Analyzing Artificial Intelligence (AI) Resources for Cancer Research Data Programs

Juergen A. Klenk[1], Dina Mikdadi[1], Chelsea A. Owens[1], Bhavani S. Singh[1], Mary G. Sears[1], Ross Campbell[1], Eric Barner[1], Ina Felau[2], Erika Kim[2], Tanja M. Davidsen[2]

[1]Deloitte Consulting LLP, Arlington, VA,

[2]Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD, Frederick National Laboratory for Cancer Research, Frederick, MD

Introduction

The National Cancer Institute's (NCI) Cancer Research Data Commons (CRDC) is a cloud-based ecosystem facilitating the sharing of critical cancer research datasets. The CRDC holds over 10 petabytes of data, including clinical trial, genomic, proteomic, and imaging data essential for comprehensive research. **CRDC's Mission** is to empower researchers by providing a cancer data ecosystem with state-of-the-art visualization, analysis, and interoperability tools in a flexible, cloud-based computational environment, and **CRDC's goals are to**:

1. Preserve the long-term value of NCI-funded data
2. Improve data submission, access, & interoperability
3. Accelerate cancer research through integrative analysis of multi-modal data
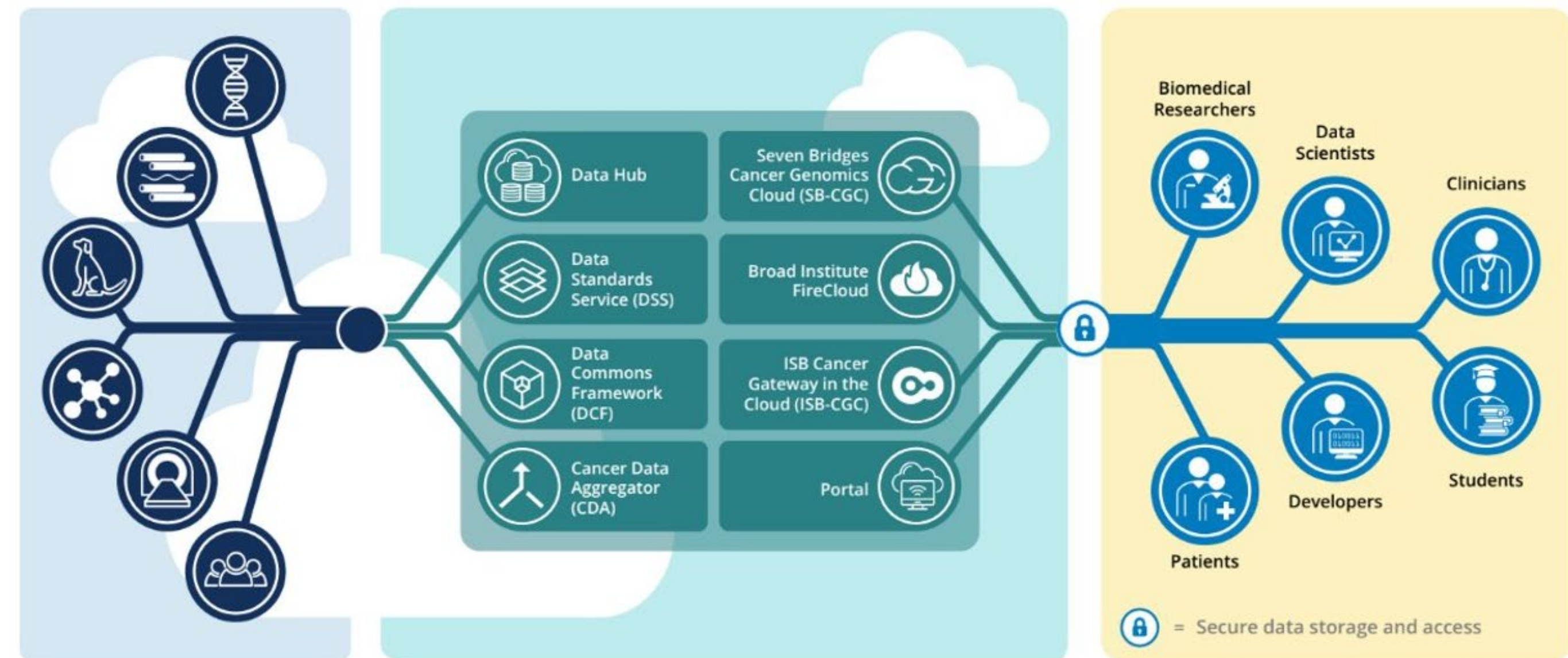

Figure 1: Overview of NCI's CRDC. Current and future resources the CRDC provides to researchers.
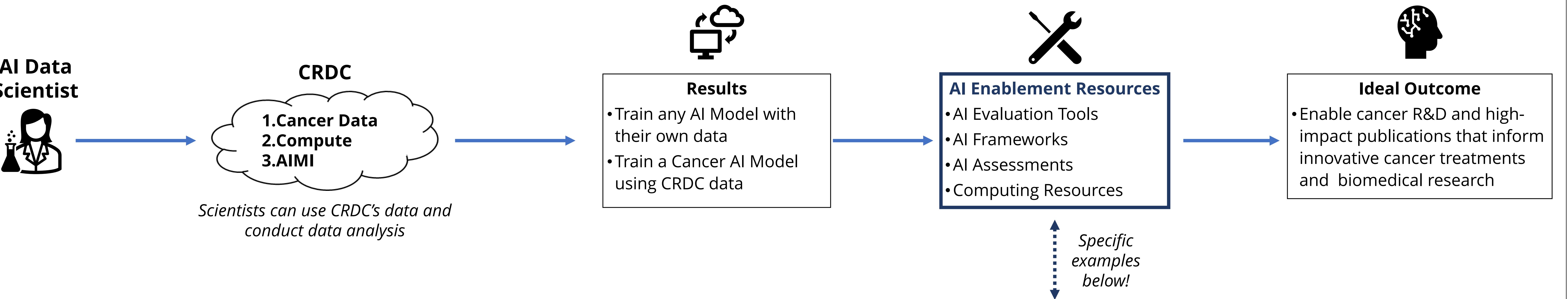
To support CRDC's mission to empower researchers with state-of-the art analysis tools, CRDC conducted an AI resource analysis to identify, track, and analyze AI resources and technology activities occurring in the US government, private industry, and international community to direct AI enablement within the CRDC ecosystem.
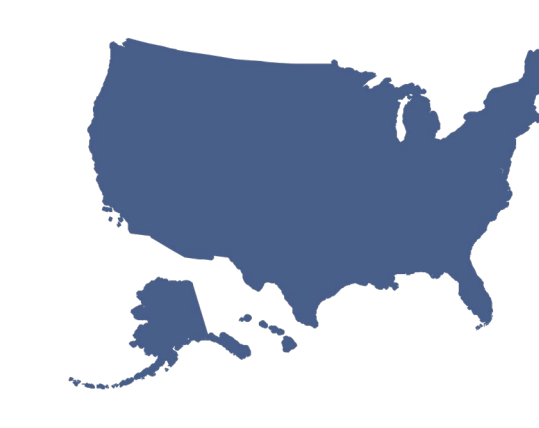
Methods

- CRDC conducted a resource analysis to inform CRDC of current AI use and policies across
  - Three branches of the Federal government
  - Private industry
  - International agencies
- CRDC identified and analyzed policies from health-focused organizations focused on
  - AI evaluation tools and frameworks
  - AI guidance
  - AI strategy

Results

Our team conducted AI resource and technology analysis across the federal government, private industry, and international community and identified the following AI enablement tools to support AI data scientists.

Scientists can use CRDC's data and conduct data analysis

Results
- Train any AI Model with their own data
- Train a Cancer AI Model using CRDC data

AI Enablement Resources
- AI Evaluation Tools
- AI Frameworks
- AI Assessments
- Computing Resources

Ideal Outcome
- Enable cancer R&D and high-impact publications that inform innovative cancer treatments and biomedical research

Specific examples below!

Source	Actors	AI Enablement Tools
US Government	**White House & Federal Agencies** **Senate AI Working Group & House of Representatives AI Task Force** **Supreme Court & Federal Courts**	• NIST's AI Risk Framework, GenAI Risk Framework, and Secure Software Development Framework • AISI Pre-deployment test results with OpenAI and Anthropic • AI Guidelines, such as " Managing Misuse Risk for Dual-Use Foundation Models" • Dioptra (AI Evaluation Tool) • Federal Agency's AI Strategies, Use Cases, and Pilots • Roadmap for AI Policy • AI-related copyright litigation rulings
Private Industry	**Stargate Project** **Chan-Zuckerberg Initiative** **US Technology Companies**	• AI Model Templates • Access to Nvidia compute resources
International Community	**UK AI Security Institute** **UK Information Commissioner's Office**	• EU AI Act • AI Accountability Frameworks • AI Evaluation Criteria • AI Risk Assessments

Conclusion

As AI guidance emerges and evolves, it is important that CRDC continues to analyze requirements to enable AI in cancer research. This ongoing analysis will inform CRDC, enhance trust within their user community, increase timely access to AI tools, and contribute to the advancement of innovative and cutting-edge cancer therapies.

Acknowledgements