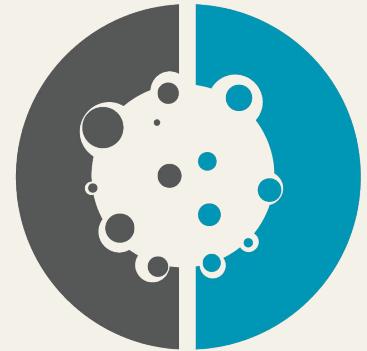


# Center for Cancer Data Harmonization: *Bringing harmony to cancer data across the CRDC*



CENTER *for*  
CANCER DATA  
HARMONIZATION

CCDH Update at CRDC All Hands  
October 26 & 27, 2020

These slides: [bit.ly/ccdh-crdc-f2f-2](https://bit.ly/ccdh-crdc-f2f-2)

<https://datascience.cancer.gov/data-commons/center-cancer-data-harmonization-ccdh>

# CCDH Workstreams



**Community Development**

Node, engagement, Website, help desk, etc.



**Data Model harmonization**

Create a consensus data model: CRDC-H



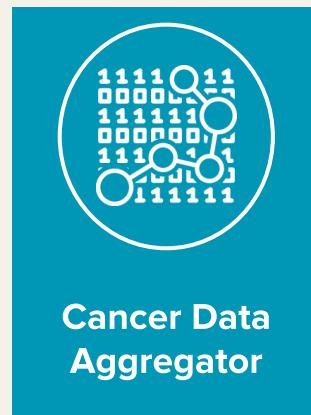
**Ontology & Terminology Ecosystem**

Mapping terminologies and value sets; Terminology development; Lookup services



**Tools & Data Quality**

Help nodes map, annotate, and transform data



**Cancer Data Aggregator**

Partnership to support:  
Query across nodes data + CRDC-H + terminologies

# Vision

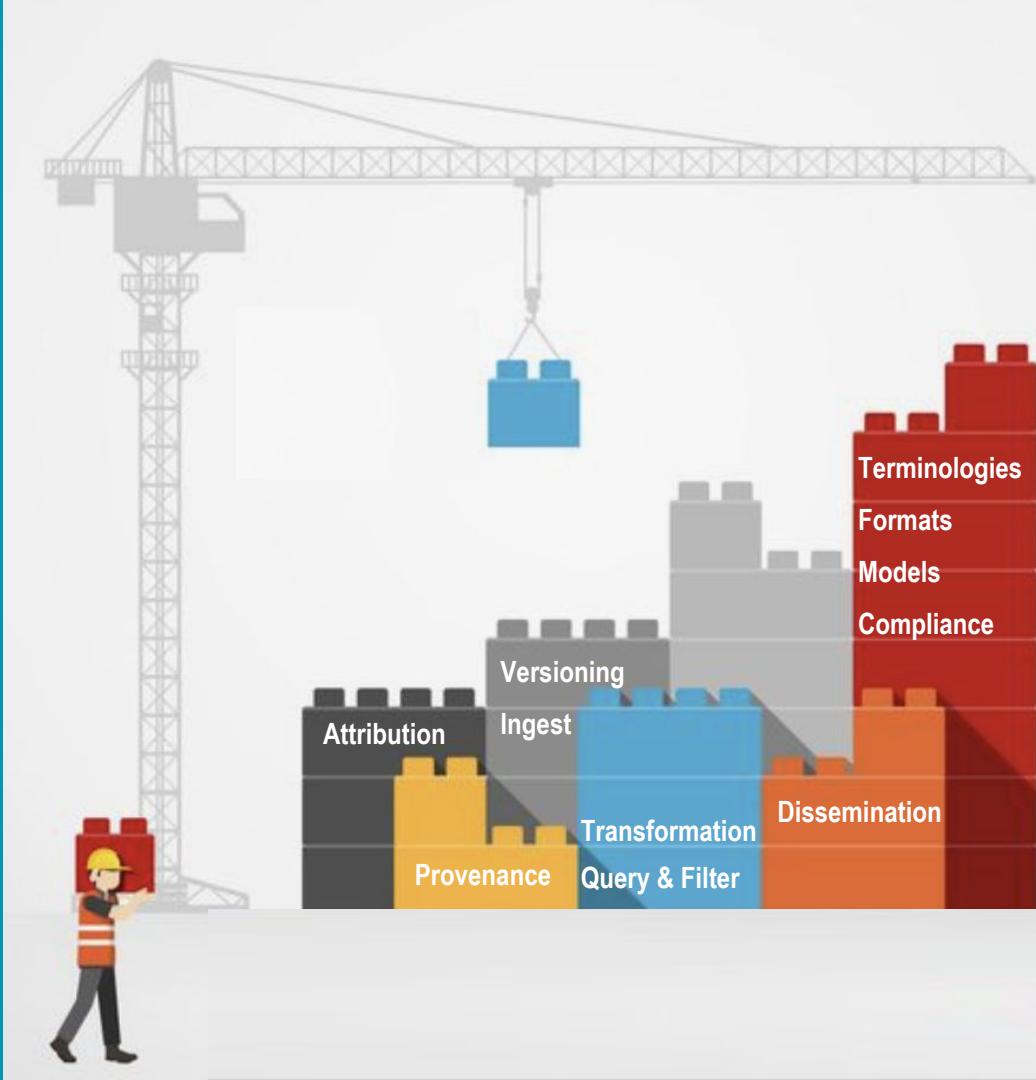
The CCDH will provide vision and guidance to nodes and other stakeholders throughout the CRDC:

- 1 : For each group, the CCDH will advise about and assist with best practices and methods for data ingestion, aggregation, harmonization, and sharing
- 2 : The CCDH can help the CRDC reach consensus regarding the standardization of workflows and data processes to ensure that the entire data lifecycle adheres to best practices and standard procedures
- 3 : Specification gathering is a key aspect of this work but will gradually be replaced with end-to-end recommendations about information architecture and data workflow management

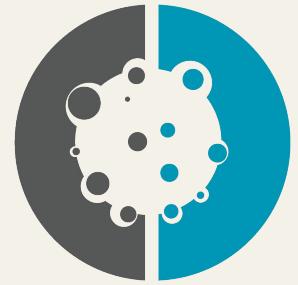
# Realizing an End-to-End CRDC Information Architecture

CCDH is coordinating a 3-month pan-CRDC requirements analysis to understand data flow and management within and across the nodes

Note that the Data and Metadata Submission Focus Group is one “end” and we are working together



# Engaging the CRDC Community



CENTER *for*  
CANCER DATA  
HARMONIZATION

# Concierge Services

Help Desk & Office Hours



Email: centerforcancerdh@gmail.com

Slack: cancerdhc.slack.com

Office hours:

Thursdays,  
11:30-12pm PT / 1:30-2pm CT / 2:30-3pm ET



Are there other  
ways you would  
like us to engage  
with you?

# Web Portal

## Web Portal *Under development*

### About

- Team organization
- Project timeline
- Presentations
  - CCDH Overview for HL7 BRR team; by Chris Chute ([slides](#))

### Standards and Tooling

Markdown file available [here](#)

### Support

### FAQs

► <https://github.com/cancerDHC/community-development/issues>

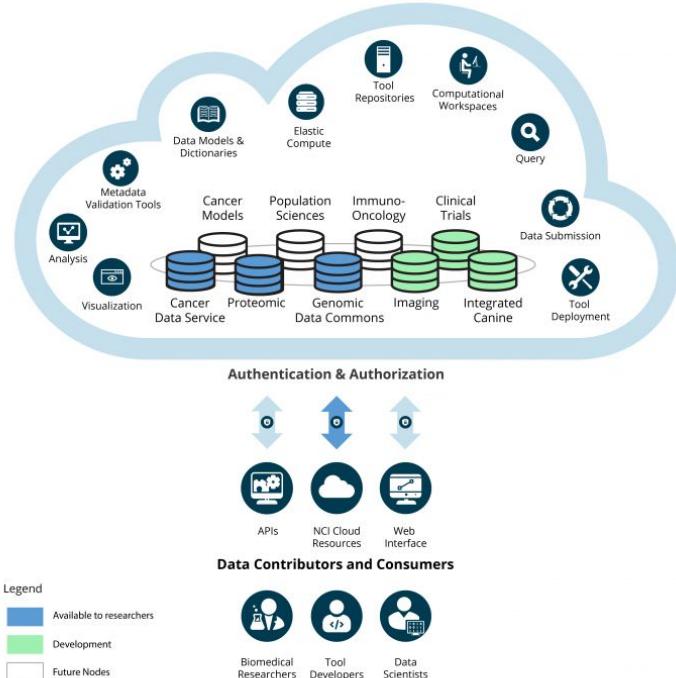
## About CCDH

The CCDH drives interoperability across NCI's Cancer Research Data Commons (CRDC) through improved data sharing capacity and data harmonization resources.

CCDH development is driven by a multi-disciplinary experts specializing in cancer, data standards, and technology at the following institutions:

- Oregon State University
- Oregon Health & Science University
- The University of Chicago
- Johns Hopkins University
- The University of North Carolina

## NCI Cancer Research Data Commons (CRDC)



# CCDH Newsletter



Gaurav Vaidya

Laura Christopherson

Quarterly Newsletter

Work products and highlights

Get to know CCDH

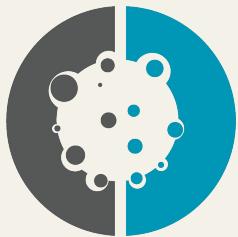
Recent presentations  
Slides and recordings

How to engage with us



Is there anything you  
would like to see in  
the newsletter?

# Development of a Harmonized CRDC-H Data Model



CENTER *for*  
CANCER DATA  
HARMONIZATION

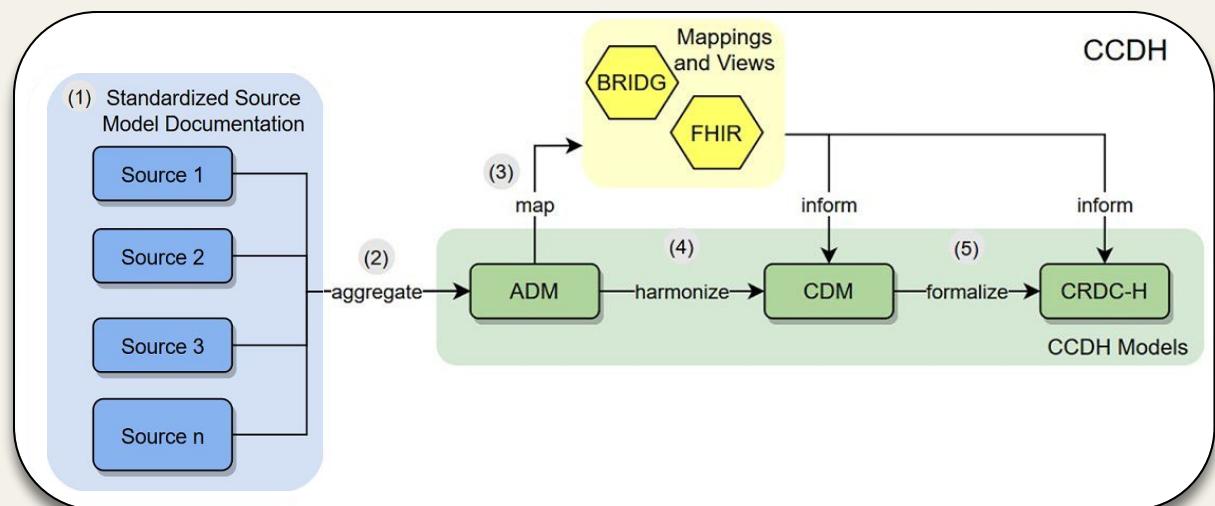
# Data Model Harmonization Workstream

## Goals:

- Provide a harmonized data model (CRDC-H) to the CRDC community
- Offer data modeling guidance and support to the CRDC community through office hours and consultation
- Support implementation of the CRDC-H through documentation, transform specifications, and services
- Evolve the CRDC-H to address changes in source models and community standards

# CCDH Model Harmonization Workflow

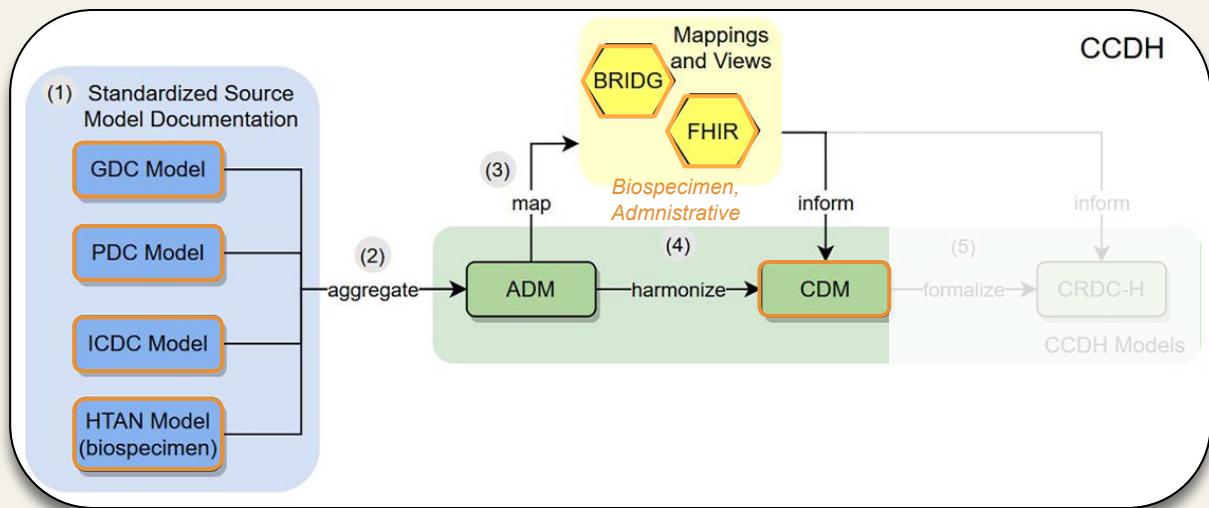
1. Standardize Source Data Model Documentation
2. Generate an Aggregated Data Model (ADM)
3. Map ADM elements to Community Standard Data Models
4. Refactor the ADM into a more deeply harmonized Conceptual Domain Model (CDM)
5. Refactor the CDM to a formal/implementable data model (CRDC-H)



An iterative process through which source model content is evaluated, aggregated, mapped, and refactored into a standards- aligned and harmonized data model.

# Phase I Harmonization Efforts (June 2020)

- Targeted four source models - GDC, PDC, ICDC, HTAN. (1)
- Harmonization focused on the Biospecimen and Administrative , and prioritized entities and attributes. (3, 4)
- Informed by mappings to BRIDG and FHIR models (3)
- Harmonized model represented informally at a conceptual level (CDM) (4)



Work products include standardized source documentation, ADM data dictionary, BRIDG/FHIR mappings, and the CDM model, and can be accessed [here](#) and [here](#).

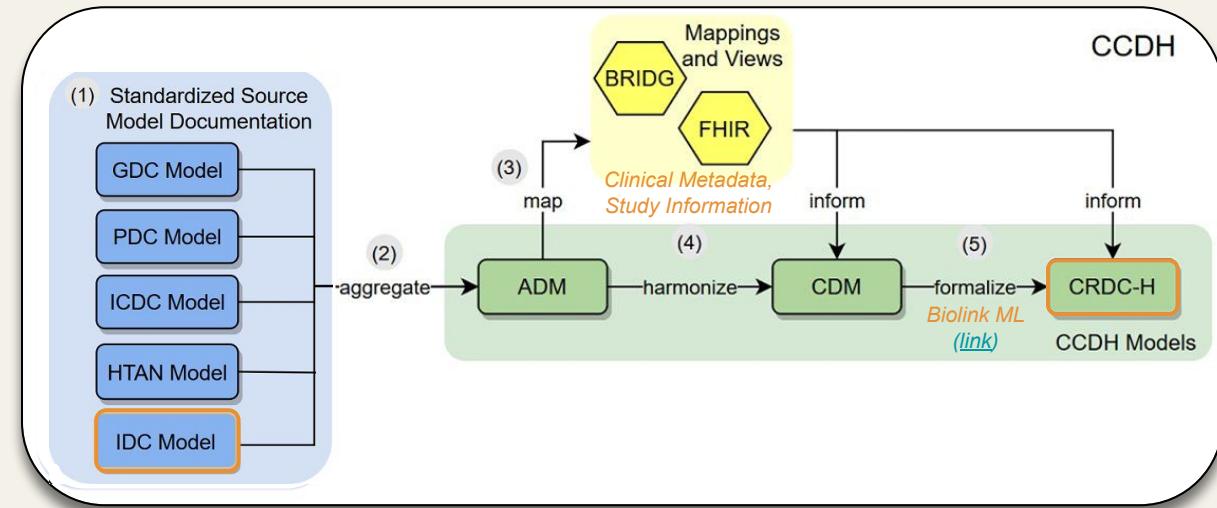
# Phase II Harmonization Efforts (ongoing)

## Iteration:

- Review and refinement of Phase I processes (1-4)
- Bring additional subdomains (**Clinical Data**, **Study Data**), and nodes (**IDC**) into the CDM. (1-4)

## Formalization:

- Evaluation of formal modeling frameworks and selection of **BiolinkML** (5)
- Development of tooling to support a version-controlled model development pipeline (1-5)



Focus on iterative improvement and expansion of Phase I harmonization efforts, and formalizing models to support implementation in CRDC systems

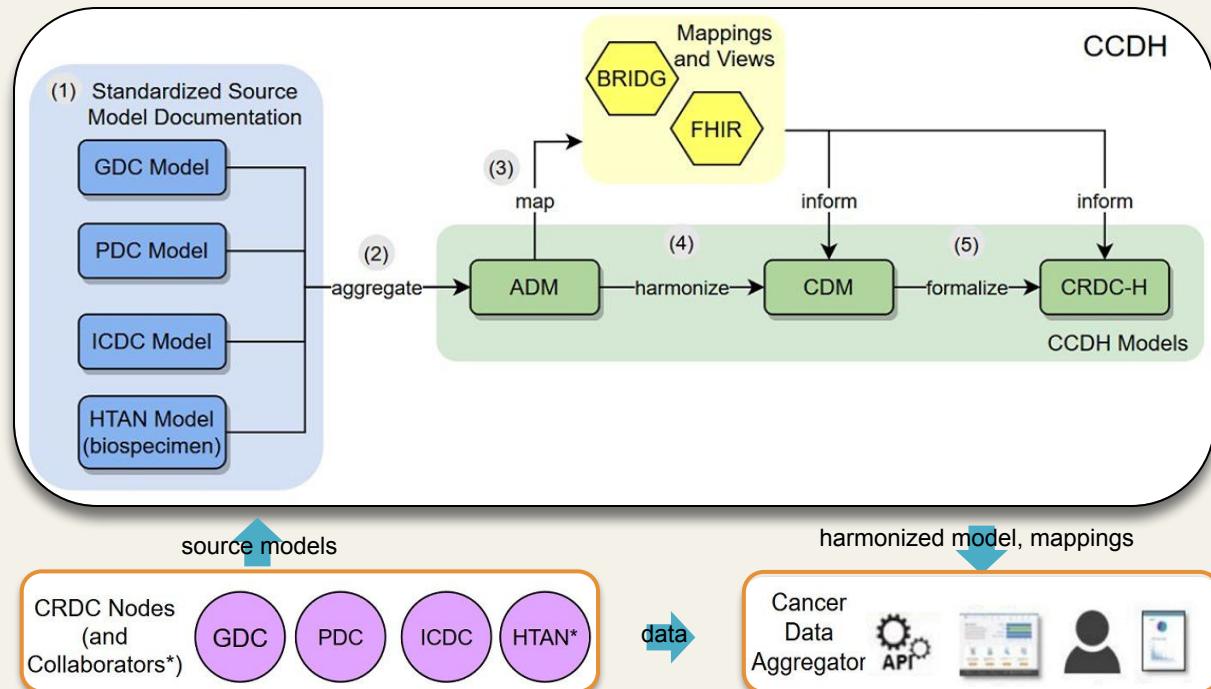
# Phase II Harmonization Efforts (ongoing)

## Node Engagement:

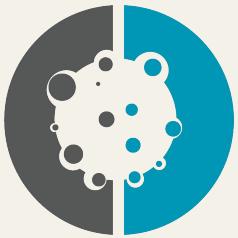
- Analysis of data model change management processes within the CRDC
- Profiling of CRDC node data to understand property usage and content

## CDA Coordination:

- End-to-end MVP to demonstrate integration and cross-node queries
- Develop formal mapping and transform specifications



# Terminologies



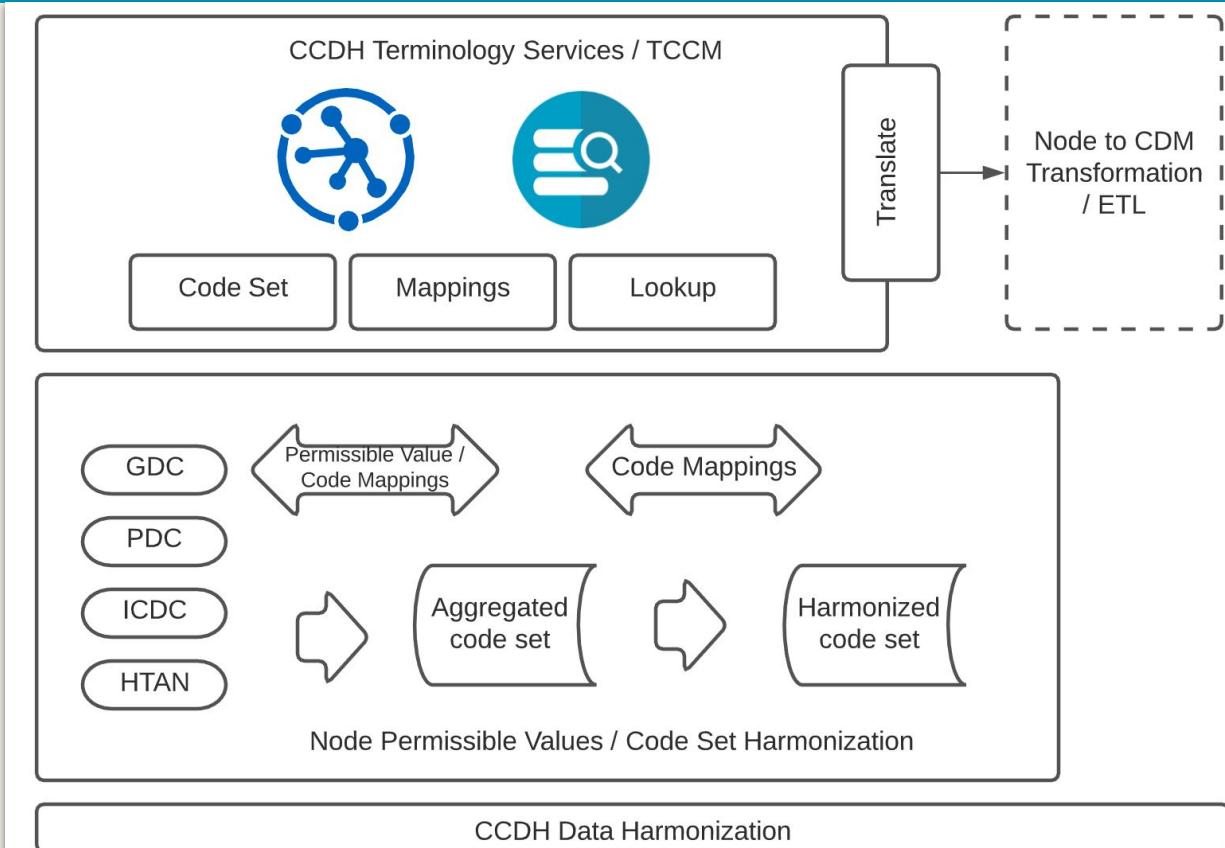
CENTER *for*  
CANCER DATA  
HARMONIZATION

# Ontology & Terminology Ecosystem Workstream

Since last we met, we have:

- 1 : Created an aggregated terminology value sets for the Conceptual Domain Model (CDM) to support the end-to-end MVP process
- 2 : Created mappings between the enumerated values in node data dictionaries to the aggregated terminology value sets
- 3 : Started the design of models and specifications for terminology services to meet the terminological requirements of the data harmonization process and terminological needs of CCDH and the NCI nodes

# Permissible Values / Code Set Harmonization

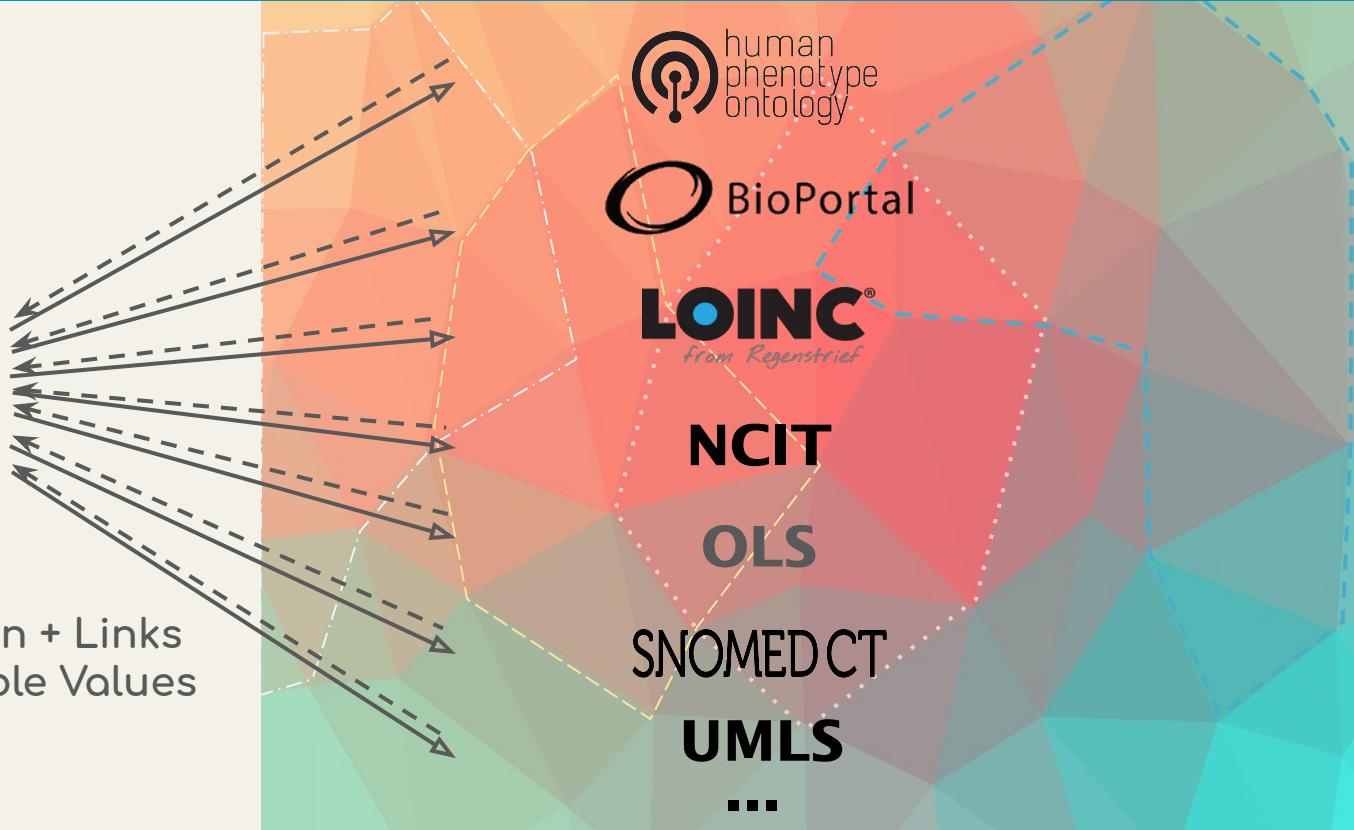


# Ontology and Terminology EcoSystem Workstream

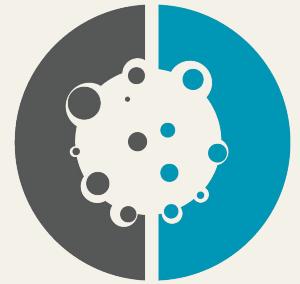
## TCCM: Terminology Common Core Model



URI + Core Information + Links  
Code Sets + Permissible Values  
Maps



# Harmonizing clinical data with caDSR and Ptolemy.V



CENTER *for*  
CANCER DATA  
HARMONIZATION

# The Problem

Two Microsoft Excel spreadsheets are shown side-by-side, illustrating a clinical data entry problem.

**Left Spreadsheet: I-SPY 1 All Patient Clinical and Outcome Data**

L1	A	F	G	H	I	J	K	L
1	SUBJECTID	PgRpos	HR Pos	Her2MostPos	HR_HER2_CATEGORY	HR_HER2_STATUS	BilateralCa	Lateralit
2	1001	0	1	0	1	HRposHER2neg	0	1
3	1002	1	1	0	1	HRposHER2neg	0	2
4	1003	1	1	0	1	HRposHER2neg	0	1
5	1004	0	0	0	3	TripleNeg	0	1
6	1005	1	1	0	1	HRposHER2neg	0	1
7	1007	1	1	0	1	HRposHER2neg	0	1
8	1008	1	1	0	1	HRposHER2neg	1	2
9	1009	0	0	0	3	TripleNeg	0	1
10	1010	0	0	0	3	TripleNeg	0	1
11	1011	0	0	0	3	TripleNeg	0	2
12	1012	1	1	0	1	HRposHER2neg	0	1
13	1013	1	1	0	1	HRposHER2neg	0	2
14	1015	0	1	1	2	HER2pos	0	1
15	1016	1	1	0	1	HRposHER2neg	0	1
16	1017	0	0	0	3	TripleNeg	0	1
17	1018	1	1	0	1	HRposHER2neg	0	2
18	1019	0	0	0	3	TripleNeg	0	2
19	1021	1	1	1	2	HER2pos	0	1
20	1022	0	1	0	1	HRposHER2neg	0	1
21	1024	0	1	0	1	HRposHER2neg	0	1
22	1025	1	1	0	1	HRposHER2neg	0	1
23	1026	0	1	1	2	HER2pos	0	1
24	1027	0	0	0	3	TripleNeg	0	1
25	1028	0	0	0	3	TripleNeg	0	2
26	1029	1	1	1	2	HER2pos	0	2
27	1030	0	0	1	2	HER2pos	0	2
28	1031	0	1	0	1	HRposHER2neg	0	2
29	1032	1	1	0	1	HRposHER2neg	0	2
30	1033	1	1	0	1	HRposHER2neg	0	1

**Right Spreadsheet: I-SPY 1 All Patient Clinical and Outcome Data**

C46	X	✓	f(x)	3-level HR/Her2 status pre-treatment	C	D
52					A	B
53						1=Yes
54						Index Tumor Laterality
55						1=Left
56						2=Right
57					C	
58						Imaging Data
59					MRI LD:	LD spans all disease present (inv & DCIS) even if there is intervening normal tissue, in mm
60					Baseline	Timepoint 1= baseline
61					1-3d AC	Timepoint 2= 1-3days after start of AC
62					InterReg	Timepoint 3= Inter-regimen
63					PreSurg	Timepoint 4= Pre-Surgery
64						Blank=not available
65						
66						
67						
68						
69						

**\*Of Note: Clinical Response & Evaluation is to coincide with MRI schedule.**

**T1 = Pre-Treatment/Baseline**

**T2 = Early Treatment Day1, cycle 2**

**T3 = Inter-Regimen**

# The Cancer Data Standards Registry and Repository (caDSR)

## Data Element Details

Public ID:	827
Version:	3.0
Long Name:	Primary Tumor Laterality Type
Short Name:	PRM_TUMOR_LATERAL_TP
Preferred Question Text:	Tumor Laterality
Definition:	for a tumor in a paired organ, a designation for the side of the body on which the tumor or cancer first developed.
Value Domain:	Tumor Laterality Type
Data Element Concept:	Primary Tumor Laterality



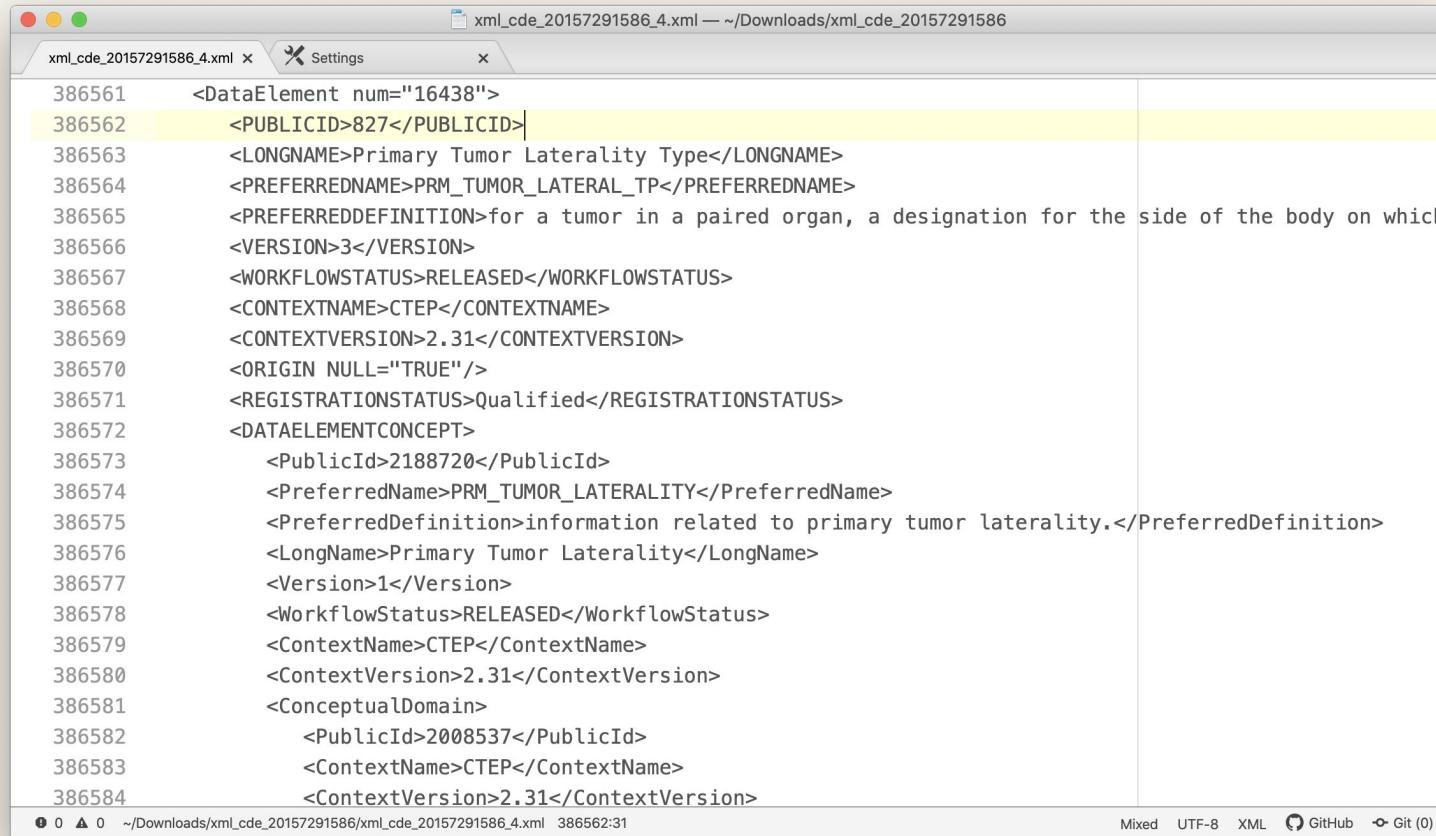
## Permissible Values

(Total number of PVs = 11)

PV	PV Meaning	PV Meaning Concept Codes	PV Meaning Description	PV Begin Date	PV End Date	VM Public ID	VM Version
Bilateral	Bilateral	C13332	Affecting both the right and left sides of the body.	2002-02-11		2558284	1.0
Both	Both	C49636	Two entities considered and identified together.	2002-02-11		2558322	1.0
Left	Left	C25229	Left; being or located on or directed toward the side of the body to the west when facing north.	2002-02-11		2574089	1.0
Midline	MIDLIN	C0549183	No Value Exists	2008-12-15		2559223	1.0
Not a Paired site	Not a Paired site	C25594:C48046:C13717(Primary)	An operation in which a term denies or inverts the meaning of another term or construction.: A set of two similar things considered as a unit.: Named locations of, or within, the body.	2002-02-11		2570492	1.0

<input type="checkbox"/> CITN (Cancer Immunotherapy Trials Ne	<input type="checkbox"/> Diagnos
<input type="checkbox"/> COG (Children's Oncology Group)	<input type="checkbox"/> Laterality
<input type="checkbox"/> CTEP (NCI Cancer Therapy Evaluation I	<input type="checkbox"/> Head an
<input type="checkbox"/> DCI (Duke Cancer Institute)	<input type="checkbox"/> Side Typ
<input type="checkbox"/> DCP (NCI Division of Cancer Preventior	<input type="checkbox"/> Involved
<input type="checkbox"/> ECOG-ACRIN (ECOG-ACRIN)	<input type="checkbox"/> Type
<input type="checkbox"/> EDRN (NCI Early Detection Research P	<input type="checkbox"/> Lesion L
	<input type="checkbox"/> Lymph N
	<input type="checkbox"/> Type

# The Cancer Data Standards Registry and Repository (caDSR)



A screenshot of a Mac OS X application window titled "xml\_cde\_20157291586\_4.xml — ~/Downloads/xml\_cde\_20157291586". The window contains XML code for a DataElement concept, specifically for Primary Tumor Laterality Type. The code is organized into numbered lines, each starting with a line number and followed by XML tags and their values. The XML structure includes DataElement, PUBLICID, LONGNAME, PREFERREDNAME, PREFERREDDEFINITION, VERSION, WORKFLOWSTATUS, CONTEXTNAME, CONTEXTVERSION, ORIGIN, REGISTRATIONSTATUS, and DATAELEMENTCONCEPT. The PUBLICID value "827" is highlighted in yellow.

```
386561 <DataElement num="16438">
386562 <PUBLICID>827</PUBLICID>
386563 <LONGNAME>Primary Tumor Laterality Type</LONGNAME>
386564 <PREFERREDNAME>PRM_TUMOR_LATERAL_TP</PREFERREDNAME>
386565 <PREFERREDDEFINITION>for a tumor in a paired organ, a designation for the side of the body on which
386566 <VERSION>3</VERSION>
386567 <WORKFLOWSTATUS>RELEASED</WORKFLOWSTATUS>
386568 <CONTEXTNAME>CTEP</CONTEXTNAME>
386569 <CONTEXTVERSION>2.31</CONTEXTVERSION>
386570 <ORIGIN NULL="TRUE"/>
386571 <REGISTRATIONSTATUS>Qualified</REGISTRATIONSTATUS>
386572 <DATAELEMENTCONCEPT>
386573 <PublicId>2188720</PublicId>
386574 <PreferredName>PRM_TUMOR_LATERALITY</PreferredName>
386575 <PreferredDefinition>information related to primary tumor laterality.</PreferredDefinition>
386576 <LongName>Primary Tumor Laterality</LongName>
386577 <Version>1</Version>
386578 <WorkflowStatus>RELEASED</WorkflowStatus>
386579 <ContextName>CTEP</ContextName>
386580 <ContextVersion>2.31</ContextVersion>
386581 <ConceptualDomain>
386582 <PublicId>2008537</PublicId>
386583 <ContextName>CTEP</ContextName>
386584 <ContextVersion>2.31</ContextVersion>
```

0 ▲ 0 ~/Downloads/xml\_cde\_20157291586/xml\_cde\_20157291586\_4.xml 386562:31

Mixed UTF-8 XML GitHub Git (0)

# Mapping columns to CDEs

Columns in TCIA spreadsheets

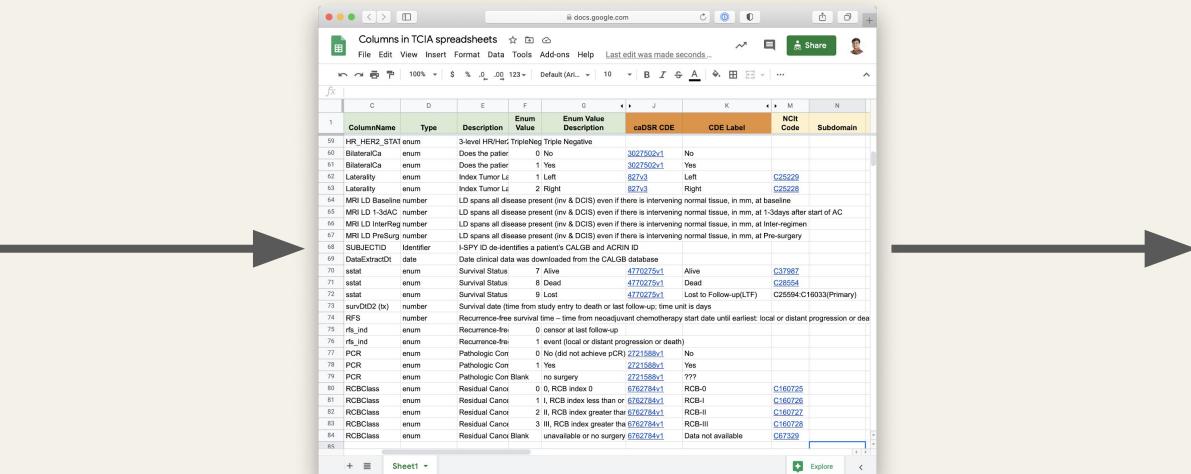
File Edit View Insert Format Data Tools Add-ons Help Last edit was made seconds... Share

	C	D	E	F	G	J	K	M	N
1	ColumnName	Type	Description	Enum Value	Enum Value Description	caDSR CDE	CDE Label	NCIt Code	Subdomain
59	HR_HER2_STAT	enum	3-level HR/Her, TripleNeg	Triple Negative					
60	BilateralCa	enum	Does the patient	0 No	3027502v1	No			
61	BilateralCa	enum	Does the patient	1 Yes	3027502v1	Yes			
62	Laterality	enum	Index Tumor Le	1 Left	827v3	Left	C25229		
63	Laterality	enum	Index Tumor Le	2 Right	827v3	Right	C25228		
64	MRI LD Baseline	number	LD spans all disease present (inv & DCIS) even if there is intervening normal tissue, in mm, at baseline						
65	MRI LD 1-3dAC	number	LD spans all disease present (inv & DCIS) even if there is intervening normal tissue, in mm, at 1-3days after start of AC						
66	MRI LD InterReg	number	LD spans all disease present (inv & DCIS) even if there is intervening normal tissue, in mm, at Inter-regimen						
67	MRI LD PreSurg	number	LD spans all disease present (inv & DCIS) even if there is intervening normal tissue, in mm, at Pre-surgery						
68	SUBJECTID	Identifier	I-SPY ID de-identifies a patient's CALGB and ACRIN ID						
69	DataExtractID	date	Date clinical data was downloaded from the CALGB database						
70	sstat	enum	Survival Status	7 Alive	4770275v1	Alive	C37987		
71	sstat	enum	Survival Status	8 Dead	4770275v1	Dead	C28554		
72	sstat	enum	Survival Status	9 Lost	4770275v1	Lost to Follow-up(LTF)	C25594:C16033(Primary)		
73	survDfD2 (tx)	number	Survival date (time from study entry to death or last follow-up; time unit is days						
74	RFS	number	Recurrence-free survival time – time from neoadjuvant chemotherapy start date until earliest: local or distant progression or dea						
75	rfs_ind	enum	Recurrence-fre:	0 censor at last follow-up					
76	rfs_ind	enum	Recurrence-fre:	1 event (local or distant progression or death)					
77	PCR	enum	Pathologic Con	0 No (did not achieve pCR)	2721588v1	No			
78	PCR	enum	Pathologic Con	1 Yes	2721588v1	Yes			
79	PCR	enum	Pathologic Con	Blank	no surgery	2721588v1	???		
80	RCBClass	enum	Residual Canc	0 0, RCB index 0	6762784v1	RCB-0	C160725		
81	RCBClass	enum	Residual Canc	1 I, RCB index less than or	6762784v1	RCB-I	C160726		
82	RCBClass	enum	Residual Canc	2 II, RCB index greater than	6762784v1	RCB-II	C160727		
83	RCBClass	enum	Residual Canc	3 III, RCB index greater tha	6762784v1	RCB-III	C160728		
84	RCBClass	enum	Residual Canc	Blank	unavailable or no surgery	6762784v1	Data not available	C67329	
85									

<http://bit.ly/columns-in-tcia>

# Executing the mappings

Laterality
1
2
1
1
1
2
1
1
2
1
1
2
1
1
2
1
1
2
1
1



# Harmonized data

ISPY1 patient clinical subset - harmonized

File Edit View Insert Format Data Tools Help Last edit was seconds ago

Share

Left

	A	V	W	X	Y	Z	AA	AB	AC	AD
1	SUBJECTID	BilateralCa_caD	BilateralCa_ncit	Laterality	Laterality_caDS	Laterality_ncit_1	MRI LD Baseline	MRI LD 1-3dAC	MRI LD InterReq	MRI LD PreSurg
3	1002	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	2 Right	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	29	26	66	16	
4	1003	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	50	64	54	46	
5	1004	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	91	90	99	43	
6	1005	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	98	109	60	42	
7	1007	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	100	100		54	
8	1008	Yes	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	2 Right	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	45	49	47	32	
9	1009	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	75	66	57	7	
10	1010	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	60	52	20	3	
11	1011	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	2 Right	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	68	68	56	47	
12	1012	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	78	84	56	57	
13	1013	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	2 Right	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	71	70	8	5	
14	1015	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	79	50	11	3	
15	1016	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	73	60	57	36	
16	1017	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	1 Left	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	54		27	16	
17	1018	No	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	2 Right	<a href="https://ncit.nci.nih.gov">https://ncit.nci.nih.gov</a>	50	37	33	26	

<https://bit.ly/2HCcu05>

# An Easier Way

The image displays three screenshots of the Ptolemy.V meta-data & data manager application, illustrating the process of mapping data elements between source datasets and standard data sets.

**Screenshot 1: Source Data Set**

This screenshot shows the "Source Data Set" page for "ISPY1". It lists data elements such as SUBJECTID, DataExtractDt, age, race\_id, and ERpos. A histogram indicates the distribution of values for the "age" field, with the highest frequency at 0.

**Screenshot 2: Column Mapping**

This screenshot shows the "Column: Lat" page for the "ISPY1" source name. It displays a mapping table where the source value "Lat" is mapped to standard values: 1 (Left), 2 (Right), and -[NULL]-. The table also includes columns for "Select to relate [Rank]" and "Description Value".

**Screenshot 3: Standard Data Set**

This screenshot shows the "Values for standard data element: Primary Tumor Laterality Type 827" page. It lists standard values: Bilateral, Both, Left, Midline, Not a Paired site, Other, Right, Unilateral, Unilateral, side not specified, and Unknown. The "Left" value is currently selected for mapping.

**Common UI Elements:**

- Header:** Ptolemy.V meta-data & data manager, Revision: 30300 Thu Jul 30 09:21:29 EDT 2020, Welcome, Getting Started, About, Glossary, Contact Us, Account.
- Buttons:** Upload Files, Source Data Sets, Translation Builders, Translation Definitions, Data Standards, Unmap All, Save Mapping, Save Mapping & Lock, Mark as unmappable.

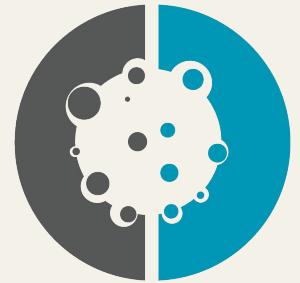
# Conclusion

- CCDH will continue to investigate possible data harmonization workflows, and would love feedback from other nodes on what works for them.
- Outstanding questions:
  - Who will be responsible for this harmonization effort?
  - Where will the resulting harmonized data reside?
  - How will search and integration of the clinical data with the CRDC nodes be enabled?

=> To be answered with end-to-end requirements analysis!

# CCDH + CDA: Defining a Minimum Viable Product

*Enabling query across the CRDC*



CENTER *for*  
CANCER DATA  
HARMONIZATION

# Minimum Viable Product (MVP)

## Goal:

Demonstrate aggregation and harmonization of data from multiple nodes using the emerging CRDC-H model

For CCDH, this presents an opportunity to test numerous aspects of the end-to-end harmonization process:

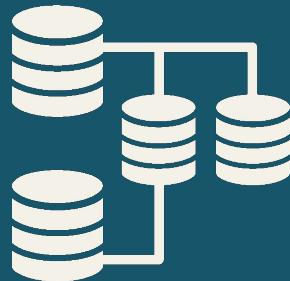
Data Model

Modeling framework / language

Element and value set level mapping

Degree of overlap in node model population of like attributes

# Initial CCDH-provided MVP Deliverables



## Minimal CDM Test Model:

A small subset of the CDM consisting of a few elements from the Biospecimen and Administrative Data domains, specified in a formal modeling language: BiolinkML

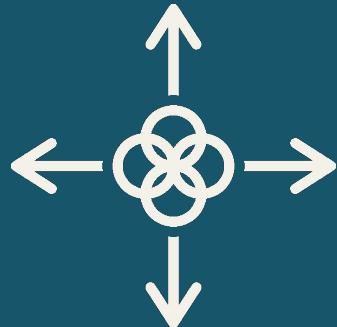
## Minimal Class and Attribute Mappings:

Mappings from these CDM elements to source node elements (GDC, PDC, HTAN), specified in a computable language/format, that can support data transformations

- Mapping language/syntax based on features/syntax of GraphQL and FHIR Mapping Language



# Initial CCDH-provided MVP Deliverables



## Minimally-Harmonized Value Sets:

Aggregations of value sets across equivalent source node attributes that take coded data. These will be deduplicated where there are clear equivalencies between values, and mappings to values in source value sets will be provided

- Utilizing the SSSOM format

## Simple Test Queries:

A set of queries to test the model and tooling, designed to demonstrate retrieval of data/records from >1 node

- Captured alongside CDA-provided queries



# MVP development: Next Steps

1

Refinement of initial proposed specifications through iterative feedback with CDA modeling and ETL experts

2

Profiling of source data to inform query development and ongoing model / value set harmonization

3

Testing of model, mappings, and value sets through implementation and query using CDA infrastructure

# Acknowledgments

## CRDC Nodes

- CDS: Cancer Data Services
- CIDC: Cancer Immunology Data Commons
- GDC: Genomic Data Commons
- ICDC: Integrated Canine Data Commons
- IDC: Imaging Data Commons
- PDC: Proteomics Data Commons

## CRDC Infrastructure

- DCF: Data Commons Framework

## CRDC Cloud Resources

- Broad Institute FireCloud
- ISB: Institute for Systems Biology Cancer Genomics Cloud
- Seven Bridges Genomics Cancer Genomics Cloud

## Center for Biomedical Informatics & Information Technology

- Allen Dearry
- Sherri de Coronado
- Erika Kim
- Denise Warzel
- Melissa Cook

## Cancer Data Aggregator

- Brian O'Connor
- Alex Baumann
- David Pot
- Jack DiGiovanna
- Cara Mason
- Annie Kuan

## Collaborators

- HTAN: Human Tumor Atlas Network
- Gabriella Miller Kids First Data Resource Center
- SEER Virtual Tissue Repository

## Frederick National Laboratory for Cancer Research

- Todd Pihl
- Resham Kulkarni

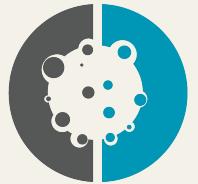
## Samvit Solutions

- Smita Hastak
- Wendy Ver Hoef
- Charles Yaghmour



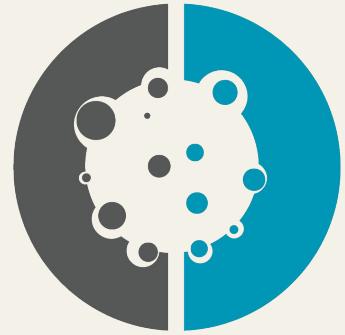
CENTER for  
CANCER DATA  
HARMONIZATION

# Questions & Answers



CENTER *for*  
CANCER DATA  
HARMONIZATION

These slides:  
[bit.ly/ccdh-crdc-f2f-2](https://bit.ly/ccdh-crdc-f2f-2)



CENTER *for*  
CANCER DATA  
HARMONIZATION

CCDH Update at CRDC All Hands  
October 26 & 27, 2020