THE FUTURE OF CANCER DATA: UNLOCKING INSIGHTS WITH PATHOLOGY REPORTING

New Frontiers:
Connecting Patient Care, Public Health, and Cancer Research

Jaime Guidry Auvil, PhD

OCTOBER 6 | 1:15–2 PM CT

COLLEGE of AMERICAN PATHOLOGISTS
Laboratory Quality Solutions

CAP23 | CHICAGO

#PATHDATA

New Frontiers: Connecting Patient Care, Public Health, and Cancer Research

Jaime M. Guidry Auvil, Ph.D.
Director, Office of Data Sharing

NIH NATIONAL CANCER INSTITUTE

October 6, 2023

Disclosures

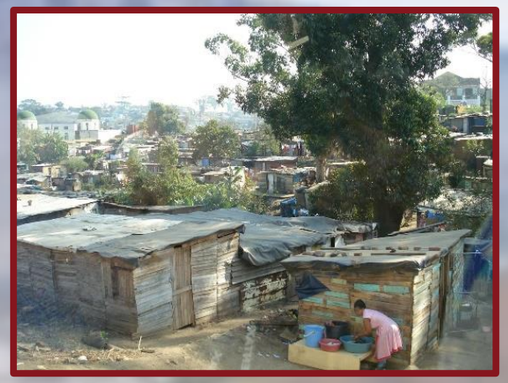
Agenda

1. Integrating Data to See the Big Picture
2. Open Science and Collaboration to Answer Critical Questions
3. Establishing Policies for Impactful Sharing of High Value Data
4. Developing Infrastructure to Support a Learning Health Model

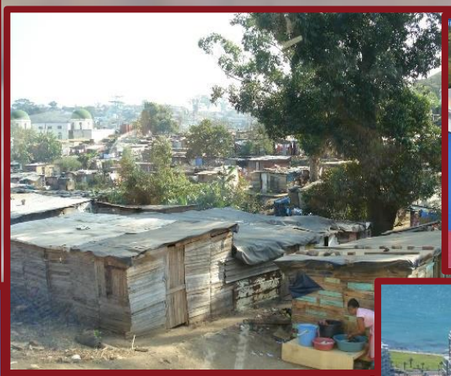
How do we “see” data?

What Are We Going to Study?

Cape Town, South Africa

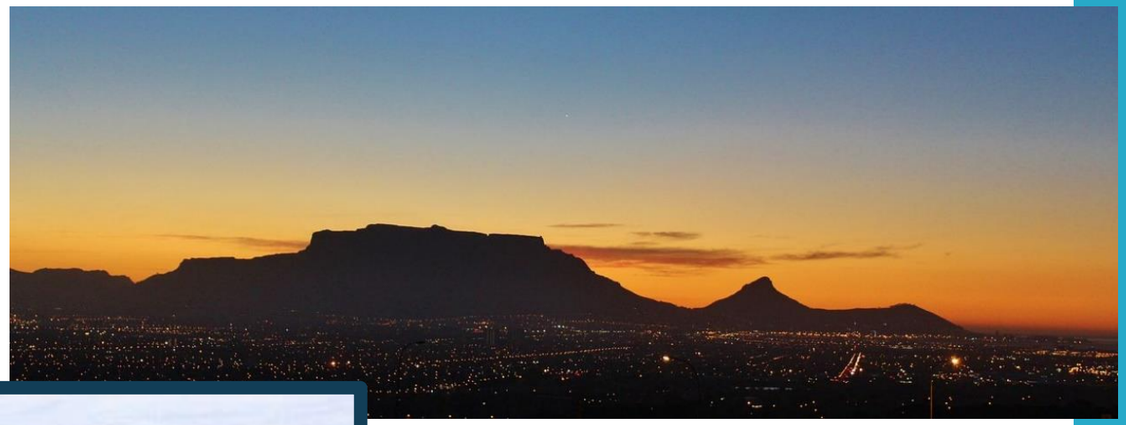


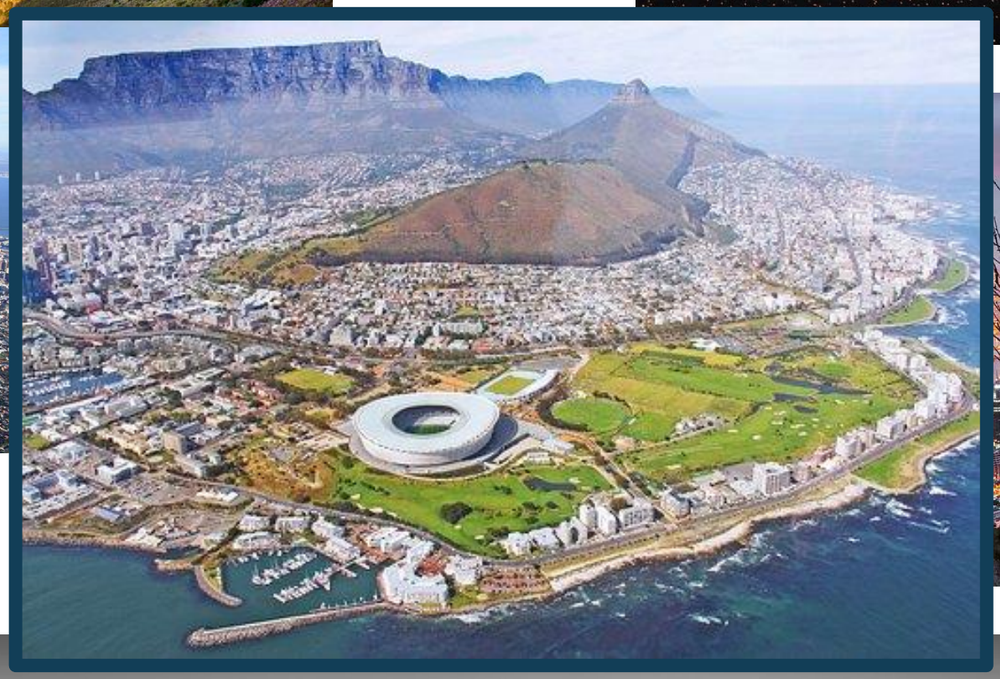


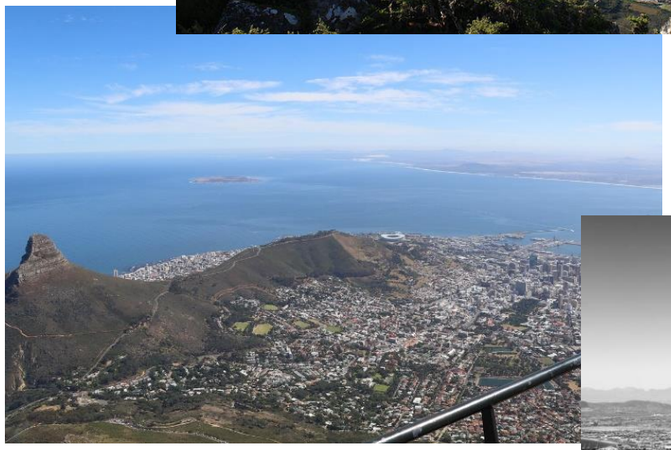


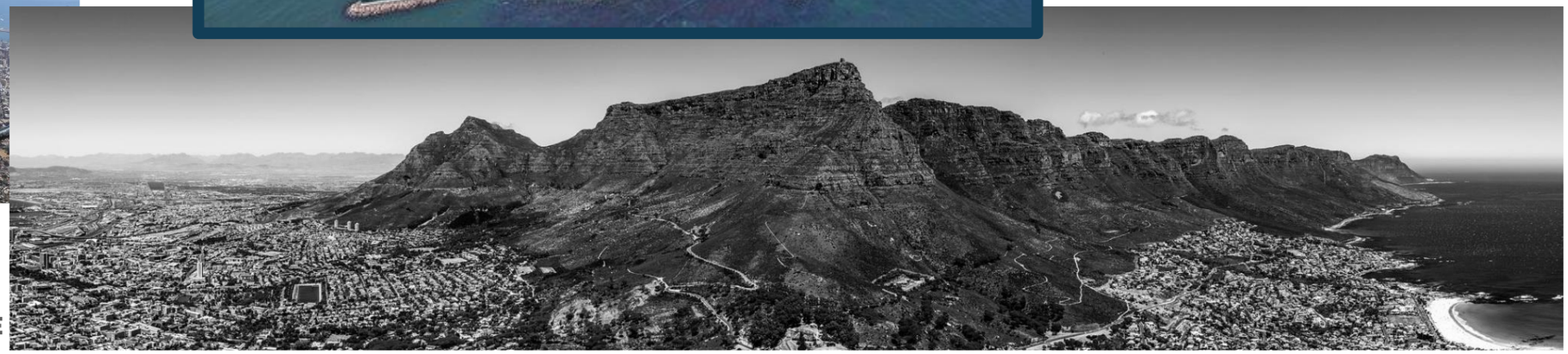




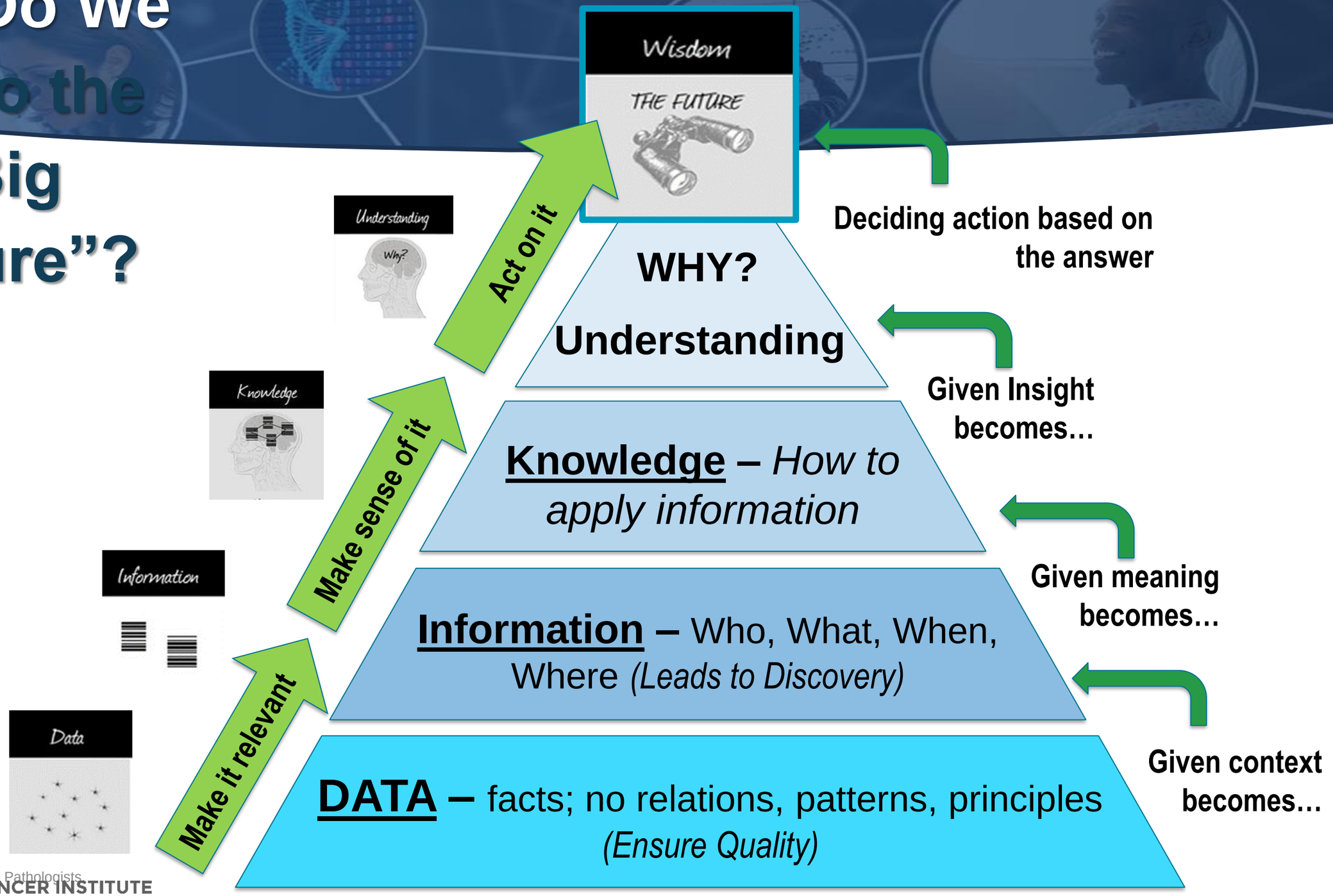




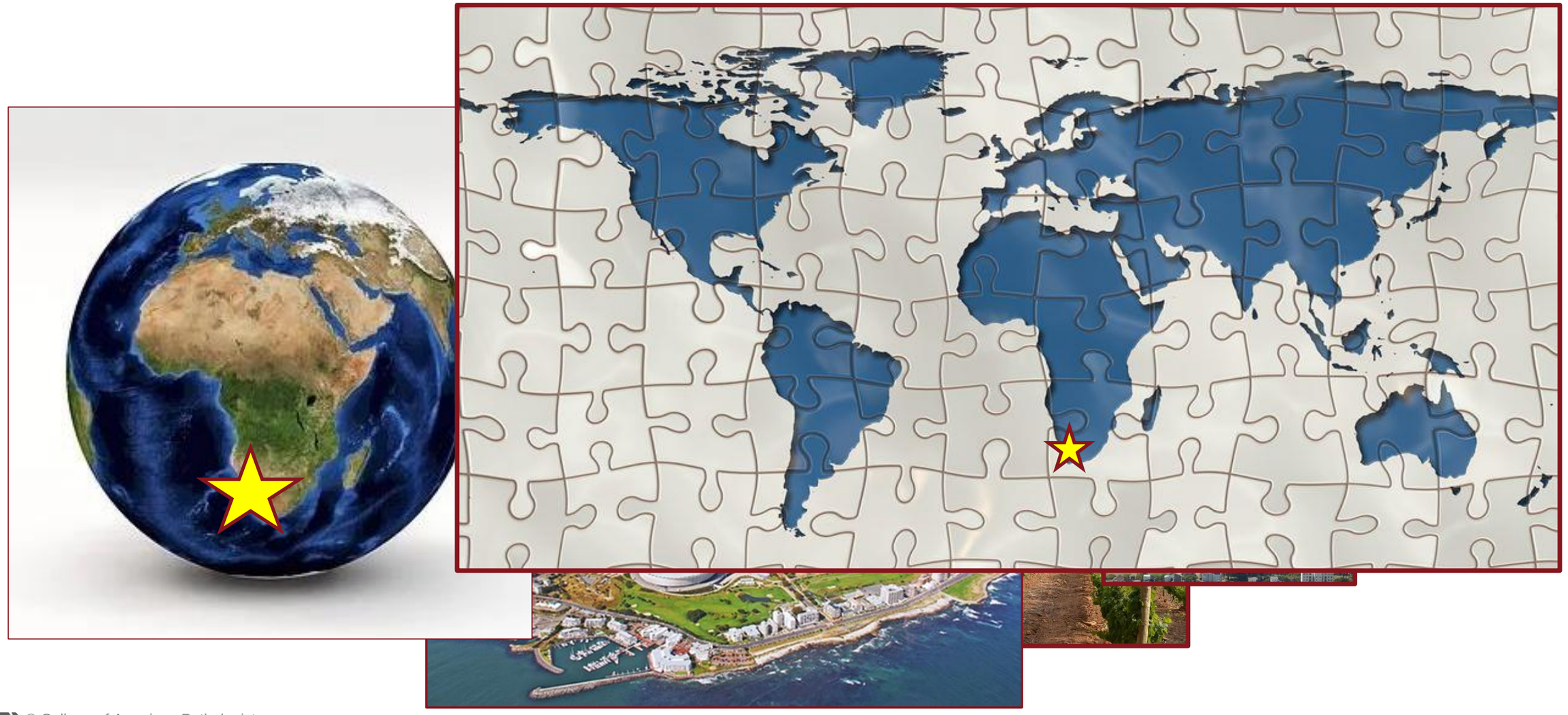




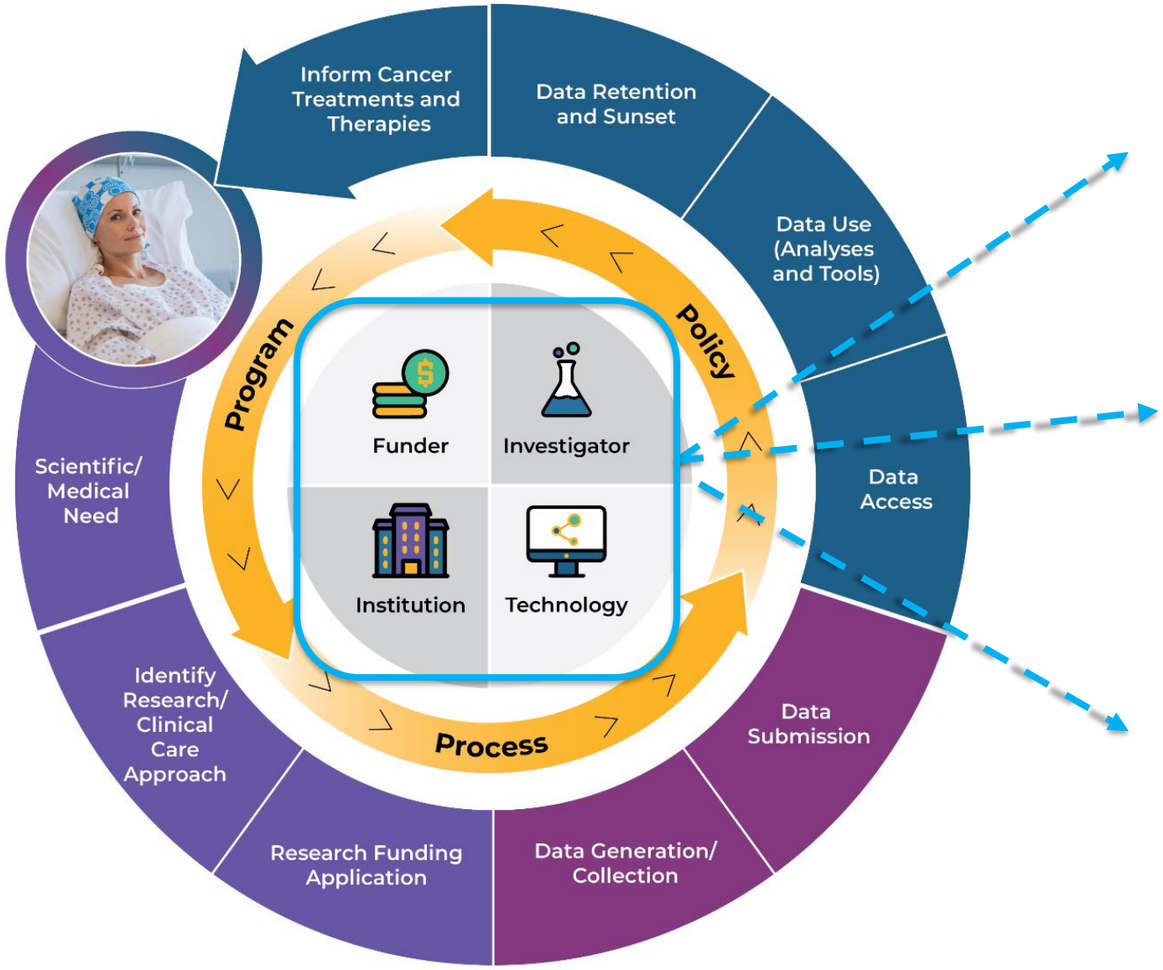
How Do We Get to the "Big Picture"?



Are We Really Seeing the Big Picture?



Scientific Data Lifecycle: Keys to Impactful Discovery



Critical Questions to Answer

Research that defines therapeutic needs and essential scientific gaps to be filled using structured datasets.

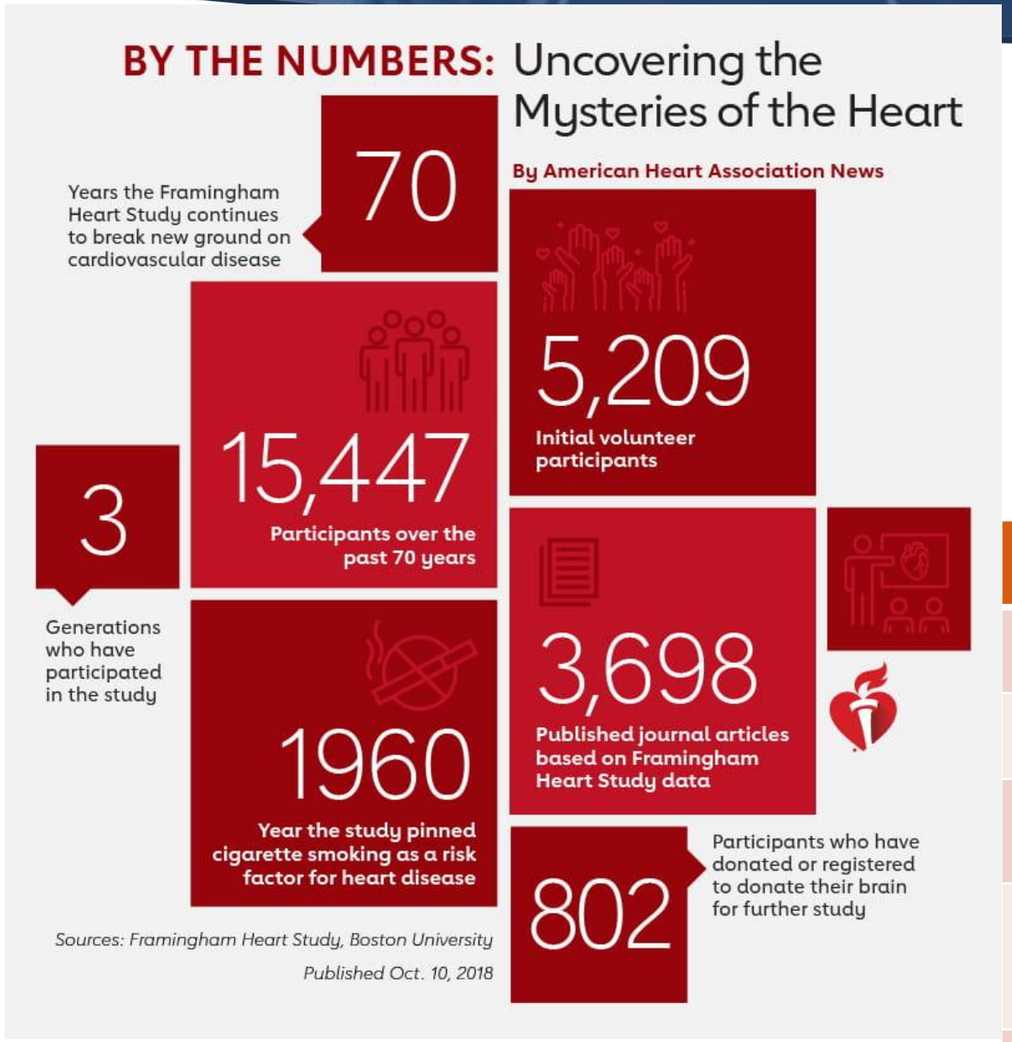
Policies to Promote Broad Use

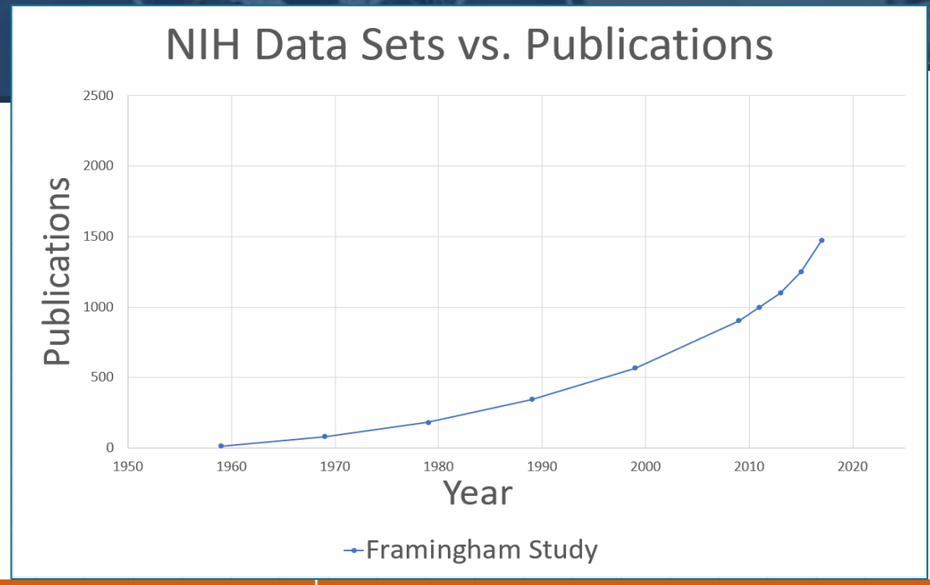
Implementation of aggressive data management, sharing and access policies that ensure rapid, free and broad access to all types of data.

Infrastructure to Support FAIR Principles

Technology platforms and tools that employ standards to make **data** findable, accessible, interoperable and reusable.

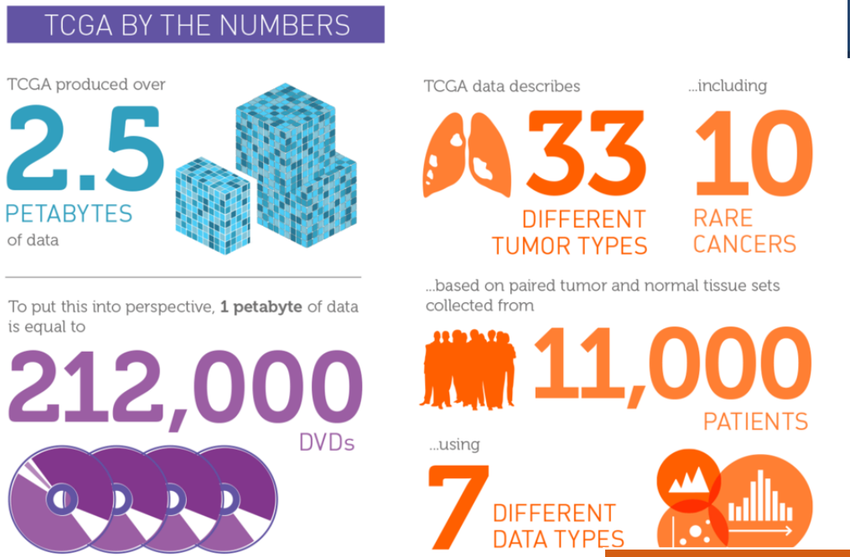
Framingham Heart Study: Success in Focused Data Collection

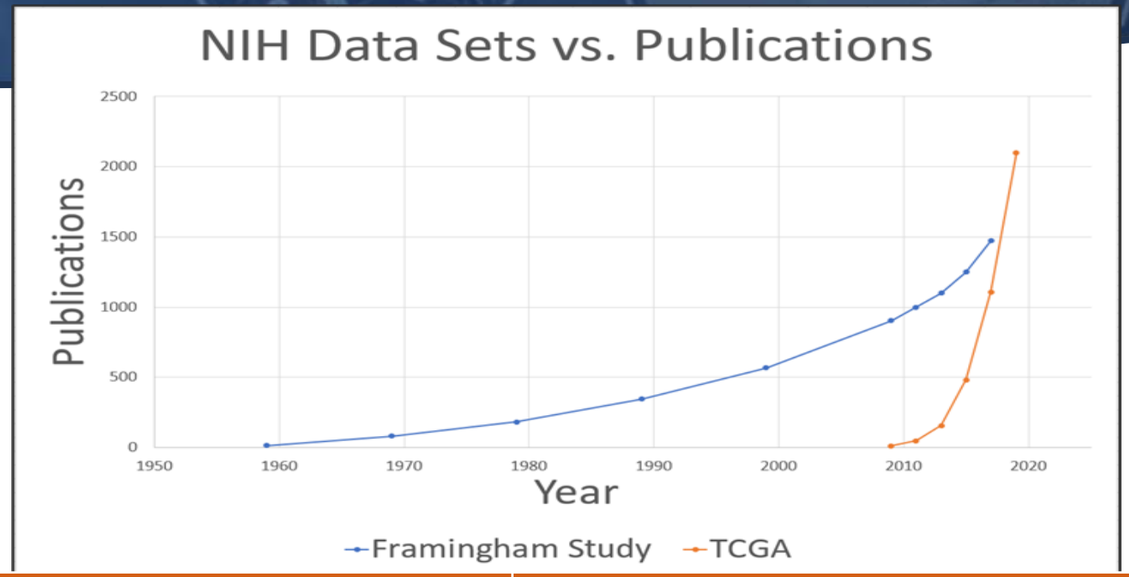




	Framingham Heart Study
Study Length	70 years
Cases Studied	15,144
Publications	**3,698**
Data Use	Consortia-based; most data available on publication
Approved Users	715 (Individual Level Data)

The Cancer Genome Atlas: Success in Open Team Science





	Framingham Heart Study	The Cancer Genome Atlas
Study Length	70 years	12 years
Cases Studied	15,144	11,429
Publications	**3,698**	**3,747**
Data Use	Consortia-based; most data available on publication	Collaborative Teams & Public Use of Data; All data immediately available
Approved Users	715 (Individual Level Data)	3,335 (Individual Level Data)

The Cancer Moonshot: Success in Mission-Driven Science

Cancer Moonsh℠:
Accelerate discovery, increase collaboration, and expand data sharing

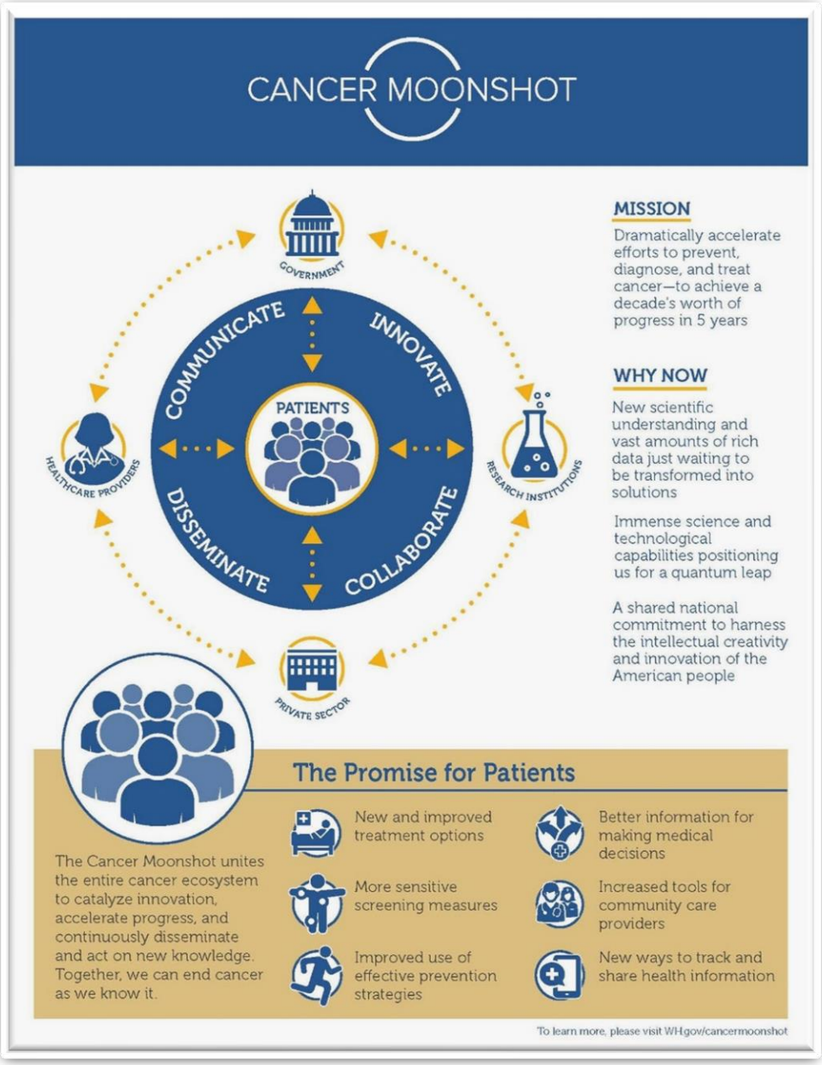
In the Cancer Moonshot's first 4 years (2017–2021):

>2,000	49	>30
Publications	Clinical Trials	Patent Filings

CANCER MOONSHOT

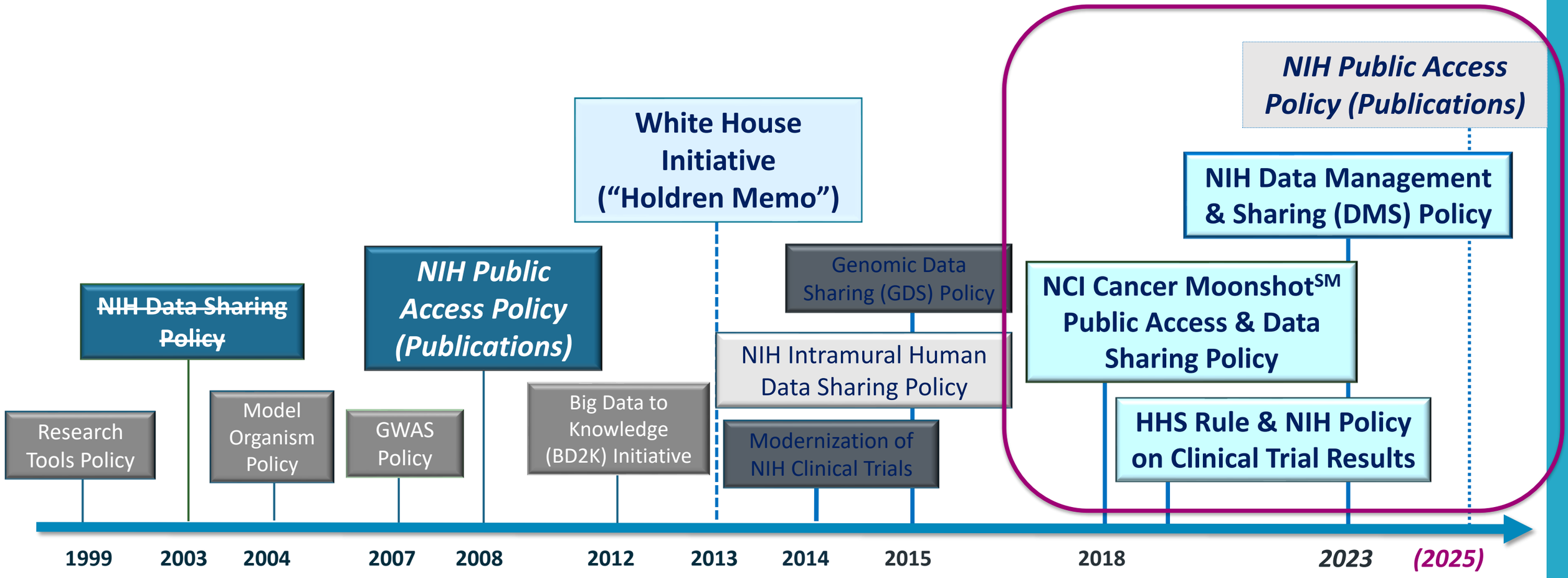
INITIATIVES 2017–2022

OVER 70	OVER 240
CONSORTIUMS OR PROGRAMS	RESEARCH PROJECTS



Take Home Message: purposeful, broad, early access to data leads to much faster and impactful outcomes

NIH Data Sharing & Public Access Policies



Investigators must share any information necessary to understand, develop or reproduce published research (raw data, statistical methods, tools, source code)

Key Messages of NIH Data Policies

NIH expects all funded research to have ***a plan to manage and share scientific data*** generated, and to make publications broadly available

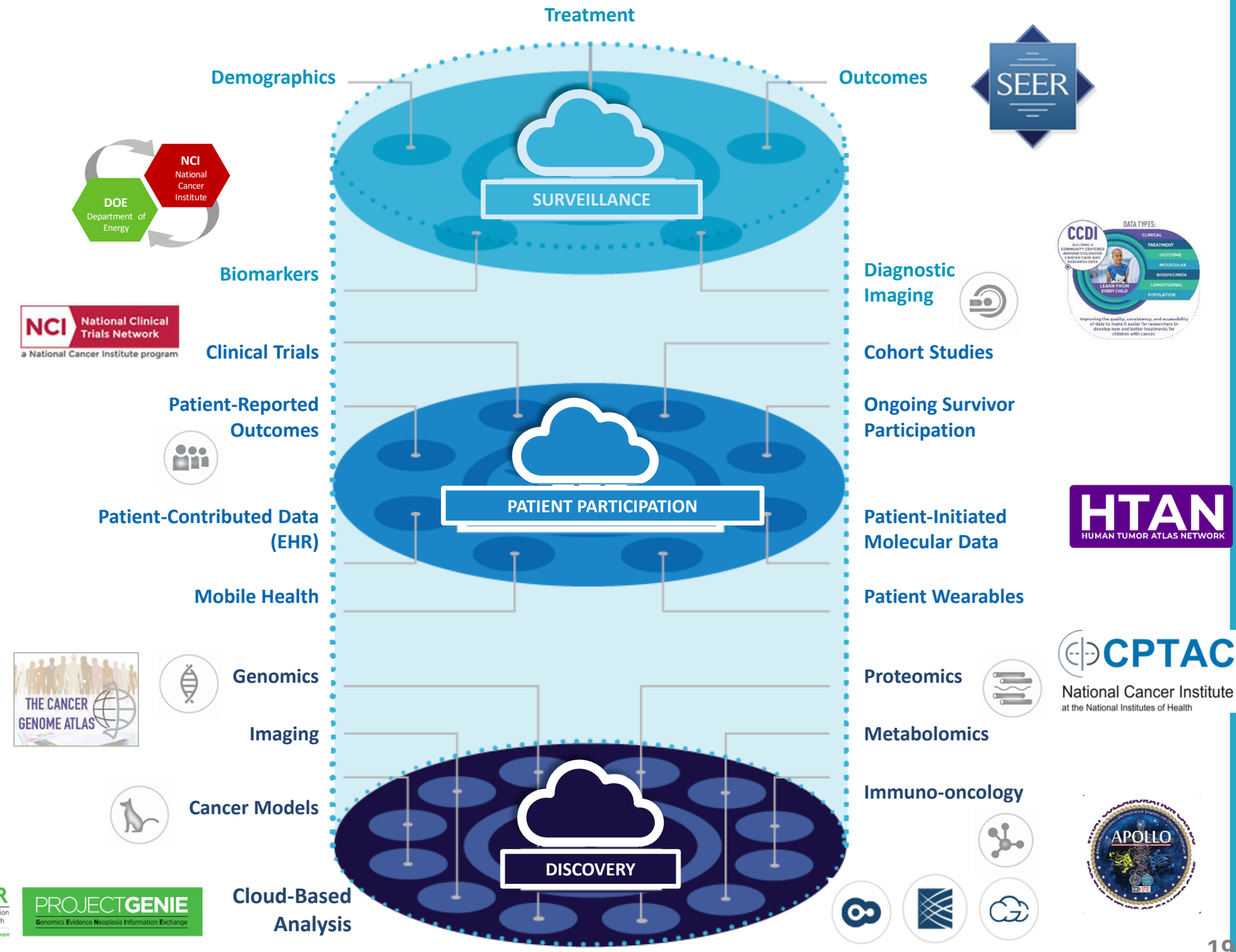
Promote ***open science***, stimulate new ***discovery***, enable ***rigor*** & ***reproducibility***, and provide ***transparency to maximize data utility*** for the ***public good***

Driving A Cultural Shift through planning for consistent, collaborative & impactful data management and sharing as a critical part of all research

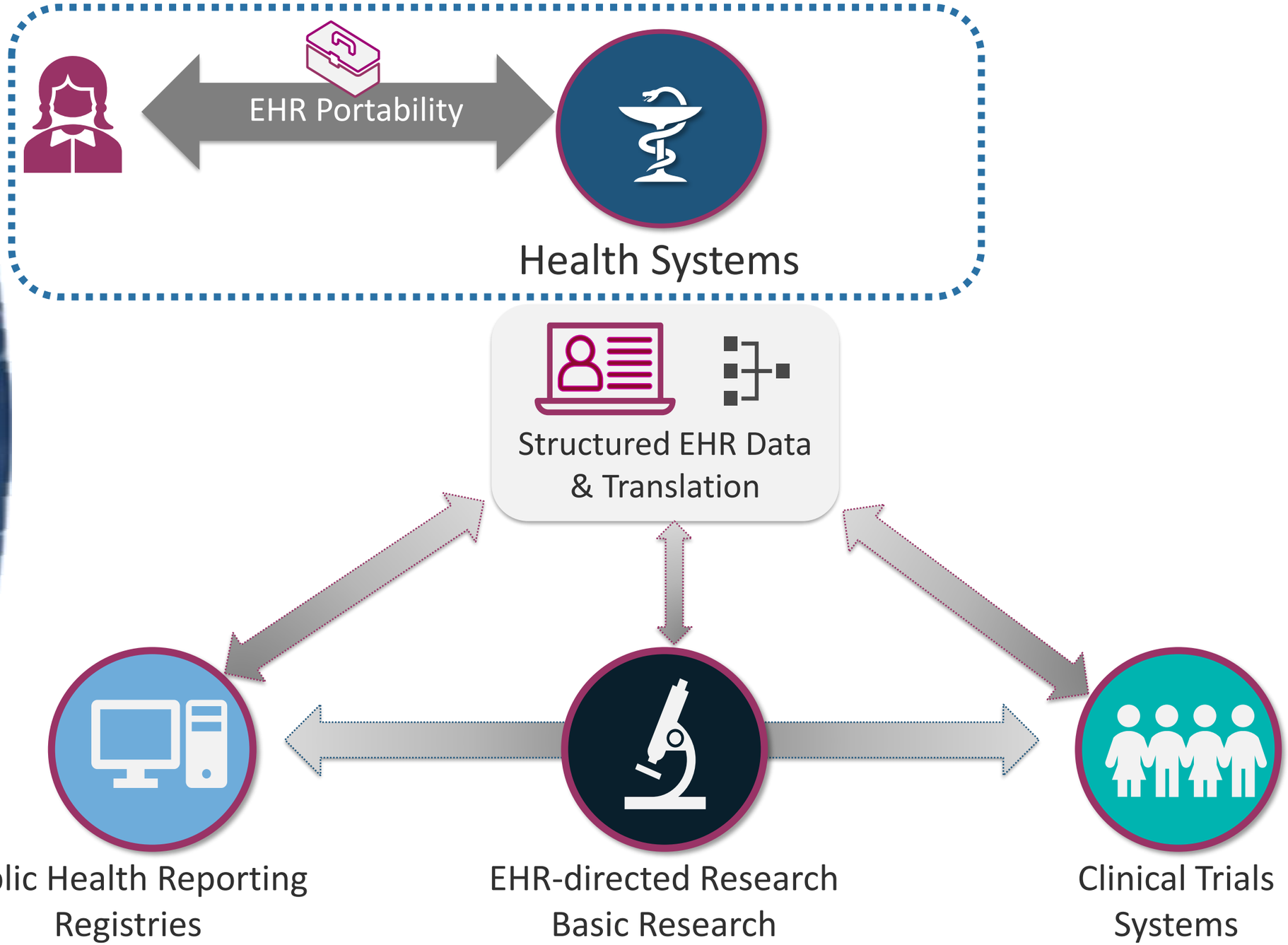
NIH is taking a "***learning approach***" (i.e., phased and iterative implementation in the years to come)

Over time, thoughtful DMS plans will inform ***clear guidance*** on the ***highest value data*** types beyond genomics (repository, timelines, etc.)

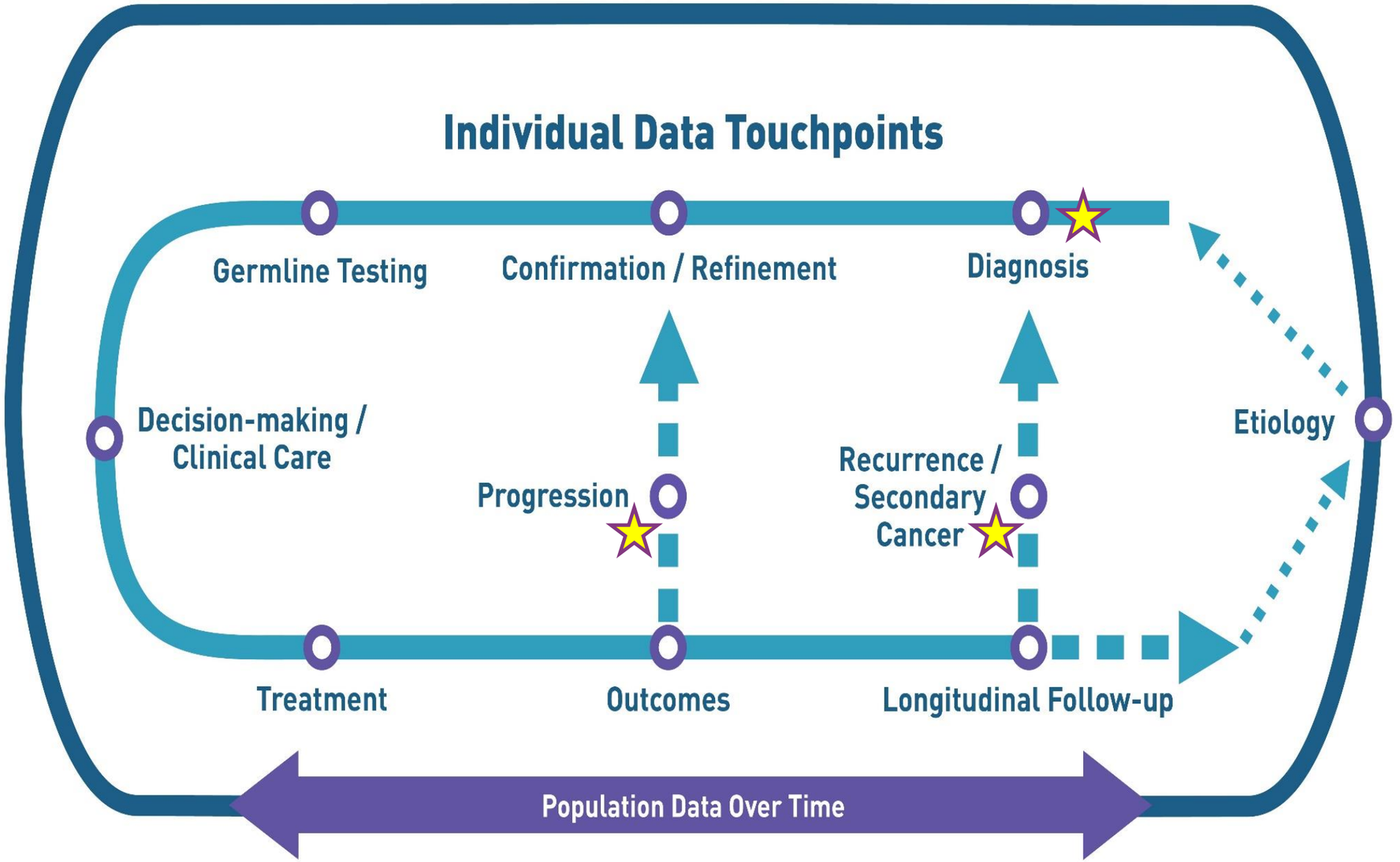
National Data Ecosystem: Integrating Cancer Research

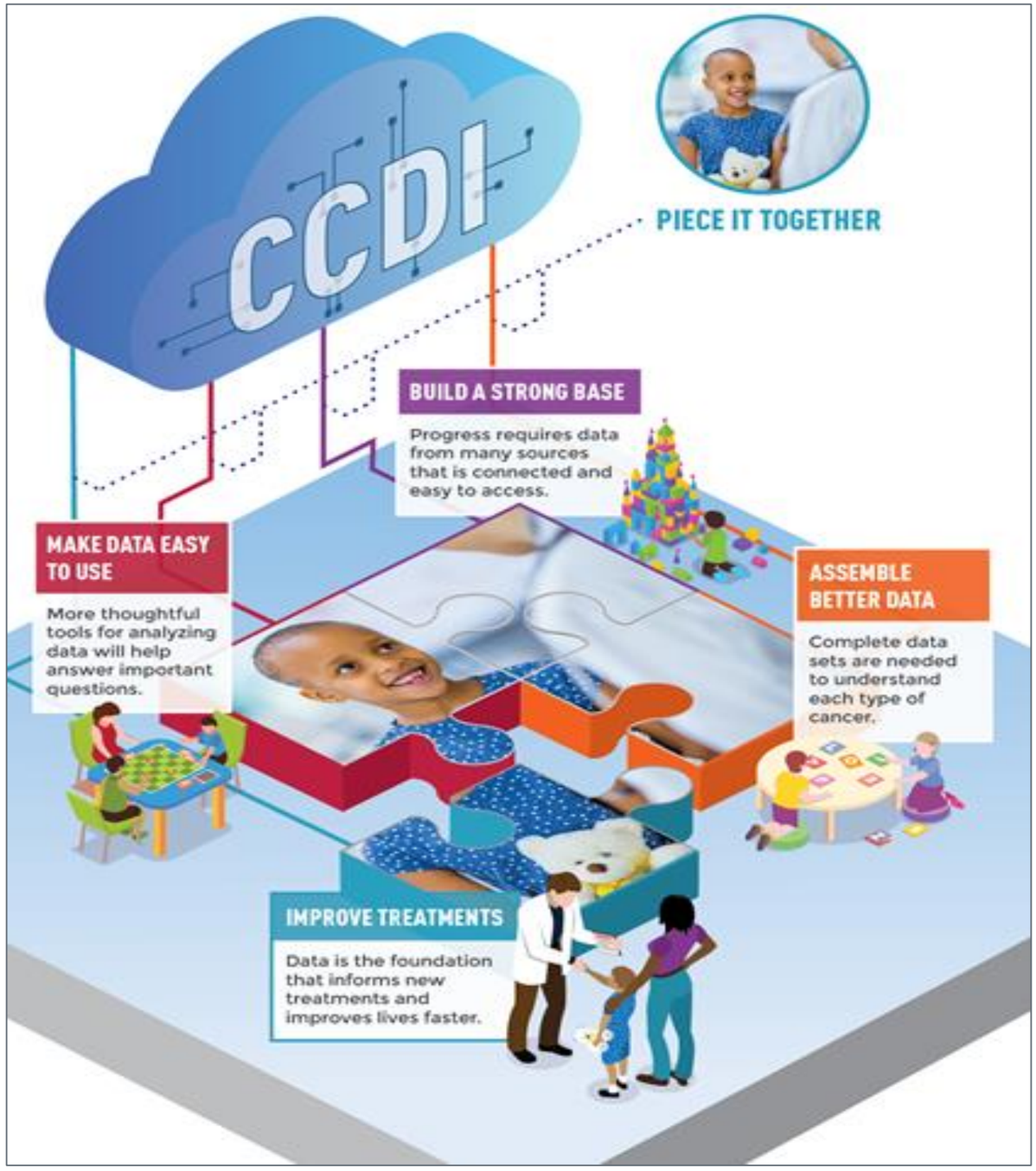


Clinical Care is the Source of Cancer Research Data



Data Touchpoints Along a Participant Journey

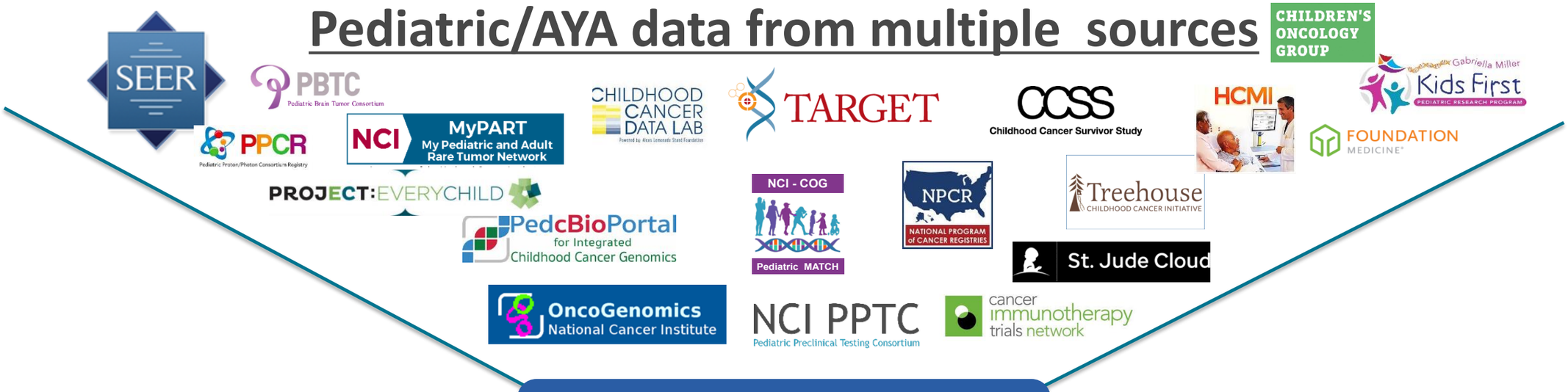




Launched in FY20, the Childhood Cancer Data Initiative (CCDI) is a 10-year, trans-NCI Program funded by Congress ($50M/yr)

CCDI supports the community of pediatric cancer researchers, advocates, families, hospitals, and networks committed to generating, using and sharing data to improve treatments, quality of life, and survivorship of every child with cancer

Pediatric/AYA data from multiple sources



CCDI

Improved understanding of why some cancers develop resistance or don't respond to treatment

Generation of new ideas for intervention

Development of new research and analytical tools

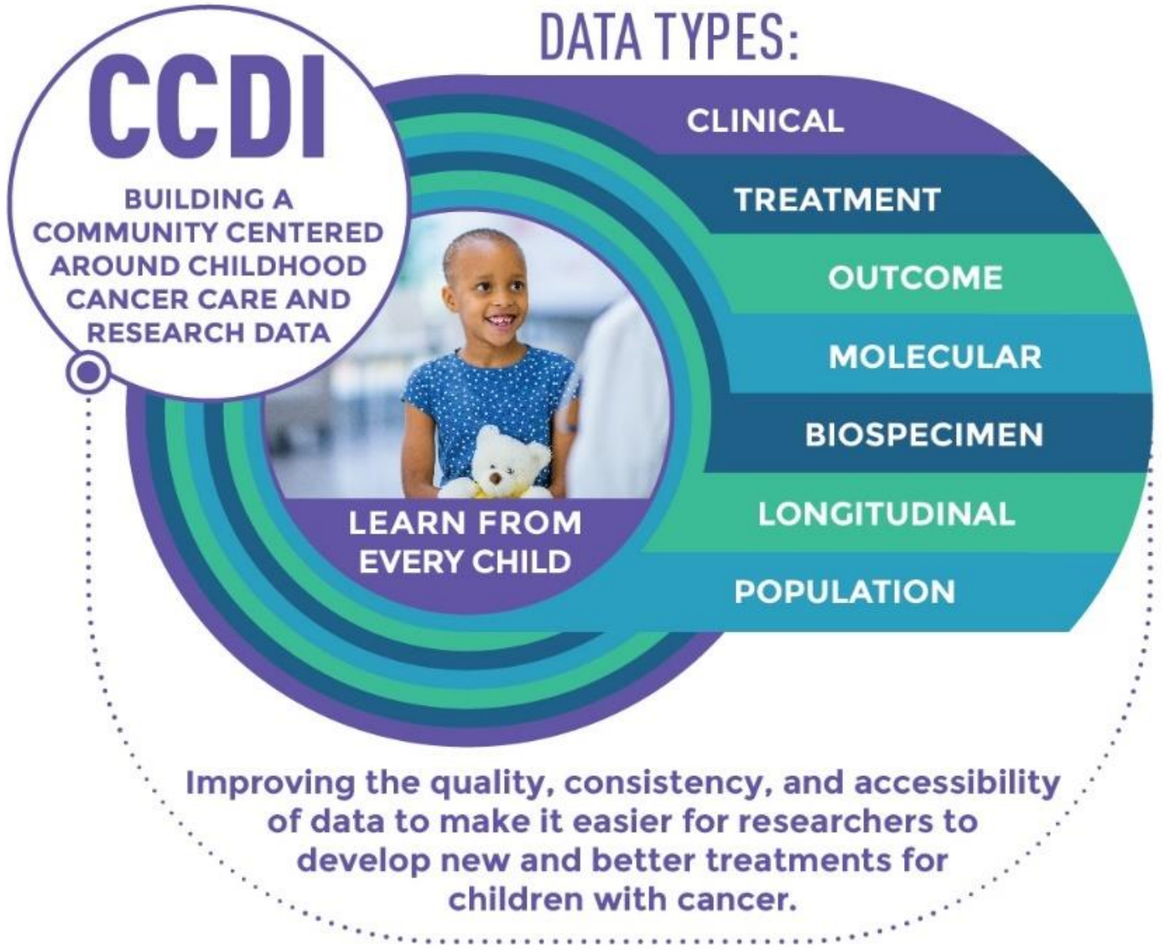
Identification of less toxic treatments and strategies for management

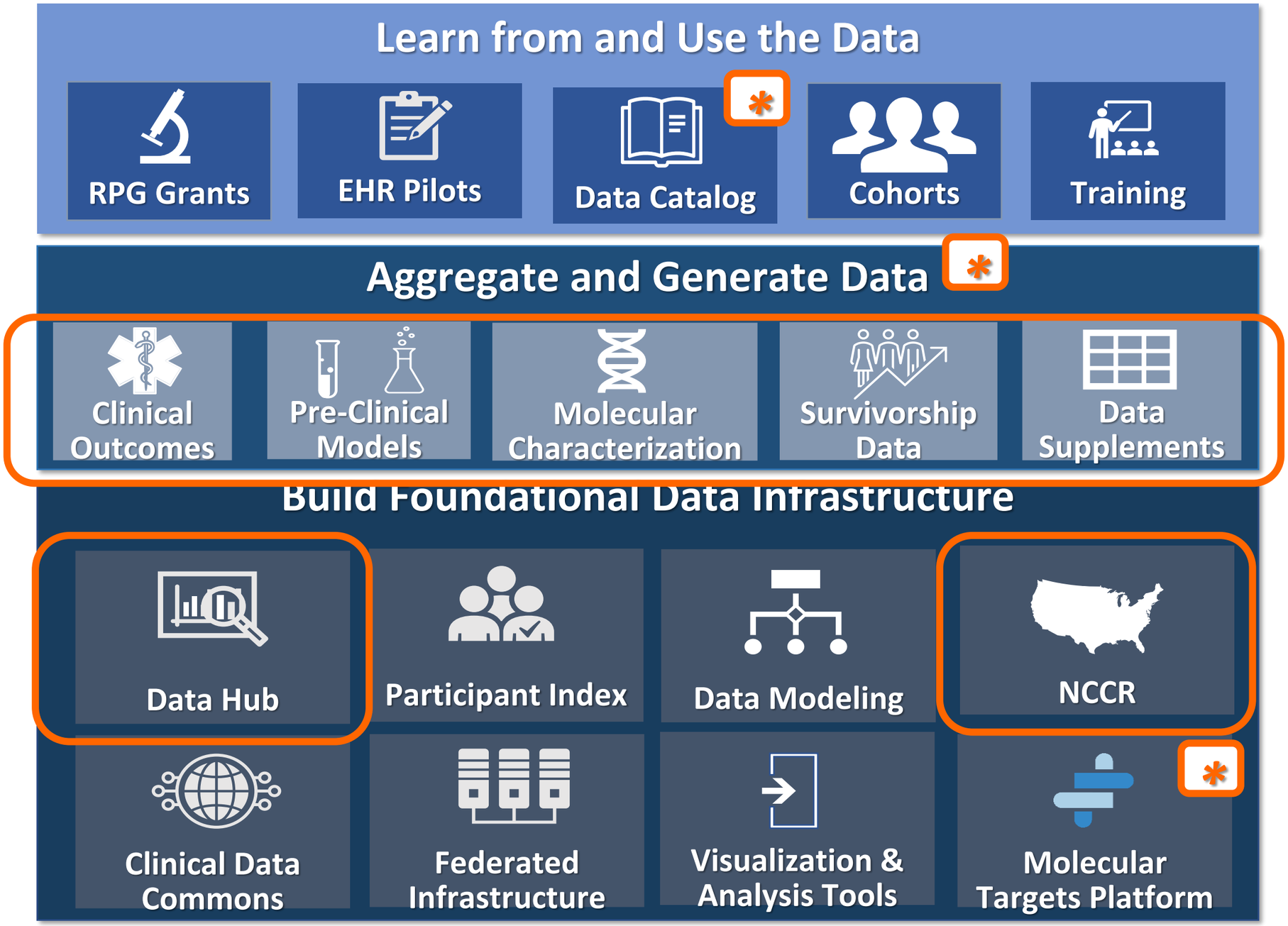
Culture change towards improved collaboration and data sharing

New therapies for childhood/AYA cancers

Foundational Goals for CCDI

- **Gather data** from every child, adolescent, and young adult diagnosed with a childhood cancer, regardless of where they receive their care
- Create a national strategy of appropriate **clinical and molecular characterization** to speed diagnosis and inform treatment for all types of childhood cancers
- **Develop a platform and tools** to bring together clinical care and research data that will improve preventive measures, treatment, quality of life, and survivorship for childhood cancers

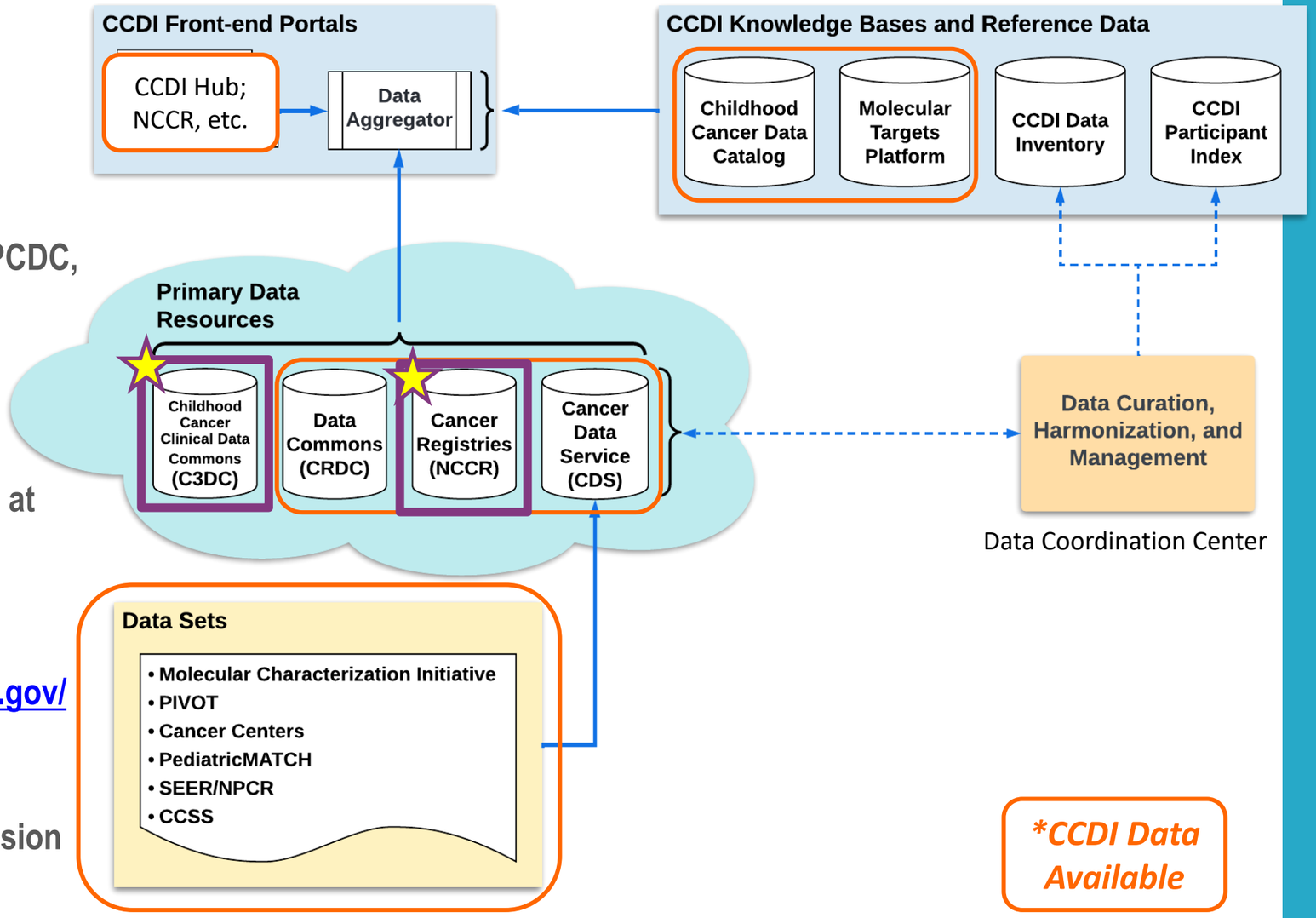




- CCDI is collecting and generating multiple types of data that are accessible through various components of CCDI Ecosystem Resources (Genomics, Imaging, Proteomics, Clinical Outcomes, Survivorship, Pre-clinical model screens, Surveillance)
- Find CCDI supported Data (Data Hub, Data Catalog)
- Access CCDI datasets, analytic and visualization tools through the National Childhood Cancer Registry, CCDI Data Hub, dbGaP/Cancer Research Data Commons, & NCTN Archive

CCDI Data Ecosystem Components: Connecting Data

- Primary databases (holding CCDI Data)
  - Cancer Research Data Commons – CDS, GDC, PDC, TCIA/IDC
  - National Childhood Cancer Registry (NCCR)
  - NCTN Archive/Clinical Trials Data Commons
  - CCDI Data Federation – dbGaP, TreeHouse, St. Jude, PCDC, KFDRC
- Knowledge Bases & Reference data
  - CCDI Data Catalog – https://datacatalog.ccdi.cancer.gov/resource/CCDI
  - Molecular Targets Platform – harmonized/aligned data at PedcBioPortal
- Data access
  - CCDI Data Hub – https://ccdi.cancer.gov
  - NCCR PedsExplorer - https://nccrexplorer.ccdi.cancer.gov/
  - dbGaP/CRDC portals
- Data processing & harmonization
  - Data Coordination Center – to assist with data submission and harmonization



National Childhood Cancer Registry

Approximately 16,000 childhood cancer patients are diagnosed in the United States annually, compared with 1.8 million new cancer cases among all ages

All childhood cases from registries (<40 at diagnosis; 1995+)

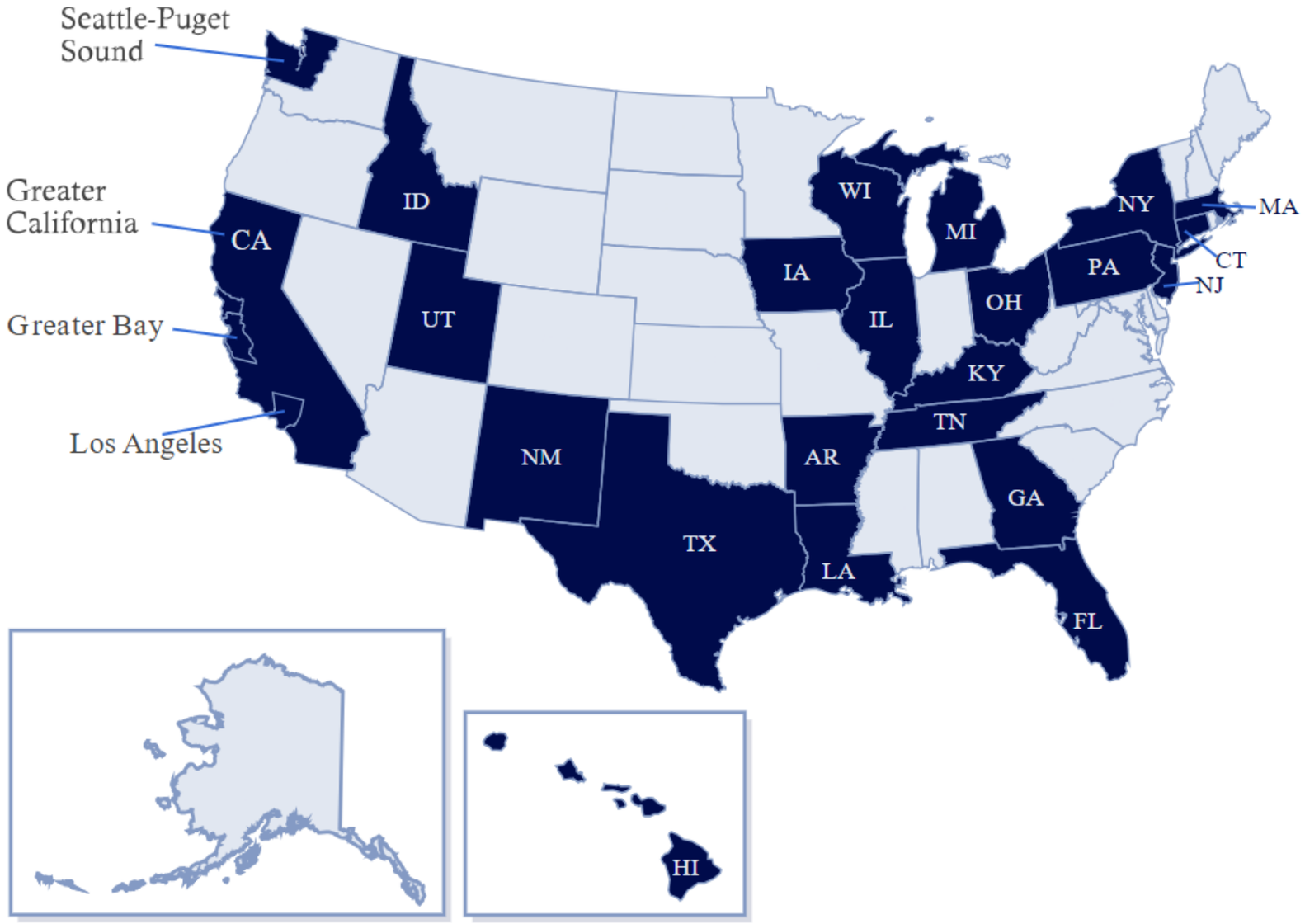
Consolidate and standardize data in a single infrastructure

Analyze and share data, in a controlled access environment, to gain insight

Data Domains

- **Longitudinal Treatment, Procedures, Outcomes (pharmacy data, radiation oncology, claims, radiology, vital status)**
- **Clinical Trials and Survivorship Studies**
- **Social Determinants of Health (including financial toxicity, residential history)**
- **Germline Molecular Characterization**

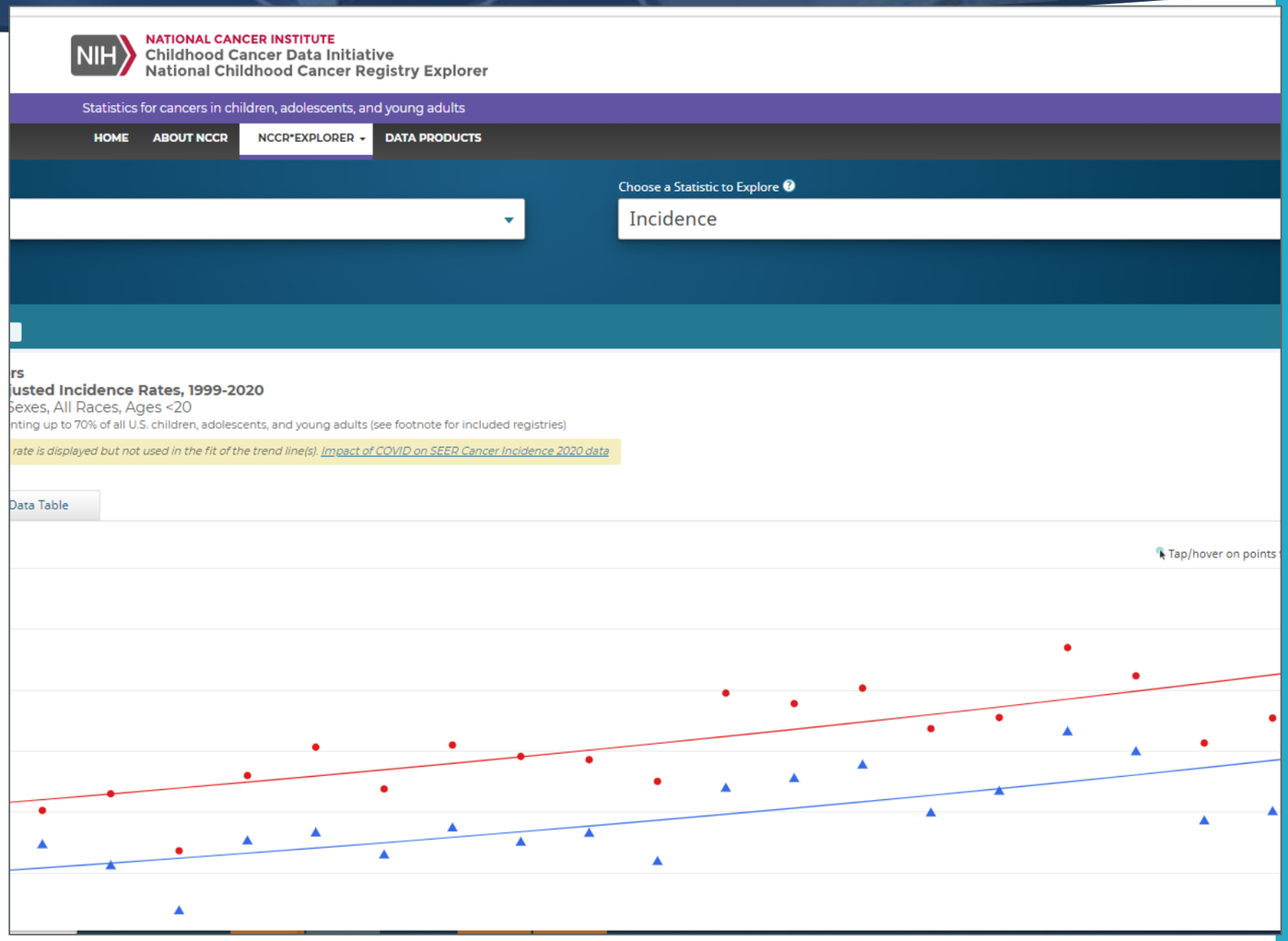
National Childhood Cancer Registry



- 25 Central Cancer Registries participate in the NCCR
- Represents 70% of the US population
- 1,700,440 reported cases under age 40 (1995-2020)
- 9 VPR registries helped identify additional 6,230 cases

Access Registry Data: NCCR*Explorer

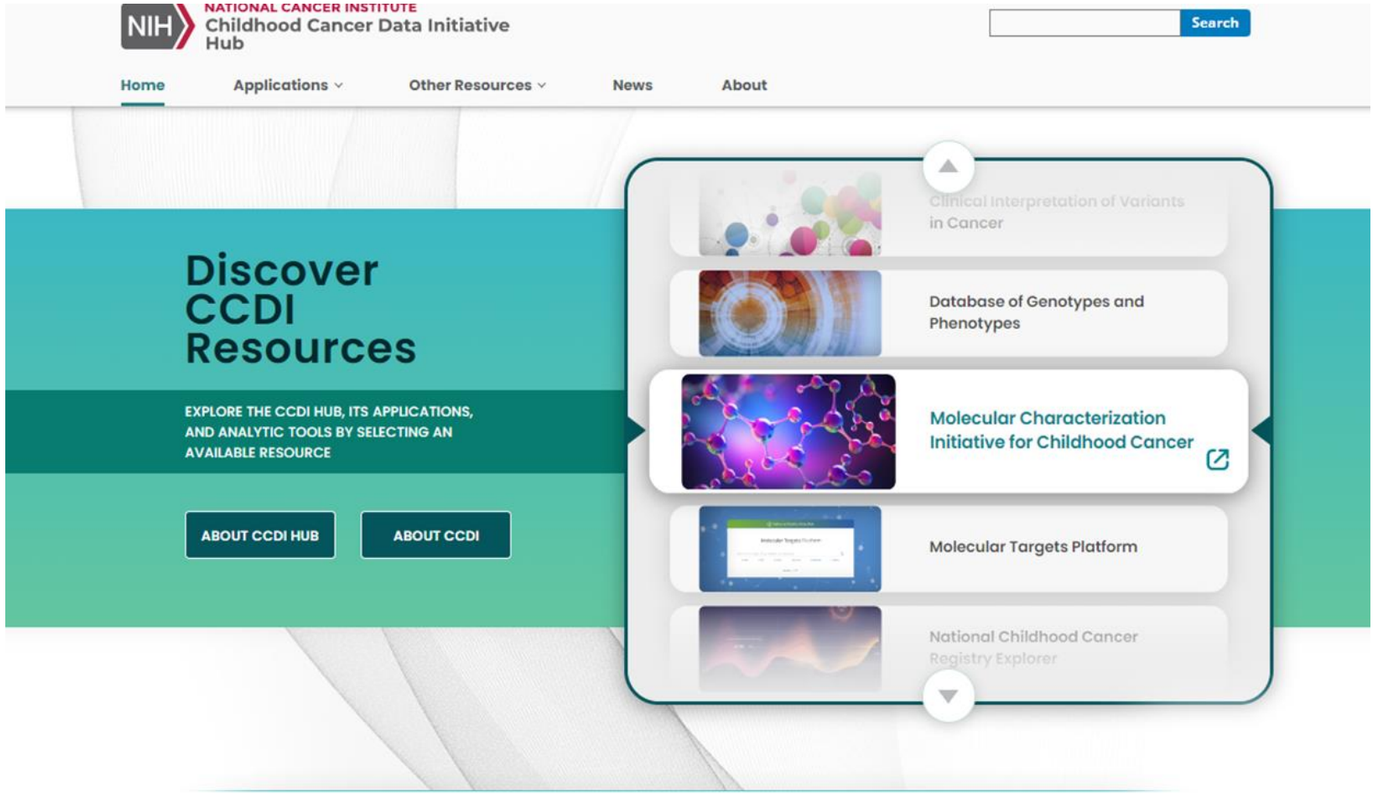
- Average of 8,000 unique visitors annually. Over the past three months, 1,200 unique visitors created 13,000 graphs
- Pre-calculated statistics in dynamic tables and plots based on user criteria for patients diagnosed under age 40
- Site-specific age groups based on clinical significance
- Histology-based groupings
- No geographic identifiers to minimize risk of reidentification of small numbers

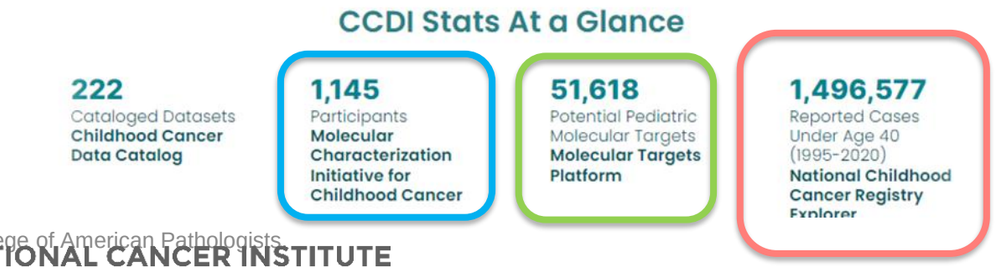


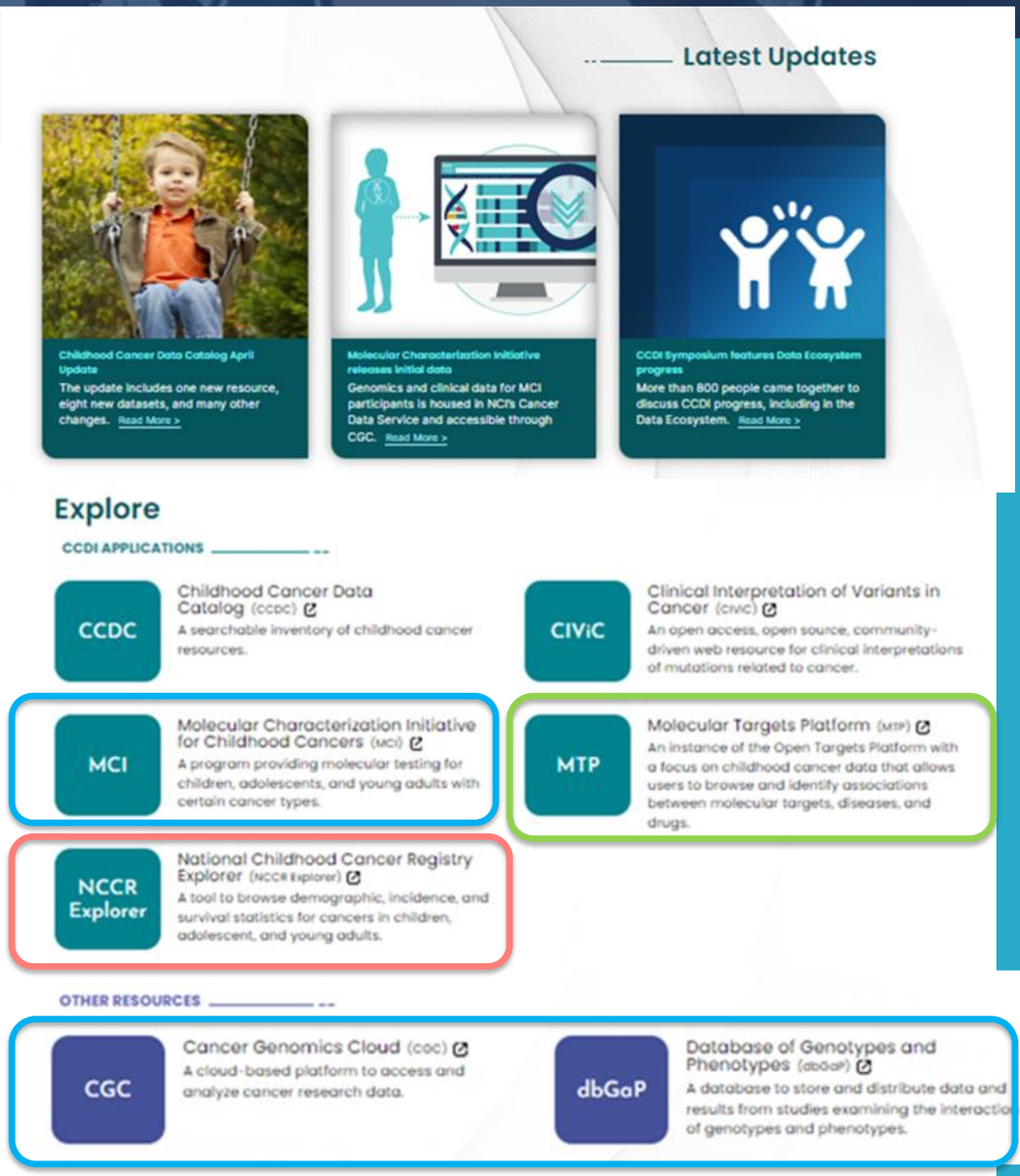
https://nccrexplorer.ccdi.cancer.gov/about/nccr.html

Access to CCDI Data & Tools

- CCDI Data Hub provides links to data, knowledge bases and tools (*https://ccdi.cancer.gov*)







learn from every child.

cancer.gov/CCDI

#data4childhoodcancer

Contact Us About Data Sharing

nciofficeofdatasharing@mail.nih.gov

#NCIODS

datasharing.cancer.gov

NATIONAL CANCER INSTITUTE

www.cancer.gov www.cancer.gov/espanol

The Future of Cancer Data:
Unlocking Insights With Pathology Reporting Summit

October 6, 2023