

# Developing Pediatric Cancer Data Standards

*CCDI Webinar Series*

# Introductions

*Dr. Gregory Reaman*

## Agenda

1. *Pediatric Data Commons*
2. *PCDC Consensus Data Modeling*
3. *D4CG, NCI Semantic Infrastructure, and CCDI Collaboration*
4. *Pediatric Cancer Core Elements*
5. *Childhood Cancer Clinical Data Commons (C3DC)*
6. *Q&A*

# Today's Speakers



**Dr. Sam Volchenbom**  
Principal Investigator and  
Pediatric Oncologist  
Data for the Common Good



**Dr. Michael Watkins**  
Manager of Data Standards  
and Modeling  
Data for the Common Good



**Brian Furner**  
Senior Director of Data and  
Technology  
Data for the Common Good



# Pediatric Cancer Data Commons (PCDC)

*Dr. Sam Volchenboum*

# CCDI Harmonization Objectives

- Establish consistent and interoperable standards for collecting, storing, and sharing pediatric cancer data.
- Facilitate data integration across studies for streamlined analysis.
- Develop core pediatric elements for clinical and research purposes.
  - Curate key data elements for retrospective harmonization and prospective collection.



Image Source: Shutterstock

# D4CG: Who We Are



DATA FOR THE  
COMMON GOOD

PEDIATRIC CANCER  
DATA COMMONS

GEARBOX



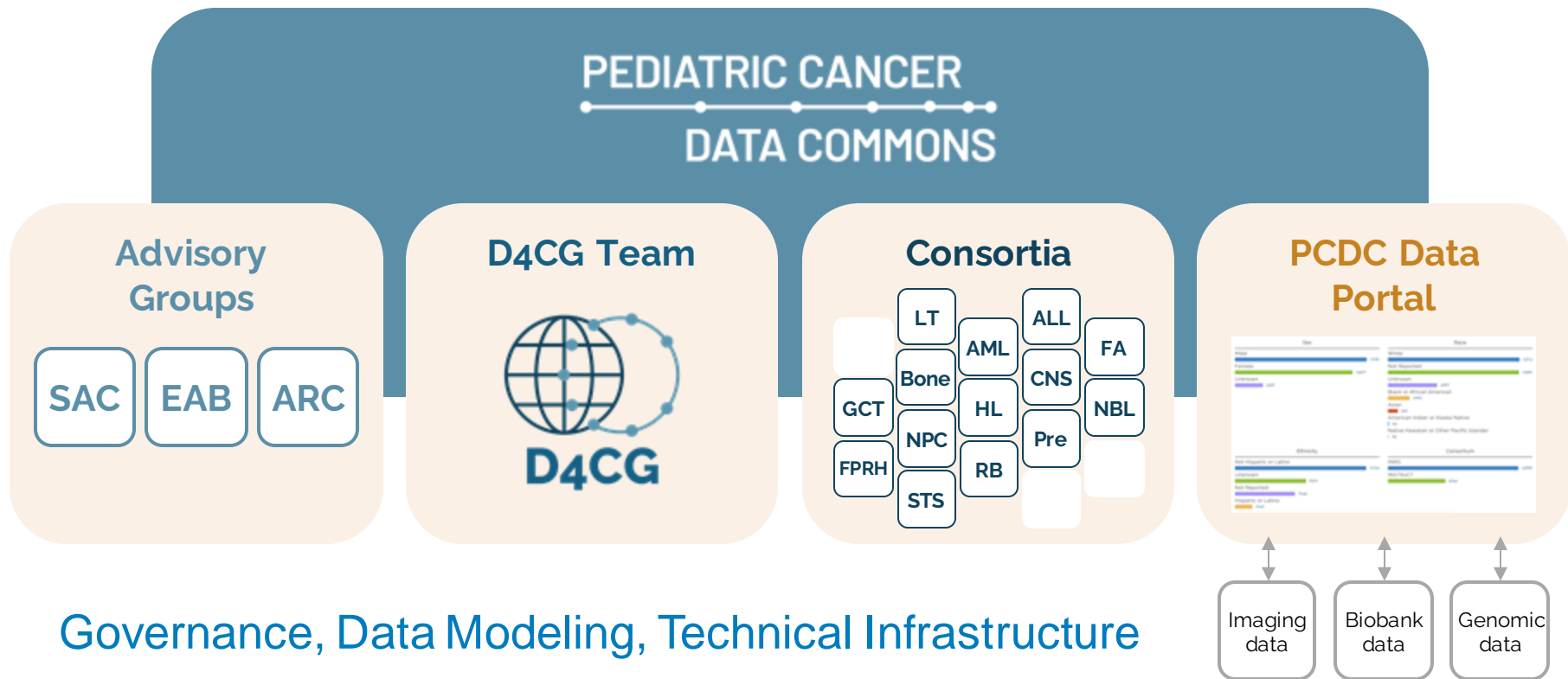
Sociome

PREDICT  
*monogenic  
diabetes*

other rare  
disease  
commons

*...and more to come*

# Service Providers

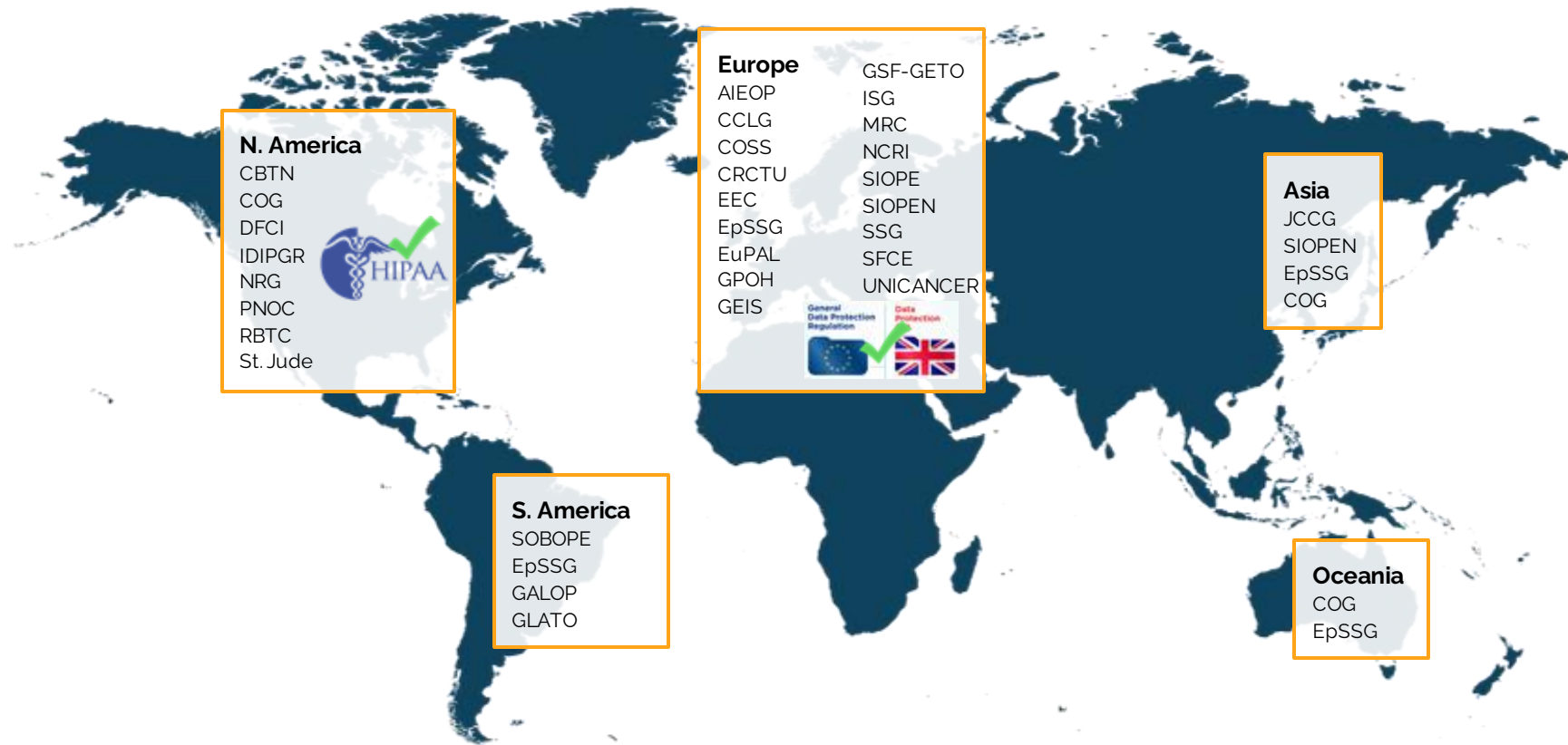




# PCDC Consensus Data Modeling

*Dr. Michael Watkins*

# An International Collaboration



# Data Dictionary Meetings

- Prioritize the re-use of existing data elements to enable cross-disease research.



NBL (INRG) File Edit View Insert Format Data Tools Extensions Help

50% \$ % .0\_ .00 123 Raleyway

A1 RowType

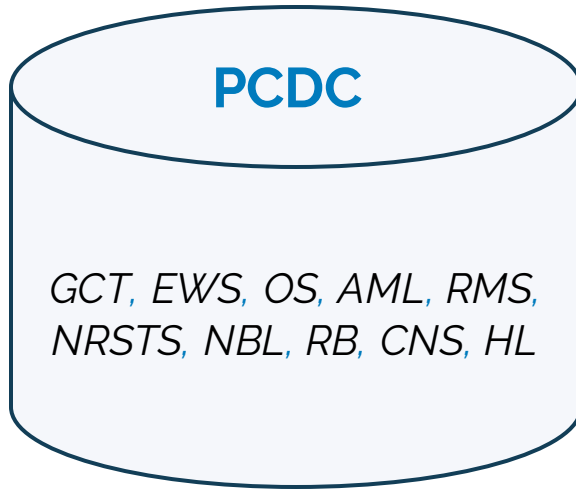
1	RowType	VariableName	Data Type	Tier	VariableSource	VariableDescription	VariableCode	ParentSubst/InSubst
2	TD	Subject Identifier						
3	TG	will be included in every table						
4	VD	PCDC_SUBJECT_ID	String	1	ALLIAMLJGCTH-A	A unique subject id assigned to a c not:C179f00		undefined,
5	PD							
6	DD	Preferred						
7	TD	Subject Characteristics						
8	TG	One row per subject per study						
9	VD	HONEST_BROKER_SUBJECT_ID	String	5	ALLIAMLJGNSG	The identifier assigned to a not:C18f949		undefined,
10	VD	CONSORTIUM	Code	1	ALLIAMLJGNSG	The consortium under whic		undefined,
11	PD							INRG
12	VD	DATA_CONTRIBUTOR_ID	Code	1	ALLIAMLJGNSG	An identifier assigned to a c not:C18f950		undefined,
13	PD							COG
14	PD							GPOH
15	PD							JCOG
16	PD							JINCS
17	PD							INRG
18	PD							SOPEN
19	PD							
20	VD	STUDY_ID	Code	1	ALLIAMLJGNSG	A sequence of characters \ not:C18f9f2		undefined,
21	PD							ANBL0030
22	PD							ANBL0032
23	PD							ANBL1232
24	PD							ANBL1201
25	VD	AGE_AT_ENROLLMENT	Integer	3	ALLIAMLJGNSG	The age (in days) when the not:C18f9f3		undefined,
26	VD	YEAR_AT_ENROLLMENT	Integer	3	GCTNBL	The year at which a subject not:C177953		undefined,
27	VD	TREATMENT_ARM	Code	2	ALLIAMLJGNSG	A specific treatment plan w not:C15538		undefined,
28	PD							"three treatment arms"
29	PD							Unknown
30	PD							Not Reported
31	VD	EFS_CENSOR_STATUS	Code	2	NBLINRG1520M	The length of time after pri		undefined,
32	PD							Censored
33	PD							Not Censored
34	PD							Unknown
35	PD							Not Reported
36	VD	AGE_AT_CENSOR_STATUS	Integer	2	NBL	Age (in days) of the subject		undefined,
37	PD							
38	DD	Preferred						
39	TD	Time Period						
40	TG	One row per subject per time perio						
41	VD	SUBMITTER_ID	String	2	ALLIAMLJGNSG	An ID provided by the data		undefined,
42	VD	PARENT_SUBMITTER_ID	String	2	ALLIAMLJGNSG	The SUBMITTER_ID of the 1		undefined,
43	VD	TIME_PERIOD_TYPE	Code	2	ALLIAMLJGNSG	The type of time period bel		undefined,
44	PD							Disease Phase
45	PD							Course
46	PD							Course
47	VD	DISEASE_PHASE	Code	2	ALLIAMLJGNSG	The phase of the cancer in not:C18f9f9		undefined,

# Cross-Domain Stakeholders

We address the challenge of complex rare diseases by convening focused communities of diverse experts:

- Pediatric oncologists
- Surgeons
- Pathologists
- Radiation oncologists
- Genetic counselors
- Statisticians
- Database engineers
- Terminologists
- Regulatory and governance experts

# A Commons of Commons



COG  
UCL  
EURO-EWING  
CRCTU  
EORTC  
FSG/SFCE  
GPOH  
ISG  
SSG  
UK

**Ewing Sarcoma**

AIEOP  
COG  
MRC  
CCLG  
NRG-Oncology  
SFCE  
SOPOBE  
DFCI

**Germ Cell Tumors**

AIEOP  
BFM-SG  
COG  
DCOG  
MRC  
NOPHO  
PPLLSG  
SJCRH  
JACLS  
JPLSG

**AML**

BOCG  
COG  
COSS-GPOH  
CRCTU  
FSG  
GALOP  
GEIS  
GPOH  
ISG  
MRC  
SSG  
UCL  
UK

**Osteosarcoma**

**Neuroblastoma**

**Soft-Tissue Sarcoma**

**Retinoblastoma**

**Embryonal CNS**

**Hodgkin Lymphoma**

COG  
GPOH  
JCCG  
JINCS  
JNBSG  
SIOPEN

AIEOP  
COG  
EpSSG  
SIOP MMT  
CWS  
STSC

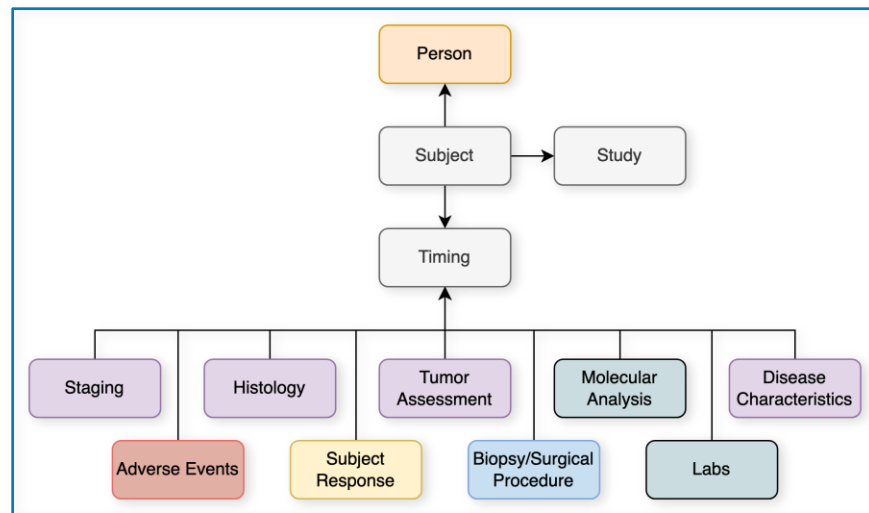
CHLA  
COG  
DFCI  
DEPICT  
EuRBG  
GALOP

CBTN  
COG  
IDIPGR  
PNOC  
RBTR  
SIOPE

COG  
SJCRH  
PHC

# Changes to the PCDC model

- Adjustments across 16+ groups are frequent.
- Versioning and change management are critically important but complex.
- High volume of changes over the years as new groups have joined.
- The rate of change requests is now slowing—an indication of maturity in the consensus elements.



Subset of the PCDC Model

# D4CG, NCI Semantic Infrastructure, and CCDI Collaboration

*Dr. Michael Watkins*

# NCI's Enterprise Vocabulary Services: Promote Semantic Interoperability

- The Cancer Data Standards Registry and Repository (caDSR): A structured repository for clinical and research metadata, which semantically defines data through controlled terms and vocabularies.
- Common Data Elements (CDEs) are standardized metadata elements used to describe specific pieces of information in clinical and research settings.
- NCI Thesaurus (NCIt): A controlled vocabulary of biomedical concepts including terms related to diseases, anatomy, procedures, and more.
- CDEs incorporate terms or concepts from NCIt, which ensures consistency and interoperability across different data collection (clinical and research) efforts.
- Semantic interoperability enables easy sharing and understanding of data across systems and research domains.



# PCDC Disease-Specific Dictionaries Alignment

Domain	Tables	ALL	AML	CNS	CP	EWS	FA	FPRH	GCT	HL	LT	NBL	NPC	NRSTS	OS	RB	RMS
Protocol	Subject Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Protocol	Time Period	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Protocol	Off Protocol Therapy Or Study	✓	✓	✓	□	✓	✓	□	□	✓	✓	□	✓	□	✓	✓	□
Demographics	Demographics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	Medical History	✓	✓	□	✓	✓	✓	✓	✓	✓	✓	✓	□	✓	✓	✓	✓
Demographics	Survival Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	Family Medical History	□	□	□	□	□	✓	□	□	□	□	□	✓	□	□	✓	□
Testing	Vitals And Anthropometrics	✓	✓	✓	✓	✓	✓	✓	✓	□	□	□	✓	□	□	□	□
Testing	Laboratory Test	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	□	✓	□	✓
Testing	Genetic Analysis	✓	✓	✓	✓	✓	□	✓	✓	□	□	✓	□	✓	✓	✓	✓
Testing	Function Test	□	✓	□	□	□	□	□	□	□	□	□	✓	□	✓	□	□
Testing	Immunohistochemistry	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Testing	Imaging	□	□	□	□	□	□	□	□	✓	□	□	□	□	□	✓	□
Disease Attributes	Diagnosis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Disease Attributes	Disease Characteristics	✓	✓	□	□	✓	✓	✓	✓	✓	✓	✓	✓	□	□	✓	□
Disease Attributes	Disease Site Assessment	□	□	✓	□	□	□	□	□	✓	✓	✓	✓	✓	□	✓	✓
Disease Attributes	Staging	□	□	✓	✓	□	✓	✓	□	✓	✓	✓	✓	□	□	✓	✓
Treatment	Radiation Therapy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	□	✓	✓	✓	✓	✓
Treatment	Stem Cell Transplant	✓	✓	□	□	✓	□	□	✓	✓	✓	□	□	□	□	✓	□
Treatment	Medication	□	✓	✓	□	✓	✓	✓	✓	□	□	□	✓	□	□	□	□
Treatment	Transfusion Medicine Procedures	□	✓	□	□	□	□	□	□	✓	□	□	□	□	□	□	□
Treatment	Cellular Immunotherapy	□	✓	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Treatment	Biopsy And Surgical Procedures	□	□	✓	□	✓	✓	✓	✓	✓	✓	□	✓	✓	✓	✓	✓
Treatment	Protocol Treatment Modifications	□	□	□	✓	□	□	□	□	□	□	□	✓	□	□	□	□
Response	Minimal Residual Disease	✓	✓	□	□	□	□	□	□	□	✓	□	□	□	□	□	□
Response	Subject Response	□	✓	✓	□	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Events	Adverse Events	✓	✓	□	✓	✓	✓	✓	✓	✓	✓	□	✓	□	✓	✓	□
Events	Subsequent Malignant Neoplasm	✓	✓	✓	□	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Events	Patient Reported Outcomes Metadata	□	□	□	□	□	□	□	✓	✓	□	□	□	□	✓	□	□
Events	Late Effects	□	□	□	□	□	□	□	□	✓	□	□	□	□	□	□	□

Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Central Nervous System Tumors (CNS), Ewing Sarcoma (EWS), Fanconi Anemia (FA), Fertility Preservation and Reproductive Health (FPRH), Germ Cell Tumors (GCT), Hodgkin's Lymphoma (HL), Liver Tumors (LT), Neuroblastoma (NBL), Nasopharyngeal Carcinoma (NPC), Non-rhabdomyosarcoma Soft-Tissue Sarcomas (NRSTS), Osteosarcoma (OS), Retinoblastoma (RB), Rhabdomyosarcoma (RMS)

[https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI\\_Thesaurus&version=23.12d&ns=ncit&code=C168547&k ey=183215209&b=1&n=null](https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&version=23.12d&ns=ncit&code=C168547&k ey=183215209&b=1&n=null)

<https://evs.nci.nih.gov/ftp1/PCDC/>

# Disease-Specific Dictionaries Published to NCI Thesaurus



- PCDC Terminology
  - ALL Project Terminology
  - AML Project Terminology
    - AML Table Terminology
      - AML Adverse Events Table
      - AML Cellular Immunotherapy Table
      - AML Concomitant Medication Table
      - AML Course Timing Table
      - AML Demographics Table
      - AML Disease Characteristics Table
      - AML Disease Phase Timing Table
      - AML Function Test Table
      - AML Lab Table
      - AML Medical History Table
      - AML Minimal Residual Disease Table
      - AML Molecular Analysis Table
      - AML Myeloid Sarcoma Involvement Table
      - AML Off Protocol Therapy/Study Table
      - AML Radiation Therapy Table
      - AML Secondary Malignant Neoplasm Table
      - AML Stem Cell Transplant Table
      - AML Subject Characteristics Table
      - AML Subject Response Table
      - AML Survival Characteristics Table
      - AML Total Dose Table
      - AML Transfusion Medicine Procedures Table
      - AML Vitals Table
- FWS Project Terminology
- GCT Project Terminology
- HL Project Terminology
- OS Project Terminology

- AML Adverse Events Table

[Export Excel](#) [Export CSV](#)

Project	NCIt Code of Table	PCDC Table PT	NCIt Concept Code	NCIt PT	NCIt SY	PCDC PT	PCDC SY	NCIt Definition	PCDC
AML	<a href="#">C173217</a>	Adverse Events Table	<a href="#">C175600</a>	PCDC Subject Identifier	Pediatric Cancer Data Commons Unique Subject Identifier	PCDC_SUBJECT_ID		A unique identifier for a subject in a Pediatric Cancer Data Commons (PCDC) study.	
AML	<a href="#">C173217</a>	Adverse Events Table	<a href="#">C172677</a>	Age in Days at Onset Adverse Event		AGE_AT_AE		Age of subject (in days) at the onset of the adverse event.	
AML	<a href="#">C173217</a>	Adverse Events Table	<a href="#">C168878</a>	Disease Phase		DISEASE_PHASE		The stage or period of an individual's disease.	
AML	<a href="#">C173217</a>	Adverse Events Table	<a href="#">C156813</a>	Initial Diagnosis		Initial Diagnosis		The first diagnosis of the individual's condition.	

# PCDC Data Elements Published in the caDSR

PCDC common data elements (CDEs) are published alongside other pediatric cancer CDEs in caDSR, fostering interoperability.

The screenshot displays the caDSR interface. On the left, a tree view shows the hierarchy of data elements. A callout box highlights the 'Pediatric Cancer (Pediatric Cancer) (427)' folder, which contains several sub-folders, including 'PCDC Aggregated Dictionary (240)'. An arrow points from this callout to the 'Stem Cell Transplant Conditioning Type|13378563|1 (Data Element)' folder in the main tree. The main tree lists various sub-elements under this folder, such as '1. Question Texts', '2. Alternate Names', '3. Alternate Definitions', '4. Associated Concepts', '5. Classifications', '6. Permissible Values', '7. Forms/Modules', '8. Reference Documents', and '9. Derivation Components'. The '6. Permissible Values' sub-element is selected, and its data is displayed in a table on the right. The table has columns for 'View', 'VM Public ID', 'Permissible Value', and 'VM'. The table contains six rows of data.

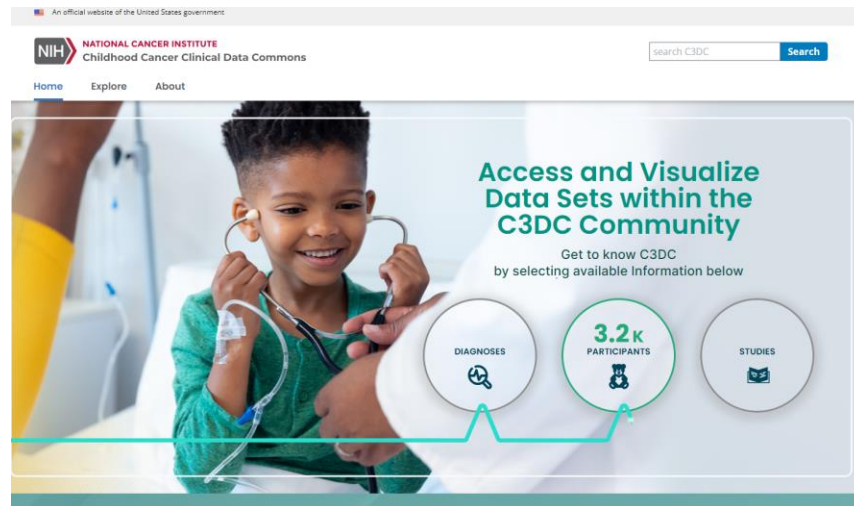
<input type="checkbox"/>	View	VM Public ID	Permissible Value	VM
<input type="checkbox"/>		2568580	Myeloablative	Myeloablative
<input type="checkbox"/>		13328603	Non-Myeloablative	Nonmyeloablative Allogeneic T
<input type="checkbox"/>		5612322	Not Reported	Not Reported
<input type="checkbox"/>		6431768	Other	Other
<input type="checkbox"/>		4383037	Reduced Intensity Conditioning/Reduced Toxicity Conditioning	Reduced Intensity
<input type="checkbox"/>		5682953	Unknown	Unknown

# Pediatric Cancer Core Elements

*Dr. Michael Watkins*

# Pediatric Cancer Core CDEs

- ~75 concepts (in active development)
- Includes CDEs from a number of high-level clinical domains:
  - Administrative, Demographics, Diagnosis, Treatment, Specimen, Molecular, and Imaging
- Importance will be indicated through tiering:
  - Tier 1—must include, regardless of the resource cost
  - Tier 2—include if resources are available
  - Tier 3—include if resources permit



# Core Data Elements Overview—Molecular Features

CDE Public ID	CDE Long Name	Alt Name	PCDC Variable	Closest FHIR Path(s)	In mCODE?
<a href="#">13362328</a>	Subject Gene Mutation Assessment Status	Status	STATUS	GenomicVariant.value	Y
<a href="#">11379445</a>	Gene Human Chromosome Name	Chromosome	CYTOGENETIC_LOCATION	GenomicVariant.cytogeneticLocation	Y
<a href="#">11280318</a>	Gene Occurrence HGNC Symbol Name	Gene Symbol	GENE	GenomicVariant.geneStudied	Y
<a href="#">13367968</a>	Gene Mutation Second Gene HGNC Symbol Name	Second Gene Symbol	GENE2	GenomicVariant.geneStudied	Y
<a href="#">13367930</a>	Gene Mutation Detection Type	The type of gain, loss or alteration	ALTERATION_TYPE	GenomicVariant.genomicDNAChangeType	Y
<a href="#">13367935</a>	Gene Mutation Abnormality Type	Gene Mutation Abnormality Type	ALTERATION_EFFECT	GenomicVariant.molecularConsequence	Y
<a href="#">13367965</a>	Gene Mutation Copy Number Variation Assessment Type	Copy Number Variation	COPY_NUMBER_STATUS	GenomicVariant.molecularConsequence	Y
<a href="#">13367961</a>	Gene Mutation ISCN Karyotype Text	ISCN	ISCN	GenomicVariant.cytogeneticNomenclature	Y
<a href="#">13367959</a>	Gene Mutation HGVS Protein Text	HGVS Protein	HGVS_PROTEIN	GenomicVariant.aminoAcidChange	Y
<a href="#">13367956</a>	Gene Mutation HGVS Coding Text	HGVS Coding	HGVS_CODING	GenomicVariant.genomicDNAChange	Y
<a href="#">6142510</a>	Molecular Analysis Genetic Zygosity Type	Zygosity	ALLELIC_STATE	GenomicVariant.allelicState	Y

- Closely aligns with HL7 Genomic Reporting Profiles used in mCODE.
- Excluded concepts: sample type, method, clinical significance, external IDs, allele frequencies.
- National Childhood Cancer Registry (NCCR) elements:
  - Core concepts present but are spread across many lab-specific fields.

# Core Data Elements Overview–Diagnosis

CDE Public ID	CDE Long Name	Alt Name	PCDC Variable	Closest FHIR Path(s)	In mCODE?
<a href="#">12318179</a>	Subject Age at Histology Assessment Day Count	Age at Hist Assessment	AGE_AT_DIAG_ASSESSMENT	PrimaryCancerCondition.assertedDate	Y
<a href="#">14714145</a>	Diagnosis Level of Certainty Status	diagnosis_certainty	-	PrimaryCancerCondition.verificationStatus	Y
<a href="#">13606067</a>	Disease Morphology ICD-O-3 Label Text	Diagnosis Classification	DIAGNOSIS	PrimaryCancerCondition.code	Y
<a href="#">13606067</a>	Disease Morphology ICD-O-3 Label Text	Diagnosis Classification	MORPH_CODE	PrimaryCancerCondition.histologyMorpholo	Y
-	-	Diagnosis Classification System	MORPH_CODE_SYSTEM	PrimaryCancerCondition.histologyMorpholo	Y
<a href="#">13279286</a>	Diagnosis Primary Anatomic Site	Tumor Site	SITE	PrimaryCancerCondition.bodySite	Y
<a href="#">13381583</a>	Biospecimen Histologic Grade Type	Histology Grade	HISTOLOGY_GRADE	-	N
<a href="#">12137353</a>	Imaging Technology DICOM Modality Type	Imaging Modality	DETECTION_METHOD	TumorSize.method	Y
<a href="#">12922545</a>	Tumor Classification Category	Tumor Classification	CLASSIFICATION	CancerDiseaseStatus.value	Y
<a href="#">13382770</a>	Disease or Disorder Staging System Name	Stage System	STAGE_SYSTEM	PrimaryCancerCondition.stage.type	Y
-	-	Stage System Version	STAGE_SYSTEM_VERSION	PrimaryCancerCondition.stage.type.version	Y
<a href="#">13382767</a>	Disease Tumor Stage Name	Stage	STAGE	PrimaryCancerCondition.summary	Y
<a href="#">3123069</a>	Longest Tumor Surgeon Diameter Measurement	Tumor size longest diameter (as c	LONGEST_DIAM_DIM1	TumorSize.tumorLongestDimension	Y

- Most concepts present and mappable to mCODE.
- Diagnosis design complications:
  - Many new rare disease definitions not published as ICD-O morph codes.
  - Hierarchical disease definitions difficult without ontology bindings.
  - Biopsies can be diagnostic, prognostic, and therapeutic.

# Pediatric Cancer Core CDEs: Draft

- These CDEs have a dedicated space within the caDSR.
- Developed in concert with curators from CCDI, NCI Semantic Infrastructure (SI) team, D4CG, and others.

Links to Favorites

- Data Elements - View/Browse
- NCI Standard Data Elements
- CRDC Standard Data Elements
- OMOP Data Elements
- Browse CDEs by Classifications
- Browse CDEs by Data Element Concept
- Browse CDEs by Protocol Forms
- Forms
- NCI Standard CRFs
- Guest User Cart
- Standard Conceptual Domains
- Standard Representation Terms
- Download Collection
- VM Match - Guest User
- Pediatric Cancer Core CDEs**

Home>>Favorites>>Pediatric Cancer Core CDEs

Display Values View - Delivery Options -

Pediatric Cancer Standard Data Elements						
<input type="checkbox"/>	View	CDE Public ID	Version	▲2 Pediatric Cancer Name	▲1 Domain	VD Type
<input type="checkbox"/>		12960571	1.00	Study ID	Administrative	Enumerated
<input type="checkbox"/>		12988910	1.00	Treatment Arm	Administrative	Enumerated
<input type="checkbox"/>		2847330	1.00	Vital Status	Clinical Observation	Enumerated
<input type="checkbox"/>		2192217	2.00	Ethnicity	Demographics	Enumerated
<input type="checkbox"/>		2192199	1.00	Race	Demographics	Enumerated
<input type="checkbox"/>		2690165	2.00	Relationship	Demographics	Enumerated
<input type="checkbox"/>		11341616	1.00	Anatomic Site	Diagnosis	Enumerated
<input type="checkbox"/>		14714145	1.00	diagnosis_certainty	Diagnosis	Enumerated
<input type="checkbox"/>		12217251	1.00	Disease Phase	Diagnosis	Enumerated
<input type="checkbox"/>		13209646	1.00	Medical History Condition Category	Diagnosis	Enumerated
<input type="checkbox"/>		13212141	1.00	Medical History Condition Type	Diagnosis	Enumerated
<input type="checkbox"/>		13279286	1.00	Tumor Site	Diagnosis	Enumerated

<https://cadsr.cancer.gov/onedata/Home.jsp>

<https://cadsr.cancer.gov/onedata/dmdirect/NIH/NCI/CO/PCDC%20Browser>

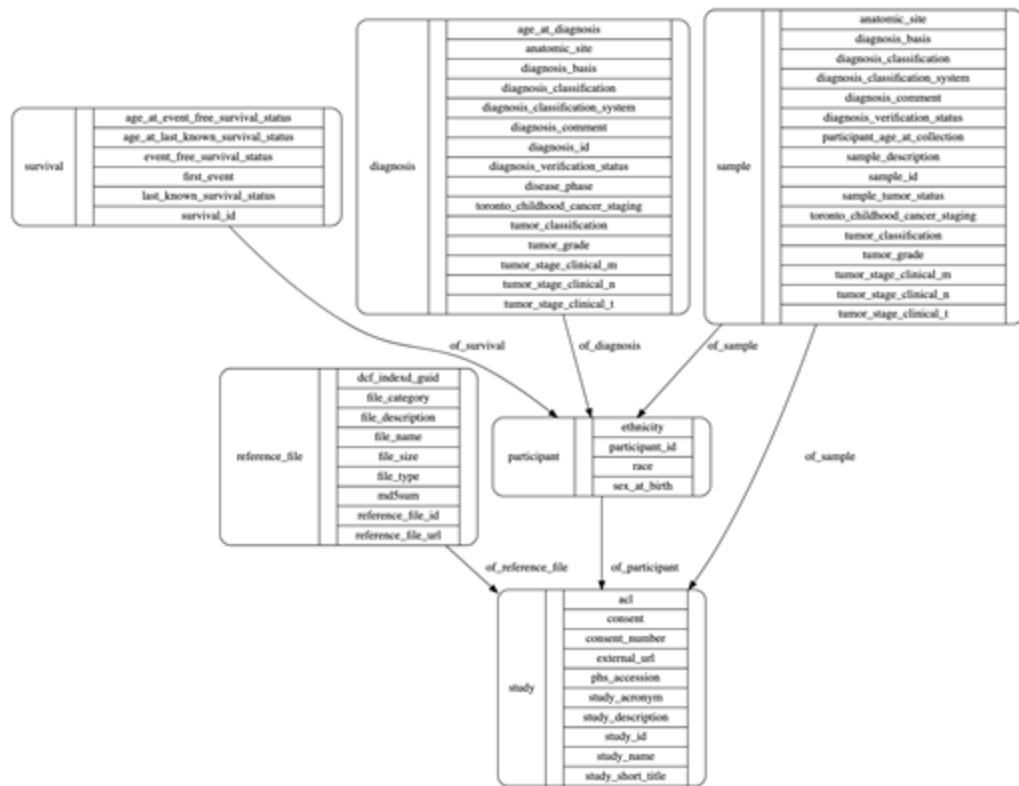


# Childhood Cancer Clinical Data Commons (C3DC)

*Brian Furner*

# C3DC Model

- C3DC includes coverage of study, participant, diagnosis, sample, and survival data.
- Model will grow as more studies are included.
- Reference file node contains metadata that describe lineage and provenance of data during mapping and extract, transform, and load (ETL) process.

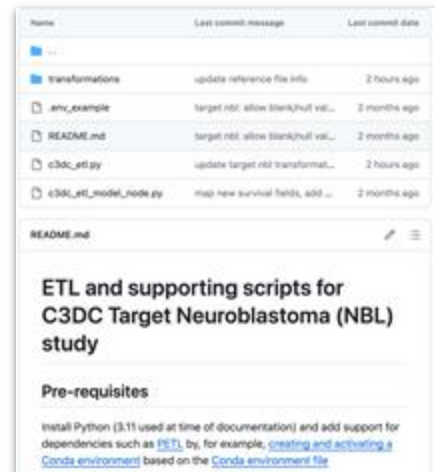


<https://github.com/CBIIT/c3dc-model>

# C3DC ETL

- Source data and harmonized data are kept separate to simplify transformations and downstream use of data.
- Mapping and transformation logic is documented and publicly available through GitHub, **ensuring transparent provenance and lineage**.
  - Source-to-destination mappings are encoded in JSON.
  - Transformation logic encapsulated in Python.
- To date, completed ETL and delivery of data for prioritized studies (TARGET NBL, MCI CNS, OncoKids Cancer Panel).

[https://github.com/chicagopcdc/c3dc\\_etl](https://github.com/chicagopcdc/c3dc_etl)



[GitHub repository for ETL](https://github.com/chicagopcdc/c3dc_etl)

```
{
  "output_field": "diagnosis.anatomic_site",
  "source_field": "ICD0",
  "type_group_index": "*",
  "default_value": {
    "Not Reported"
  },
  "replacement_values": [
    {
      "old_value": "C16.9",
      "new_value": [
        "C16.9 : Stomach, N05"
      ]
    },
    {
      "old_value": "C22.0",
      "new_value": [
        "C22.0 : Liver"
      ]
    }
  ]
},
```

Mapping example

# Childhood Cancer Clinical Data Commons (C3DC)

Allows researchers to search for harmonized participant-level data.

3147 DIAGNOSIS RECORDS 3150 PARTICIPANTS 2 STUDIES

Clear all filtered selections

DEMOGRAPHICS

Participant ID Search

UPLOAD PARTICIPANTS SET

SEX AT BIRTH

RACE

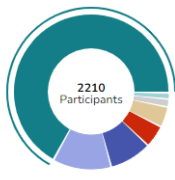
ETHNICITY

DIAGNOSIS

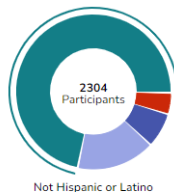
SURVIVAL

STUDY

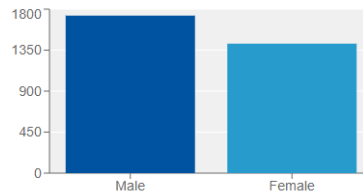
Race



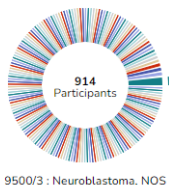
Ethnicity



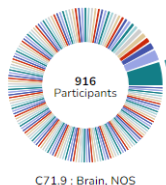
Sex at Birth



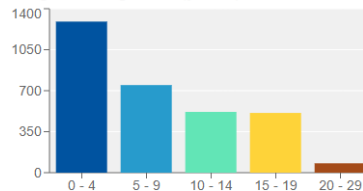
Diagnosis (ICD-O)



Genomic Site



Age at Diagnosis (years)



Participants (3150) | Diagnosis (3147) | Survival (1221) | Studies (2)

# Find Out More About CCDI

**Learn about CCDI and subscribe to our monthly newsletter.**

[cancer.gov/CCDI](https://cancer.gov/CCDI)

**Questions? Email us.**

[NCIChildhoodCancerDataInitiative@mail.nih.gov](mailto:NCIChildhoodCancerDataInitiative@mail.nih.gov)



# Q&A

# CCDI March Community Forum

Monday, March 18, 12:00 – 1:00 pm ET

This session aims to gather community input on the Coordinated Pediatric and Young Adult Rare Cancer Initiative.

Topics include:

- Outcomes from February's Genomic Harmonization Task Force meeting
- Expansion of the Molecular Characterization Initiative
- Update on the Coordinated Pediatric and Young Adult Rare Cancer Initiative

Register Here: <https://events.cancer.gov/ccdi/webinar/registration>

**Thank you!**





**NATIONAL  
CANCER  
INSTITUTE**

[cancer.gov](https://cancer.gov)

[cancer.gov/espanol](https://cancer.gov/espanol)