Cancer Research Data Commons Fall 2024

Impact of Data Science on Cancer Research

Warren A. Kibbe, Ph.D., FACMI
Deputy Director, Data Science & Strategy

NIH NATIONAL CANCER INSTITUTE

October 17, 2024

Agenda

1. *Maximizing the Impact of Shared Data*
2. *Data in Cancer Research*
3. *AI and Cancer Research*

Maximizing the Impact of Shared Data

Some basic principles

Simple

Empowering

Usable

These need to permeate everything we do and build!

FAIR Principles

Maximizing the Impact of Shared Data

Data Visibility

Making data **Findable**

Maximizing the Impact of Shared Data

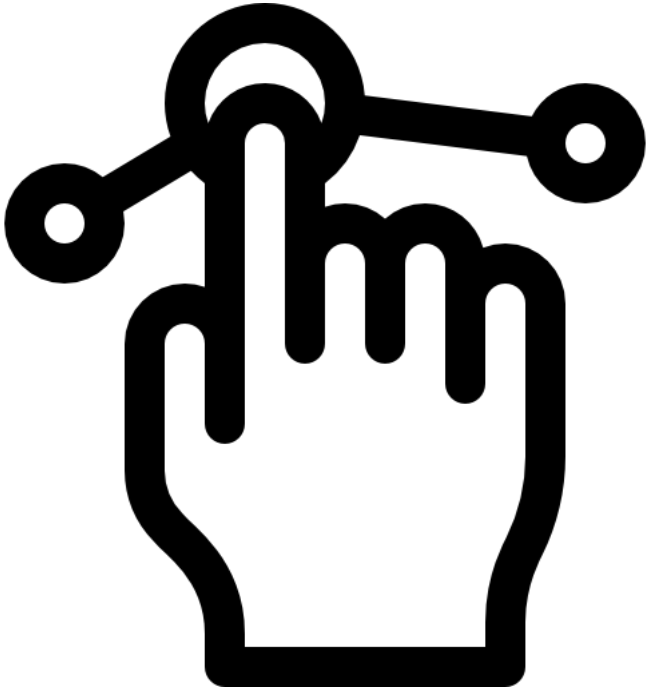
Access through open or clear data use agreements

Making data **Accessible**

Maximizing the Impact of Shared Data

Usability of the digital infrastructure

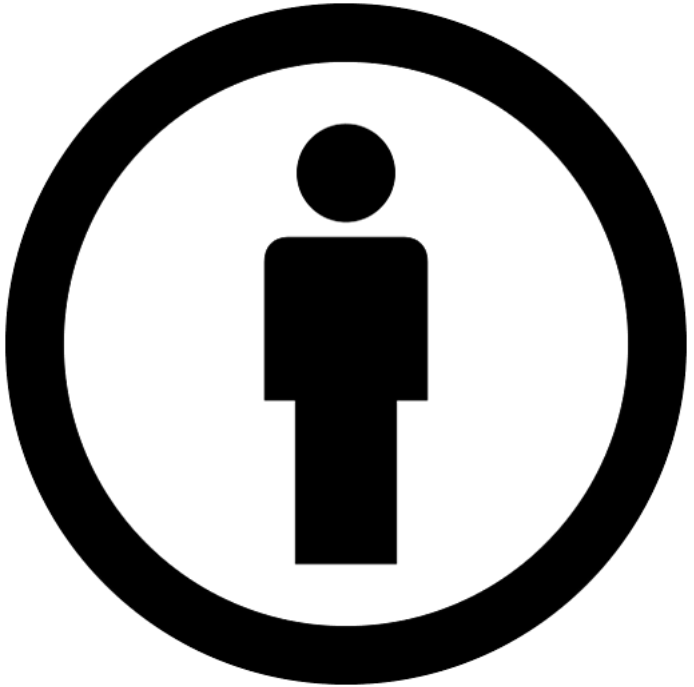
Make the environment **Usable**



Not a FAIR principle!

Maximizing the Impact of Shared Data

- Attribution - give credit to data generators, infrastructure builders, tool builders, analysts



Not a FAIR principle!

Maximizing the Impact of Shared Data

Interoperable – clear standards for connecting multiple resources, tools, teams, research

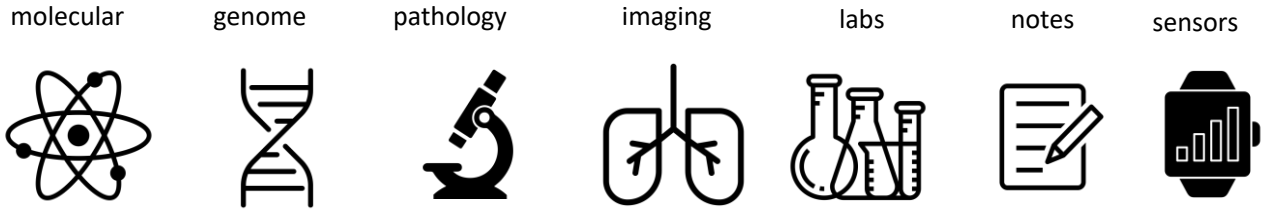
Maximizing the Impact of Shared Data

Reusable

Make the data, tools, workflows easy to manage, share, datasets and analyses easy to reproduce

Use of Data in Cancer Research

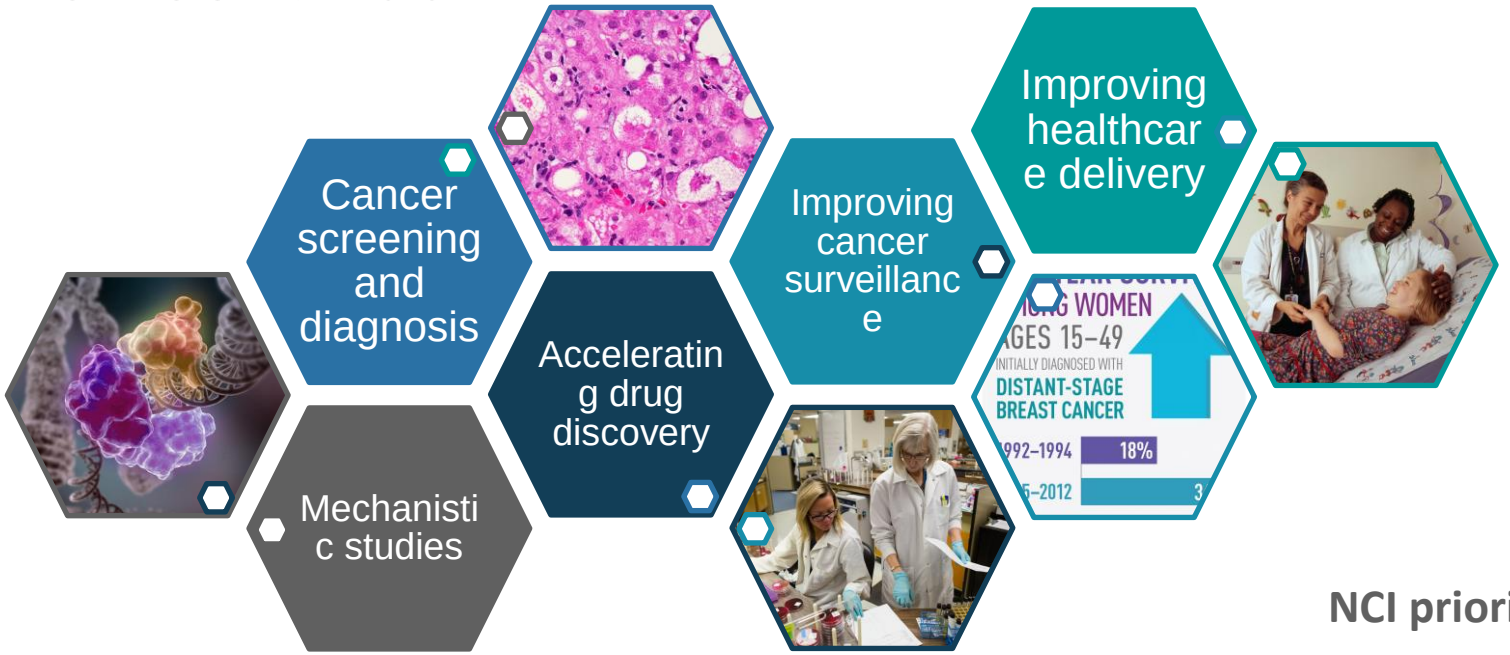
Current sources of data



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

icons by the Noun Project

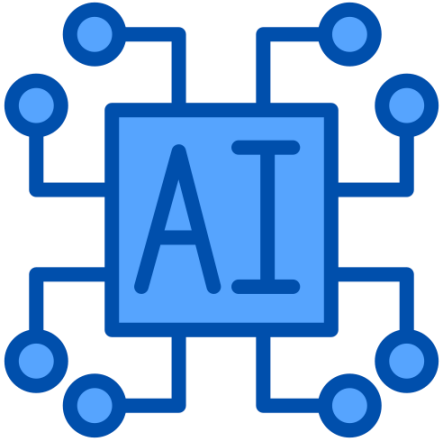
Data across the cancer research continuum



NCI priority!

Understanding the whole person in their full context

Data Science Training



AI READY

DATA SCIENTISTS & CANCER RESEARCHERS

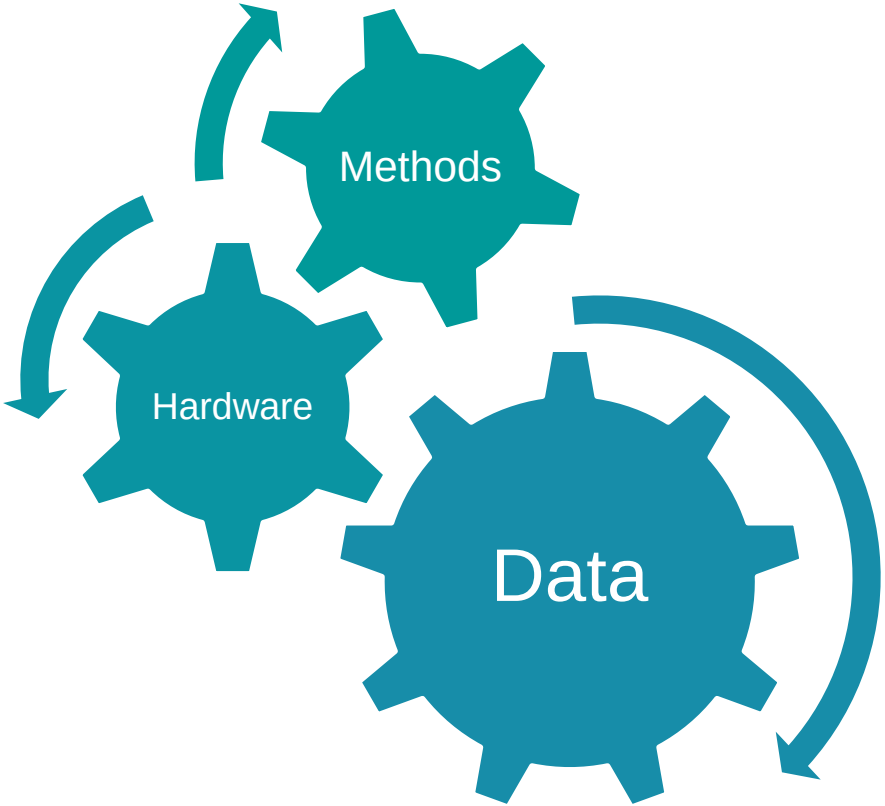
DIVERSE WORKFORCE

NCI priority!

icons by Assurant.com & Iconfinder.com

AI and Cancer Research

Recent advances have led to promising new applications of AI to cancer research



Rise of Artificial Intelligence

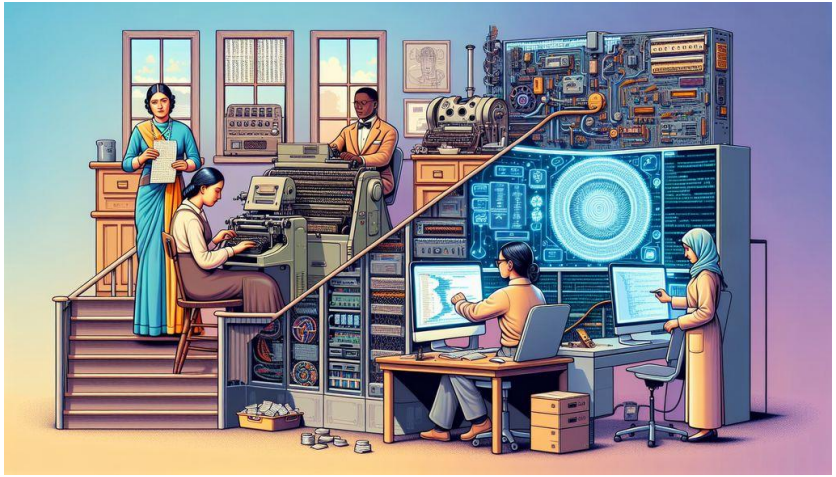


*Image generated by AI (DALL*E3)*

Key AI milestones throughout history

1800s

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

1960s

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

2018

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

2022

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

Principles for thinking about AI

NCI priority!

Leveraging the Investment in AI – Opportunities

- **Integrate generalist AI models** and tools to benefit cancer research and care
- **Coordinate** AI research activities, **integrate training, workforce development, adoption of AI** into cancer research

Image from beyond prompting

Thank You

cancer.gov

cancer.gov/espanol