Office of the National Coordinator for Health Information Technology

Transforming Cancer Data Collection and Use

December 14, 2023

ONC 2023 ANNUAL MEETING

@ONC_HealthIT

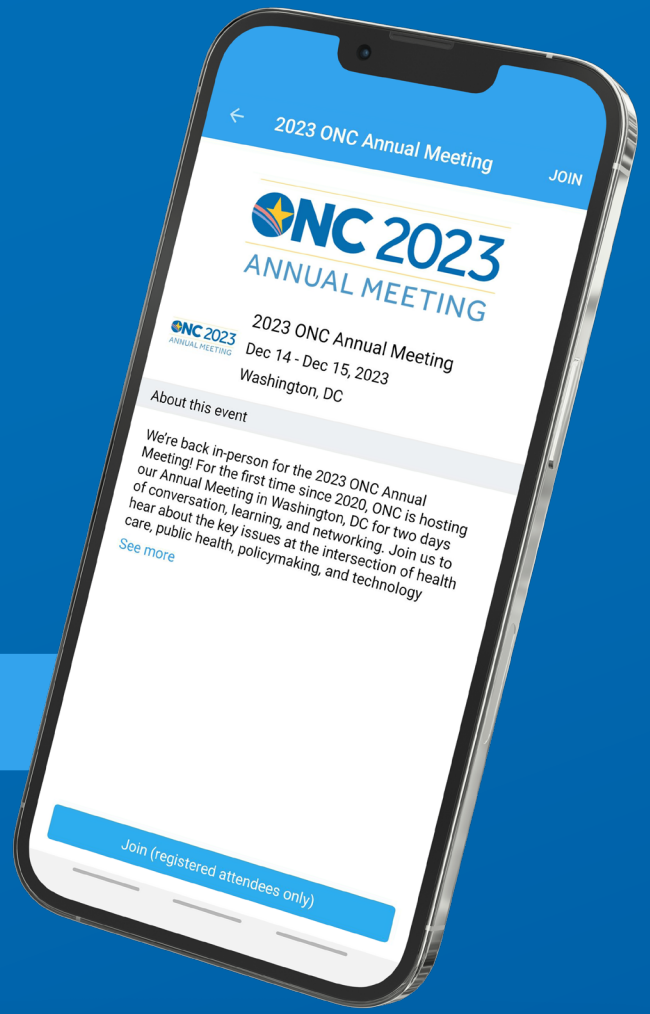
Share your content on X and don't forget to use the hashtag **#ONC2023**

ONC 2023 ANNUAL MEETING

Check out the agenda, speaker bios, venue layout, and more!

Download the event app today.

https://whova.com/portal/oncan_202312



Meet Our Moderator & Panelists

- **Aneesh Chopra (Moderator)**
  - CareJourney
- **Ali B. Abbasi, MD**
  - US Food and Drug Administration
- **Su Chen, MD**
  - CodeX FHIR Accelerator
- **Jennifer Goldsack, MChem, MA, MBA, OLY**
  - Digital Medicine Society
- **Joseph D. Rogers (Joe), MS**
  - Centers for Disease Control and Prevention
- **Umit Topaloglu, PhD, FAMIA**
  - National Institute of Cancer

Agenda

- Introductions
- Setting the stage on cancer data:
  - Aneesh Chopra (CareJourney)
  - Adi Abbasi (FDA)
  - Umit Topaloglu (NCI)
  - Joe Rogers (CDC)
- Panel Discussion

Transforming Clinical Evidence Generation through federal data standards

Ali Abbasi MD

Senior Policy Advisor

Office of the Commissioner, US Food and Drug Administration

ONC Annual Meeting

12/14/2023

Disclaimer

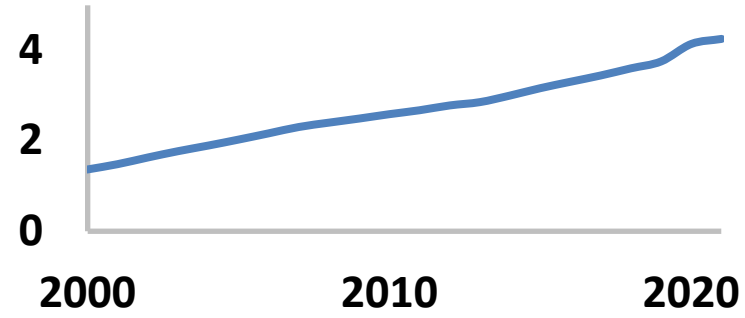
The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance from the Food and Drug Administration (FDA) or the Department of Health and Human Services (HHS).

Our model of generating evidence needs improvement

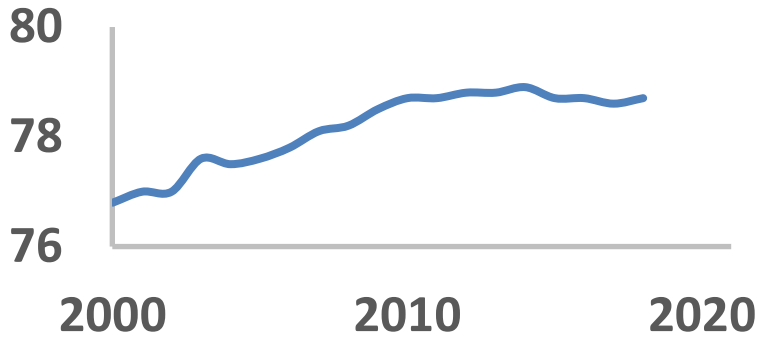
- Strong record of innovation and early-phase trials but late-phase and post approval trials are increasingly complex and expensive
- Low quality trials fail to deliver useful evidence on the safety-efficacy balance after approval
- Many questions that can only be answered in real-world clinical practice remain unanswered
- COVID-19 highlighted both challenges and opportunities for streamlined trials

DRAFT

US National Health Expenditure ($ Tr)[1]

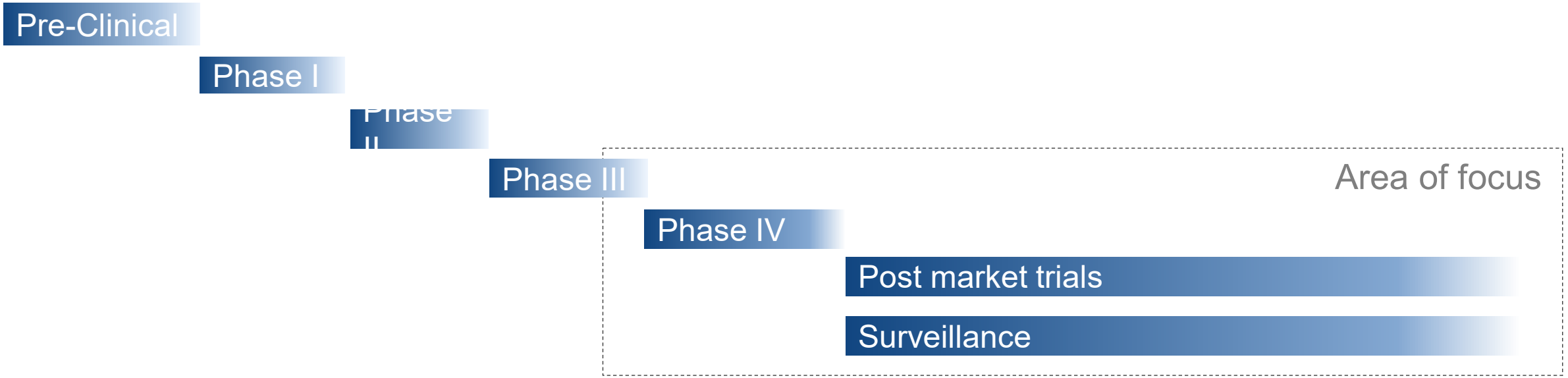


US Life Expectancy[2]



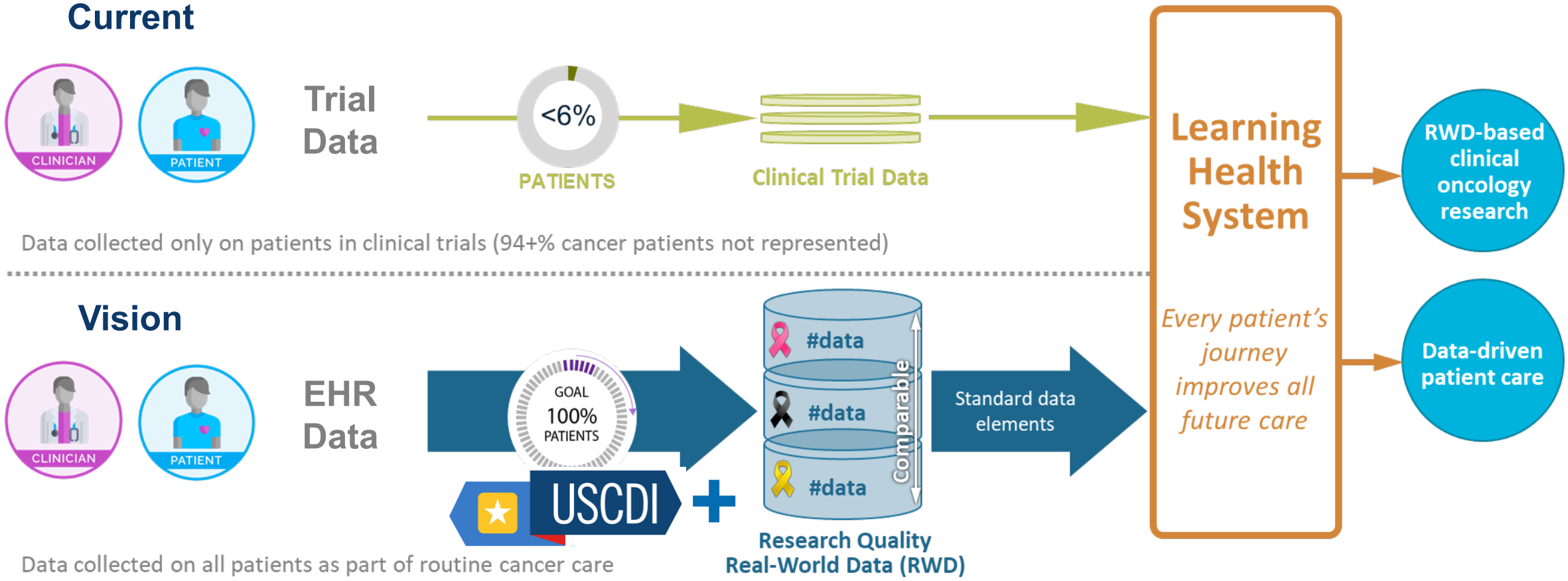
1 Source: CMS 2 Source: CDC

FDA is working to accelerate evidence generation in the late phase and post market setting



Can we streamline the use of EHR data for clinical trials and

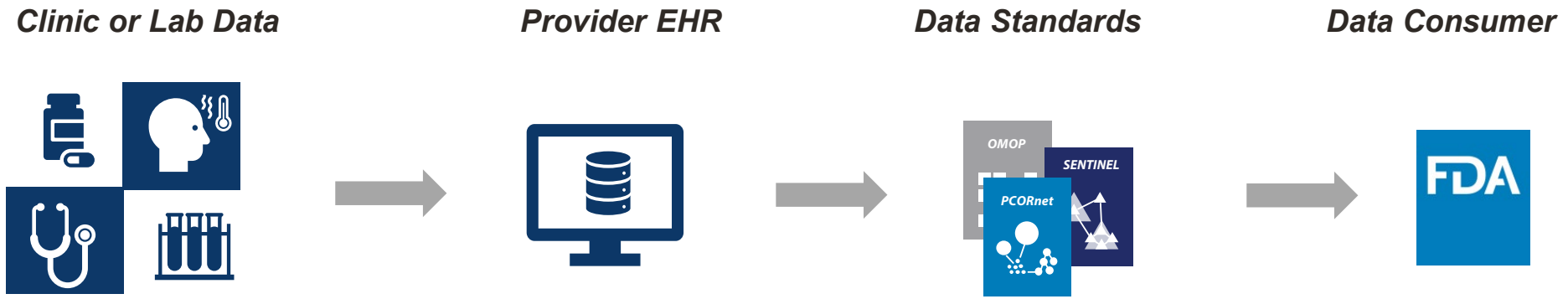
EHR data can help us learn from all patients



DRAFT

Adapted from ICAREdata® Project

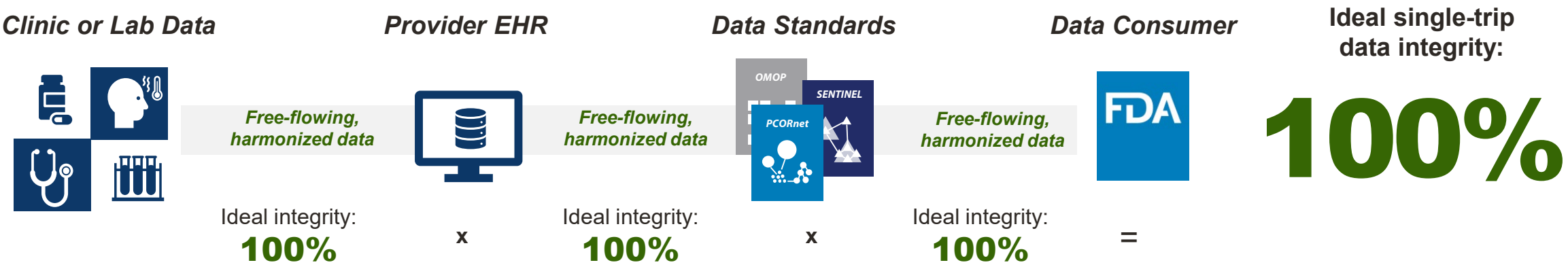
The journey of a data element



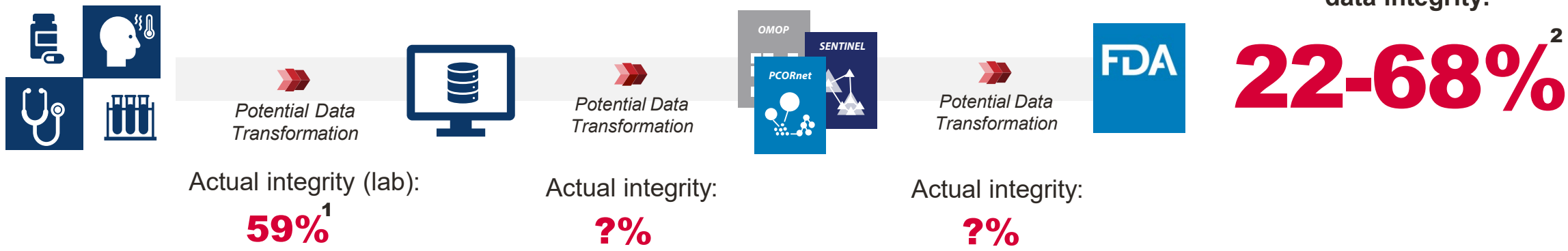
DRAFT

Interoperability challenges span the entire journey of data elements

Expectations:

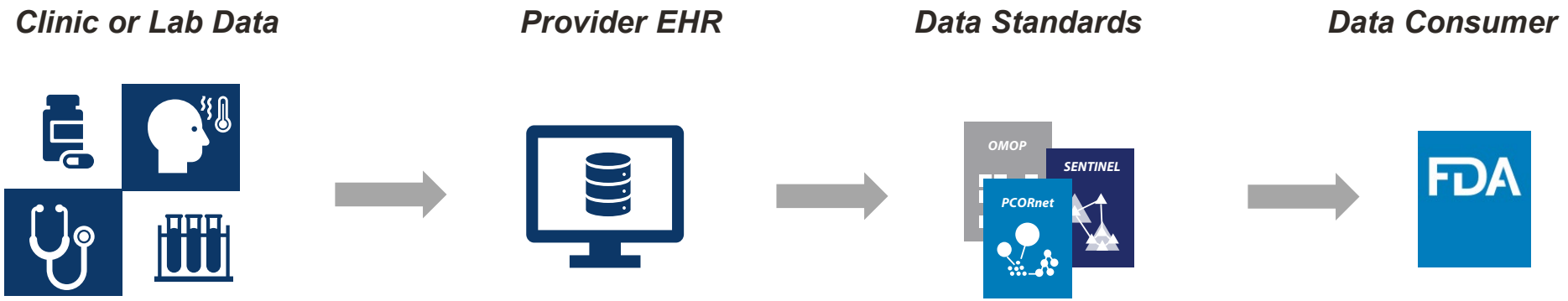


Reality:



Source: FDA/CDRH/SHIELD. *Citations: [1]Encoding laboratory testing data: case studies of the national implementation of HHS requirements and related standards in five laboratories. [2]Quantitating and assessing interoperability between electronic health records*

HL7 FHIR and USCDI+ can help solve “last mile” interoperability challenges



First Mile

Last Mile

HL7 FHIR USCDI +

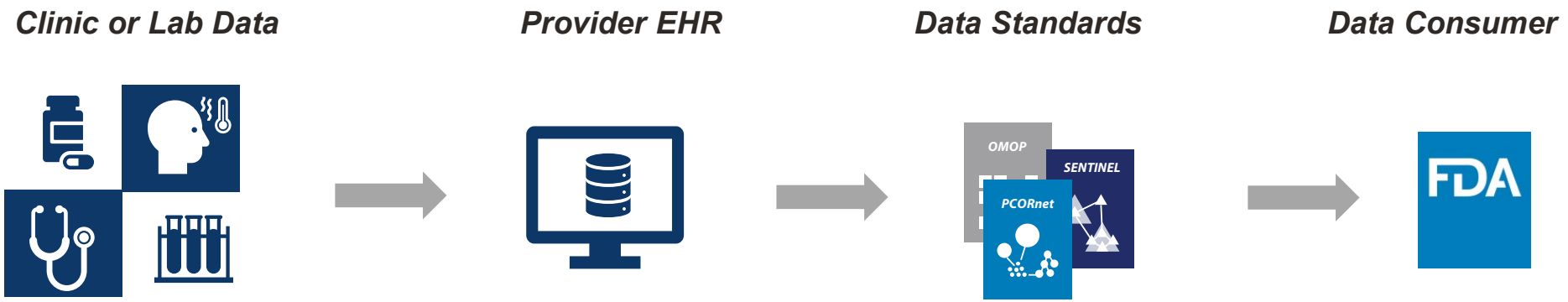
DRAFT

HL7 FHIR and USCDI+ can help solve “last mile” interoperability challenges

DRAFT

Source: Mitra Rocca, FDA/CDER/OTS

USCDI+ can help address first mile interoperability when combined with subject specific data models



First Mile

Last Mile

USCDI + mCODE™

DRAFT

Outstanding challenges for USCDI+

Scale

Scale up and streamline process of creating USCDI+ beyond oncology

Provenance

Verification of source data, maintenance of audit trail, and quality control are key for regulatory uses of data

Adaptability

Create pathways for rapid adaptation of USCDI+ standards

Incentives

Leverage incentives from federal partners to drive widespread adoption

Clinical Trial Matching Pilot

Umit Topaloglu PhD FAMIA

Chief, Clinical and Translational Research Informatics Branch

Informatics and Data Science Program, CBIIT

NIH NATIONAL CANCER INSTITUTE

1/17/2024

Agenda

1. *RWD Program and its goals*
2. *CTRP and Structured Eligibility*
3. *Clinical Trial Matching*
4. *Collaboration and Engagement Opportunities*

RWD Program and its Goals

Why EMR Data Quality is an Interoperability Problem

Need both Syntactic and Semantic Interoperability

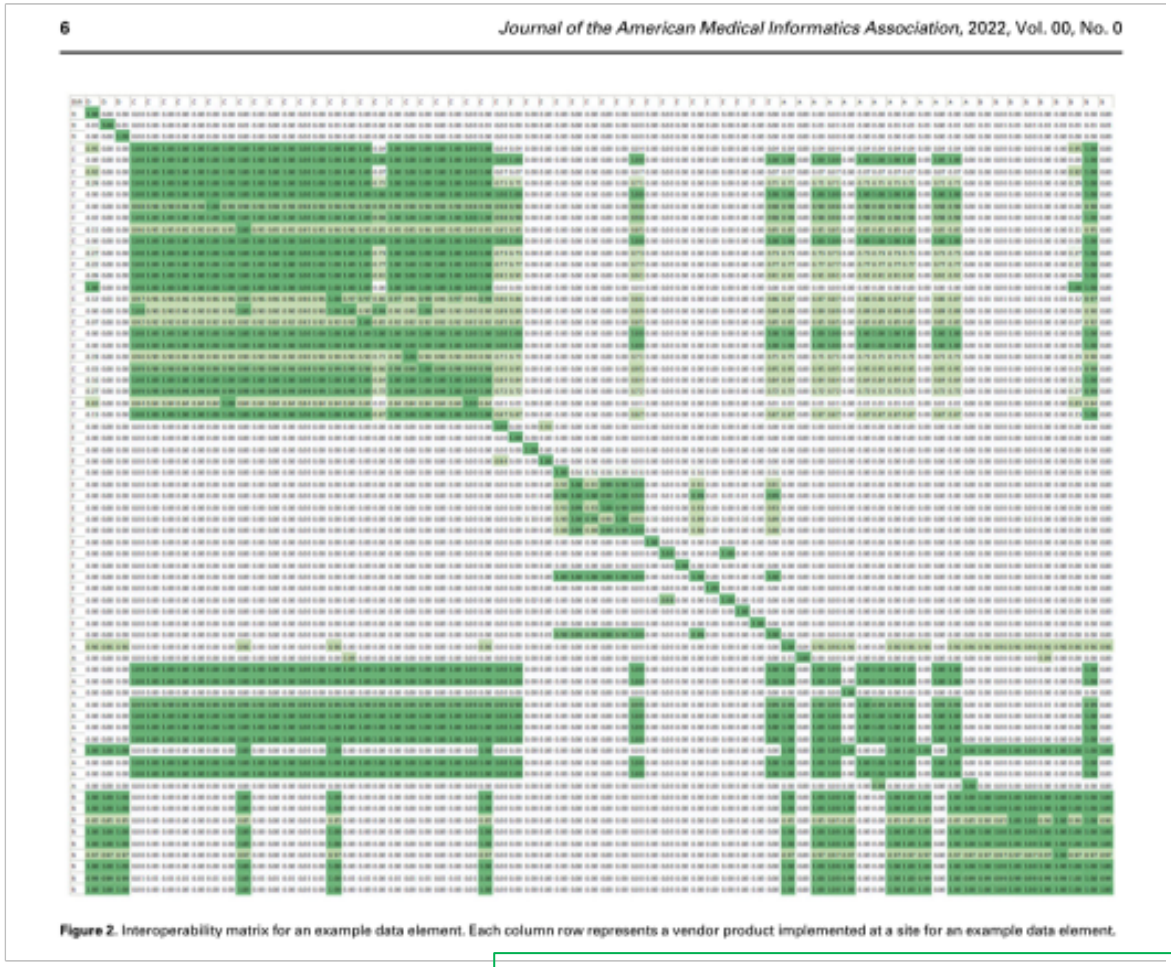
Syntactic = Clustering within EMRs = Structure of where variable is located

Semantic = Clustering of match/choice of LOINC, RxNorm representation (or not!)

Overall **Intra**-vendor interoperability score was 0.68, compared to a mean of 0.22 for **inter**-vendor interoperability (weighted by number systems)

"In the most favorable case'...approximately two-thirds of data types will be "understood" by a receiving site."*

CBIIT RWD Program working to understand relative impacts of syntactic/FHIR/export model vs. Semantic infrastructure/models



Bernstam et al, 2022

This is CancerLinQ data from 68 clinical sites representing all major EMRs

The scientific goals of the NCI's Real-World Data Program

Goal 1 Identify key components required for embedding RWD into the institutional ecosystem (i.e., infrastructure, data, people, inst. support)

Goal 2 Develop NCI Framework for Assessing RWD Data Quality and Identify Gaps for Development and Innovation

Goal 3 Define key requirements for building robust informatics tools to support high-quality RWD

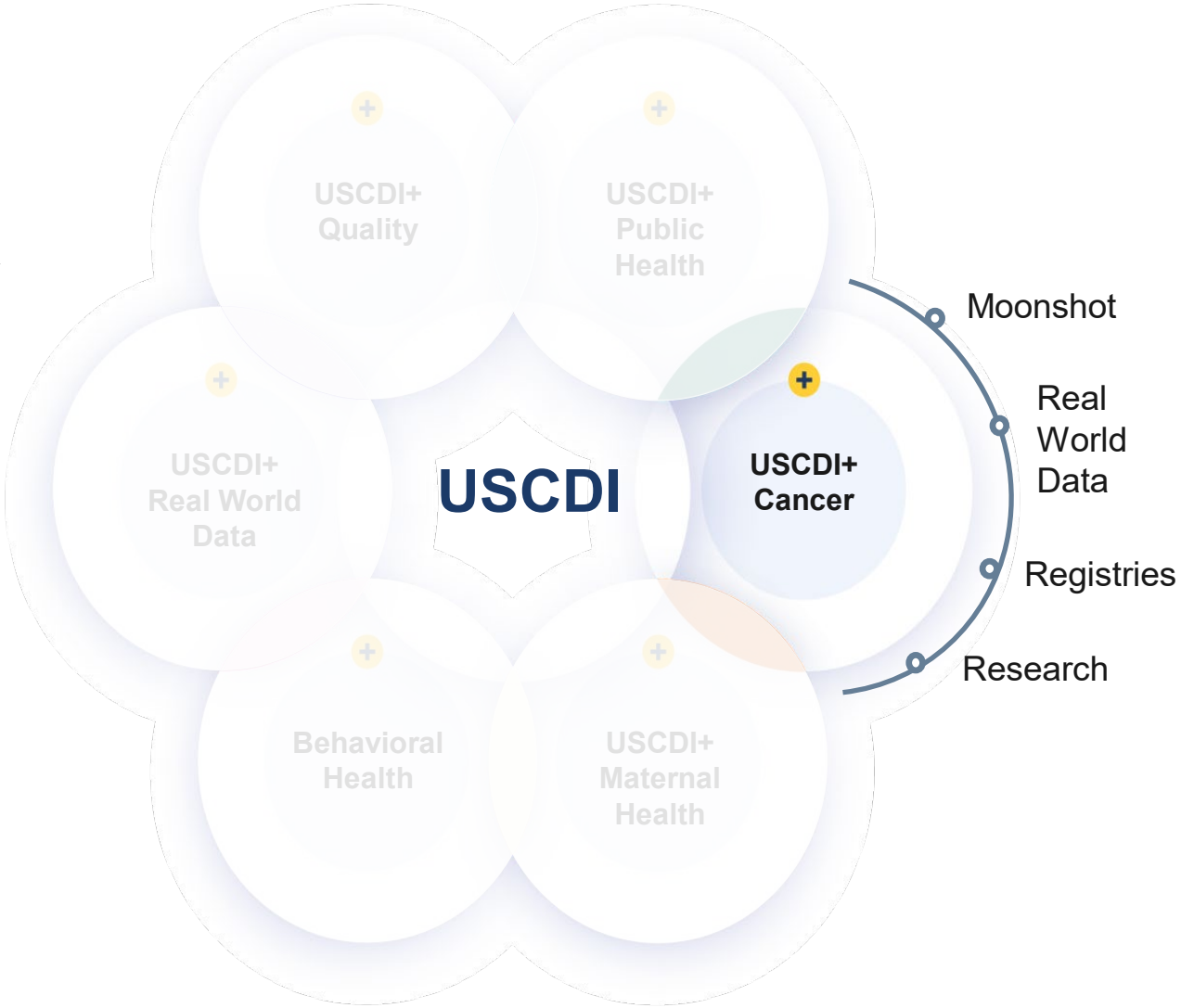
Goal 4 Engage, federal, academic, industry, and patient stakeholders to strengthen and support the RWD in the NCI for high-quality RWD across the Oncology Research Ecosystem

RWD, LLM and Federated Learning P30 NCI Admin Supplements

1. Scholar: Taxiarchis Botsis - Johns Hopkins University
2. Scholar: Karthik Natarajan - Columbia
3. LLMs: Travis Zack, Madhumita Sushil – UCSF
4. LLMs: Kushal Dey, Pulkit Jain – MSKCC
5. LLMs: Thanh Thieu – Moffitt
6. FLAMMAI: Adam Resnick – Upenn
7. FLAMMAI: Shannon McWeeney - OHSU

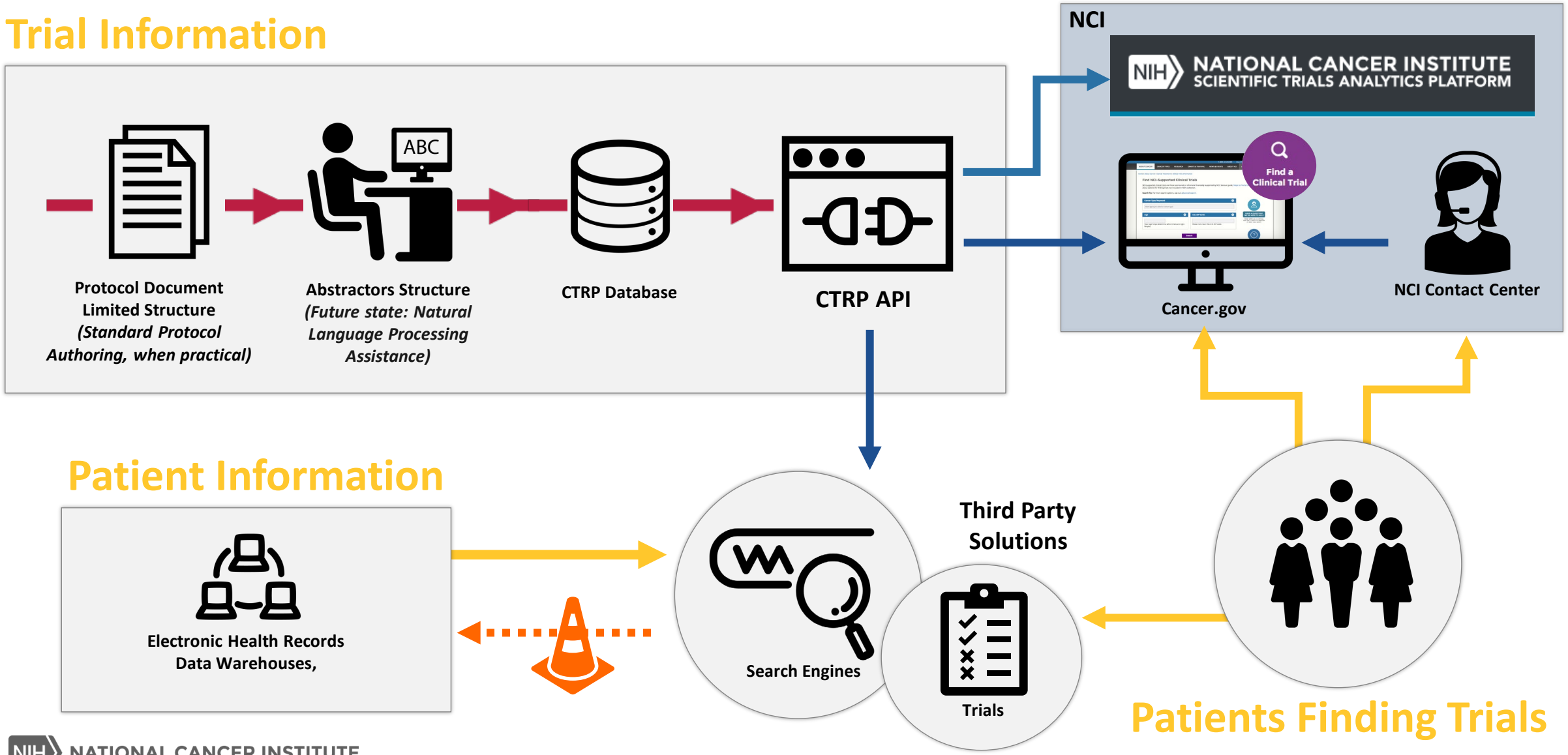
USCDI+: Cancer

- Capture the Cancer-Specific data needs that fall outside the scope of USCDI
- Harmonize Cancer data elements into a common data element list
- Support NCI and Cancer Moonshot real-world data use cases for:
  1. Reduced time for Clinical Trial Recruitment / Clinical Trial Matching
  2. Timely identification and capture of Immune-related Adverse Events for the upcoming Immunotherapy studies
  3. Increase completeness and improved recency of the cancer registry data elements collection

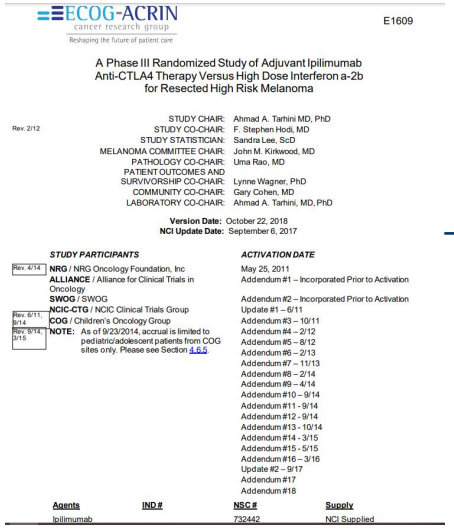


CTRP and Structured Eligibility Criteria

NCI Cancer Clinical Trials Search – Multiple Interrelated Parts



Clinical Trial Information Flow



ECOG-ACRIN cancer research group
Reshaping the future of patient care

E1609

A Phase III Randomized Study of Adjuvant Ipilimumab Anti-CTLA4 Therapy Versus High Dose Interferon a-2b for Resected High Risk Melanoma

Rev. 2/12

STUDY CHAIR: Ahmad A. Tarhini MD, PhD
STUDY CO-CHAIR: F. Stephen Hodi, MD
STUDY STATISTICIAN: Sandra Lee, ScD
MELANOMA COMMITTEE CHAIR: John M. Kirkwood, MD
PATHOLOGY CO-CHAIR: Uma Rao, MD
PATIENT OUTCOMES AND SURVIVORSHIP CO-CHAIR: Lynne Wagner, PhD
COMMUNITY CO-CHAIR: Gary Cohen, MD
LABORATORY CO-CHAIR: Ahmad A. Tarhini, MD, PhD

Version Date: October 22, 2018
NCI Update Date: September 6, 2017

STUDY PARTICIPANTS	ACTIVATION DATE
Rev. 4/14 **NRG** / NRG Oncology Foundation, Inc	May 25, 2011
ALLIANCE / Alliance for Clinical Trials in Oncology	Addendum #1 – Incorporated Prior to Activation
SWOG / SWOG	Addendum #2 – Incorporated Prior to Activation
Rev. 6/11, 9/14 **NCIC-CTG** / NCIC Clinical Trials Group	Update #1 – 6/11
COG / Children's Oncology Group	Addendum #3 – 10/11
Rev. 9/14, 3/15 **NOTE:** As of 9/23/2014, accrual is limited to pediatric/adolescent patients from COG sites only. Please see Section 4.6.5.	Addendum #4 – 2/12
	Addendum #5 – 8/12
	Addendum #6 – 2/13
	Addendum #7 – 11/13
	Addendum #8 – 2/14
	Addendum #9 – 4/14
	Addendum #10 – 9/14
	Addendum #11 - 9/14
	Addendum #12 - 9/14
	Addendum #13 - 10/14
	Addendum #14 - 3/15
	Addendum #15 - 5/15
	Addendum #16 – 3/16
	Update #2 – 9/17
	Addendum #17
	Addendum #18

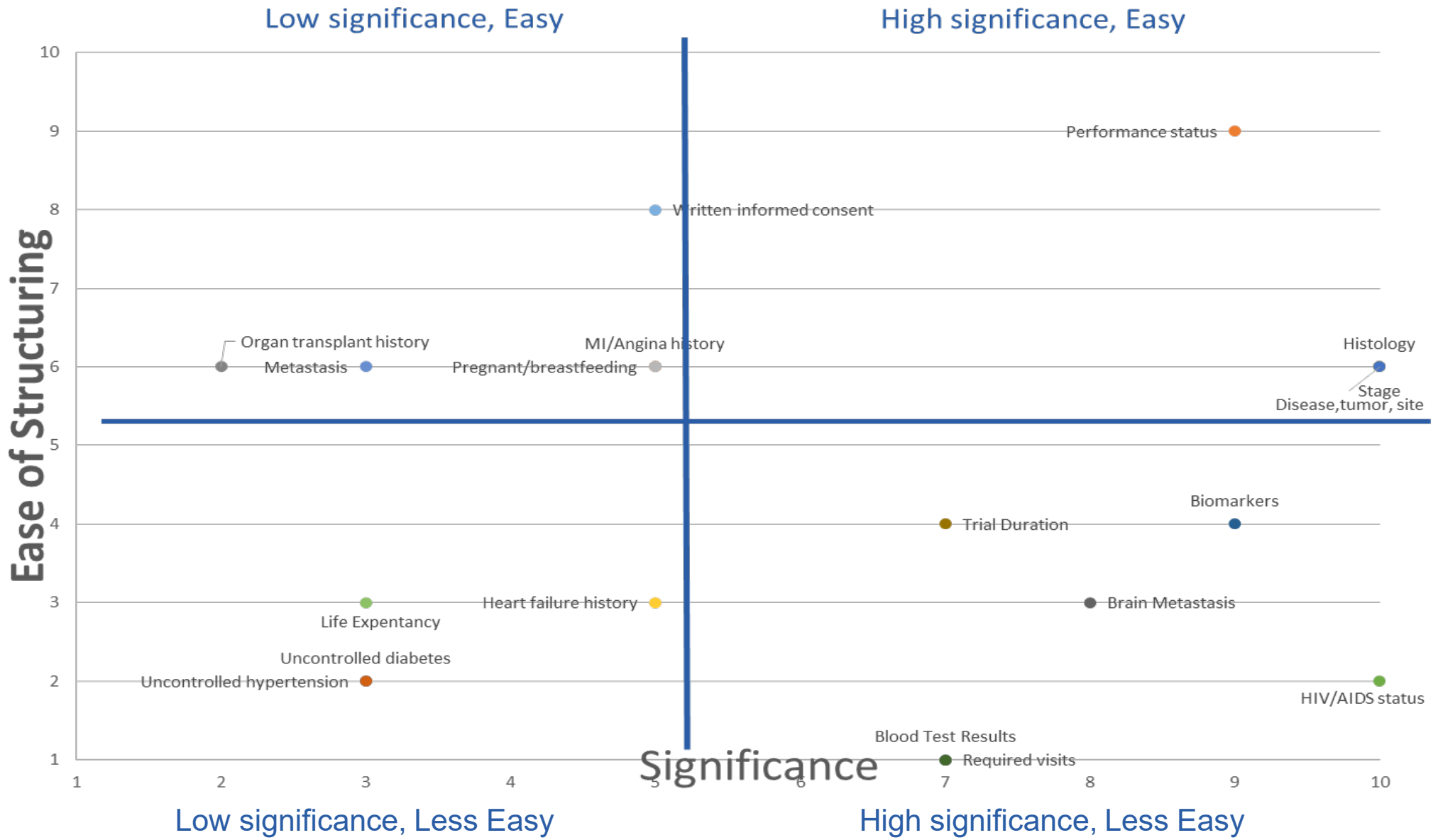
Agents	IND #	NSC #	Supply
Ipilimumab		732442	NCI Supplied

Clinical Trials Reporting Office CTRO

CTRP Staff abstract and add structure and coding terms, e.g., disease and intervention

Free Text in Protocol	Standardized Text	Structured and Coded
NRG-GY028, NCT05538897: HIV-infected patients on effective anti-retroviral therapy with undetectable viral load within 6 months of registration are eligible for this trial	HIV positive on antiretroviral therapy and undetectable viral load included	(C15175 = NO) OR ((C15175=YES) AND (C94631 = YES) AND (C51952) AND (C111568)
MSKCC - NCT06017258: Positive serologic test results for HIV	HIV positive excluded	(C15175 =NO)
CCR - NCT05960773: History of human immunodeficiency virus (HIV) infection or acquired immunodeficiency syndrome (AIDS)-related illness.	HIV positive or Acquired Immunodeficiency Syndrome excluded	(C15175 =NO) OR (C2851 = NO)
Yale - NCT05313243: Exclusion: Has a known history of Human Immunodeficiency Virus (HIV).	HIV Positive excluded	(C15175 = NO)

Ease



CCCT PMF analysis - 2016 - unpublished

Structured Eligibility Criteria and LLM Pilot

- Evaluate the use of LLM technology to facilitate the abstraction of clinical trial eligibility criteria (e.g. disease, prior therapy, and biomarkers)
- ChatGPT (GPT-3.5), GPT-4, Llama2, Mistrial-7B, Clinical Longformer, and Clinical Bigbird are being tested

3.1 Randomization Eligibility Criteria
3.1.15 Patients must be adequately recovered from surgery at the time of randomization.
3.1.6 Positive for translocation or inversion events involving the ALK gene locus (e.g. resulting in EML4-ALK fusion…

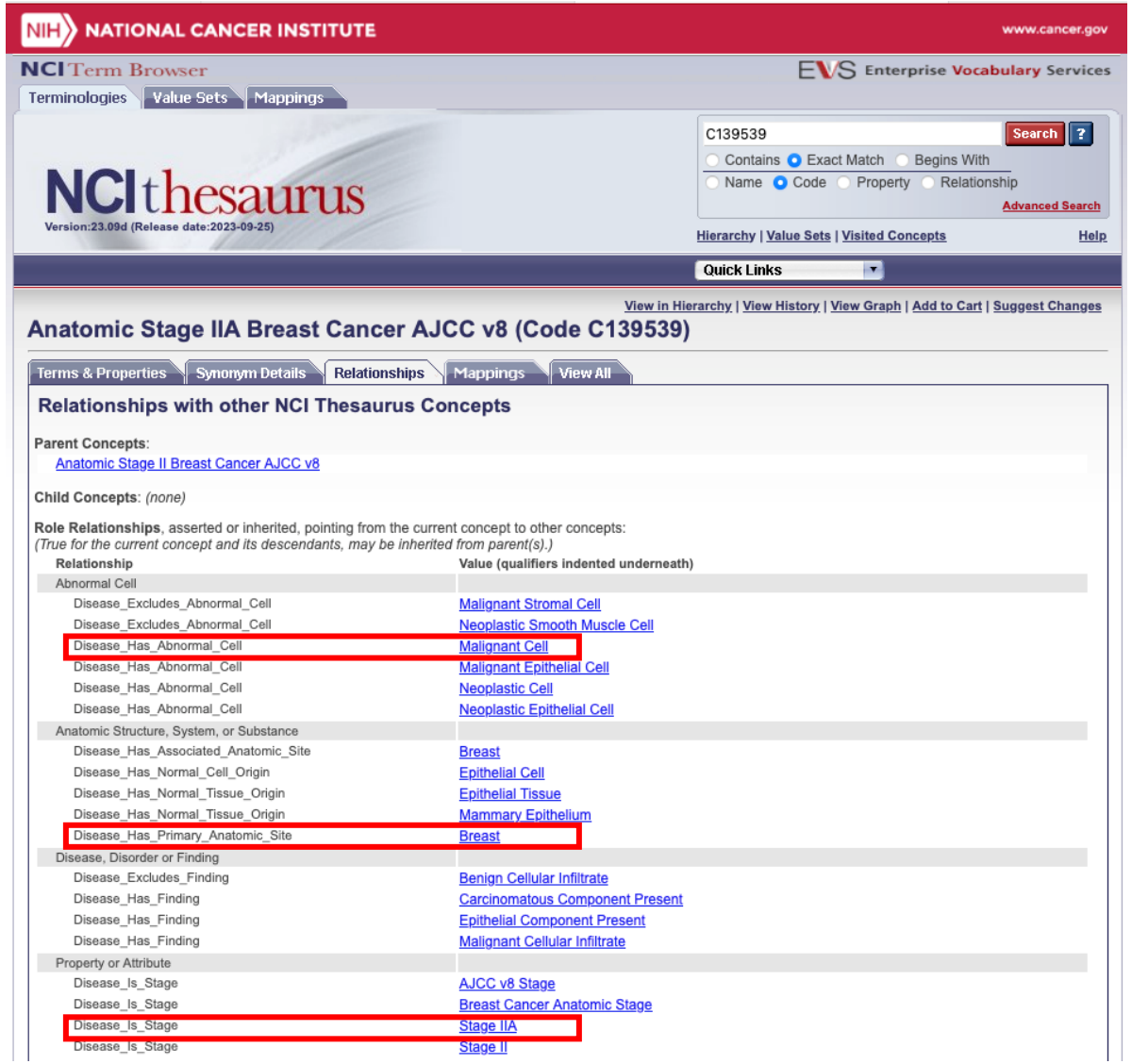
Trial ID	Dis-ease	Inclusion Criteria		Line # in the Protocol
		Prior Therapies	Bio-markers	
Trial 1	Breast		HER2 neg ER pos	19
Trial 2	Lung NSCLC	Surgery	EML4-ALK fusion	34
Trial 3	Ovary	Chemo	ER neg PR neg	6

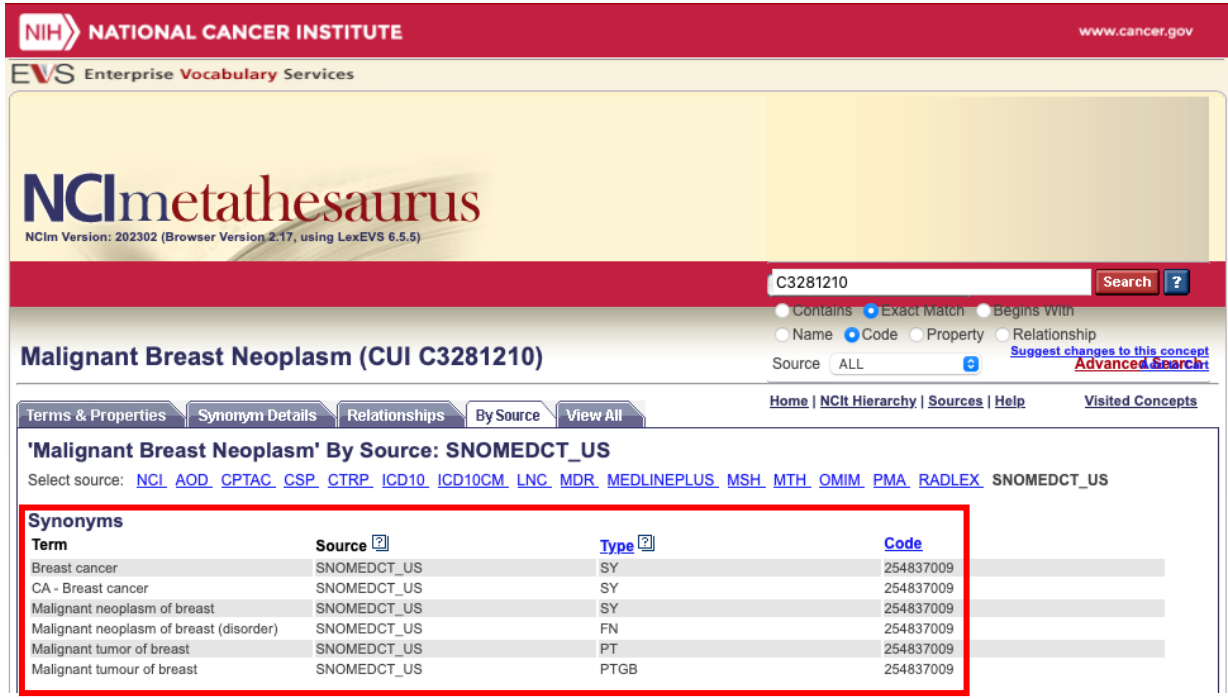
Clinical Trial Matching

Clinical Trial Matching

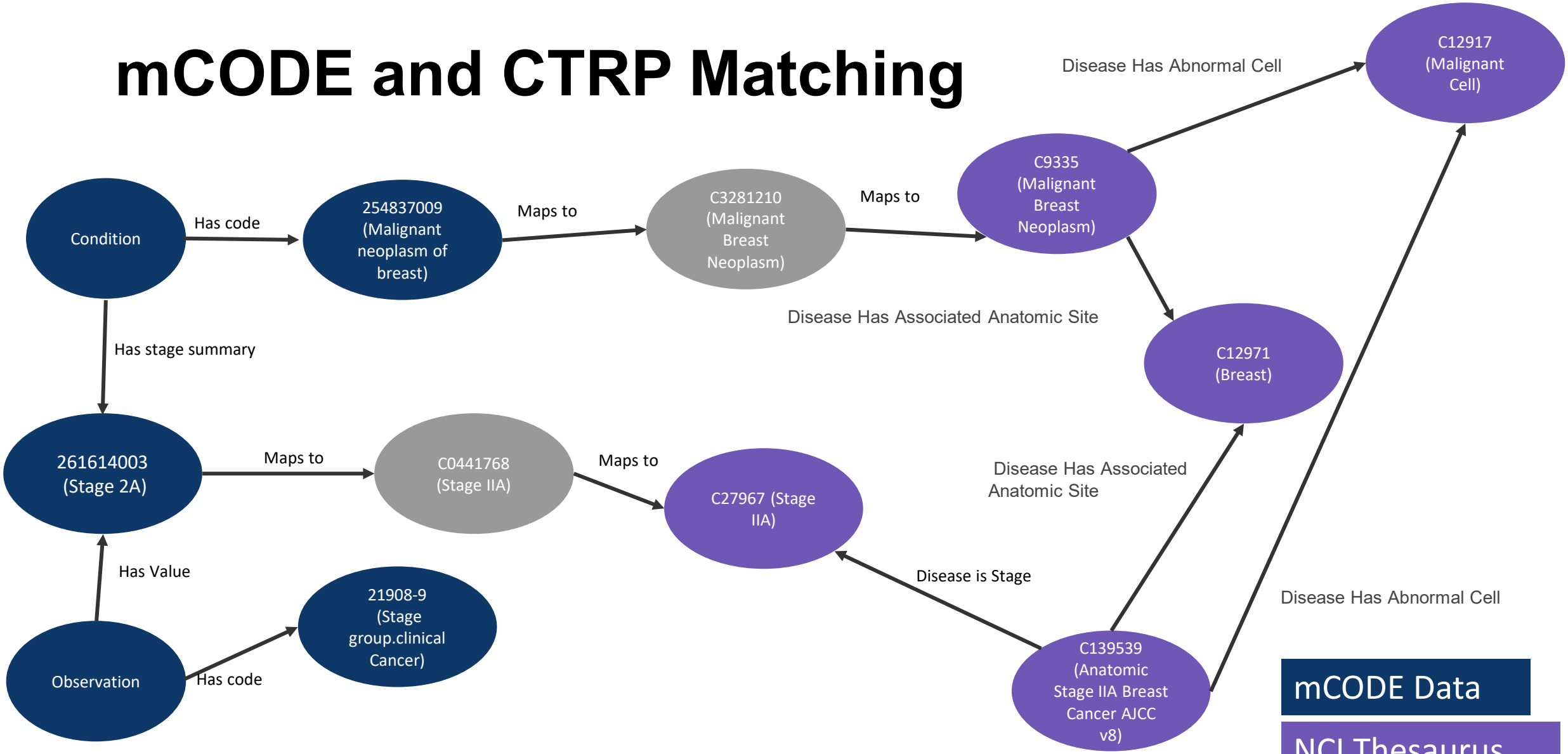
- Clinical Trial Matching expensive & error prone
  - Select solutions on extracting structured eligibility criteria
    - Criteria2query https://github.com/OHDSI/Criteria2Query
    - The Leaf Clinical Trials Corpus https://www.nature.com/articles/s41597-022-01521-0
    - MatchMiner/CTML https://github.com/dfci/matchminer
    - Deep6 https://deep6.ai/ , GenomOncology https://www.genomoncology.com/
  - Accessing computable clinical data
    - FHIR- limited data, terminology mapping
    - Unstructured data
  - Clinical Trial Recruitment workflows

CTRP Annotations with NCIt and NCI Metathesaurus





mCODE and CTRP Matching



mCODE Data

NCI Thesaurus

NCI Metathesaurus

Uses mCODE Primary Cancer Condition and group. Clinical mCODE TNM Clinical Stage Group profiles

Collaboration and Engagement Opportunities

Collaboration and Engagement Opportunities

- Federal Government
  - ONC, FDA, CDC, and CMS
  - HHS Data Strategy
  - WH Cancer Cabinet | Data and Innovation Task Force
- Non-Profit and Standards Organizations
  - CancerX
  - FHIR Accelerators: CodeX, Vulcan
- For-Profit and FNIH

Acknowledgments

CBIIT Informatics and Data Science Program

Jill Barnholtz-Sloan, PhD
Robinette Renner, PhD
Shannon Silkensen, PhD
Lyubov Remennik, MD, PhD
Gilberto Fragoso, PhD
Denise Warzel, MSc.
Brenda Duggan, RN
Goutham Reddy, MD
Mel Nisonger, BS
Anne Marie Meyer, PhD
Hannes Neidner, MD

- *Data Ecosystem Branch*
- *Computational Biology and Bioinformatics Branch*

CBIIT Leadership Team

Tony Kerlavage, PhD
Jill Barnholtz-Sloan, PhD
Jeff Shilling, CIO
Jaime Auvil-Guidry, PhD
Marcos Munozramos

CCCT

Sheila Prindiville, MD
Gisele Sarosy, MD
Polly Dhond, PhD
CTRP Team

All partners throughout NCI/NIH and externally in CBIIT programs

CBIIT Contractors

Thanks for Listening

Umit.Topaloglu@nih.gov

NATIONAL CANCER INSTITUTE

www.cancer.gov **www.cancer.gov/espanol**

Cancer Prevention Works

Transforming Cancer Data Collection and Use: Aligning Cancer Health IT Standards for Use Across Research, Health Care, and Public Health

Cancer Registries

Joseph D. Rogers (Joe), MS

Lead Health Scientist (Informatics) / Team Lead
Informatics, Data Science, and Applications Team
Email: JRogers@cdc.gov Tel: 770-488-4701

National Program of Cancer Registries (NPCR)
Cancer Surveillance Branch (CSB)
Division of Cancer Prevention and Control (DCPC)
Centers for Disease Control and Prevention (CDC)

December 14, 2023

Cancer Data Driving Action | www.cdc.gov/cancer/npcr/

National Program of Cancer Registries (NPCR)

1.8 million	New cancer cases each year
200+ data items for each case	Cancer site and histology Patient demographics Stage at diagnosis First course of treatment

WHO is getting cancer (for instance, by race, age, or sex)?

WHAT types of cancer are increasing or decreasing?

WHERE will prevention efforts have the biggest impact?

WHEN are screening or prevention strategies working?

HOW far has the cancer spread, and are we catching cancer early?

What is a cancer registry?

What is the value of cancer registries?

Traditional Public Health Focus

Monitor
Cancer incidence and trends over time

Evaluate
Cancer patterns in populations and to identify high-risk groups

Guide
Planning and evaluation of cancer control programs

Inform
Priorities for allocating health resources

Research
Cancer cause and prevention strategies to see which work well

Epidemiological and Clinical Focus

Epidemiological Studies
Cancer data for special studies

Precision Medicine
Targeted therapies based on genomics/biomarkers

Survivorship
Measuring what works best during and after treatment

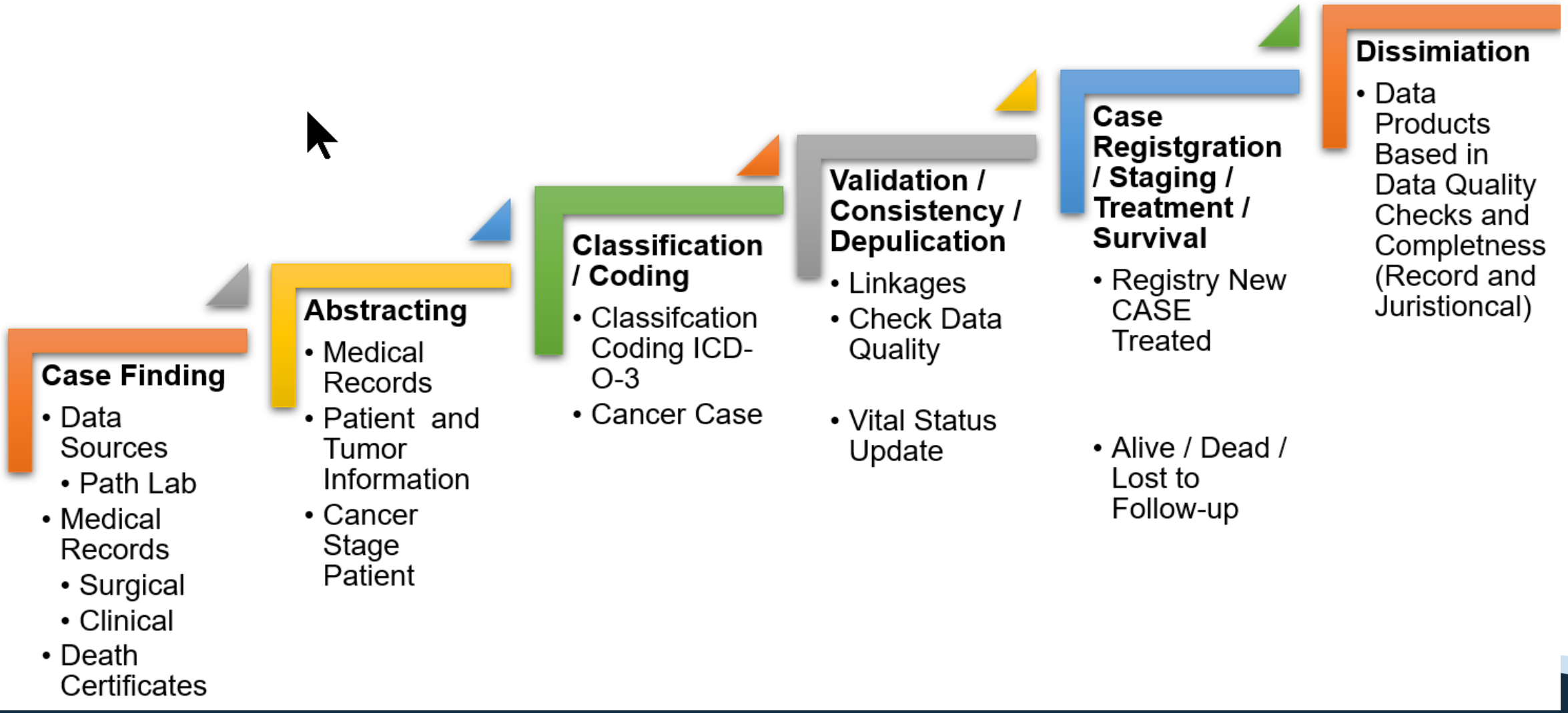
Treatment and Outcomes Assessment
Data for evaluating what works best

Linkages and Clinical Trails
Real-time and longitudinal data.

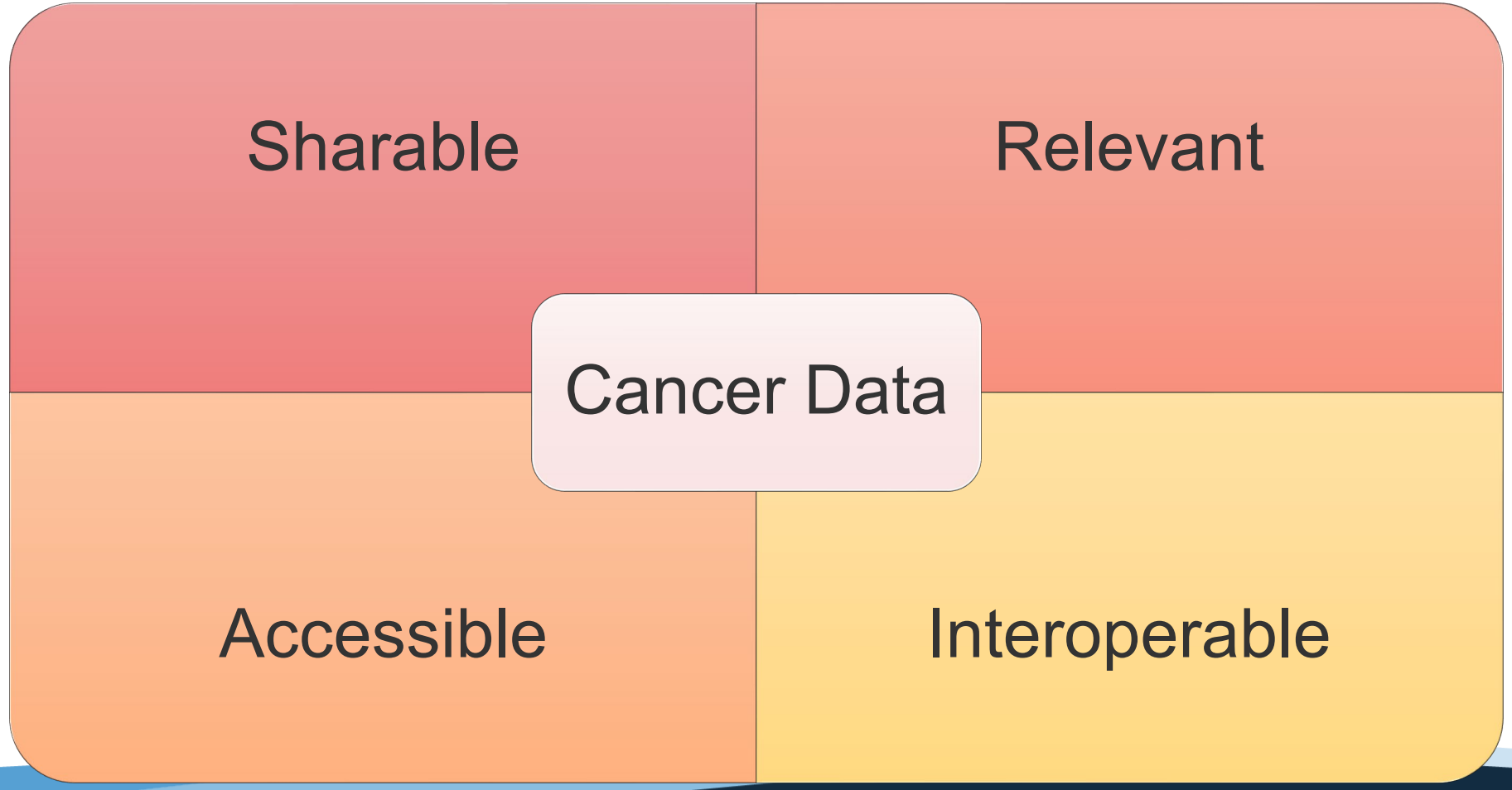
What are the Data Sources for Cancer Registries? (1 of 2)

- Electronic Health Records
- Disease Indices
- Surgery Schedules
- Admission & Discharge Documents
- Pathology Reports
- Cytology Reports
- Cancer-Related Biomarkers Tests
- Nuclear Medicine & Radiation Oncology Logs
- Medical Oncology Logs & Autopsy Documents
- Mortality data: State and National Database
- Demographic data: Master Patient Index, DMV, Voting Records, and so
- Claims Data: State and Private Databases
- Drug Prescription Databases

What are the Data Sources for Cancer Registries? (2 of 2)



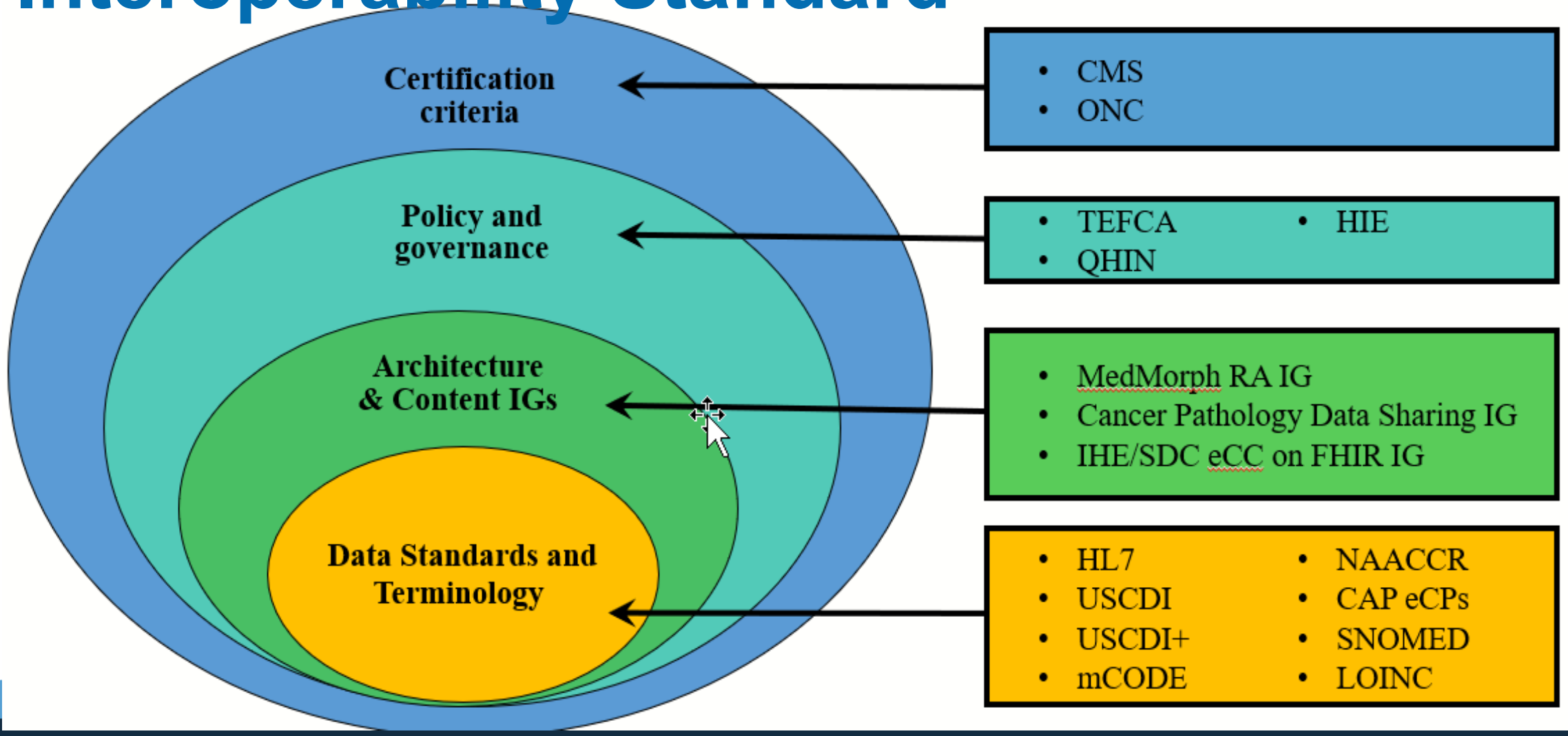
What is the Ultimate Goal of Cancer Registries?



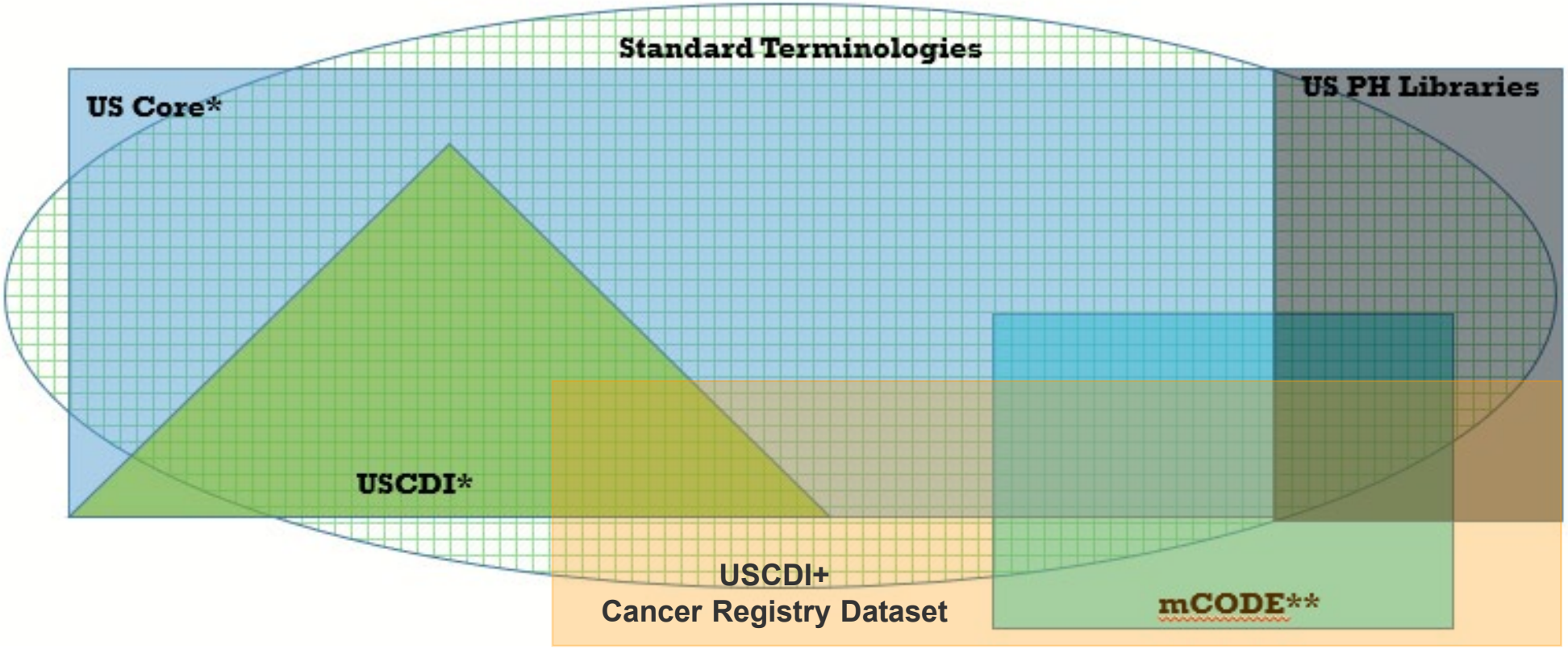
Challenges and Opportunities

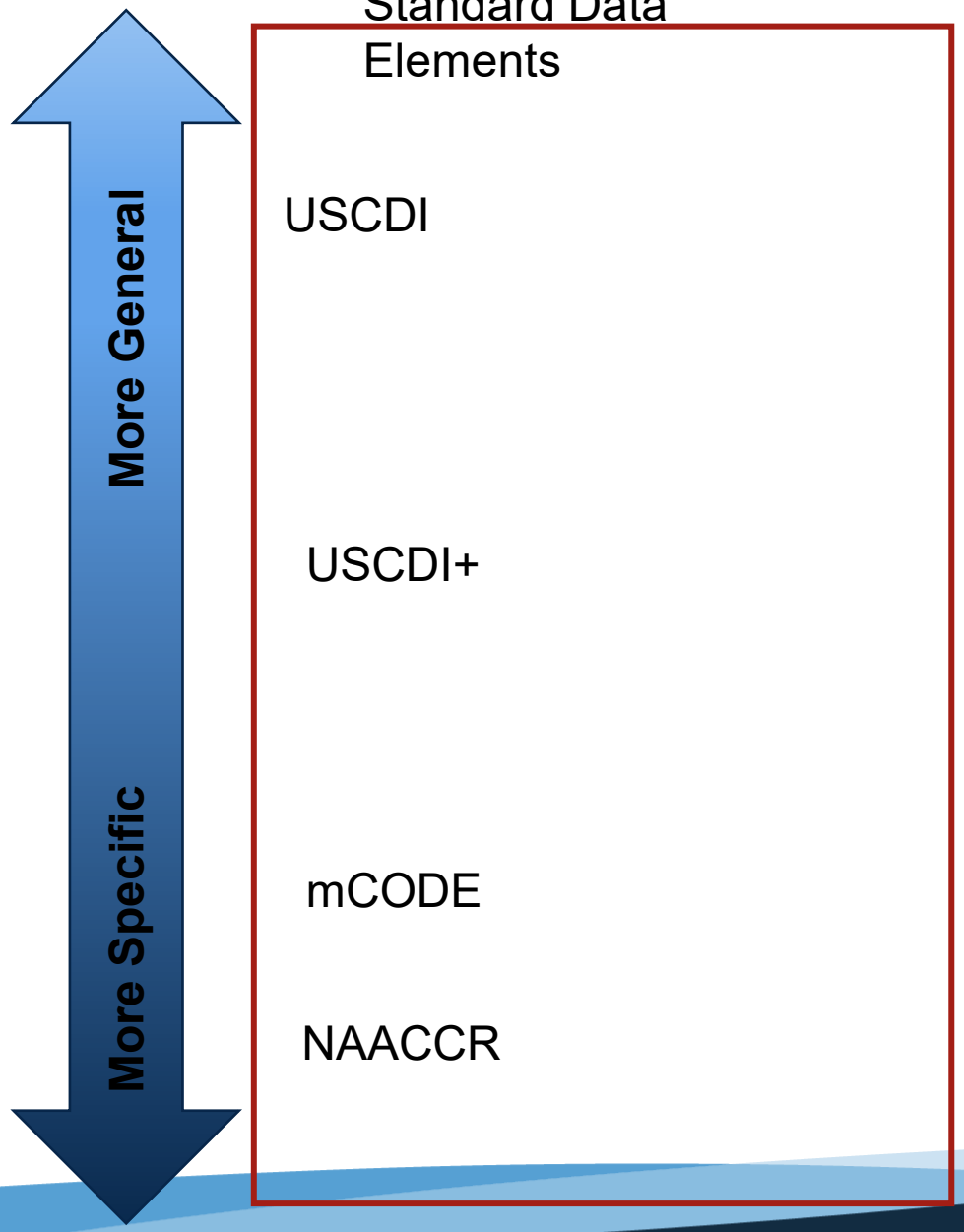
Timeliness	Realtime Reporting
	Direct Reporting from EHRs and Pathology Labs
High Quality	Require Standardized Edits From Data Collection to Reporting
	Crowdsourcing Curation
Case Finding / Identification / Auto Coding / Completeness	Natural Language Processing (NLP) Case Ascertainment and Autocoding
	Crowdsourcing Accuracy and Precision
Cost	Investment in Uniform National Platforms for Reporting and Processing of Data
	Automation Across the Data Pipeline (i.e., auto consolidation)
	Crowdsourcing Quality Control
Security and Privacy	Federated Cancer Surveillance Cloud Computing Platform
	Invest in Data Governance Standards and Agreements
	FedRamp High/Moderate NIST Controls
Data that is Sharable, Accessible, Relevant, and Interoperable	Invest in Data Governance Standards and Agreements
	Invest in Interoperability Standards: FHIR IGs for EHRs and Pathology Reporting
	Invest in ONC Interoperability Standards and Initiatives for Health IT: USCDI, +, and TEFCA

CDC/NPCR Investment in National Interoperability Standard



Health IT Interoperability Data Standards





Standard Data Elements

USCDI

USCDI+

mCODE

NAACCR

Implementation Guides

Foundational Content

US Core
US Public Health Profile Library

Architecture

MedMorph Reference Architecture

Domain-specfic

mCODE

Central Cancer Registry Reporting
Cancer Pathology Data Sharing

IHE SDC on FHIR

Comparison of USCDI and NAACCR Data Elements

Code System Information

Code System OID 2.16.840.1.113883.6.238

Code System Name Race & Ethnicity - CDC

Code System Code PH_RaceAndEthnicity_CDC

Code System Concepts | Code System Details

966 Code System Concepts found 1 2 3 4 5 6

Concept Code	Concept Name	Preferred Concept Name	Code System	
1006-6	Abenaki	Abenaki	Race & Ethnicity - CDC	Details
1579-2	Absentee Shawnee	Absentee Shawnee	Race & Ethnicity - CDC	Details
1490-2	Acoma	Acoma	Race & Ethnicity - CDC	Details
2126-1	Afghanistani	Afghanistani	Race & Ethnicity - CDC	Details
2060-2	African	African	Race & Ethnicity - CDC	Details
2058-6	African American	African American	Race & Ethnicity - CDC	Details
1994-3	Agdaagux	Agdaagux	Race & Ethnicity - CDC	Details
1212-0	Agua Caliente	Agua Caliente	Race & Ethnicity - CDC	Details
1045-4	Agua Caliente Cahuilla	Agua Caliente Cahuilla	Race & Ethnicity - CDC	Details
1740-0	Ahtna	Ahtna	Race & Ethnicity - CDC	Details
1654-3	Ak-Chin	Ak-Chin	Race & Ethnicity - CDC	Details
1993-5	Akhiok	Akhiok	Race & Ethnicity - CDC	Details
1897-8	Akiachak	Akiachak	Race & Ethnicity - CDC	Details
1898-6	Akiak	Akiak	Race & Ethnicity - CDC	Details

01	White
02	Black or African American
03	American Indian or Alaska Native
04	Chinese
05	Japanese
06	Filipino
07	Native Hawaiian
08	Korean
10	Vietnamese
11	Laotian
12	Hmong
13	Cambodian
14	Thai
15	Asian Indian, NOS or Pakistani, NOS (code 09 prior to Version 12)
16	Asian Indian
17	Pakistani
20	Micronesian, NOS
21	Chamorro
22	Guamanian, NOS
25	Polynesian, NOS
26	Tahitian
27	Samoan
28	Tongan
30	Melanesian, NOS
31	Fiji Islander
32	Papua New Guinean
96	Other Asian, including Asian, NOS and Oriental, NOS
97	Pacific Islander, NOS
98	Some other race
99	Unknown by patient

Comparison of USCDI v2 and NAACCR v24

USCDI v2			NAACCR v24
Sex (Assigned at Birth)	**Sexual Orientation (SNOMED CT codes)**	**Gender Identify (SNOMED CT codes)**	**Sex**
M – Male	Lesbian, gay or homosexual - 38628009	Male - 446151000124109	1 – Male
F – Female	Straight or heterosexual - 20430005	Female - 446141000124107	2 – Female
UNK – Unknown	Bisexual - 42035005	Female-to-Male (FTM)/Transgender Male/Trans Man - 407377005	3 – Other (intersex, disorders of sexual development/DSD). The word hermaphrodite formerly classified under this code is an outdated term.
	Something else, please describe - nullFlavor OTH	Male-to-Female (MTF)/Transgender Female/Trans Woman - 407376001	4 – Transsexual, NOS
	Don't know - nullFlavor UNK	Genderqueer, neither exclusively male nor female - 446131000124102	5 – Transsexual, natal male
	Choose not to disclose - nullFlavor ASKU	Additional gender category or other, please specify - nullFlavor OTH	6 – Transsexual, natal female
		Choose not to disclose. nullFlavor ASKU	9 – Not stated/Unknown

Thank you!

U.S. Cancer Statistics now includes the latest data about new cancer cases through 2020.

Use, analyze, and visualize the data with our Data Visualizations tool at cdc.gov/cancer/dataviz/

Questions? Please contact us at uscsdata@cdc.gov.

Cancer Prevention Works

Division of Cancer Prevention and Control

Reliable. Trusted. Scientific.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Panel Discussion

Moderator & Panelists

Aneesh Chopra (Moderator)
- CareJourney

Ali B. Abbasi, MD
- US Food and Drug Administration

Su Chen, MD
- CodeX FHIR Accelerator

Jennifer Goldsack, MChem, MA, MBA, OLY
- Digital Medicine Society

Joseph D. Rogers (Joe), MS
- Centers for Disease Control and Prevention

Umit Topaloglu, PhD, FAMIA
- National Institute of Cancer

Kyle Cobb

Kyle.cobb@hhs.gov

HealthIT.gov