HOTAIR pathway, BRAFV600E and TERT promoter mutations

Impact:

- Genomic classifiers outperformed the American Thyroid Association's clinical risk stratification system

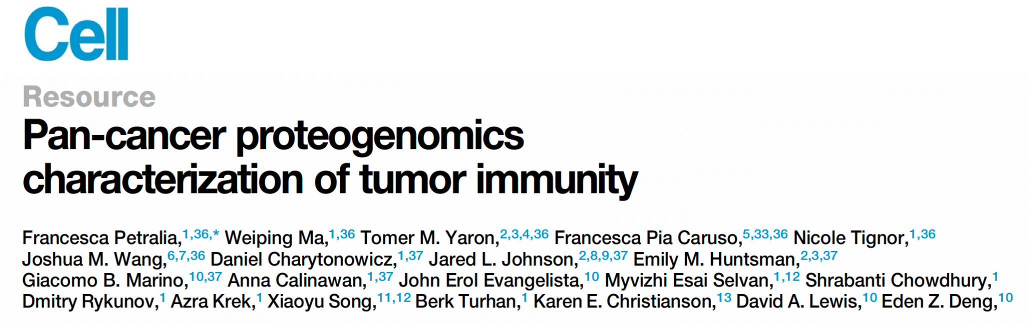
Success Story 2: Leveraging CRDC to decipher the Pan-Cancer Immune Landscape

Problem:

- Immunotherapy only successful in a small proportion of cancer cases

Goal:

- Develop comprehensive understanding TME across cancers
- Reveal immune cell surveillance and tumor immune evasion mechanisms



Cell

Resource

Pan-cancer proteogenomics characterization of tumor immunity

Francesca Petralia,[1,36,*] Weiping Ma,[1,36] Tomer M. Yaron,[2,3,4,36] Francesca Pia Caruso,[5,33,36] Nicole Tignor,[1,36] Joshua M. Wang,[6,7,36] Daniel Charytonowicz,[1,37] Jared L. Johnson,[2,8,9,37] Emily M. Huntsman,[2,3,37] Giacomo B. Marino,[10,37] Anna Calinawan,[1,37] John Erol Evangelista,[10] Myvizhi Esai Selvan,[1,12] Shrabanti Chowdhury,[1] Dmitry Rykunov,[1] Azra Krek,[1] Xiaoyu Song,[11,12] Berk Turhan,[1] Karen E. Christianson,[13] David A. Lewis,[10] Eden Z. Deng,[10]

Petralia, F., et al. (2024) Cell, 187:1255-1277, 2024

NCI **C**linical **P**roteomic **T**umor **A**nalysis **C**onsortium (CPTAC)

- Government, academia, industry partnership
- 1500+ patients, 10 tumor types
- Mass spectrometry proteomic data (PDC)
  - Protein expression
  - Post-translational modifications
  - Protein-protein interactions
- Genomic Data (GDC, CDS)
  - WGS, WXS
  - RNA-seq
  - CNV
- Clinical Data (GDC, PDC)
  - Treatment outcomes
  - More
- Imaging Data (IDC)

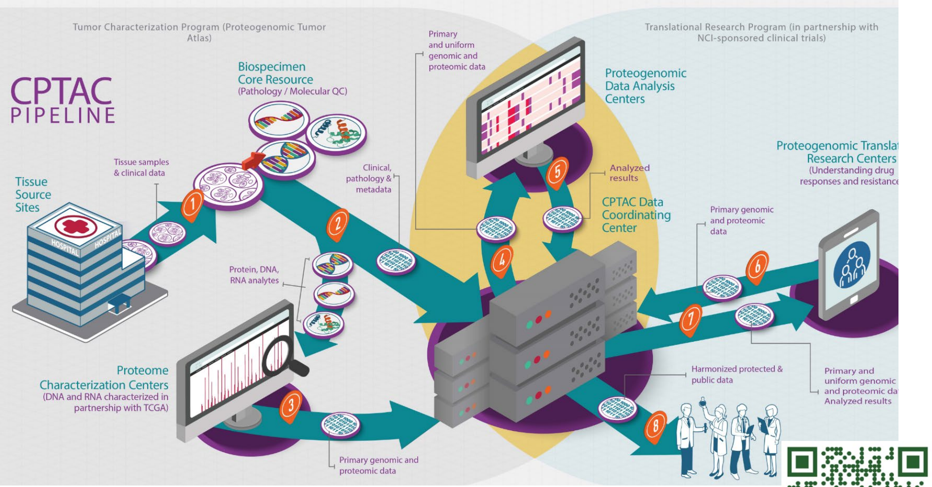

Tumor Characterization Program (Proteogenomic Tumor Atlas)
Translational Research Program (in partnership with NCI-sponsored clinical trials)
CPTAC PIPELINE
Tissue Source Sites
HOSPITAL
HOSPITAL
Tissue samples & clinical data
1
Biospecimen Core Resource (Pathology / Molecular QC)
Clinical, pathology & metadata
2
Protein, DNA, RNA analytes
Proteome Characterization Centers (DNA and RNA characterized in partnership with TCGA)
3
Primary genomic and proteomic data
Primary and uniform genomic and proteomic data
4
5
Proteogenomic Data Analysis Centers
Analyzed results
CPTAC Data Coordinating Center
Proteogenomic Translat
Research Centers
(Understanding drug
responses and resistance
Primary genomic and proteomic data
6
7
Harmonized protected & public data
Primary and uniform genomic and proteomic da
Analyzed results
8

Image: Clinical Proteomic Tumor Analysis Consortium, National Cancer Institute.
https://proteomics.cancer.gov/programs/cptac

Success Story 2: Leveraging CRDC to decipher the Pan-Cancer Immune Landscape

Approach:

- Combine CPTAC data from across CRDC
- Analyze genomic, epigenetic, transcriptomic, and proteomic alterations across tumors
- 1,056 tumor samples,10 cancers
- Classify tumors into immune subtypes
- Correlate with clinical outcomes

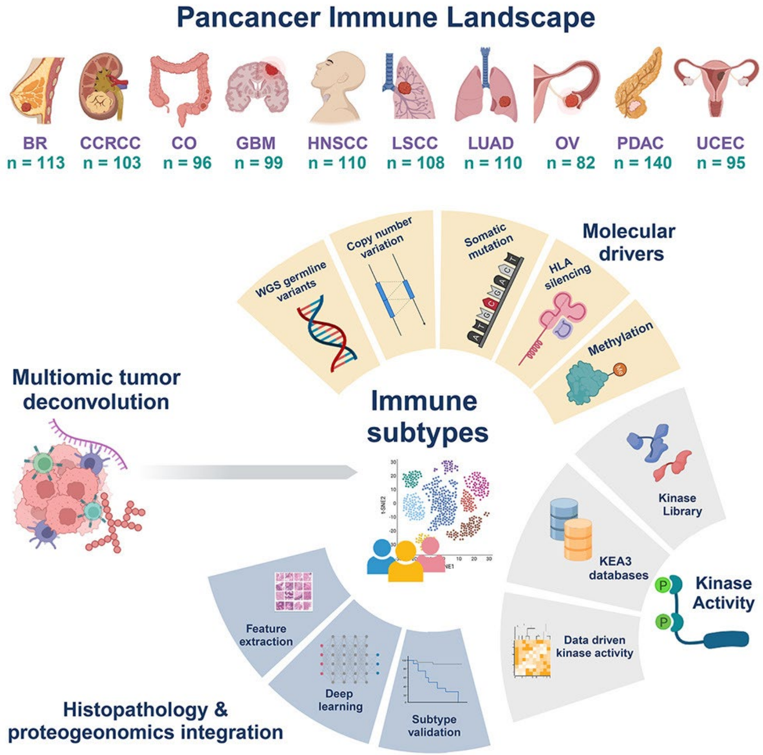


Image: Pan-cancer Immune Landscape Infographic. Derived from Petralia, F., et al., Cell 187, 2024

Success Story 2: Leveraging CRDC to decipher the Pan-Cancer Immune Landscape

Key findings:

- 7 distinct immune subtypes
  - Common immune reactions, evasion mechanisms independent of cancer type
- Correlations between PFS and immune subtypes, TME immune cell load
- Specific kinases activated in subtypes
  - Immune evasion, pathogenesis, and host immunity

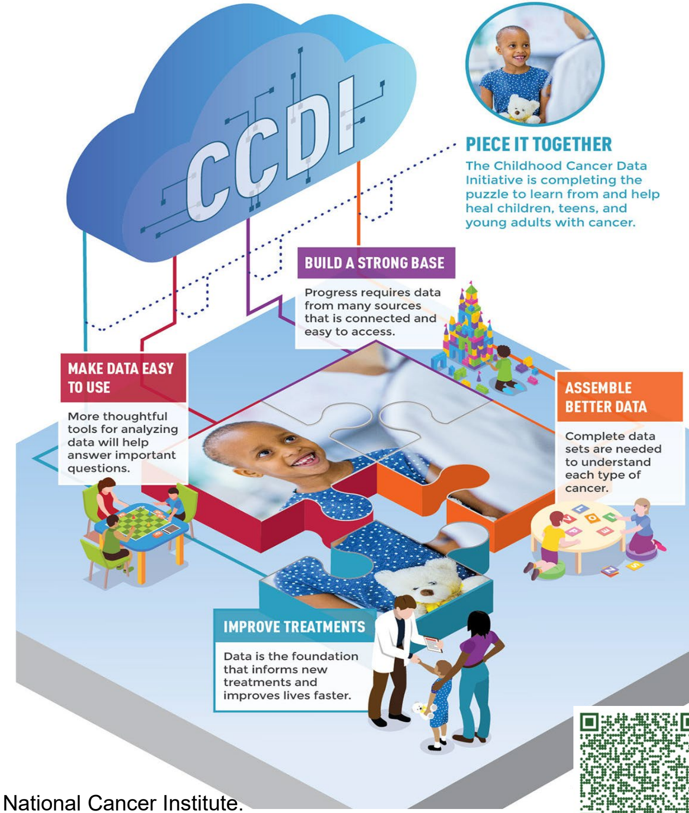
Impact:

- Multi-dimensional view of tumor biology
- Novel patient stratification, therapeutic strategies
- New interactive web portals: PhosNet Vis, ProKap
  - Leverage PDC's CPTAC pan cancer kinase and transcription factor activity score data to explore relationships with immune subtypes
  - New avenues for research and target discovery

CRDC Impact Story 3: Childhood Cancer Data Initiative (CCDI)

CCDI- Exemplar for building a learning health care system for cancer

- Community of researchers, advocates, hospitals and networks committed to sharing pediatric cancer data to accelerate research on childhood cancers.
- Federated Pediatric Cancer Data Ecosystem
  - Childhood cancer data and resources from across the nation
    - Research repositories
    - Patient registries
    - Hospitals



PIECE IT TOGETHER
The Childhood Cancer Data Initiative is completing the puzzle to learn from and help heal children, teens, and young adults with cancer.

BUILD A STRONG BASE
Progress requires data from many sources that is connected and easy to access.

MAKE DATA EASY TO USE
More thoughtful tools for analyzing data will help answer important questions.

ASSEMBLE BETTER DATA
Complete data sets are needed to understand each type of cancer.

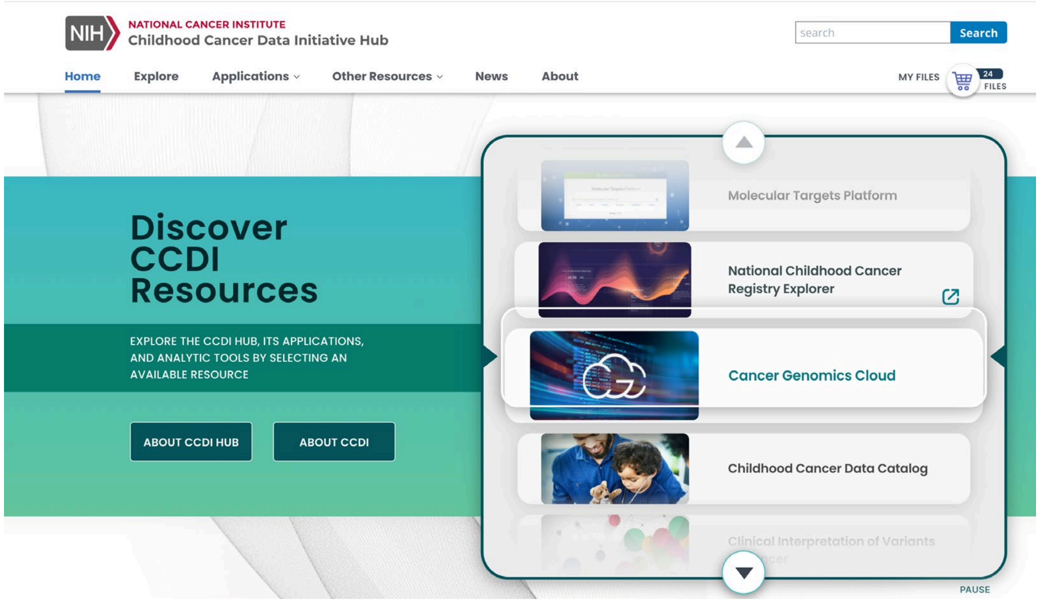
IMPROVE TREATMENTS
Data is the foundation that informs new treatments and improves lives faster.

Image: Childhood Cancer Data Initiative. National Cancer Institute.
https://www.cancer.gov/research/areas/childhood/childhood-cancer-data-initiative/about

CRDC Impact Story 3: Childhood Cancer Data Initiative (CCDI)

New CCDI Hub

- Entry point for researchers looking to use and connect with CCDI-related data and resources



Discover CCDI Resources

EXPLORE THE CCDI HUB, ITS APPLICATIONS, AND ANALYTIC TOOLS BY SELECTING AN AVAILABLE RESOURCE

NIH
NATIONAL CANCER INSTITUTE
Childhood Cancer Data Initiative Hub
search
Search
Home
Explore
Applications
Other Resources
News
About
MY FILES
24 FILES
ABOUT CCDI HUB
ABOUT CCDI
Molecular Targets Platform
National Childhood Cancer Registry Explorer
Cancer Genomics Cloud
Childhood Cancer Data Catalog
Clinical Interpretation of Variants
PAUSE

Image: Childhood Cancer Data Initiative Hub. National Cancer Institute. https://ccdi.cancer.gov/

CRDC Impact Story 3: Childhood Cancer Data Initiative (CCDI)

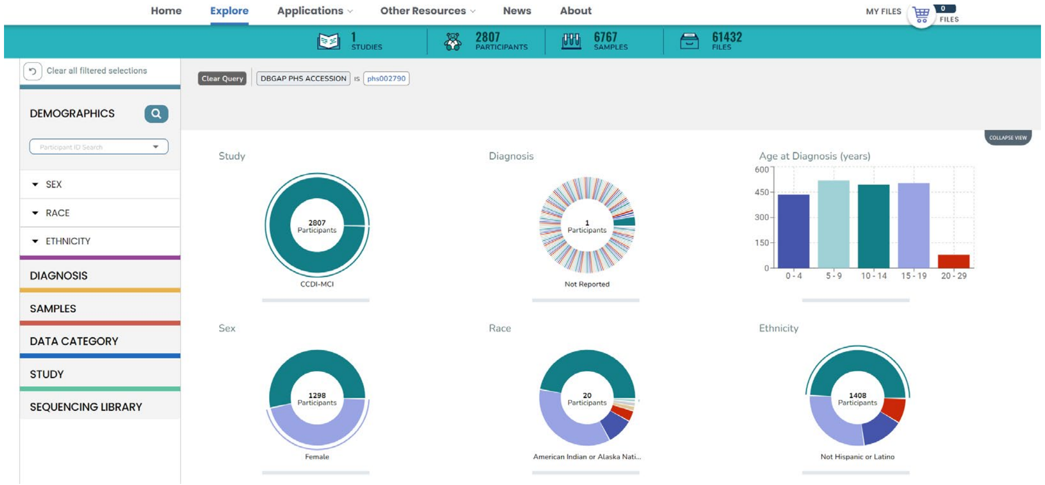
- Cancer Data Service (CDS) and Imaging Data Commons (IDC)
  - CCDI data from 1400+ participants
  - 70,000+ files
    - Genetic testing data
    - WGS and WXS
    - Full transcriptome sequencing
    - Single cell analysis
    - Imaging data

- Hosts Data for CCDI project
  - CCDI's Molecular Characterization Initiative (MCI)
    - Detailed clinical and molecular information, patients treated at academic and medical institutions around the country

CRDC Impact Story 3: Childhood Cancer Data Initiative (CCDI)

CRDC Infrastructure supports new CCDI Data Hub

- Bento framework
  - GUI for Data Exploration
- Authentication & Authorization controls
  - Integration of dbGaP A&A workflows for controlled data access



Clear Query DBGAP PHS ACCESSION IS phs002790

CRDC Impact Story 3: Childhood Cancer Data Initiative (CCDI)

CRDC Infrastructure supports new CCDI Data Hub

- Access to SB-CGC
  - Cloud-based environment
  - 500+ bioinformatics tools, workflows
  - Combine with own datasets or data across CRDC, NCI Cloud Platform Interoperability (NCPI)
  - Collaborative workspaces

CANCER GENOMICS CLOUD
SEVEN BRIDGES

NCPI

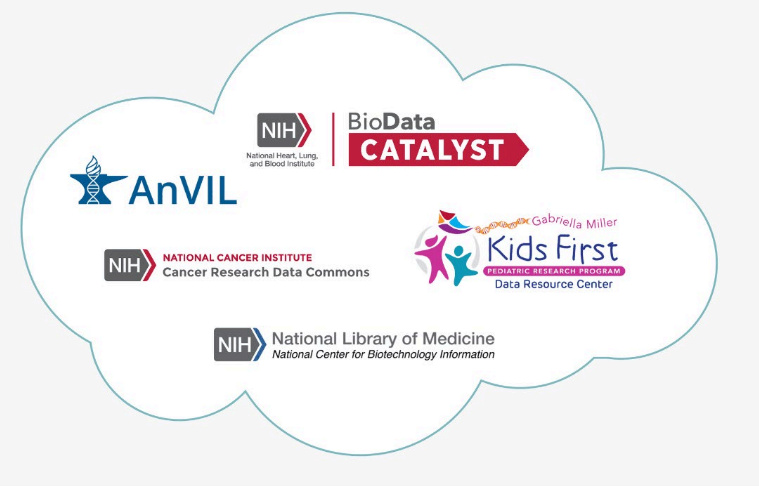

NIH National Heart, Lung, and Blood Institute
BioData CATALYST
AnVIL
NIH NATIONAL CANCER INSTITUTE Cancer Research Data Commons
Gabriella Miller Kids First PEDIATRIC RESEARCH PROGRAM Data Resource Center
NIH National Library of Medicine National Center for Biotechnology Information

Image: NIH Cloud Platform Interoperability Effort. NIH Office of Data Science Strategy. https://datascience.nih.gov/nih-cloud-platform-interoperability-effort."

AACR American Association for Cancer Research®

ANNUAL MEETING

2024 • SAN DIEGO

APRIL 5-10

#AACR24

AACR.ORG/AACR24

CRDC Lessons Learned and Future State

Anthony Kerlavage, Ph.D.
Director
Center for Biomedical Informatics and Information Technology
National Cancer Institute, NIH, US

NIH NATIONAL CANCER INSTITUTE

2302007Nn

Disclosure Information

Anthony Kerlavage, PhD

I am a full-time paid employee of the NIH/NCI.
I have no financial relationships to disclose.

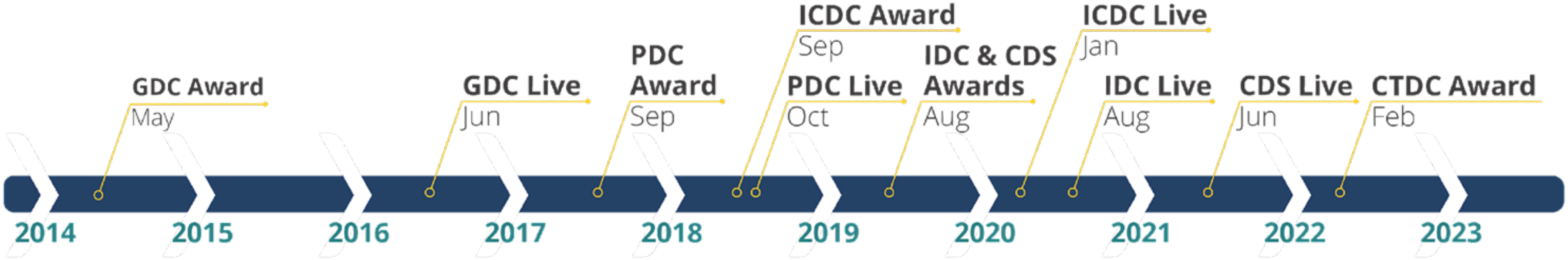
The Year of Open Science

- Advancing national open science policy
- Providing access to the results of the nation’s taxpayer-supported research
- Accelerating discovery and innovation
- Promoting public trust
- Driving more equitable outcomes

NIH’s Data Management and Sharing Policy went into effect on January 25, 2023, fulfilling the memorandum’s provisions around public access to scientific data.

CRDC: Celebrating 10 Years of Data Sharing

Data Commons



GDC Award May

GDC Live Jun

PDC Award Sep

ICDC Award Sep

PDC Live Oct

IDC & CDS Awards Aug

ICDC Live Jan

IDC Live Aug

CDS Live Jun

CTDC Award Feb

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023

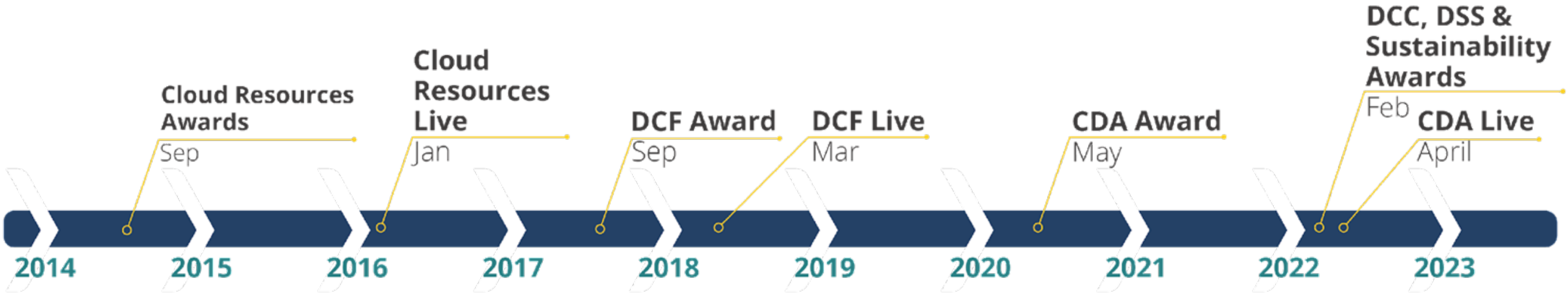
DATA COMMONS

GDC: Genomic Data Commons
PDC: Proteomic Data Commons
ICDC: Integrated Canine Data Commons

IDC: Imaging Data Commons
CDS: Cancer Data Service
CTDC: Clinical and Translational Data Commons
PSDC: Population Science Data Commons

CRDC: Celebrating 10 Years of Data Sharing

Data Infrastructure & Analysis



Cloud Resources Awards Sep 2015

Cloud Resources Live Jan 2016

DCF Award Sep 2017

DCF Live Mar 2018

CDA Award May 2020

DCC, DSS & Sustainability Awards Feb 2022

CDA Live April 2022

DATA INFRASTRUCTURE AND ANALYSIS

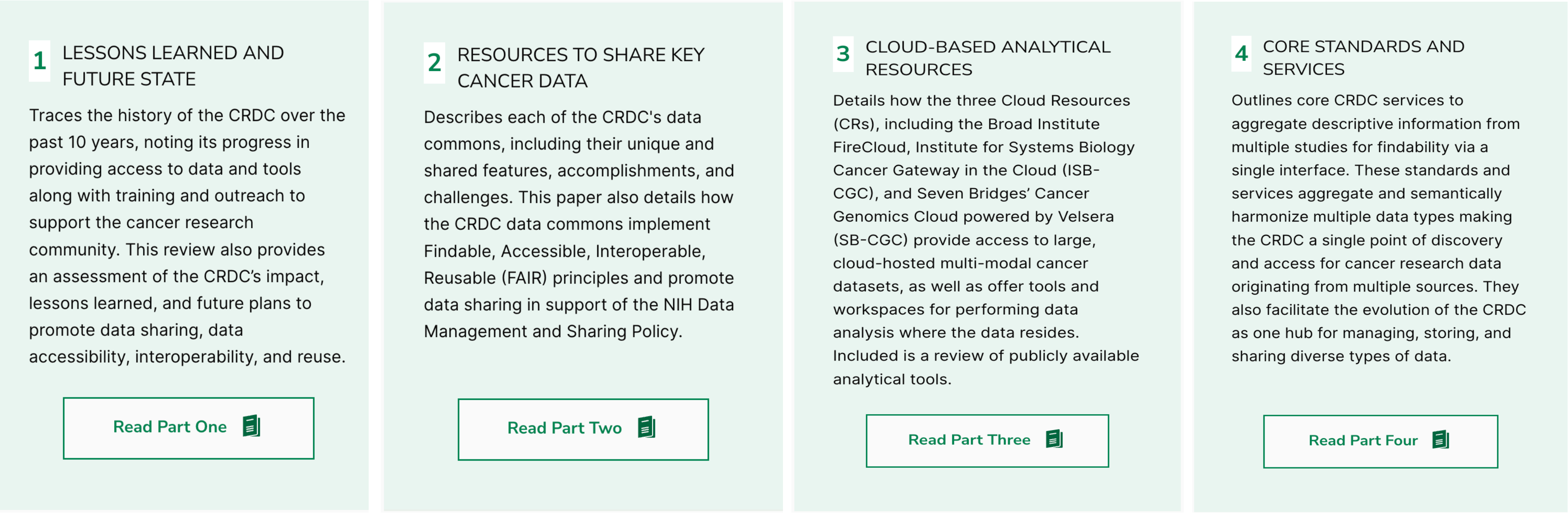
Cloud Resources: NCI Cloud Resources
DCF: Data Commons Framework
CDA: Cancer Data Aggregator

DCC: **CRDC Data Hub**
DSS: CRDC Data Standards Services
Sustainability: CRDC Sustainability Study

AACR American Association for Cancer Research®
ANNUAL MEETING
2024 • SAN DIEGO
APRIL 5-10 • AACR.ORG/AACR24 • #AACR24

Commemorating 10 Years of the CRDC

A four-part invited series in AACR *Cancer Research* journal showcasing how the CRDC empowers the cancer research community.



1 LESSONS LEARNED AND FUTURE STATE

Traces the history of the CRDC over the past 10 years, noting its progress in providing access to data and tools along with training and outreach to support the cancer research community. This review also provides an assessment of the CRDC's impact, lessons learned, and future plans to promote data sharing, data accessibility, interoperability, and reuse.

Read Part One

2 RESOURCES TO SHARE KEY CANCER DATA

Describes each of the CRDC's data commons, including their unique and shared features, accomplishments, and challenges. This paper also details how the CRDC data commons implement Findable, Accessible, Interoperable, Reusable (FAIR) principles and promote data sharing in support of the NIH Data Management and Sharing Policy.

Read Part Two

3 CLOUD-BASED ANALYTICAL RESOURCES

Details how the three Cloud Resources (CRs), including the Broad Institute FireCloud, Institute for Systems Biology Cancer Gateway in the Cloud (ISB-CGC), and Seven Bridges' Cancer Genomics Cloud powered by Velsera (SB-CGC) provide access to large, cloud-hosted multi-modal cancer datasets, as well as offer tools and workspaces for performing data analysis where the data resides. Included is a review of publicly available analytical tools.

Read Part Three

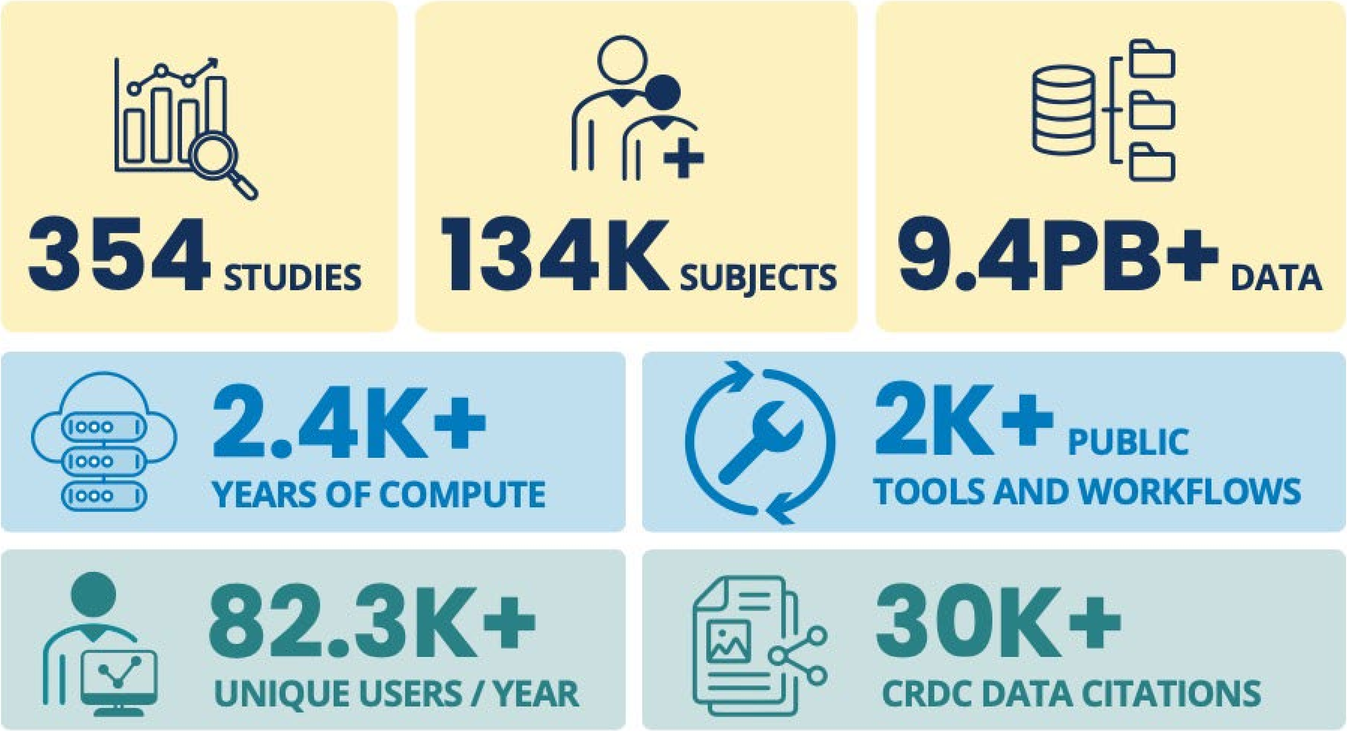
4 CORE STANDARDS AND SERVICES

Outlines core CRDC services to aggregate descriptive information from multiple studies for findability via a single interface. These standards and services aggregate and semantically harmonize multiple data types making the CRDC a single point of discovery and access for cancer research data originating from multiple sources. They also facilitate the evolution of the CRDC as one hub for managing, storing, and sharing diverse types of data.

Read Part Four

https://datacommons.cancer.gov/publications/aacr-cancer-research

CRDC Impact to Date



354 STUDIES

134K SUBJECTS

9.4PB+ DATA

2.4K+ YEARS OF COMPUTE

2K+ PUBLIC TOOLS AND WORKFLOWS

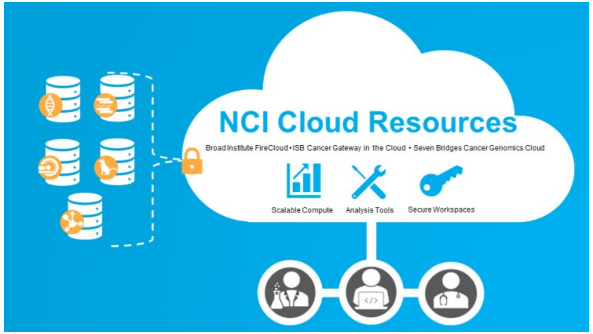
82.3K+ UNIQUE USERS / YEAR

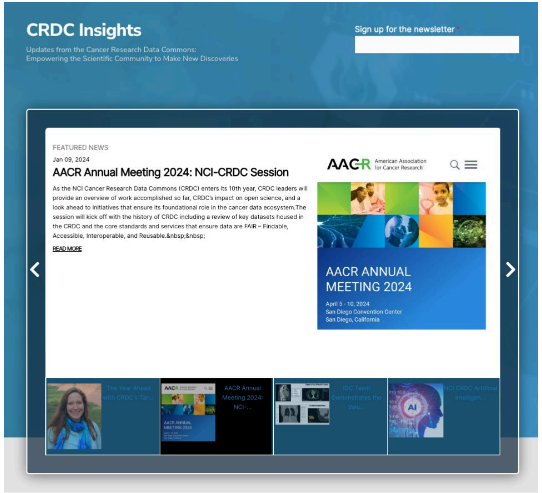
30K+ CRDC DATA CITATIONS

Lessons Learned: Community Engagement

- Increase in training and educational resources
  - Community engagement
    - Challenges, seminars, conferences, training sessions
  - Training videos, tools, and documentation
    - cloud use, cloud cost prediction tools, multi-modal analysis
  - CRDC Insights: Quarterly Newsletter

AI Data Readiness Challenge for the NCI Cancer Research Data Commons (CRDC)
AI
NIH NATIONAL CANCER INSTITUTE Center for Biomedical Informatics & Information Technology


NCI Cloud Resources
Broad Institute FireCloud • ISB Cancer Gateway in the Cloud • Seven Bridges Cancer Genomics Cloud
Scalable Compute
Analysis Tools
Secure Workspaces



CRDC Insights

Updates from the Cancer Research Data Commons:
Empowering the Scientific Community to Make New Discoveries

Sign up for the newsletter *

FEATURED NEWS

Jan 09, 2024

AACR Annual Meeting 2024: NCI-CRDC Session

As the NCI Cancer Research Data Commons (CRDC) enters its 10th year, CRDC leaders will provide an overview of work accomplished so far, CRDC's impact on open science, and a look ahead to initiatives that ensure its foundational role in the cancer data ecosystem.The session will kick off with the history of CRDC including a review of key datasets housed in the CRDC and the core standards and services that ensure data are FAIR – Findable, Accessible, Interoperable, and Reusable.

READ MORE

AACR American Association for Cancer Research
AACR ANNUAL MEETING 2024
April 5 - 10, 2024
San Diego Convention Center
San Diego, California
The Year Ahead with CRDC's Tan...
AACR Annual Meeting 2024: NCI-...
IDC Team Demonstrates the Valu...
NCI CRDC Artificial Intelligen...

Lessons Learned: Sustainability Study

- Planning for long-term sustainability of CRDC resources
  - CRDC supports the democratization of cancer research by providing **cloud-based, secure storage and analytic tools** for cancer data.
  - The Sustainability Study **supports the CRDC in planning for the financial sustainability** of its future work, **ensuring it operates efficiently and delivers value for money to the cancer research community** for years to come.

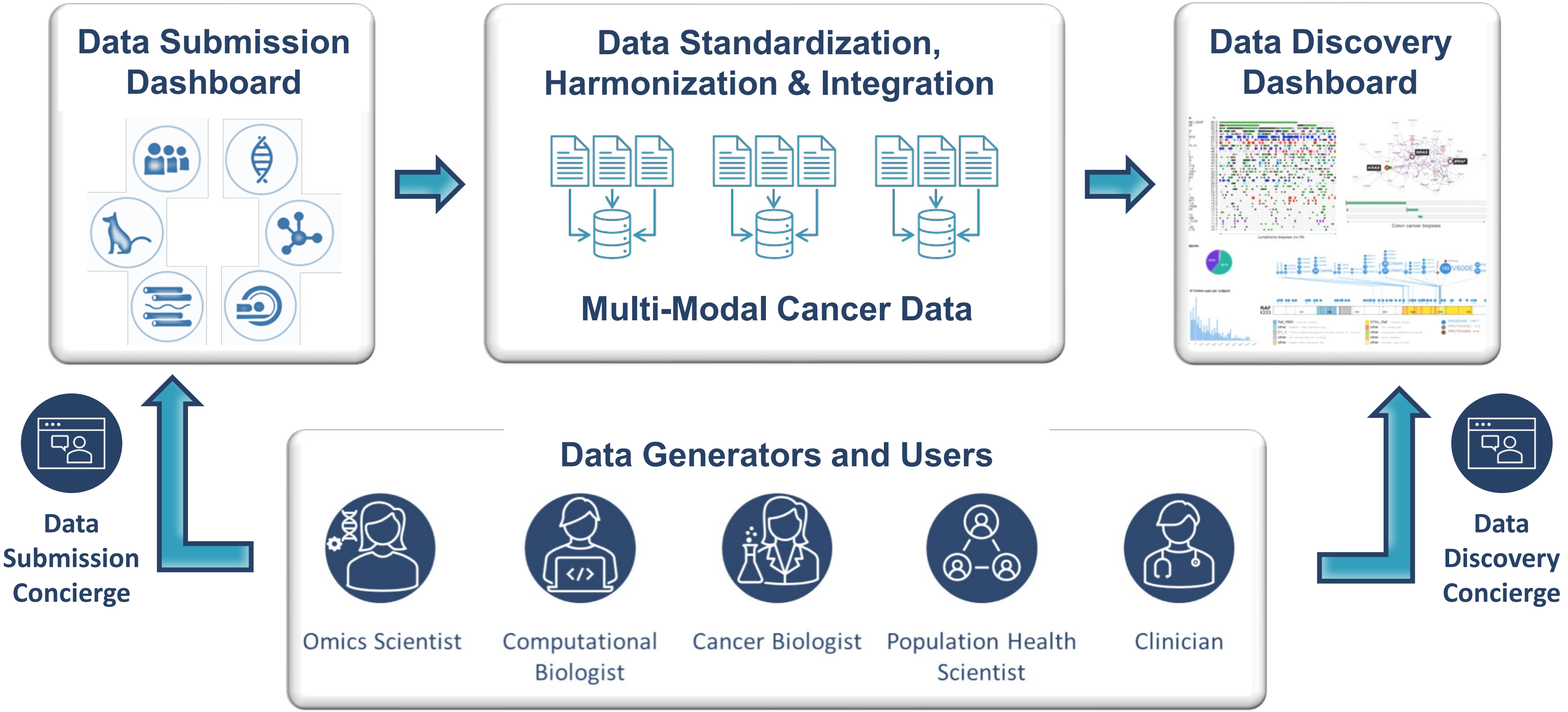
RECOMMENDATIONS FOR COST SAVINGS & OPERATIONAL OPTIMIZATION

BASELINE FOR CURRENT FINANCIAL REQUIREMENTS TO OPERATE CRDC; DEVELOP FUTURE FINANCIAL PROJECTIONS UNDER VARIOUS SCENARIOS

IDENTIFY BEST PRACTICES FROM EXTERNAL ENVIRONMENTS

Poster #3568: “Managing large-scale cancer research data programs” on April 8th, 1:30 – 5:00 pm

Lessons Learned: Lowering Barriers



Data Submission Dashboard

Data Standardization, Harmonization & Integration

Multi-Modal Cancer Data

Data Discovery Dashboard

Data Submission Concierge

Data Generators and Users

Omics Scientist

Computational Biologist

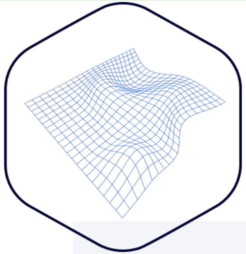
Cancer Biologist

Population Health Scientist

Clinician

Data Discovery Concierge

ARPA-H Biomedical Data Fabric* Toolbox



NCI in partnership with ARPA-H will advance the next-generation of tools to synthesize and speed use of health research data, starting with cancer

Make biomedical research data easier to use

Reduce effort for data integration

Develop new data fabric capabilities & tools

Build health data science models that can be applied across disciplines

* A data fabric provides a unified, consistent layer of data services that can work across many different systems and environments.

BDF Toolbox - Technical Areas

TA1: Automated Data Collection

Lower barriers to high-fidelity, timely, and automated data collection of research data across labs and health record systems

TA2: Machine-Assisted Curation

Prepare, connect, and harmonize multi-source data for analysis at scale

TA3: Intuitive Exploration

Enable advanced, human-centered data exploration and dashboards for use by diverse stakeholders and decision-makers

TA4: User Engagement

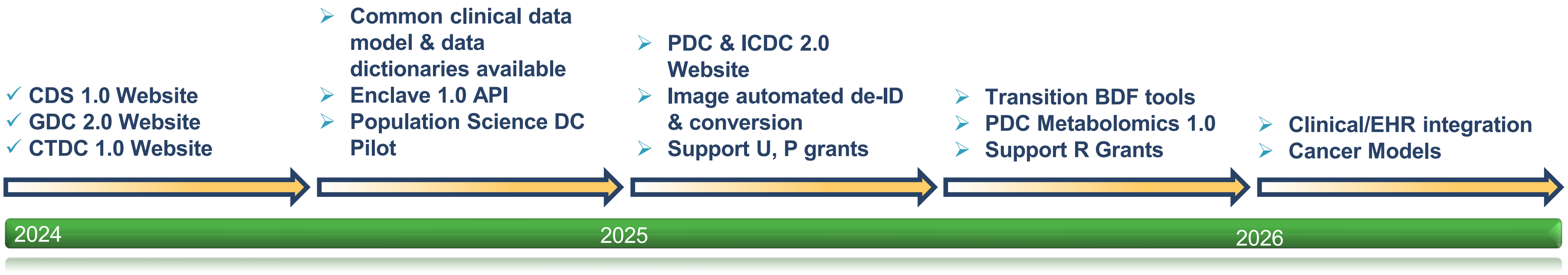
Evaluate data fabric tools across researchers, clinicians, and patients to create tools that will be enthusiastically adopted.

TA5: Cross-Domain Generalization

Leverage tools and platforms to generalize data across biomedical domains and disease types.

CRDC Roadmap

Data Commons



- ✓ CDS 1.0 Website
- ✓ GDC 2.0 Website
- ✓ CTDC 1.0 Website

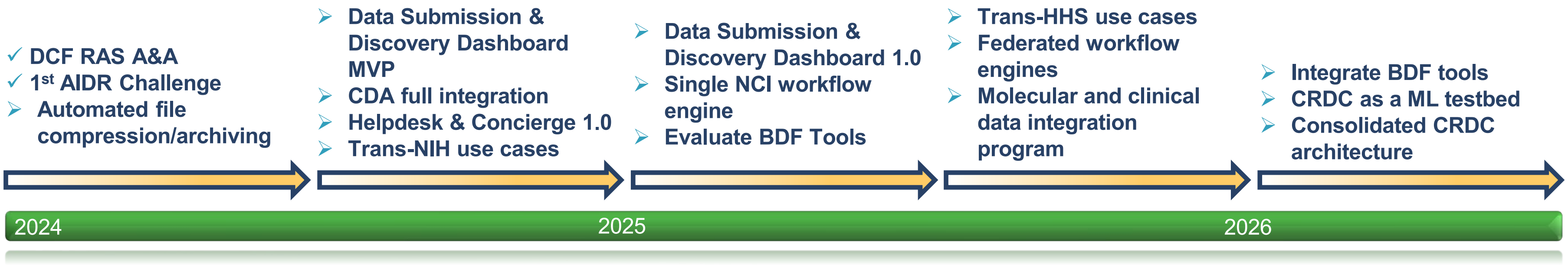
- ➢ Common clinical data model & data dictionaries available
- ➢ Enclave 1.0 API
- ➢ Population Science DC Pilot

- ➢ PDC & ICDC 2.0 Website
- ➢ Image automated de-ID & conversion
- ➢ Support U, P grants

- ➢ Transition BDF tools
- ➢ PDC Metabolomics 1.0
- ➢ Support R Grants

- ➢ Clinical/EHR integration
- ➢ Cancer Models

Infrastructure & Analysis



- ✓ DCF RAS A&A
- ✓ 1st AIDR Challenge
- ➢ Automated file compression/archiving

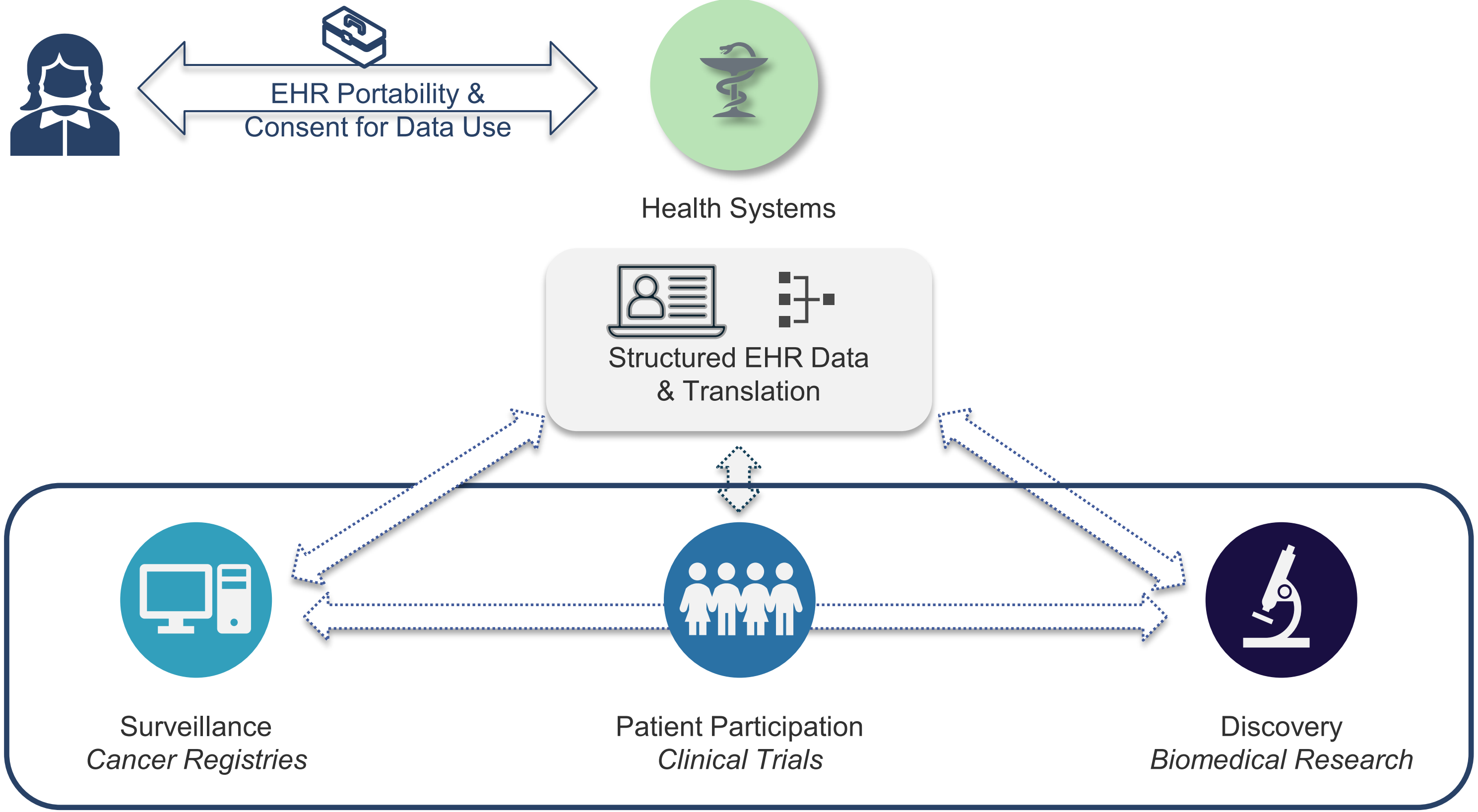
- ➢ Data Submission & Discovery Dashboard MVP
- ➢ CDA full integration
- ➢ Helpdesk & Concierge 1.0
- ➢ Trans-NIH use cases

- ➢ Data Submission & Discovery Dashboard 1.0
- ➢ Single NCI workflow engine
- ➢ Evaluate BDF Tools

- ➢ Trans-HHS use cases
- ➢ Federated workflow engines
- ➢ Molecular and clinical data integration program

- ➢ Integrate BDF tools
- ➢ CRDC as a ML testbed
- ➢ Consolidated CRDC architecture

Learning Health System for Cancer



Acknowledgement

NCI

Durga Addepalli
Jill Barnholtz-Sloan
Erin Beck
Emi Casas-Silva
Zhaoyi Chen
Heather Creasy
Tanja Davidsen
Ina Felau
Emily Greenspan
Jaime Guidry Auvil
Toby Hecht
Tony Kerlavage
Erika Kim
Henry Rodriquez
Pothur Srinivas
Louis Staudt
David Sturgill
Granger Sutton
Zhining Wang
Xu Zhang

FFRDC/FNL

Amanda Bell
Paula Darte
Hayley Dingerdissen
Sharon Gaheen
Naila Gulzar
Mark Jensen
Gina Kuffel
John Otridge
Sam Pathak
Todd Pihl
Sudha Venkatachari
Ulli Wagner
Mike Warfe

Essex

Shanthala Basavappa
Melissa Cook
Kim Gambini
Amal Ghannam
Shannon Lane
Luis Santana

ARPA-H

Andrea Bild
Julie Bletz
Jennifer Roberts
Alastair Thomson

All CRDC contractors

All partners throughout NCI/NIH and data contributors.

CANCER RESEARCH

The Foundational Cancer Journal
Driving Transformative Science

AACRJournals.org
@CR_AACR

AACR American Association for Cancer Research

Questions & Discussion

NIH NATIONAL CANCER INSTITUTE
Cancer Research Data Commons

datacommons.cancer.gov

AACR Cancer Research Series

CANCER RESEARCH

The Foundational Cancer Journal Driving Transformative Science

AACRJournals.org
@CR_AACR

AACR American Association for Cancer Research®

A four-part invited series published online in March 2024 highlighting the CRDC's accomplishments from the past 10 years.

- LESSONS LEARNED AND FUTURE STATE
- RESOURCES TO SHARE KEY CANCER DATA
- CLOUD-BASED ANALYTICAL RESOURCES
- CORE STANDARDS AND SERVICES

Learn more about the series on the CRDC Website.

Let's Connect

2024 AACR Annual Meeting: San Diego, CA

Presentations

- **Impact of the Cancer Research Data Commons (CRDC)**
  - Sunday, April 7 - 1:00pm – 2:00pm
- **NCI Artificial Intelligence (AI) Programs and Resources for Advancing Cancer Research**
  - Wednesday, April 10 - 10:15am -11:15am

Posters

- **ISB – Cancer Gateway in the Cloud**
  - Monday, April 8
- **CRDC – Sustainability Implementation Planning**
  - Monday, April 8
- **Velsera – Seven Bridges, Cancer Genomics Cloud**
  - Monday, April 8
  - Tuesday, April 9
  - Wednesday, April 10
- **Broad – FireCloud (Terra)**
  - Wednesday, April 10

View the AACR Program for more details.

NIH NATIONAL CANCER INSTITUTE

2024 AACR Annual Meeting: San Diego, CA

Posters

- **Sunday, April 7, 2024 / 1pm - 5pm**
  Session PO.RSP01.01 – Regulatory Science and Policy 920/3 – Insights from the NCI request for information on existing data sharing processes for NIH-funded research

- **Monday, April 8, 2024 / 9 am – 12:30 pm**
  Session PO.SHP01.01 – Science and Health Policy 1303/17 – Harness the power of data to improve cancer care – understanding the complex landscape of health policies and regulations

- **Tuesday, April 9, 2024 / 2:30 – 3:30 pm**
  Session NIH10 – Building on the Power of Data and Community – CCDI data ecosystem: Tools and resources

- **Wednesday, April 10, 2024 / 10:15 – 11:15 am**
  Session NIH12 - NCI Artificial Intelligence (AI) Programs and Resources for Advancing Cancer Research – NCI funding opportunities, resources and activities

Save the Date

2024 CRDC Fall Symposium: October 16-17, 2024

A one-and-a-half day event highlighting the 10th anniversary of the CRDC as well as plans for the future.

NIH MASUR AUDITORIUM, BETHESDA MD (10/16)
NCI CAMPUS, ROCKVILLE MD (10/17)

PRE-REGISTRATION REQUIRED

REGISTER & MORE INFORMATION AT DATACOMMONS.CANCER.GOV

CRDC and ODS Collaboration Session
Wednesday, October 16 @ 1:30 PM ET

- Data Sharing & Access within CRDC
- CRDC Symposium Kick-Off

Immediately following NCI Office of Data Sharing Symposium (separate event registration)

CRDC Session
Thursday, October 17 @ 9:00 AM ET

- CRDC History & Current State
- Success Stories & Impactful Programs
- Future Spotlight
- Fireside Chat

NIH NATIONAL CANCER INSTITUTE