NCI Data Sharing: Broad & Equitable Access to Research

Jaime M. Guidry Auvil, Ph.D.

Director, Office of Data Sharing

May 20, 2019

NIH NATIONAL CANCER INSTITUTE

What *IS* Data Sharing?

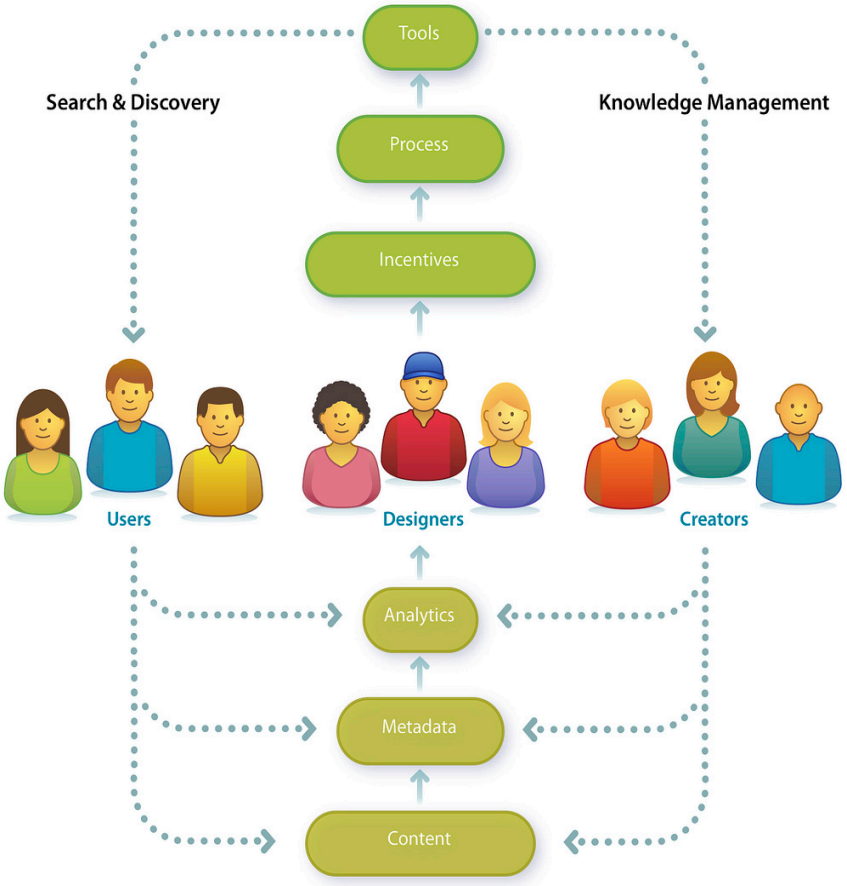
Data sharing – practice of making research data & metadata available for use by the broader community

The NEW ENGLAND JOURNAL of MEDICINE

M. Mello, et al. "Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing" NEJM, June 7, 2018

- **Translation** of research results into knowledge, products & procedures to improve human health
- **Replication** of results is key
- Transparency & **openness** are part of the scientific method
- Can be done a variety of ways

Benefits of Data Sharing



Increases **statistical power** & **scientific value** by integrating data from multiple studies

Enables & facilitates:

- **Reproducibility** & **validation** of research results
- Investigation of wide range of research questions
- **Innovation** of methods and tools for research

Reduces duplication and saves time, valuable resources & experimental costs

"To create a comprehensive data sharing vision and strategy for NCI which advocates for the proper balance between broad and equitable data sharing with the needs of the research and participant communities."

- *Office of Data Sharing Mission*

Office of Data Sharing Key Priorities

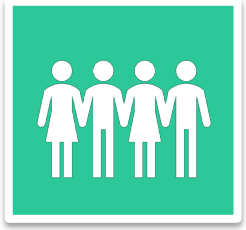
Provide leadership & guidance to enhance data sharing for NCI & cancer research community.

Guide NCI implementation & interpretation of NIH & NCI data management & sharing policies.

Advise on considerations for ethical data access and sharing for the cancer community.

Facilitate data submission & access procedures for NCI datasets/ repositories.

Encourage participation in major data sharing initiatives.



Create data sharing resources to inform & guide cancer communities.

Seeking Appropriate Balance

Natural tension between values & needs:

- Protect privacy and research integrity
- Respect broad range of participant wishes
- Promote health advances through research
- Support investigators and their ability to do good work

Lead NCI Data Sharing & Management Implementation

Establish the NCI approach to Data Sharing policies considering:

- NCI mission
- DOC programmatic priorities
- Patient/ study participant intent

Educate and incentivize stakeholders to support understanding of and compliance with NIH & NCI policies

Guidance for NCI Cancer Moonshot℠ Public Access and Data Sharing Policy

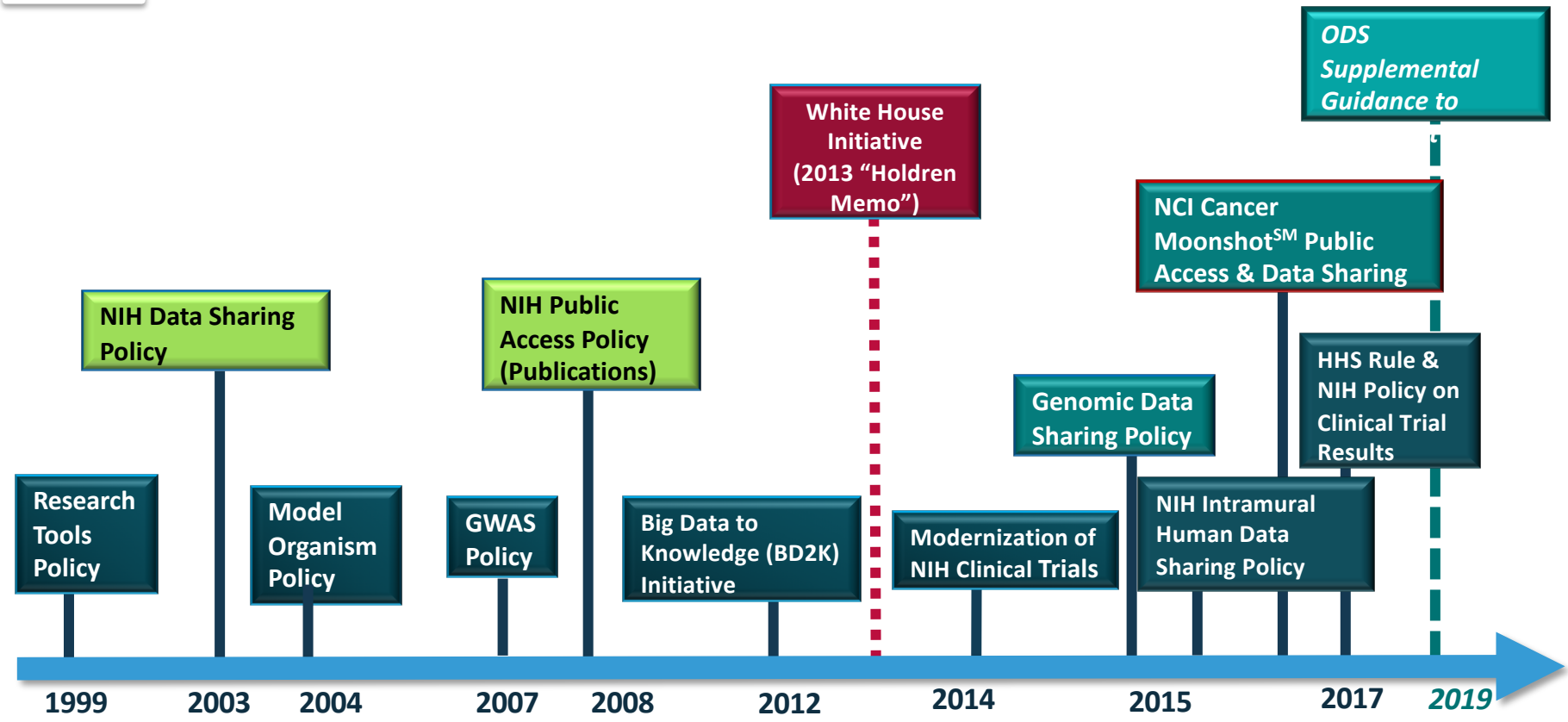
Represent NCI on NIH and external governance committees

CANCER MOONSHOT

Data Sharing Resources

Encourage Data Sharing

Relevant NIH & NCI Data Sharing Policies



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

Credit to Dina Paltoo & Laura Lyman Rodriguez, adapted

NIH Data Sharing* & Publication Policies

Goal → Make the results & accomplishments (final research data) of NIH-supported activities available in a timely fashion (publication) for use by other researchers

Data sharing – essential for expedited translation of research results into knowledge, products, & procedures to improve human health

❖*NIH expects that funded investigators may benefit from first and/or continuing use of data, but not from prolonged exclusive use.*

- **Data Sharing Plans** required for activities funded at **$500,000** or more in direct costs in any single year
- Submit final, peer-reviewed manuscripts to NLM **PubMed Central** upon publication acceptance (broadly available **within 12 months**)

**NIH Data Sharing Policy is currently being refined; hope to release by end of 2019*

The NCI Cancer Moonshot℠

Mission: "*Enable all participants across the cancer research and care continuum to contribute, access, combine and analyze diverse data that will enable new discoveries and lead to lowering the burden of cancer.*"

Overarching goals

- Accelerate progress in cancer, including prevention & screening; cutting edge basic research → wider standard of care
- Encourage greater cooperation & collaboration among academia, government, and private sector
- ***Enhance data sharing***

NCI Cancer Moonshot℠ Public Access & Data Sharing Policy

Goal → maximize availability of publications and sharing of underlying data.

- Awardees should make resulting Publications, and to the extent possible the Underlying Primary Data, immediately and broadly available to the public.
- Policy is in addition to requirements and expectations specified under other applicable NIH public access and data sharing policies (*including but not limited to*):
  - HHS Rule for Disseminating Clinical Trials
  - NIH Intramural Human Data Sharing Policy
  - NIH Genomic Data Sharing Policy

Advocate for the Appropriate Balance

- **Engage stakeholders** to refine NCI & NIH data sharing strategies.
- Economic, ethical, Legal and Social Issues (**dEELSI**) program for data sharing
- **Advocacy, outreach** and community **engagement** (health disparities among underserved populations).
- **Promote a "healthy" commercial marketplace** with less restrictive business models (e.g. not based on "controlled" access cancer research/care data).

Data Sharing Resources

Encourage Data Sharing

Current Barriers to Interoperability

- General **Inconsistency or Lack** of:
  - **Broad consent language & uniform consent processes**
  - Data formats & metadata standards
  - Searchable, interconnected data repositories with associated tools, services
  - Agreed upon ontologies, vocabularies, data models

- **Policy & Procedural Obstacles** preventing participants or researchers from sharing data
  - Mandates and legal issues from funding sources (GDPR)
  - Lack of resources to format data and metadata files, and further submit them to databases
  - Inability to effectively track & manage data and forms digitally
  - No criteria for how to choose the best database to house the data

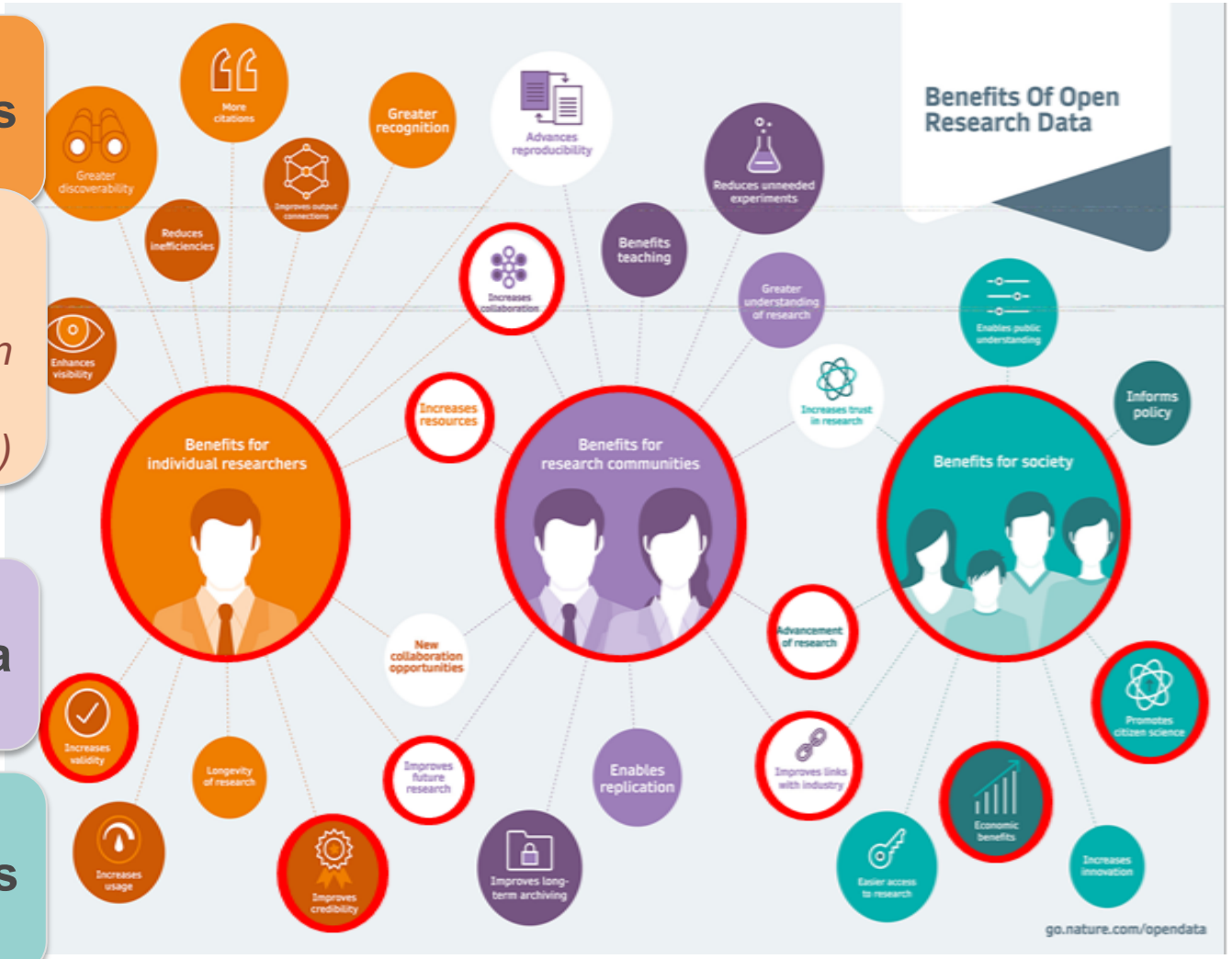
Benefits of Open Science

Open access

- Research & data
- All levels of investigators *(citizen & professional scientists, clinicians)*

Open data

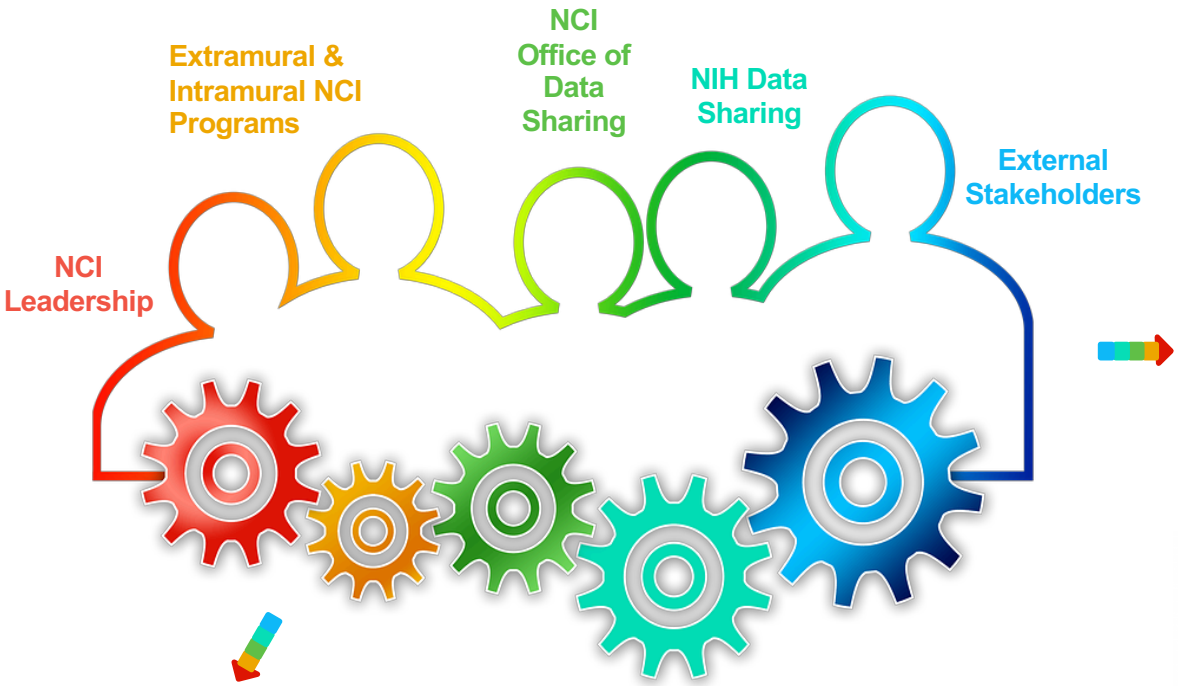
Open sources



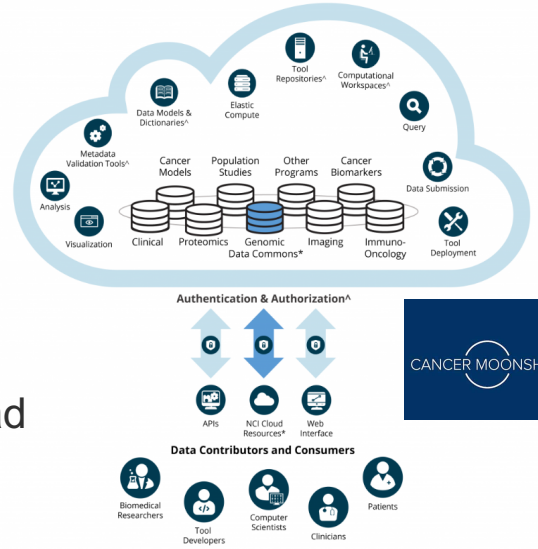
http://www.springernature.com/gp/authors/research-data-policy

National Academies of Sciences, Engineering, and Medicine. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116

Work Across NCI to Enhance Data Sharing



Work with NCI DOCs to implement data sharing policies & procedures (e.g. CIB, Cancer Research Data Commons; NCI DOCs, Moonshot[SM])



Promote open-access licenses wherever possible:

- research data & metadata
- Publications
- Annotations
- Software
- other research methodologies and/or tools

Support development of standards & metrics for broad data interoperability and sharing of cancer data and metadata.

NIH Data Management Process

Data Collection
Data Sharing Plans

Data Submission
& Management

Data Distribution &
Secondary Use

Research Participants

Submitting Investigators

dbGaP

Recipient Investigators

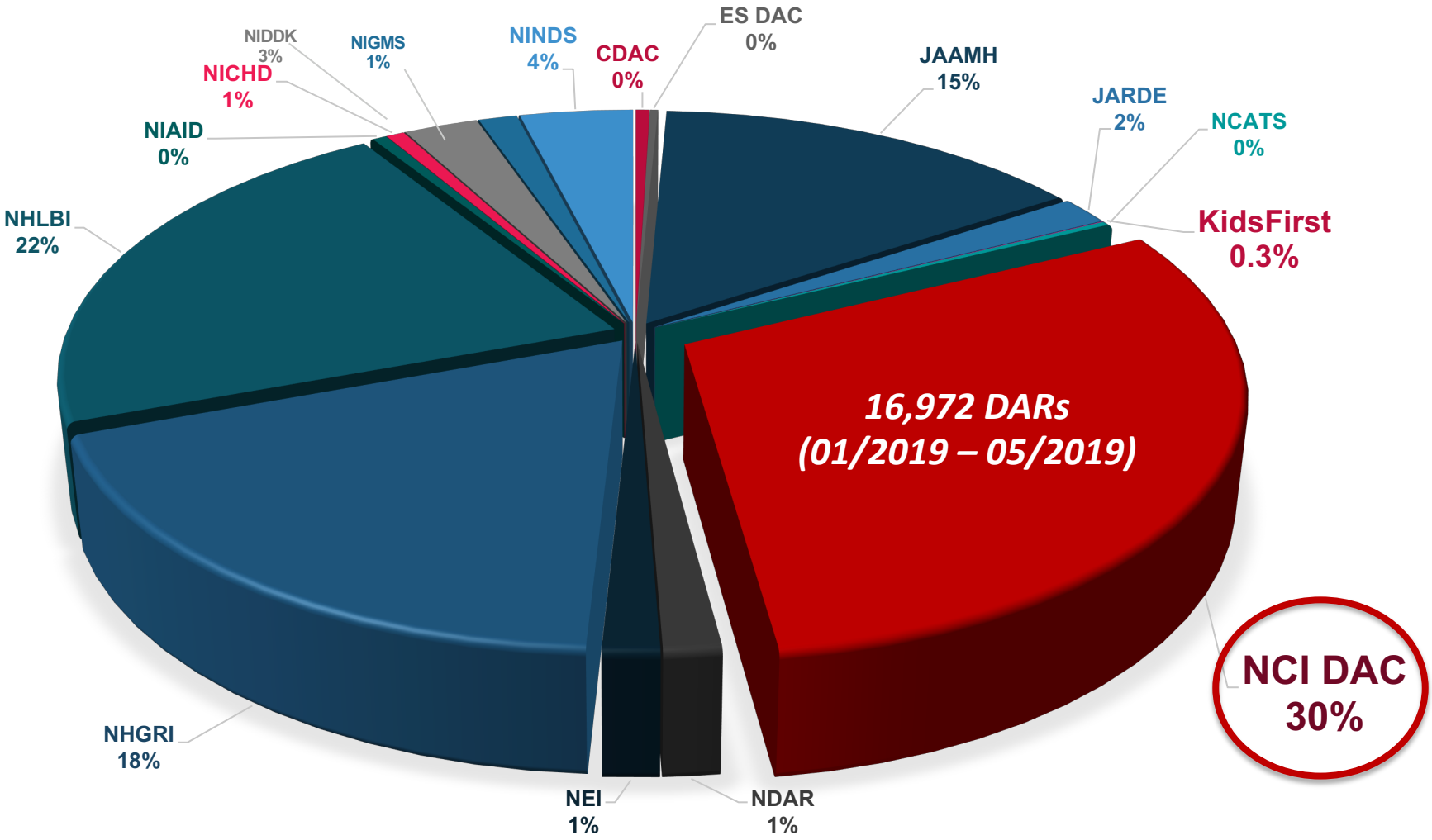
Unrestricted Access

Registered Access

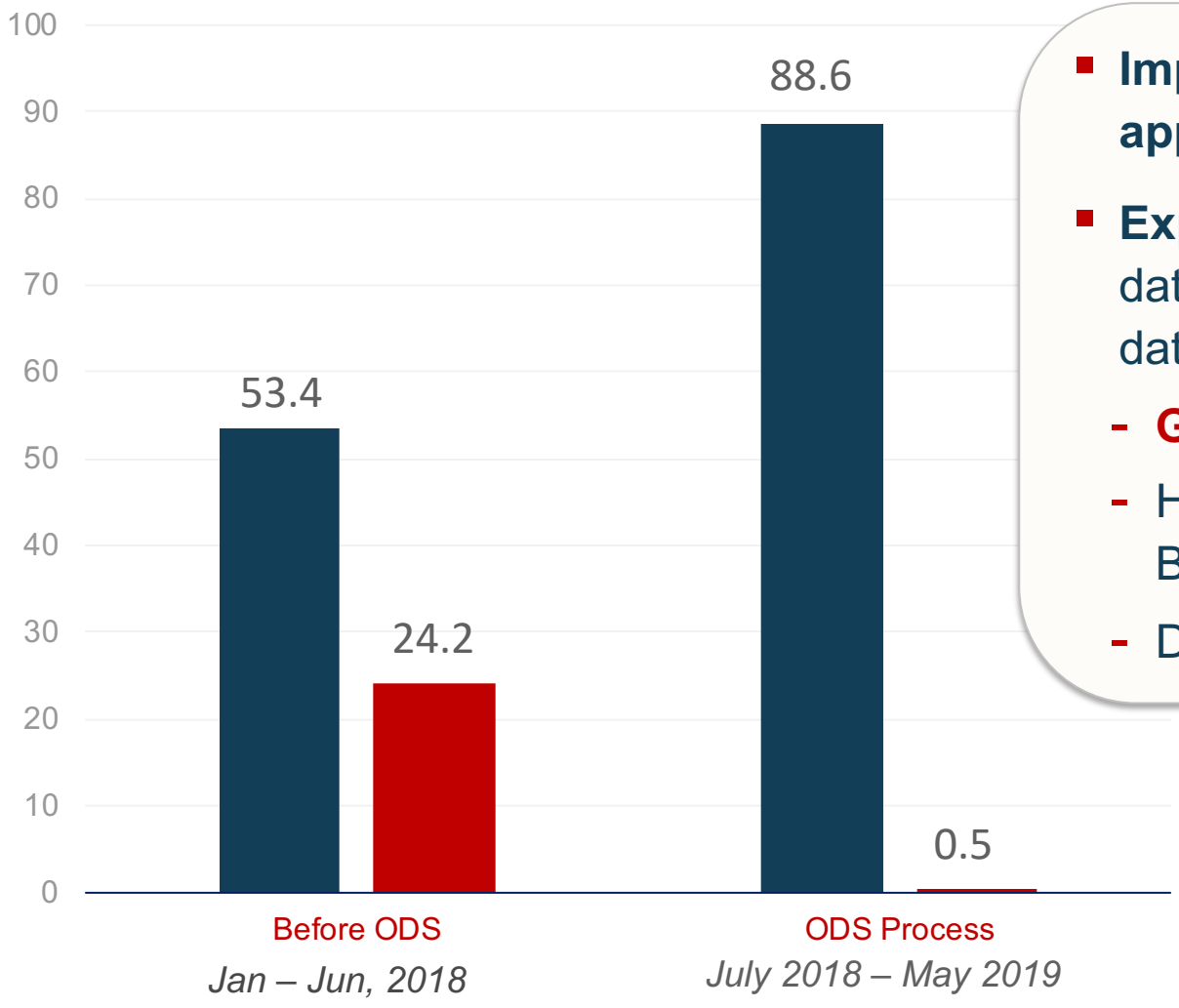
Controlled Access [Individual level data]

Credit to Vivian Ota Wang – ODS

Requests to Access Controlled Datasets



ODS Data Access Centralization



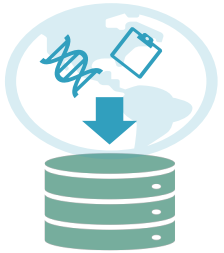
- **Implement multiple team approach**
- **Expedited reviews** for datasets with broader data uses
  - **General Research Use**
  - Health, Medical & Biomedical
  - Disease Specific Cancer

Childhood Cancer Data Initiative

Facilitate the sharing of childhood cancer data from multiple sources through a connected data infrastructure

FEDERATE

Identify opportunities to align & integrate multiple data sources to make data work better for patients, clinicians, researchers

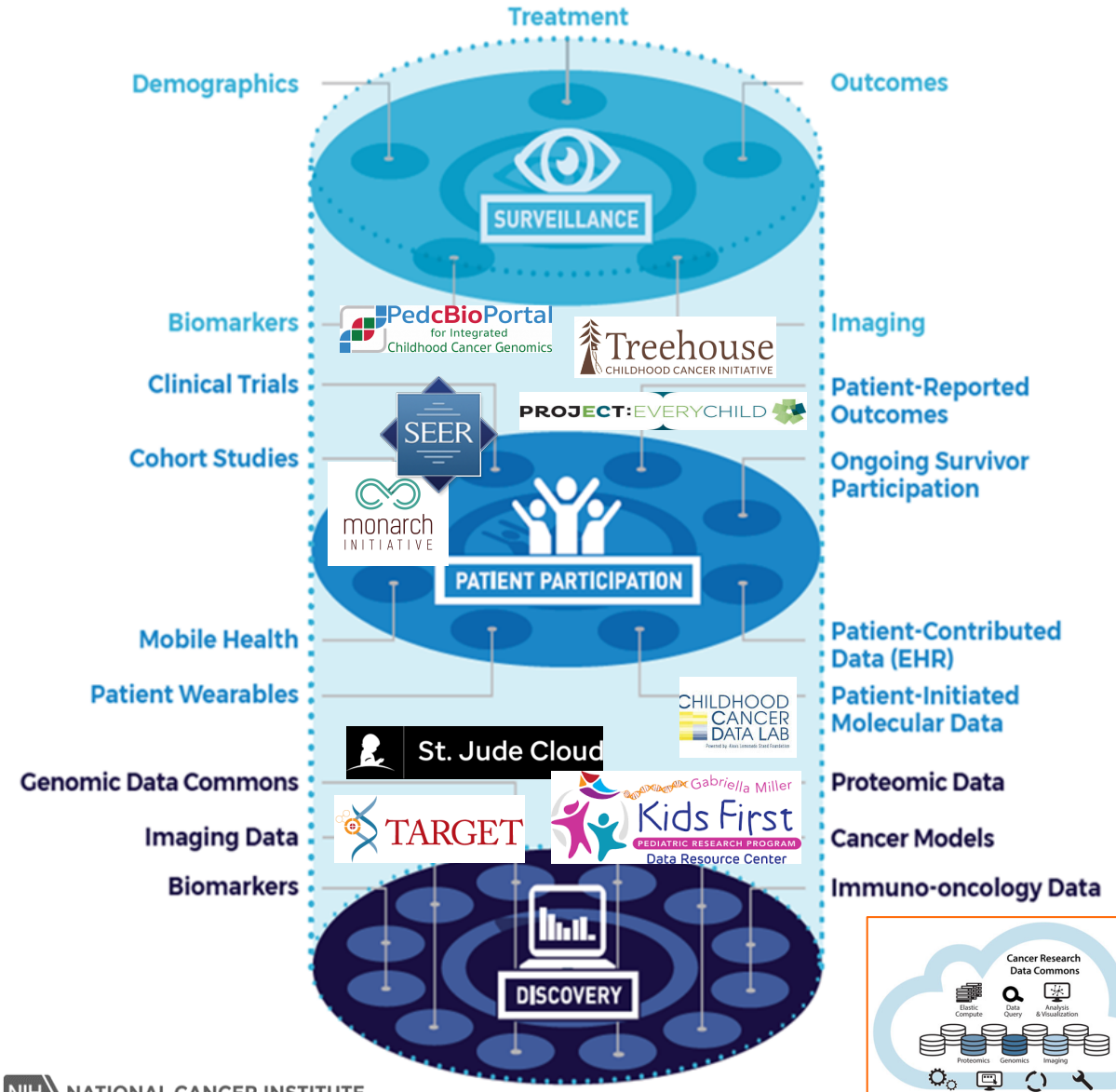


AUGMENT

Maximize every opportunity to improve treatments and outcomes for children with cancer

USE

Building a Cancer Data Ecosystem:
Basic Research → Clinical Care

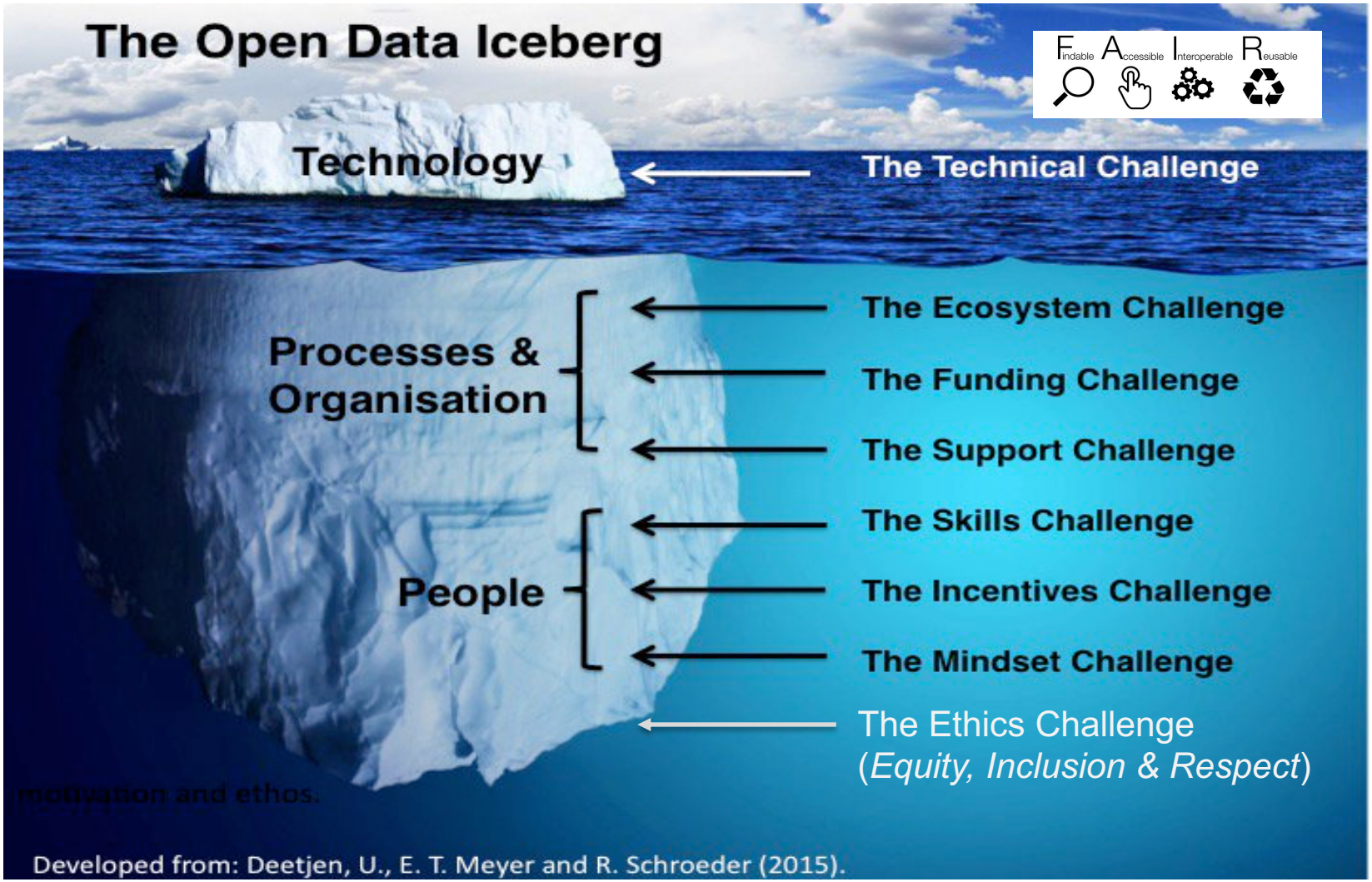


Integrating Cancer Research Data

- Enhanced cloud-computing
- Underlying data science infrastructure
- Services linking clinical, image & molecular data
- Develop standards & tools data interoperability
- Sustainability & data governance for Ecosystem

➢ A NCI **Cancer Research Data Commons** is one component

Partly FAIR, Partly Cloudy

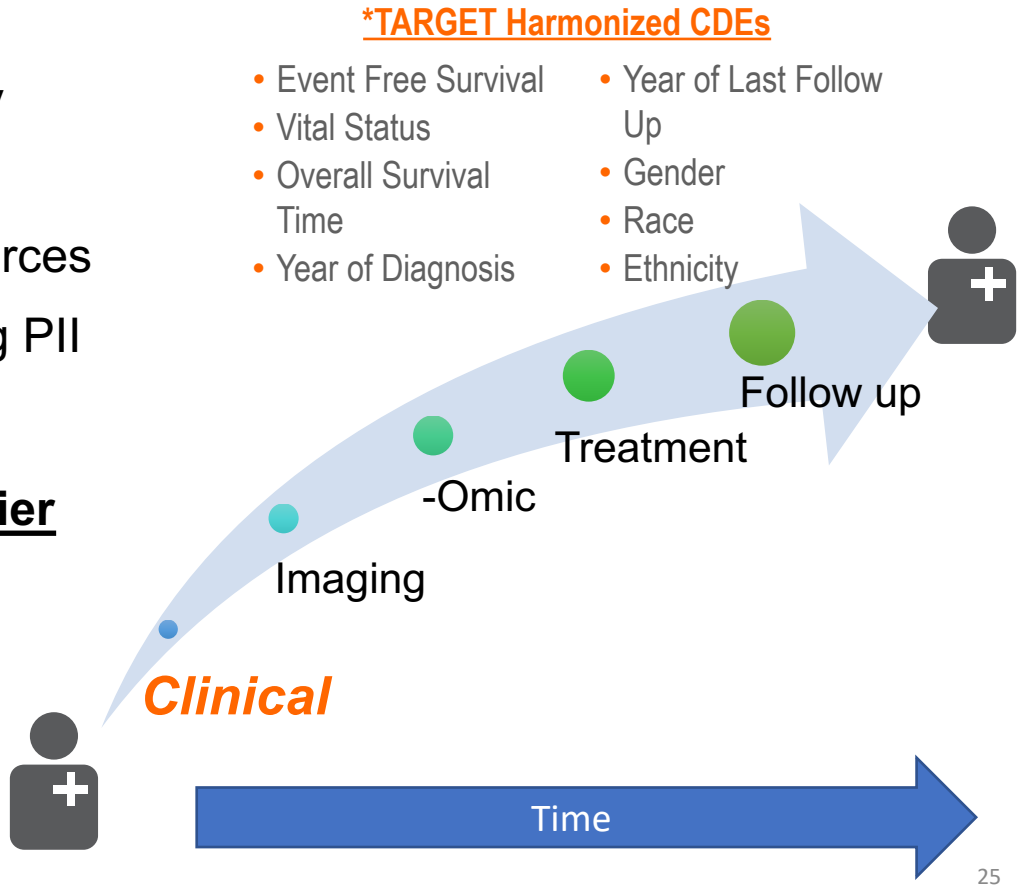


Wilkinson, M. D. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*3:160018 doi: 10.1038/sdata.2016.18

Connecting Patient-Level Data

Goal → Privacy-preserving patient record linkage across multiple data sources, data types, & research studies; update patient-level data over time

- Challenges:
  - Protecting patient confidentiality
  - Consistency of identifying data
  - PII available across diverse sources
  - Accuracy of linkage with varying PII
  - Scalability
- **Encrypted Unique Patient Identifier**
  - Allows linkage of diverse data
  - Permits data sharing across multiple sources without release of PII



Precision Oncology is a Grand Challenge

Courtesy of P. Kuhn (USC)

It Requires:

- Deep biological understanding
- Advances in scientific methods
- Advances in tools, informatics & technology
- Advances in data management and analytics

*Cancer Research and Care generate detailed **data** that are critical to create a learning health system for cancer*

Maximizing Data Utility

Goal: *To rapidly identify viable molecular targets to better understand & treat cancer.*

Initial Discovery

TCGA data describes **33** DIFFERENT TUMOR TYPES ...including **10** RARE CANCERS

...based on paired tumor and normal tissue sets collected from **11,000** PATIENTS

...using **7** DIFFERENT DATA TYPES

Preclinical Evaluation/ Data Mining

TCGA RESULTS & FINDINGS

MOLECULAR BASIS OF CANCER	Improved our understanding of the genomic underpinnings of cancer
TUMOR SUBTYPES	Revolutionized how cancer is classified
THERAPEUTIC TARGETS	Identified genomic characteristics of tumors that can be targeted with currently available therapies or used to help with drug development

Early Clinical Trials & Biology Studies

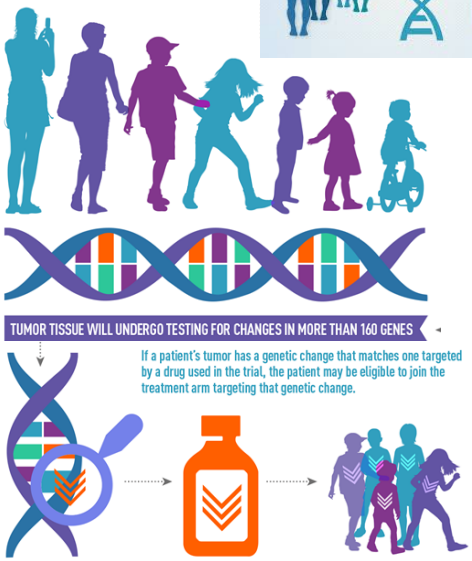
Definitive Clinical Trial/ Standard Tx

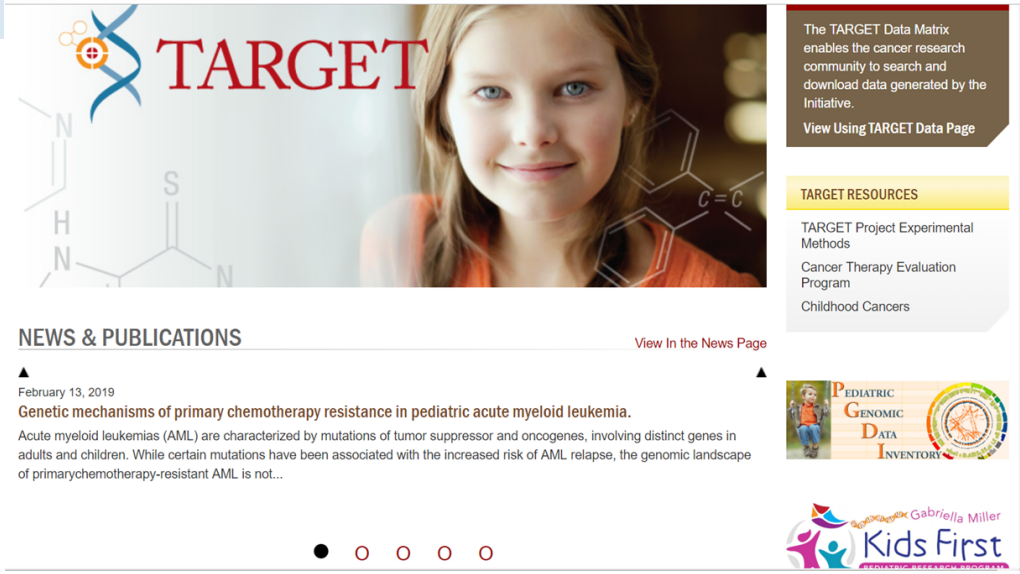
Translating Key Discoveries

Identify novel targets for hard to treat, refractory disease

Discover new ways to use existing targets or therapeutics

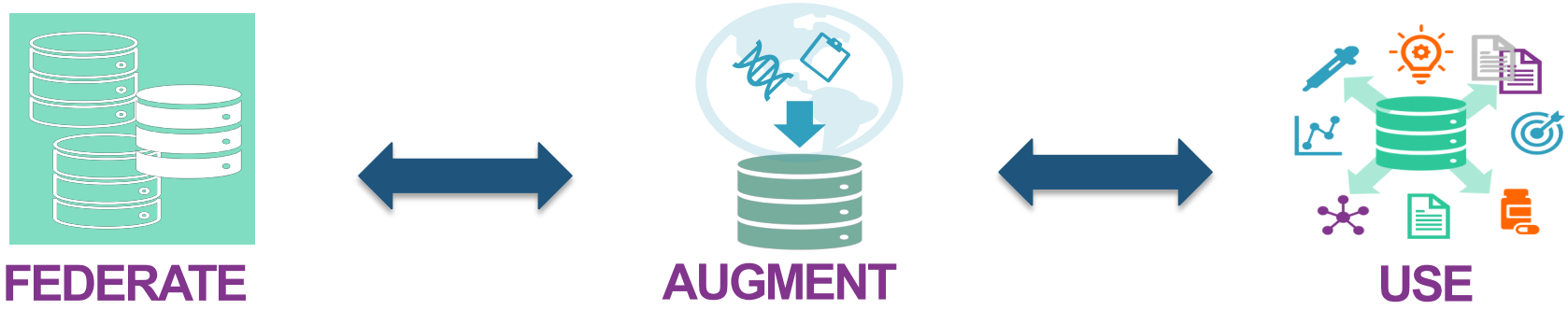
Reduce use of invasive, costly or toxic therapies based on empirical data





Application of Cancer Genomics to Clinical Research

Childhood Cancer Data Initiative Symposium



- **July 29–July 31, 2019**: Childhood Cancer Data Initiative Symposium and poster session in Washington, DC, to plan and shape the initiative

Contact Us About Data Sharing

nciofficeofdatasharing@mail.nih.gov

#NCIODS

datasharing.cancer.gov