

NCI Image Repositories for Imaging AI

Keyvan Farahani, PhD

*Center for Biomedical Informatics and
Information Technology*

farahani@nih.gov
datascience.cancer.gov

The Cancer Imaging Archive (TCIA)

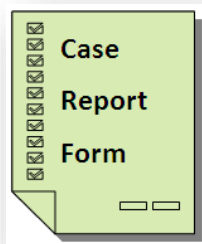
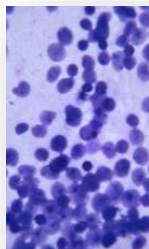
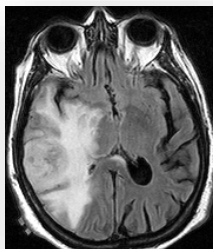


Mission: Provide image repository services to encourage data sharing for cancer imaging research

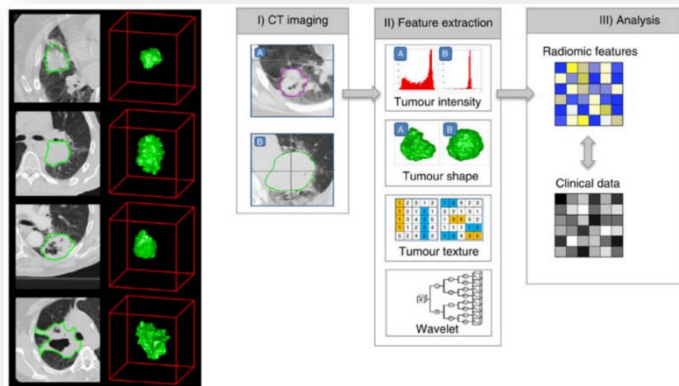


<http://cancerimagingarchive.net>

TCIA Overview



- **Collection proposals reviewed by TCIA Advisory Group**
- **110 collections** consisting of patient-level data from ~47,300 subjects available for download
- Covers **radiology, radiation therapy, and pathology**
- **Wide variety** of cancers & phantoms
- Most have **associated supporting data**
 - Demographics/outcomes/therapy
 - Image analyses (annotations, segmentations, features)
 - Genomics/Proteomics
- **REST API**
- **TCIA Data Science / AI publications: 98 manuscripts in 74 journals**



<http://cancerimagingarchive.net>

National Cancer Data Ecosystem for Sharing and Analysis

Cancer MoonshotSM

Overarching goals – Jan. 2016

- ❑ **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**

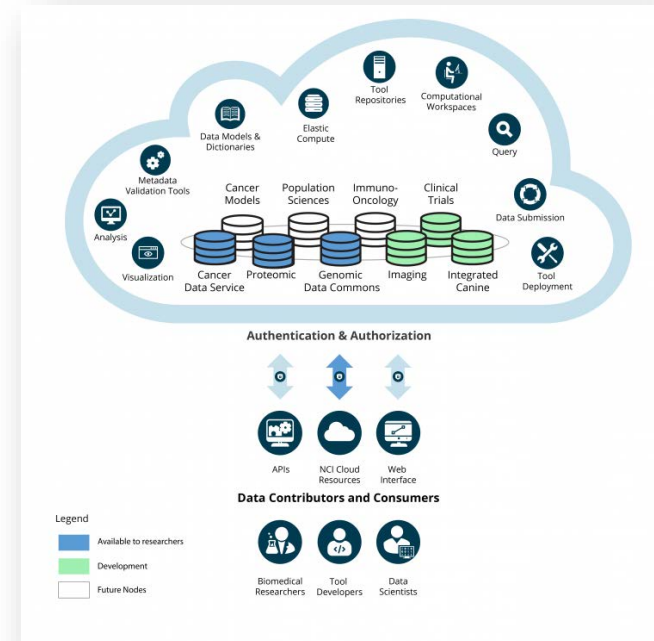


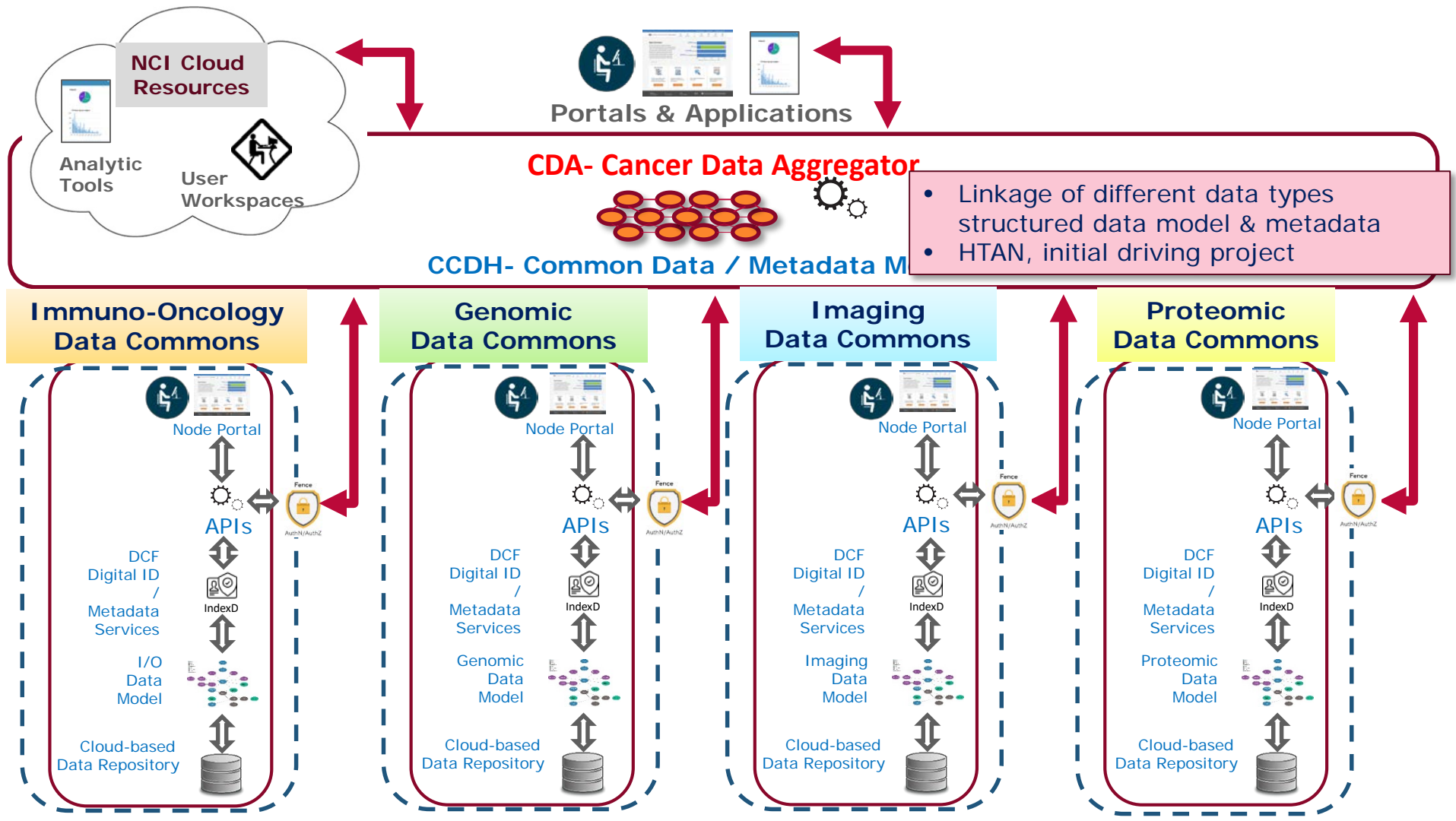
- **Build a National Cancer Data Ecosystem**
 - Enhanced cloud-computing platforms
 - Services that link disparate information, including clinical, image, and molecular data
 - Essential underlying data science infrastructure, standards, methods, and portals for the Cancer Data Ecosystem
 - Establish sustainable data governance to ensure long-term health of the Ecosystem.
 - Develop standards and tools so that data are interoperable.

Cancer Research Data Commons (CRDC)

A data science infrastructure to connect repositories, analytical tools, and knowledge bases

- Virtual, expandable, secure research infrastructure
- Storage and elastic compute
- Analysis, sharing, and archival of results
- Cross-domain analysis of large datasets



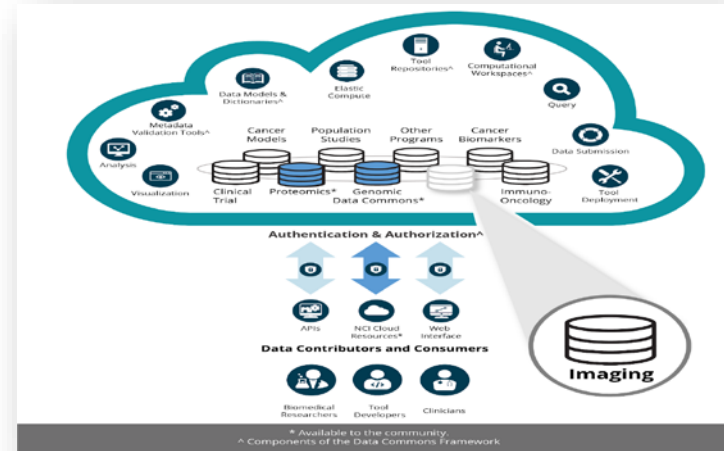


NCI Imaging Data Commons (IDC)



Cloud resource that connects researchers with

- cancer image collections
- robust infrastructure with imaging data, metadata and experimental metadata from disparate sources
- resources for searching, identifying and viewing images
- additional data types in other CRDC nodes.



Implementation:

- Google Cloud Platform
 - OHIF viewer
 - Non-restrictive Open Source
 - DICOM as prime standard
- Public release: Fall 2020*

Acknowledgements

University of Arkansas

Fred Prior
Kirk Smith
Lawrence Tarbox
Bill Bennett

Fredrick National Laboratory for Cancer Research

John Freymann
Justin Kirby
Brenda Fevrier-Sullivan
John Otridge
Todd Pihl
Ulli Wagner

BWH, Harvard

Ron Kikinis
Andrey Fedorov
Hugo Aerts

Institute for Systems Biology

William Longabaugh

General Dynamics IT

David Pot

Isomics

Steve Piper

Pixelmed

David Clunie

Cancer Imaging Program (NCI)

Paula Jacobs

Center for Biomedical Informatics and Information Technology (NCI)

Allen Dearry
Tony Kerlavage

farahani@nih.gov



**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol