

# NMDP: Harmonizing Data Through Common Data Elements

Kathleen Malum

Supervisor, Clinical Data

NMDP



The CIBMTR<sup>®</sup> (Center for International Blood and Marrow Transplant Research<sup>®</sup>) is a research collaboration between the National Marrow Donor Program<sup>®</sup> (NMDP)/Be The Match<sup>®</sup> and the Medical College of Wisconsin (MCW).

# NMDP. Find Cures. Save Lives.

---

- We save lives through cell therapy
- Global nonprofit leader in cell therapy
- Drive groundbreaking research, treatment, and support to cure blood cancer and blood disorders
- Largest registry of potential bone marrow or blood stem cell donors and umbilical cord blood units
- Commitment to research has a significant impact on blood cancer and blood disorders survival and quality of life



# CIBMTR: Leading Research for Life

---

- Center for International Blood and Marrow Transplant Research
- Research collaboration between the Medical College of Wisconsin (MCW) and NMDP
- Collects and maintains outcomes data for Clinical Research
- Mission is to improve patient outcomes, increase access to cellular therapies, and ensure donor safety
- Data from > 675,000 patients, >1,800 publications, and approximately 200 ongoing studies and clinical trials.



DATA

Acquisition, analysis, sharing, and visualization of diverse data

# SCTOD

- CIBMTR holds the contract for the Stem Cell Therapeutic Outcomes Database (SCTOD)
- Authorized the C.W. Bill Young Cell Transplantation Program and passed in December 2005
  - 5 year renewal cycle
  - Last renewed in 2022
- Mandated collection of outcomes data

One Hundred Eleventh Congress  
of the  
United States of America

AT THE SECOND SESSION

*Begun and held at the City of Washington on Tuesday,  
the fifth day of January, two thousand and ten*

An Act

To amend the Stem Cell Therapeutic and Research Act of 2005.

*Be it enacted by the Senate and House of Representatives of  
the United States of America in Congress assembled,*

**SECTION 1. SHORT TITLE.**

This Act may be cited as the “Stem Cell Therapeutic and Research Reauthorization Act of 2010”.

# Data Collection Applications

---

- **FormsNet3<sup>SM</sup>**

- Web-based application
- Data is captured using a form structure
- Allows for electronic completion and validation of recipient and donor data
- Does not allow for direct transmission of data from a transplant center's database

- **AGNIS<sup>®</sup>**

- A Growable Network Information System
- Electronic data exchange between transplant center databases and FormsNet<sup>SM</sup>
- Common Data Elements (CDEs) used to represent metadata
- Allows for unambiguous data interpretation

# CIBMTR Metadata Team

---

- Formed in 2007
- Develops a standard language for the sharing, analysis and exchange of cellular therapy data
  - Define the metadata
  - Map to data standards
- Long-standing history of using CDEs
- Helps create more consistency within and across forms



# Why CDEs?

- AGNIS needed a standard, publicly accessible “language” for electronic data exchange
  - At the time, CIBMTR needed a form-based data exchange mechanism and did not want to use a standard data model (such as the BRIDG model) for exchange
- caDSR benefits:
  - Ability to define form questions in a standard way (CDEs) and associate them with an identifier (CDE ID)
  - Both forms and CDEs would be publicly accessible
  - Ability to reuse CDEs created by other contexts
  - CDEs followed an internationally recognized format (ISO 11179)
  - CDEs were associated with definitions



# Standardizing Data Collection through Common Data Elements



**Utilize Common Data Elements (CDEs) that contain standardized terminology concepts defining the meaning of the data**



**Define the meaning of a question text (data point) by aligning to a CDE**



**Data can be collected the same way across multiple forms  
Ensures consistent data collection**



**Make data interoperable**

- Like data points between the form revisions maintain their CDE association and provide a historical and semantic definition trail
- As form content evolves, new CDEs are created to capture the semantics and define the data point
  - CDEs provide the means for the various contexts in which a particular data point is collected



# Organizing Metadata

---

- CIBMTR Form Definition Manager (FDM)
  - Allows attributes from the CDE to be pulled into a data dictionary code (DDC)
  - Each DDC is associated to a CDE
  - Extension of Metadata
- caDSR II Form Maintenance
  - Web-based application
  - Collection of CDEs used on specific CIBMTR forms



# CDE Integration with CIBMTR Form Development

- CIBMTR commitments:
  - Data Standards starts with our Scientific Leadership
  - Internal systems to integrate with caDSR II (via APIs)
    - FormsNet<sup>SM</sup> FDM
    - Knowledge Management System (Symedical<sup>®</sup>)
- Each form revision builds through the reuse of existing CDEs and the creation of CDEs to define new data points

**“... good data management and stewardship is not a goal in itself, but rather a pre-condition supporting knowledge discovery and innovation”** -- The FAIR Guiding Principles for scientific data management and stewardship  
<https://www.nature.com/articles/sdata201618#citeas>

# Vision: Utilizing CDEs for Interoperability



- CIBMTR Data Transformation Initiative
  - Vision: To optimize **the acquisition and utilization** of entrusted data assets to accelerate breakthroughs that transform patient experiences
  - Automate CIBMTR data collection to facilitate data sharing
    - Utilizes Electronic Health Record (EHR) integration engines
    - Looking for data standards such as FHIR and mCODE to be able to implement data interchange
  - Actively **influence the standardization and interoperability** for cellular therapy data
  - Standards such as mCODE are aligned with our CDEs

## minimum Common Oncology Data Elements

**A FHIR-based core set of common data elements for cancer that is standardized, computable, clinically applicable and available in every electronic health record for patients with a cancer diagnosis**

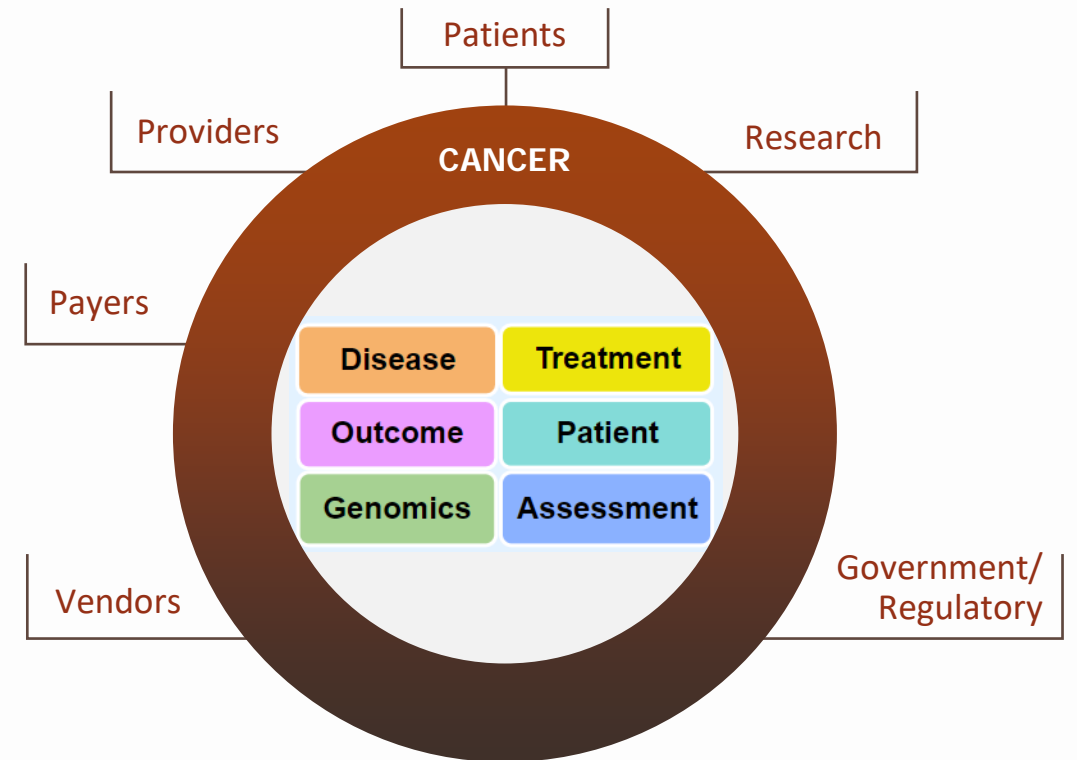
A standard health record for oncology

The minimal set of data elements applicable to all cancers, and collected for:

Standardized information exchange

Use-case driven and targeted use

Oncology data element domains: patient, disease, treatment, outcomes, genomics, lab/vital



# CDEs ensure the semantics remain the same across different collection mechanisms

## From FormsNet

80. ECOG score (at diagnosis)

Known - Go to question 81

Unknown - Go to question 82

81. ECOG score (at diagnosis)

0 - Asymptomatic (Fully active, able to carry on all pre-disease activities without restriction)

1 - Symptomatic but completely ambulatory (Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature. For example, light house or office work)

2 - Symptomatic, < 50% in bed during the day (Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours)

3 - Symptomatic, > 50% in bed, but not bedbound (Capable of only limited self-care, confined or chair 50% or more of waking hours)

4 - Bedbound (Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair)

## From EHRs using FHIR

**HL7** minimal Common Oncology Data Elements (mCODE) Implementation Guide 3.0.0 - STU3 Release

Home Content by Group Conformance FHIR Artifacts Examples Release Notes More

Table of Contents Artifacts Summary ECOC Performance Status Profile

minimal Common Oncology Data Elements (mCODE) Implementation Guide, published by HL7 International / Clinical Interoperability Council. This is not an authorized publication; it is the continuous build for version 3.0.0. This version is based on the current content of https://github.com/HL7/mCODE-IG and changes regularly. See the directory of published versions.

**18.23.1 Resource Profile: ECOG Performance Status Profile**

Official URL: http://hl7.org/fhir/us/mcode/StructureDefinition/ecoc-ecog-performance-status Version: 3.0.0

Draft as of 2023-10-25 Maturity Level: 4 Computable Name: ECOGPerformanceStatus

The Eastern Cooperative Oncology Group (ECOG) Performance Status represents the patient's functional status and is used to determine their ability to tolerate therapies in serious illness, specifically for chemotherapy. (Definition from LOINC:R)

**Conformance**

Resources associated with an in-scope patient with Observation.code LOINC 89247-1 SHALL conform to this profile. Beyond this requirement, a resource SHOULD ensure that any resource instance associated with an in-scope patient that would reasonably be expected to conform to this profile be published in this form.

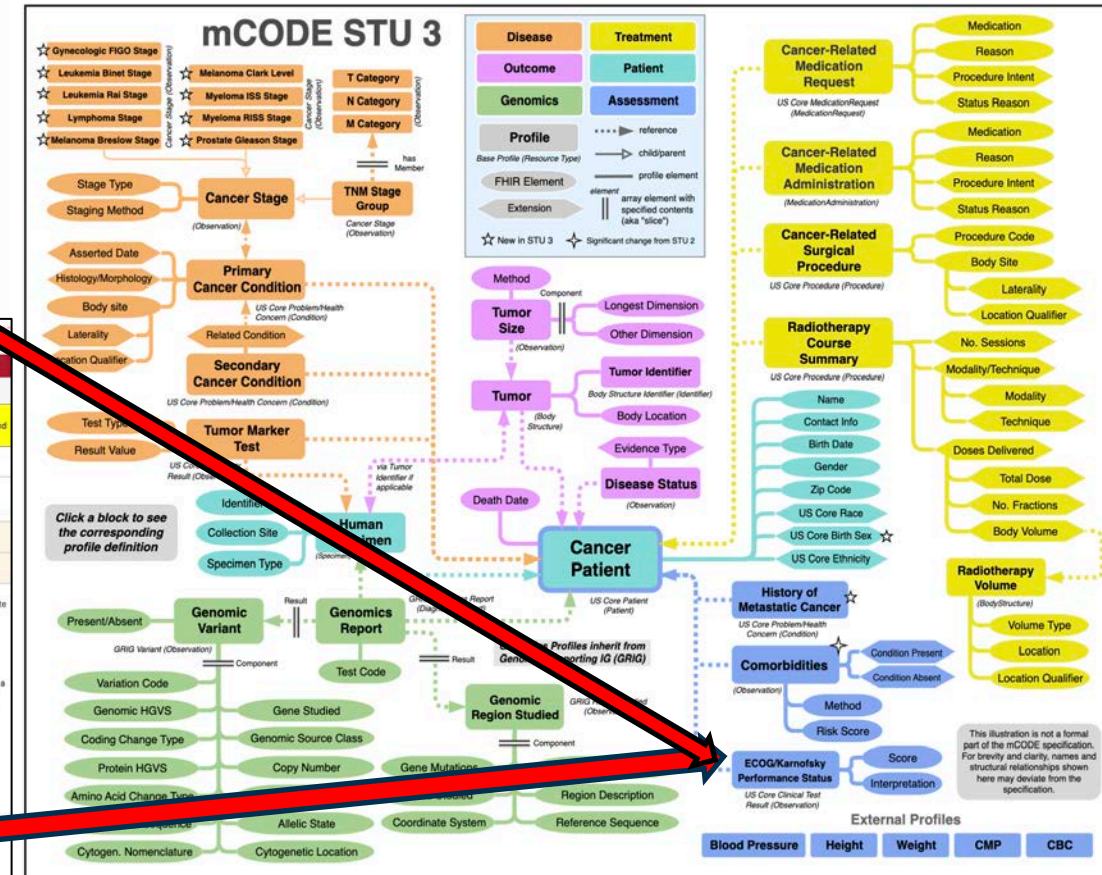
Examples for this Resource Profile: Observation[ecog-performance-status-fully-active] and Observation[ecog-performance-status-jerry-m]

**1.1.2 Formal Views of Profile Content**

Profiles, Differentials, Snapshots and how the different presentations work.

Structure is derived from USCoreObservationClinicalTestResultProfile

Element	Flags	Card.	Type	Description & Constraints
Observation	0..*	0..*	USCoreObservationClinicalTestResultProfile	Measurements and simple assertions
basedOn	0..*	0..*	Reference[ServiceRequest   CarePlan]	Fulfills plan, proposal or order
partOf	0..*	0..*	Reference[Procedure]	Part of referenced event
code	1..1	1..1	CodeableConcept	Clinical Test Name
coding	1..*	1..*	Coding	Required Pattern: At least the following Code defined by a terminology system



CDE	CDE Long Name	Definition
88	Performance Status Assessment Eastern Cooperative Oncology Group Scale	The ECOG functional performance status of the patient/participant.



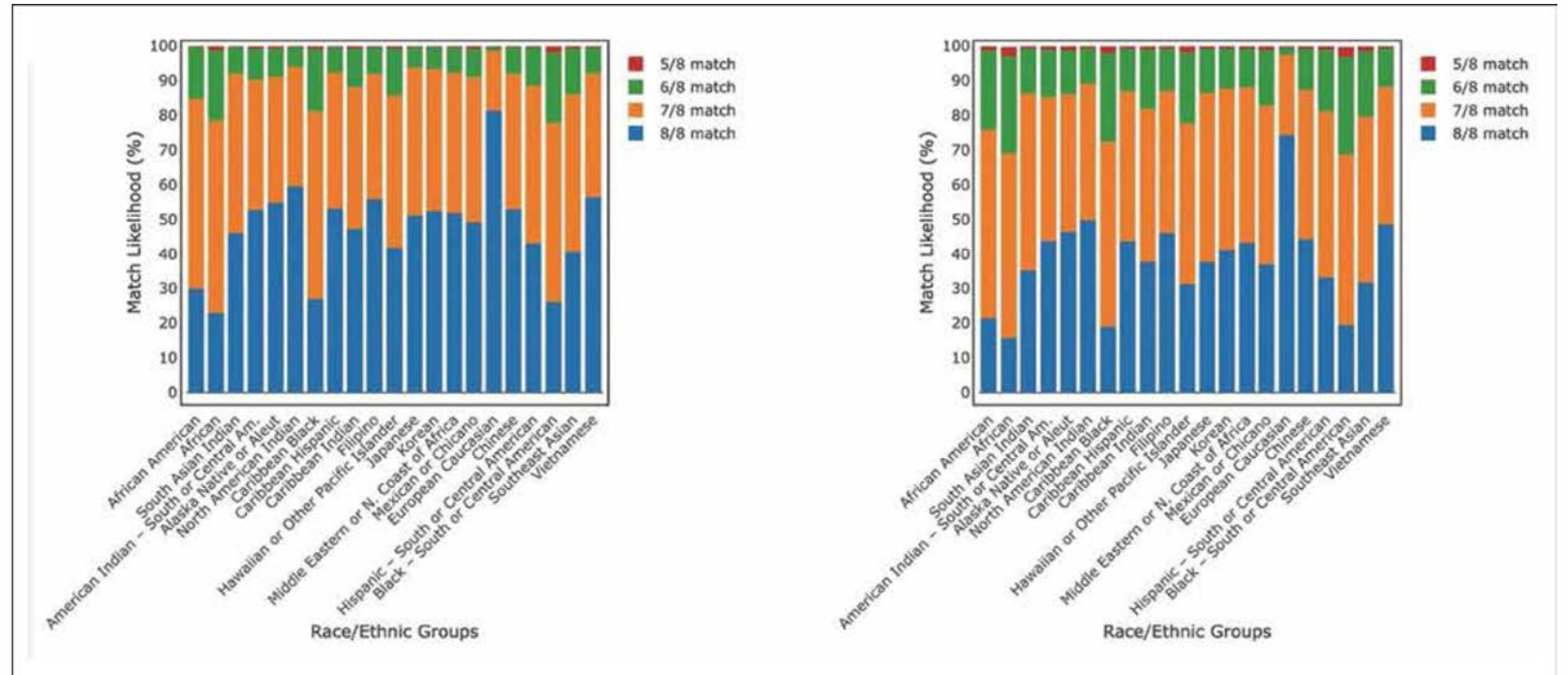
# Take Aways

---

- CIBMTR has historically led the field in standardizing data
  - [Value of CDES](#)
  - Facilitates data collection, exchange and sharing
  - Semantic data interoperability
    - Information is stored as discrete data elements and able to be extracted or submitted
  - Promotes data analysis
    - If data points are not consistently defined, data analysis may require further interpretation
  - Consistency in the way the data is collected, formatted, and described
    - Semantic alignment within and across forms
  - Data Standardization
    - Information is categorized using industry approved vocabulary and unit of measure standards
    - Clear representation of data points

# CIBMTR Research in Action: Removing Health Care disparities

- “Hematopoietic cell transplantation (HCT) is curative for hematologic disorders, but outcomes are historically inferior when using HLA-mismatched donors. Despite unrelated donor registries listing > 38 million volunteers, 25%-80% of US patients lack an HLA-matched unrelated donor, with significant disparity across ethnic groups.” -- <https://ascopubs.org/doi/10.1200/JCO.20.03502>
- The NMDP Access trial, using data collected by the CIBMTR, and using CDEs



## Breaking Human Leukocyte Antigen-related Barriers in Allogeneic Hematopoietic Cell Transplantation

Steven Devine, MD

*The Hematologist* (2024) 21 (2)

<https://doi.org/10.1182/hem.V21.2.2024211>

# Thank you

---

- Metadata Team:
  - Jenni Bloomquist: Senior Manager, Clinical Data Quality
  - Sandra Sorensen: Senior Metadata Analyst
  - Hannah Florin: Metadata Analyst
  - Hira Rizvi: Metadata Analyst
  - Kelsey Okuno: Metadata Analyst
- Mentors:
  - Robinette Renner: Program Director of Semantic Infrastructure (NIH/NCI)
  - Jane Pollack: Principle Research Informaticist

