# CCDI, Using Data, and AI

*Warren A. Kibbe, Ph.D., FACMI*
*NCI Deputy Director for Data Science and Strategy*
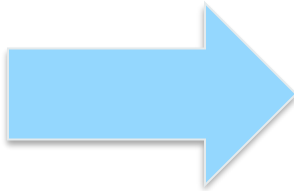
*Meeting of the National Council of Research Advocates*

NIH NATIONAL CANCER INSTITUTE

September 18, 2024

# Agenda

1. Introduction
2. Update on CCDI
3. Using Data
4. AI and Cancer Research
5. Discussion

# About My Role

- NCI's first Deputy Director for Data Science and Strategy

- Started June 30, 2024

- What does this role entail?

- Advise NCI director and other senior leaders on the **utilization, stewardship, and sharing of data**

- Provide **strategic direction to CBIIT** and oversee all aspects of data science for NCI

- Lead NCI's **implementation of the NIH Strategic Plan for Data Science**

- Provide strategic counsel for **key NCI data science initiatives**

- Serve as **senior data science liaison** on NIH and other government committees

# My (Recent) Background

2013 - 2017
Director, Center for Biomedical Informatics and Information Technology (CBIIT), **NCI**

2016 - 2017
Acting Deputy Director, **NCI**
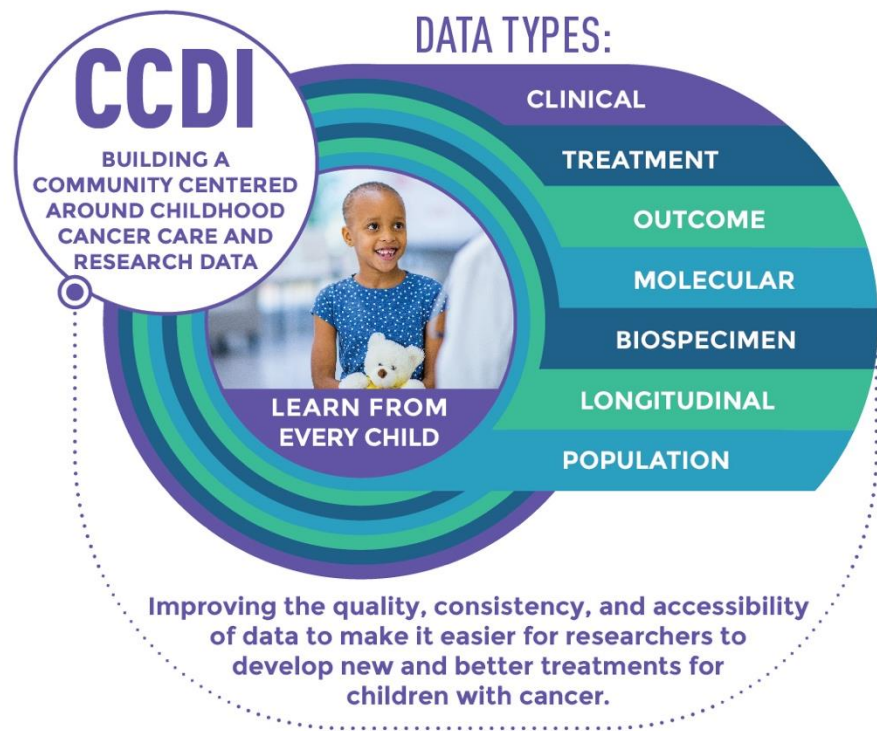
2017 - 2024
**Duke University School of Medicine**
- Chief, Translational Biomedical Informatics
- Vice Chair, Department of Biostatistics and Bioinformatics,

**Duke Cancer Institute**
- Chief Data Officer

2024 - Present
Deputy Director, Data Science and Strategy, **NCI**

# About CCDI

NIH NATIONAL CANCER INSTITUTE

# CCDI's Key Goals

- Gather data from every child, adolescent, and young adult with childhood cancer

- Create a national strategy of molecular characterization to inform diagnosis and treatment

- Develop a platform and tools for clinical and research data to improve prevention, treatment, quality of life, and survivorship

- Engaging the entire childhood cancer care and research community



**CCDI**
BUILDING A COMMUNITY CENTERED AROUND CHILDHOOD CANCER CARE AND RESEARCH DATA

LEARN FROM EVERY CHILD

DATA TYPES:
- CLINICAL
- TREATMENT
- OUTCOME
- MOLECULAR
- BIOSPECIMEN
- LONGITUDINAL
- POPULATION

Improving the quality, consistency, and accessibility of data to make it easier for researchers to develop new and better treatments for children with cancer.

*Flores-Toro JA et al., J Clin Oncol, 2023*
*Jagu S et al., Pediatr Blood Cancer, 2024*

# CCDI Hub

## CCDI Stats At a Glance

**317**
Cataloged Datasets
**Childhood Cancer
Data Catalog >**

**2,990***
Participants with
Available Genomic
and Clinical Data
**Molecular
Characterization
Initiative >**

**58,867**
Potential Pediatric
Molecular Targets
**Molecular Targets
Platform >**

**1,700,440**
Reported Cases
Under Age 40
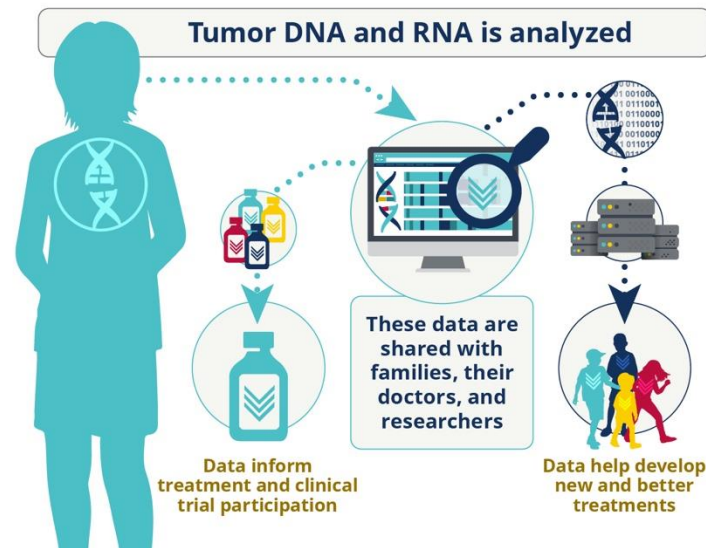(1995–2020)
**National Childhood
Cancer Registry
Explorer >**

* Counts for MCI participants in CCDI Hub and total MCI participants consented may differ.

- Entry point for researchers, data scientists, and citizen scientists

- Information and direct links to CCDI platforms, tools and resources

- New discoveries that impact patient lives

# CCDI Molecular Characterization Initiative (MCI)

- National strategy for appropriate clinical and molecular characterization for every child with cancer:

  - Minimum set of molecular diagnostics

  - Standardized molecular profiling on up to 3,000 children annually

  - Enabling discoveries as clinical data are connected with other datasets

  - Building a clinically annotated biobank for future research from remaining tissues



Tumor DNA and RNA is analyzed

These data are shared with families, their doctors, and researchers

Data inform treatment and clinical trial participation

Data help develop new and better treatments

cancer.gov/CCDI-molecular

# Pediatric Cancers

| Type | Introduced | Number |
|------|-----------|--------|
| Newly diagnosed CNS tumors | March 2022 | 3095 |
| Soft tissue sarcoma | May 2022 | 930 |
| Rare tumors | September 2022 | 416 |
| Neuroblastoma high risk | February 2024 | 174 |
| Ewing sarcoma | 2025 | - |
| Relapsed tumors | Planned | - |

**Adding research molecular characterization (2025)**

# Childhood Cancer Clinical Data Commons (C3DC)

- Enables researchers to search harmonized participant-level data across studies using a standardized set of pediatric cancer CDEs.

- Expanded from 2 to 18 harmonized datasets

- Updated UI with additional pages
  - Study details page now supports open-source data downloads

- Enhanced C3DC data model to include treatment and treatment response clinical nodes

- Improved user-friendly documentation

- V3 released on July 31
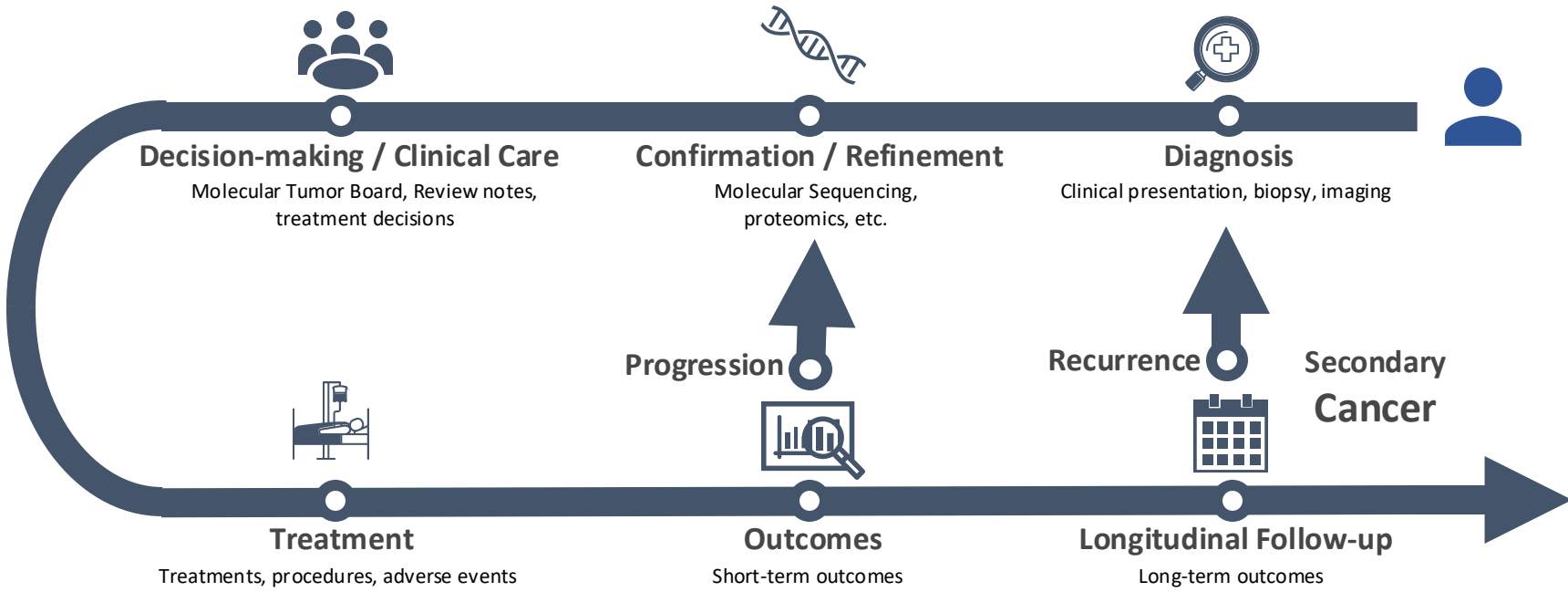
- Access harmonized childhood cancer clinical data https://clinicalcommons.ccdi.cancer.gov/home
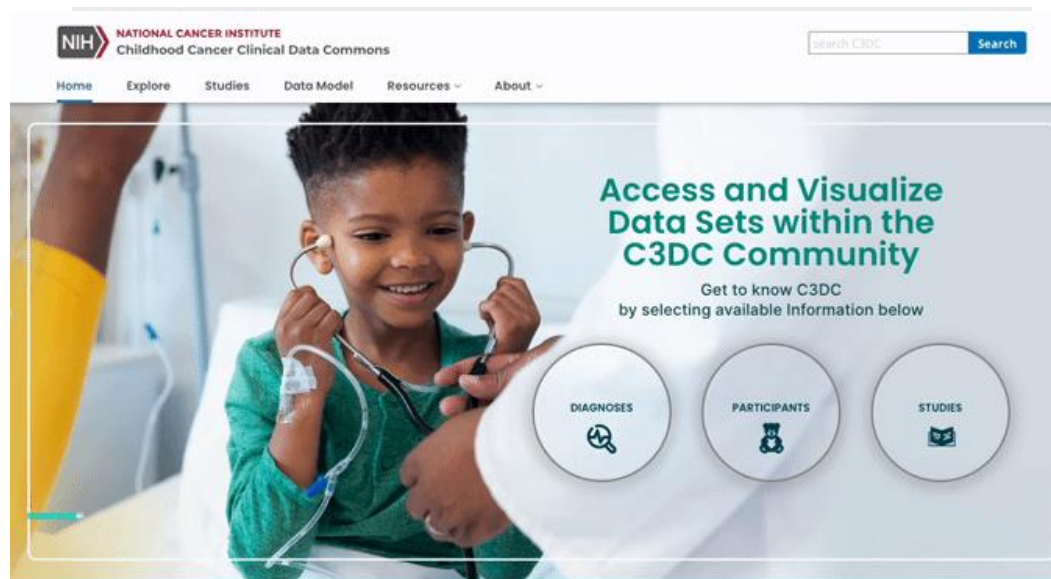
Use of Data in Cancer Research

# Patient Data Touchpoints

⊙ **Data Touchpoint**

Screening, Diagnosis, Treatment – Molecular Characterization Initiative

**Decision-making / Clinical Care**
Molecular Tumor Board, Review notes, treatment decisions

**Confirmation / Refinement**
Molecular Sequencing, proteomics, etc.

**Diagnosis**
Clinical presentation, biopsy, imaging

**Progression**

**Recurrence**

**Secondary Cancer**

**Treatment**
Treatments, procedures, adverse events

**Outcomes**
Short-term outcomes

**Longitudinal Follow-up**
Long-term outcomes

Epidemiology / Population Sciences Data –
Familial data, environmental, registry, population studies, disease cohorts

# Current sources of data

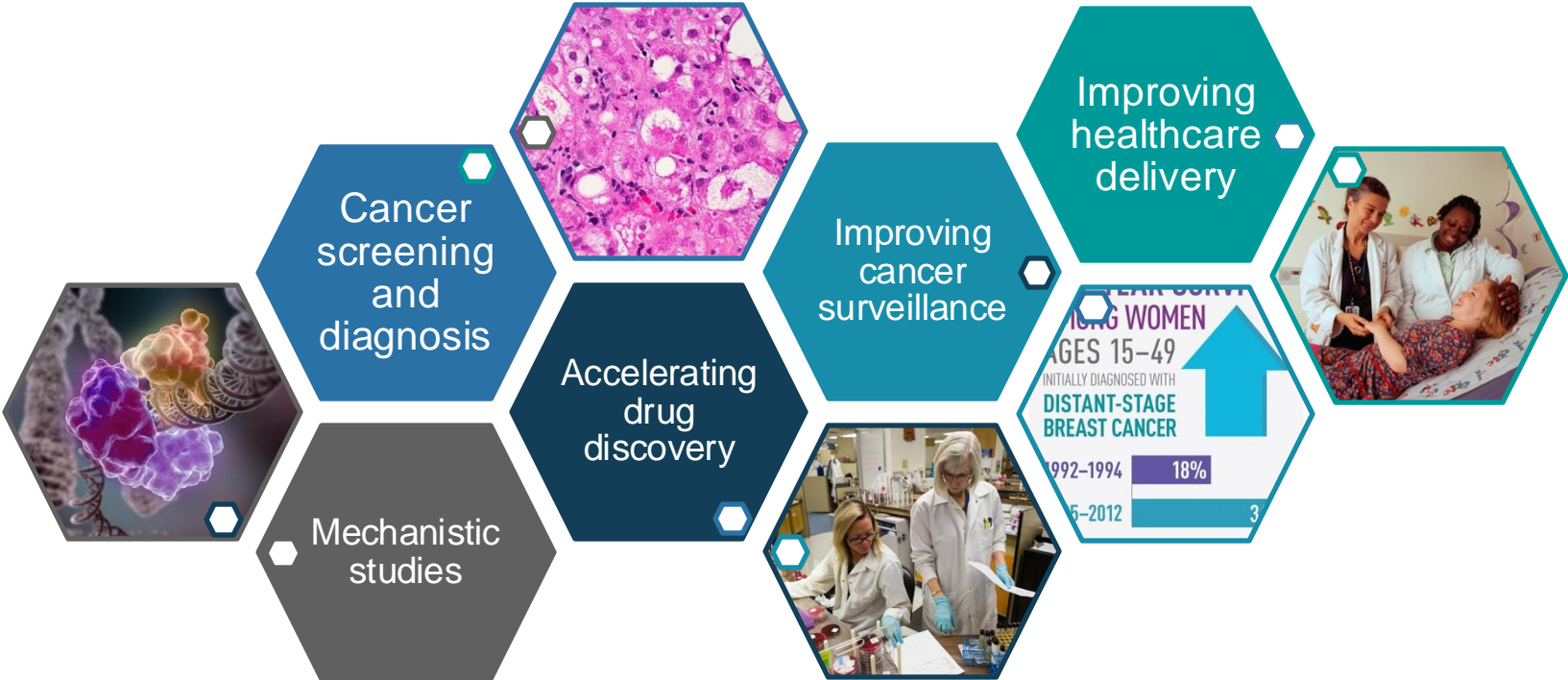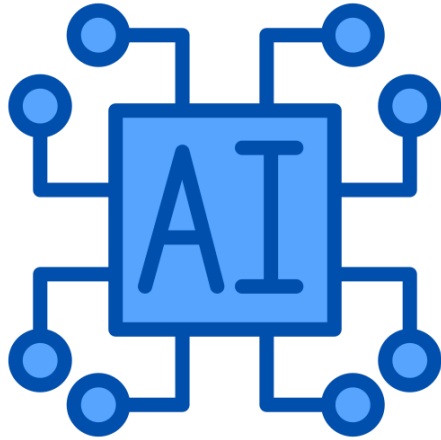molecular     genome     pathology     imaging     labs     notes     sensors

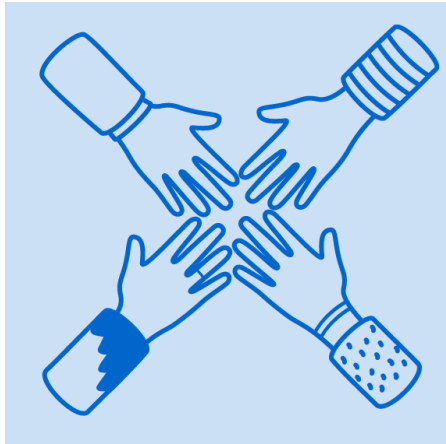# Our ability to generate biomedical data continues to grow in terms of variety and volume

# Data across the cancer research continuum



Cancer screening and diagnosis

Mechanistic studies

Accelerating drug discovery

Improving cancer surveillance

Improving healthcare delivery

# Data Science Training



AI
READY

DATA SCIENTISTS
& CANCER RESEARCHERS

DIVERSE
WORKFORCE

icons by Assurant.com & Iconfinder.com

# AI and Cancer Research

NIH NATIONAL CANCER INSTITUTE

**Recent advances have led to promising new applications of AI to cancer research**
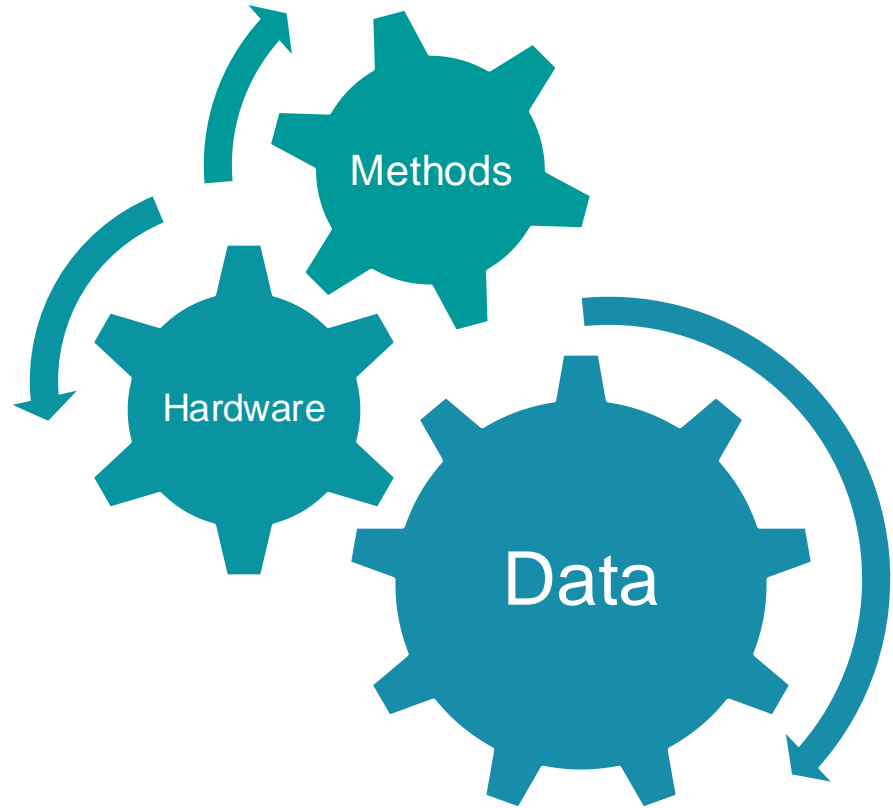
# Rise of Artificial Intelligence



*Image generated by AI (DALL*E3)*

## Key AI milestones throughout history

| 1800s | 1960s | 2018 | 2022 |
|---|---|---|---|

Pre-digital computers; conceptual / **math models** of how people think

Beginnings of **modern AI**; more math models (not yet computing power)

Researchers at Google introduce **BERT models** (Bidirectional Encoder Representations from Transformers)

**Launch of ChatGPT** (and other **large language models**); huge opportunity to embed AI in cancer research

# Principles for thinking about AI
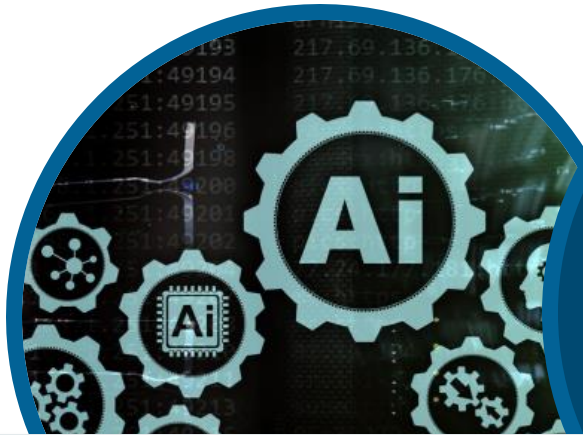
AI
READY

TRUSTWORTHY
& ETHICAL

DIVERSE
WORKFORCE