Cancer Research Data Ecosystem

Warren Kibbe, PhD

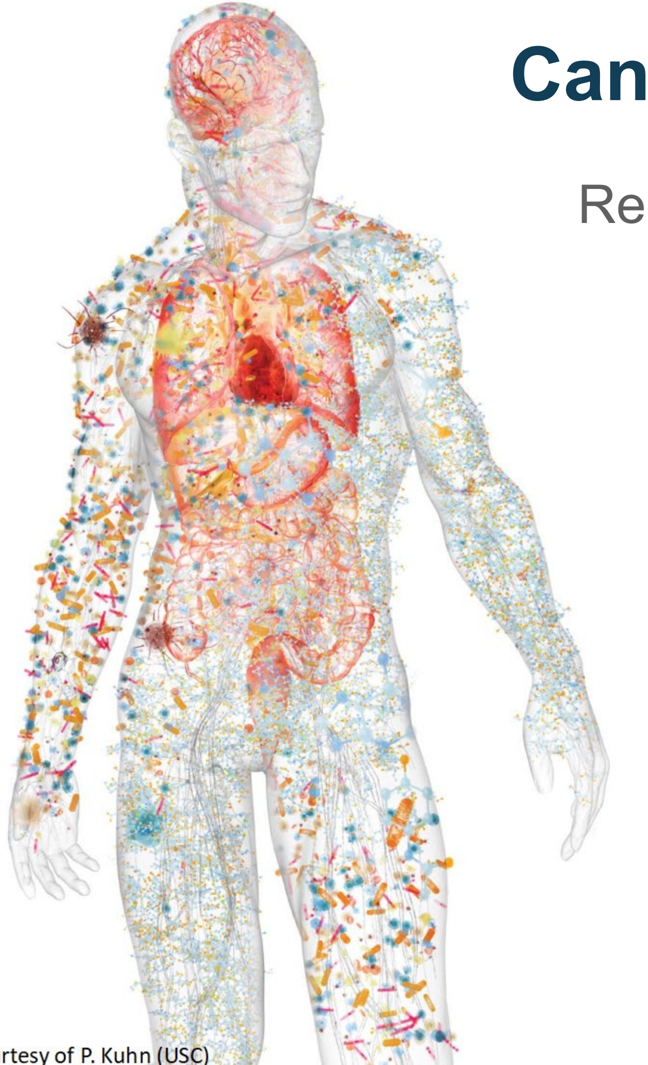
Warren.kibbe@nih.gov

@wakibbe

NIH NATIONAL CANCER INSTITUTE

NCI Imaging Community Call
January 9th, 2017

2017
HAPPY NEW YEAR!



Courtesy of P. Kuhn (USC)

Cancer is a Grand Challenge

Requires:

Deep biological understanding

Advances in scientific methods

Advances in instrumentation

Advances in technology

Data and computation

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

Courtesy of P. Kuhn (USC)

2006-2015:

A Decade of Illuminating the Underlying Causes of Primary Untreated Tumors Omics Characterization

THE CANCER IMAGING ARCHIVE

The Cancer Genome Atlas

CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM

(10,000+ patient tumors and increasing)

Cancer Statistics

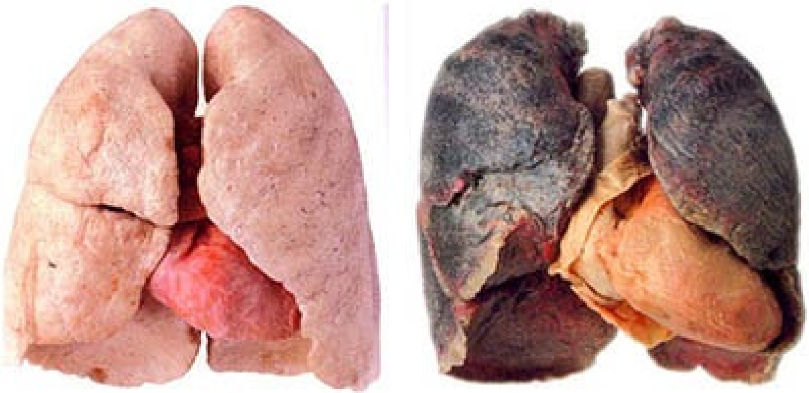
In 2016 there were an estimated

1,700,000 new cancer cases

and

600,000 cancer deaths

- American Cancer Society



Cancer remains the **second most common cause of death** in the U.S.

- Centers for Disease Control and Prevention 2015

Understanding Cancer

- **Precision medicine** will lead to **fundamental understanding** of the complex interplay between genetics, epigenetics, nutrition, environment, clinical presentation and **direct effective, evidence-based prevention and treatment**.

Changing the conversation around data sharing

NIH Data Commons

NCI Genomic Data Commons

National Cancer Data Ecosystem

- How do we find data, software, standards?
- How can we make data, annotations, software, metadata accessible?
- How do we reuse data standards?
- How do we make more data machine readable?

Data Commons co-locate data, storage and computing infrastructure, and frequently used tools for analyzing and ***sharing data*** *to create an* ***interoperable*** *resource for the research community.*

*Robert L. Grossman, Allison Heath, Mark Murphy, Maria Patterson, A Case for Data Commons Towards Data Science as a Service, to appear. Source of image: Interior of one of Google's Data Center, www.google.com/about/datacenters/.

Cancer Data Sharing and Data Commons:

A Cancer Research Data Ecosystem

- Making data available for discovery, validation, new therapies
- Working toward a learning National Cancer Data Ecosystem
- Maximizing the impact, reuse, and reproducibility of cancer research
- Changing incentives for data sharing

Reduce the risk, improve early detection, outcomes, and survivorship in cancer

NIH Genomic Data Sharing Policy

https://gds.nih.gov/

Went into effect January 25, 2015

NCI guidance:

http://www.cancer.gov/grants-training/grants-management/nci-policies/genomic-data

Requires public sharing of genomic data sets

FAIR – Making data Findable, Accessible, Attributable, Interoperable, Reusable, and provide Recognition

Force11 white paper
https://www.force11.org/group/fairgroup/fairprinciples

Cancer Research Data Ecosystem – Cancer Moonshot BRP

Discovery

Proteogenomics
Imaging data
Clinical trials

Well characterized
research data sets

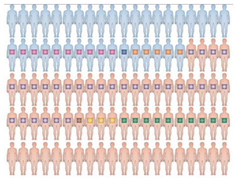
GDC

Research information
donor

Patient engaged
Research

Clinical Research
Observational studies

Cancer cohorts

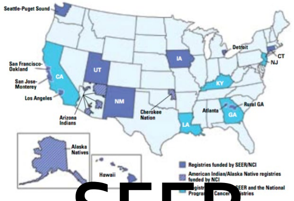


Active research
participation

Surveillance
Big Data
Implementation research

EHR, Lab Data, Imaging,
PROs, Smart Devices,
Decision Support

Patient data



SEER

Learning from every
cancer patient

The Beau Biden Cancer Moonshot

How do we enable meaningful, patient-centered and patient-level data sharing for cancer and promote access to clinical trials for all Americans?

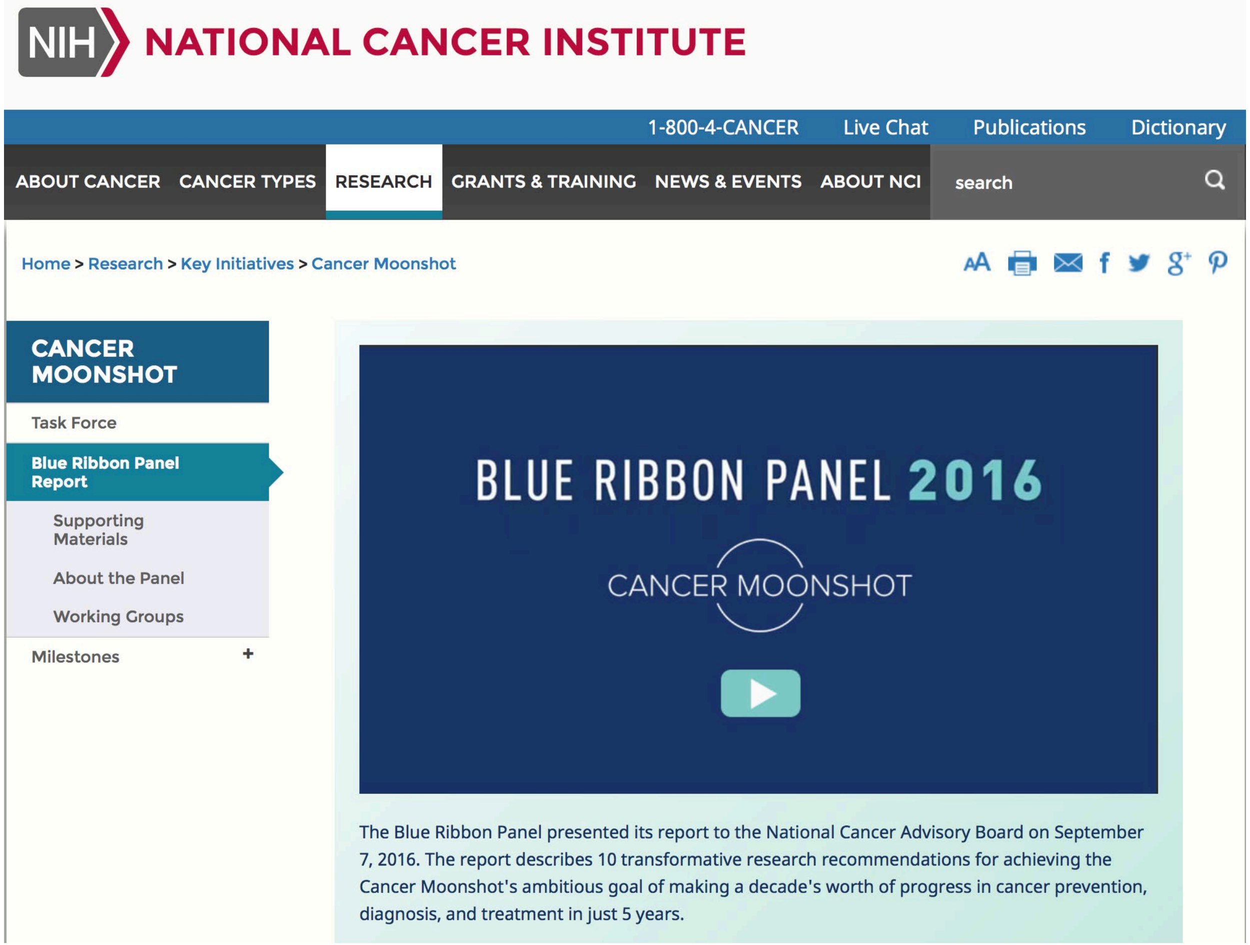
Goals of the Beau Biden Cancer Moonshot

- Accelerate progress in cancer, including prevention & screening
  - From cutting edge basic research to wider uptake of standard of care
- Encourage greater cooperation and collaboration
  - Within and between academia, government, and private sector
- Enhance data sharing

(Presidential Memo 2016)

A Few Beau Biden Cancer Moonshot Milestones

- Announced by President Obama at the State of the Union January 12, 2016
- Blue Ribbon Panel convened at AACR, April 18, 2016
- Genomic Data Commons went public June 6, 2016
- Vice President’s Cancer Moonshot Summit – June 29, 2016
- Rethinking Clinical Trial Search – Open API at https://clinicaltrialsapi.cancer.gov
- Blue Ribbon Panel recommendations – accepted by the National Cancer Advisory Board on September 7th, 2016
- Cancer Moonshot Task Force and BRP recommendations sent to President on October 17th, 2016 https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/milestones and released at https://cancer.gov/brp
- 21st Century Cures Act funding the Beau Biden Cancer Moonshot bill was passed 94-5 by the Senate on December 8 and signed by President Obama December 13, 2016.



The Blue Ribbon Panel presented its report to the National Cancer Advisory Board on September 7, 2016. The report describes 10 transformative research recommendations for achieving the Cancer Moonshot's ambitious goal of making a decade's worth of progress in cancer prevention, diagnosis, and treatment in just 5 years.

https://cancer.gov/brp

Blue Ribbon Panel Recommendations

- Network for Direct Patient Engagement
- Cancer Immunotherapy Translational Science Network
- Therapeutic Target Identification to Overcome Drug Resistance
- **A National Cancer Data Ecosystem for Sharing and Analysis**
- Fusion Oncoproteins in Childhood Cancers
- Symptom Management Research
- Prevention and Early Detection – Implementation of Evidence-based Approaches
- Retrospective Analysis of Biospecimens from Patients Treated with Standard of Care
- Generation of 4D Human Tumor Atlas
- Development of New Enabling Cancer Technologies

Cancer Research Data Ecosystem – Cancer Moonshot BRP

Discovery

Proteogenomics
Imaging data
Clinical trials

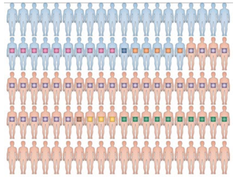
Well characterized
research data sets

GDC
Research information
donor

Patient engaged
Research

Clinical Research
Observational studies

Cancer cohorts

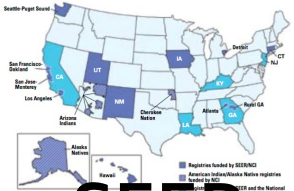


Active research
participation

Surveillance
Big Data
Implementation research

EHR, Lab Data, Imaging,
PROs, Smart Devices,
Decision Support

Patient data



SEER
Learning from every
cancer patient

Genomic Data Commons

The Cancer Genomic Data Commons (**GDC**) is an existing effort to standardize and simplify submission of genomic data to NCI and follow the principles of **FAIR** – Findable, Accessible, Attributable, Interoperable, Reusable, and Provide Recognition.

The GDC is part of the NIH Big Data to Knowledge (**BD2K**) initiative and an example of the **NIH Commons**

Microattribution, nanopublications, tracking the use of data, annotation of data, use of algorithms, supports the data /software /metadata life cycle to provide credit and analyze impact of data, software, analytics, algorithm, curation and knowledge sharing

Force11 white paper

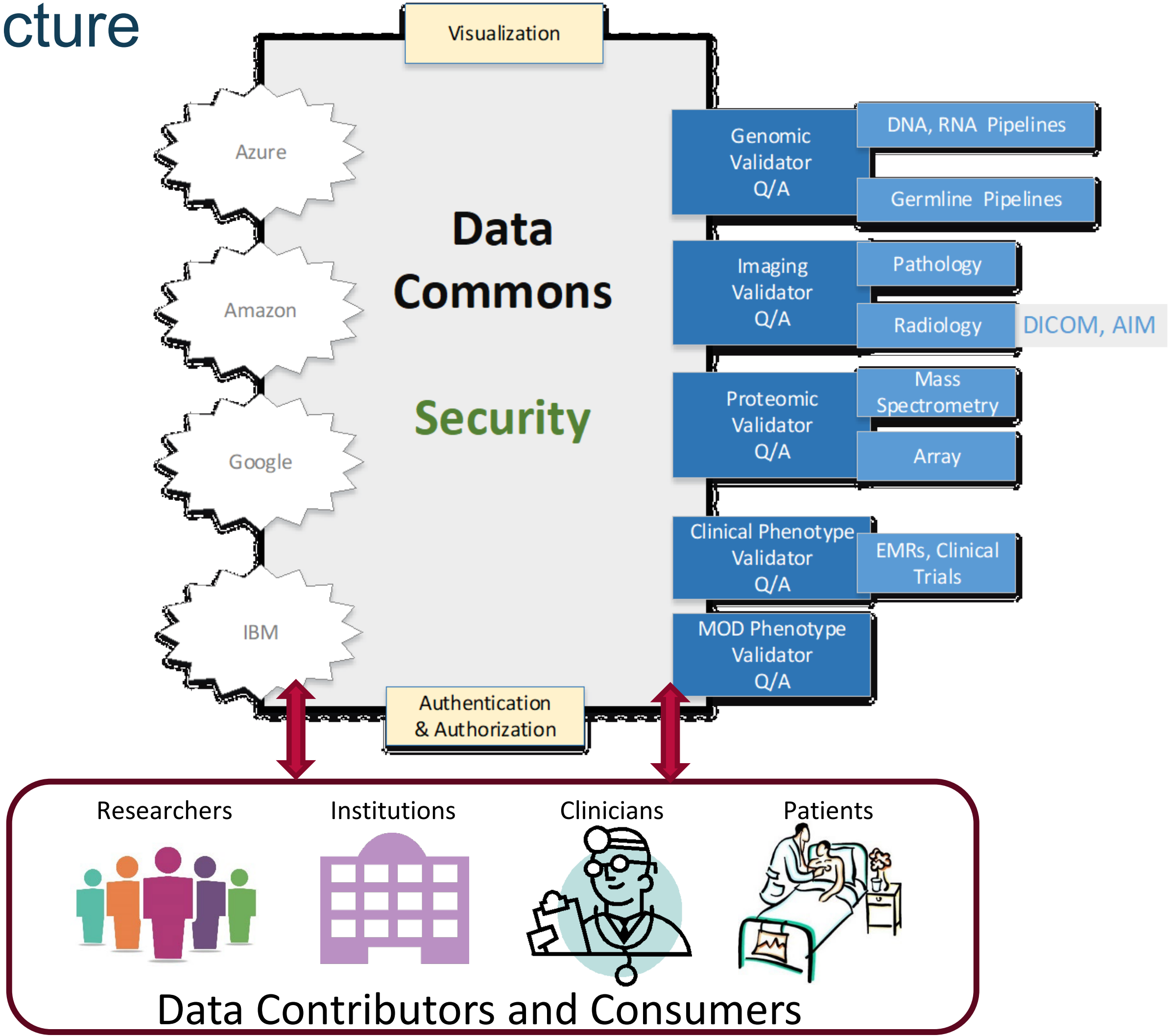
https://www.force11.org/group/fairgroup/fairprinciples

NCI Genomic Data Commons

- **The GDC went live on June 6, 2016 with approximately 4.1 PB of data.**
- This includes:
  - 2.6 PB of legacy data
  - 1.5 PB of "harmonized" data
- 577,878 files about 14194 cases (patients), in 42 cancer types, across 29 primary sites.
- 10 major data types, ranging from Raw Sequencing Data, Raw Microarray Data, to Copy Number Variation, Simple Nucleotide Variation and Gene Expression.
- Data are derived from 17 different experimental strategies, with the major ones being RNA-Seq, WXS, WGS, miRNA-Seq, Genotyping Array and Expression Array.
- **Foundation Medicine announced the release of 18,000 genomic profiles to the GDC at the Cancer Moonshot Summit.**

Data Commons Structure

NCI Thesaurus
caDSR
NLM UMLS
RxNorm
LOINC
SNOMED



Visualization

Azure
Amazon
Google
IBM

Data Commons

Security

Genomic Validator Q/A
DNA, RNA Pipelines
Germline Pipelines

Imaging Validator Q/A
Pathology
Radiology
DICOM, AIM

Proteomic Validator Q/A
Mass Spectrometry
Array

Clinical Phenotype Validator Q/A
EMRs, Clinical Trials

MOD Phenotype Validator Q/A

Authentication & Authorization

Researchers
Institutions
Clinicians
Patients

Data Contributors and Consumers

Questions?

Warren Kibbe, Ph.D.

Warren.kibbe@nih.gov

@wakibbe

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA

NIH

NATIONAL
CANCER
INSTITUTE

www.cancer.gov

www.cancer.gov/espanol