

Managing Code Lists in Longitudinal Studies

Washington, DC
Wednesday, March 28, 2012

Alexandra Shlionskaya
Sophia Kuan
Jay Greenfield

Agenda

- The Trouble with Code Lists in Longitudinal Studies and a Proposal – Jay Greenfield
- The Data Documentation Initiative (DDI) Support for the Proposal – Sophia Kuan
- Benefits and Challenges – Alexandra Shlionskaya

Section I – Jay Greenfield

THE TROUBLES WITH CODE LISTS AND A PROPOSAL

The Trouble with Code Lists in Longitudinal Studies

- The problem is that over time as the subject of each visit gets older, the answer to a question necessarily changes:

PA001. I would now like to ask about products that may have been used in your home or yard to control for mice, rats, ants, termites, cockroaches, bees, wasps, moths, or other insects and rodents during the past 6 months. When responding to the questions in this section, please think about {C_FNAME/the child}'s primary address or the place where {he/she} lives most of the time.

The Problem with Code Lists in Longitudinal Studies

- The problem is that over time as the subject of each visit gets older, the answer to a question necessarily changes:

PA004/(WHO_APPLY). Who applied this product? Was it....

You	1
A friend or family member	2
Building maintenance, or	3
A professional exterminator?	4
REFUSED	-1
DON'T KNOW	-2

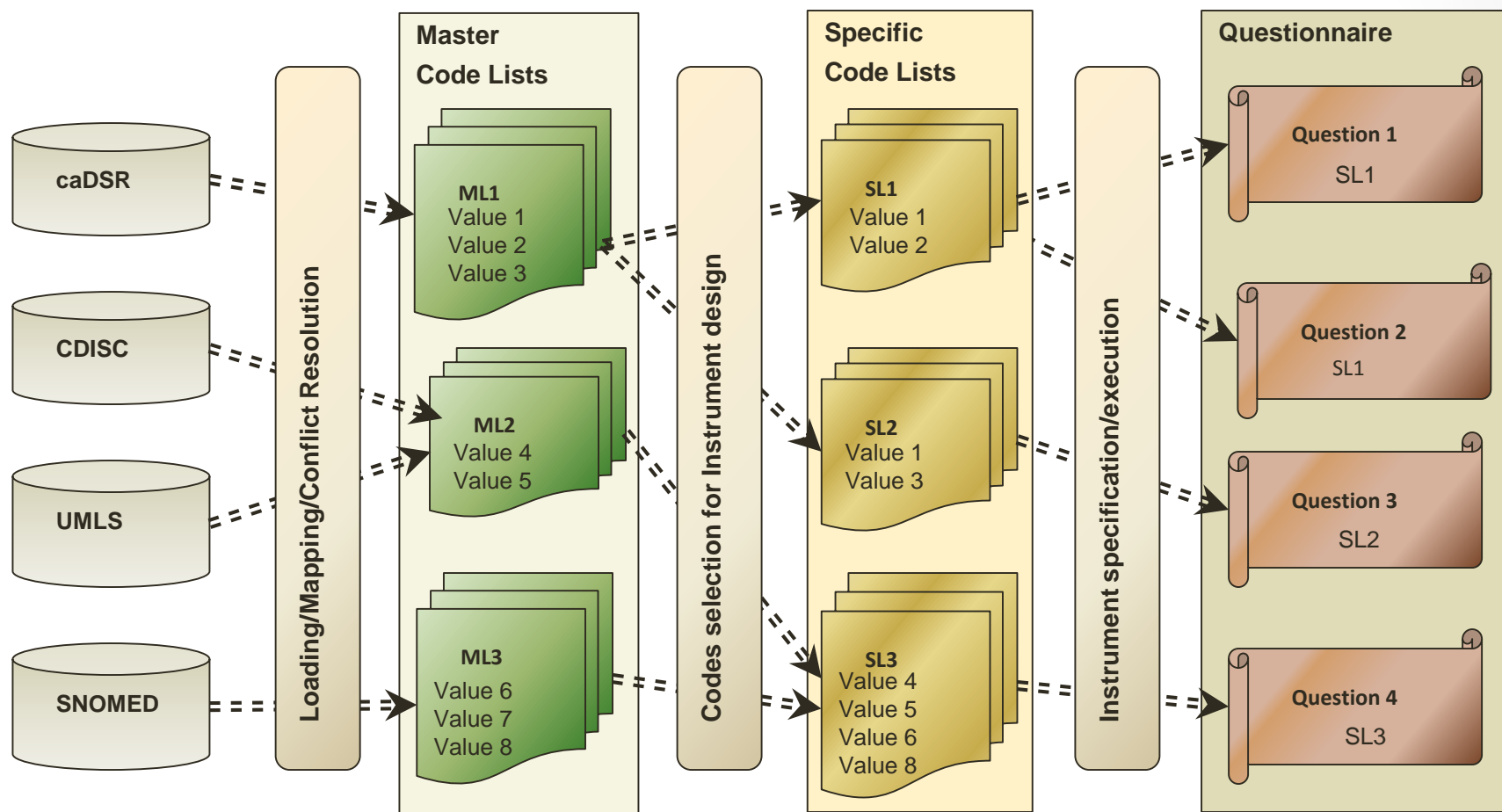
PA004/(WHO_APPLY). Who applied this product? Was it....

You	1
{C_FNAME}	2
A friend or family member	3
Building maintenance, or	4
A professional exterminator?	5
REFUSED	-1
DON'T KNOW	-2

- Infant
- Toddler
- Early Childhood

- Middle Childhood
- Late Childhood

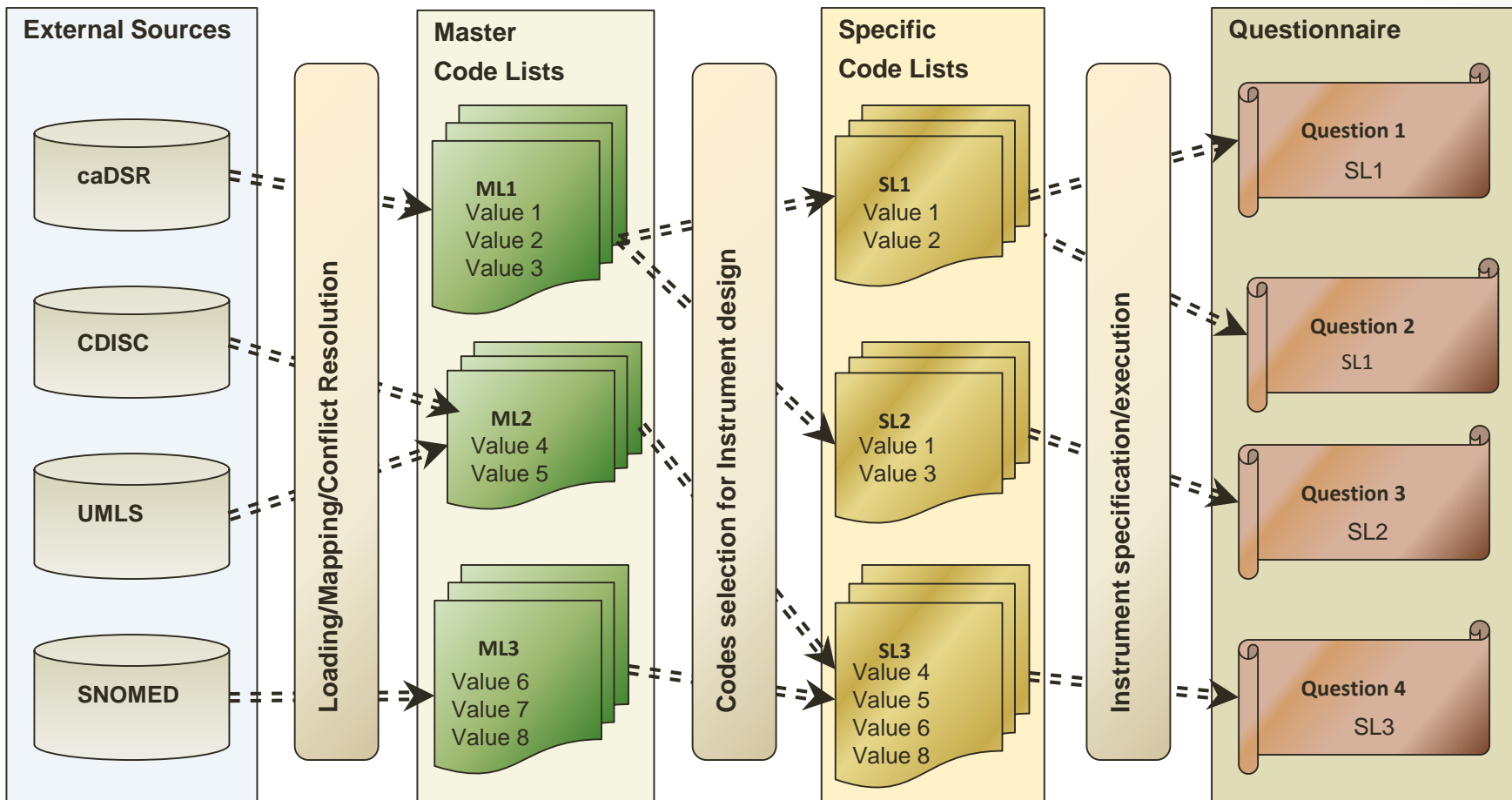
Enter an Approach called Master and Specific Code Lists



ISO 21090 NullFlavor Values

- NI: No information
- INV: Invalid
- OTH: Other
- PINF: positive infinity
- NINF: negative infinity
- UNC: unencoded
- DER: derived
- UNK: unknown
- ASKU: asked but unknown
- NAV: temporarily unavailable
- NASK: no asked
- QS: sufficient quantity
- TRC: trace
- MSK: masked
- NA: no applicable

Enter an Approach called Master and Specific Code Lists



Cancer Data Standards Registry and Repository (caDSR)



caDSR


caDSR is a database and a set of APIs and tools to create, edit, control, deploy, and find common data elements (CDEs) for use by metadata consumers, and information about the UML models and Forms containing CDEs for use in software development.

The common data elements are developed by the National Cancer Institute Center for Biomedical Informatics and Information Technology (NCI CBIIT) and [caBIG®](#) partners in the research community, to be used as metadata descriptors for research and caCORE-like applications.

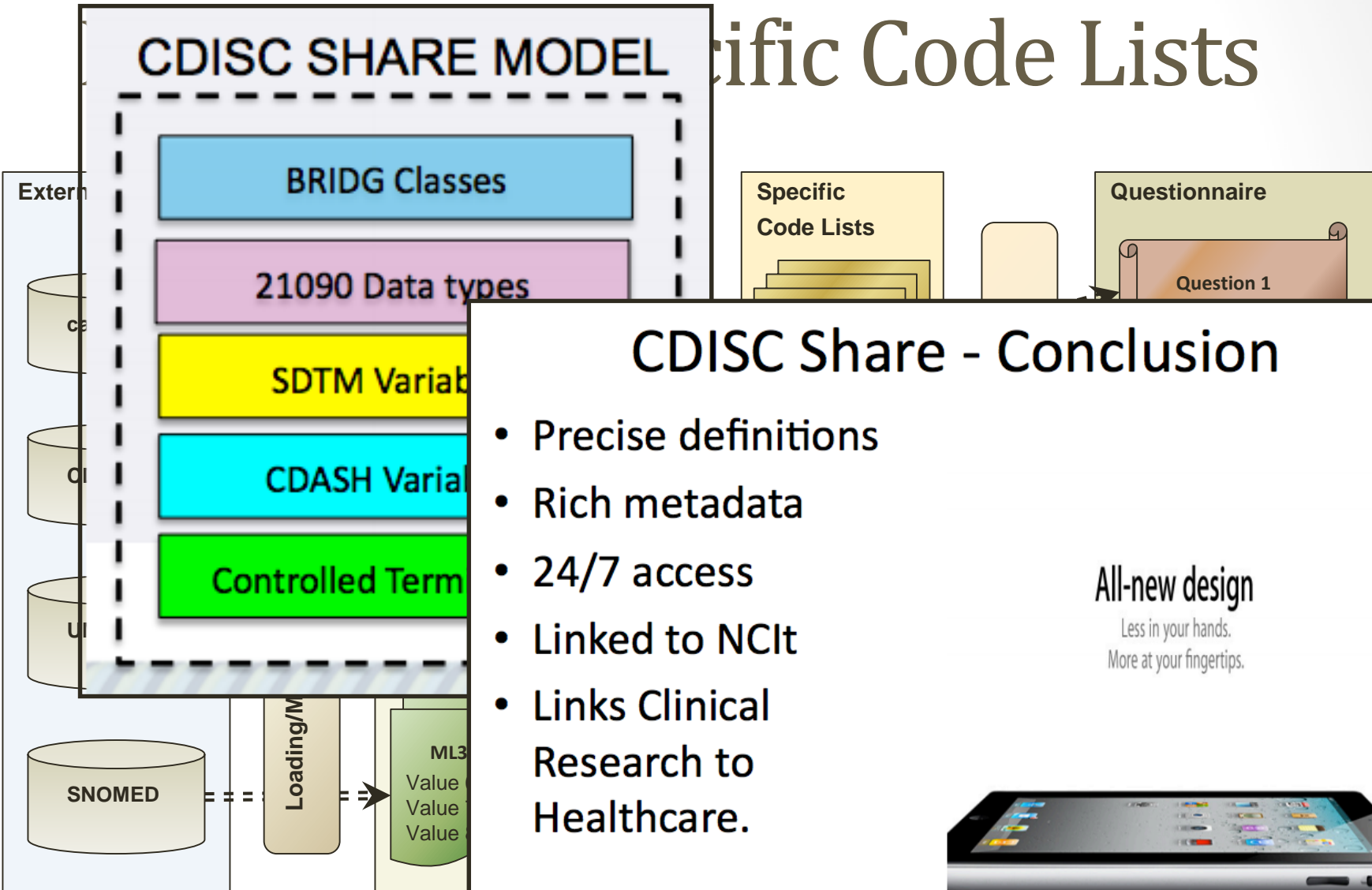
The UML models represent information domains and are developed by research partners in collaboration with CBIIT. The semantics of the data elements used in the models are extracted and transformed into administered components in the caDSR database. When a UML model is registered in caDSR, the collection of related CDEs are classified as part of the UML Model and visible as a collection in the UML Model Browser.

The Forms are developed by data managers supporting a variety of research projects and visible using the Form Builder tool or via the "Catalog of Published Forms" in the caBIG Context.

Use of CDEs addresses a biomedical data management problem, namely the many and varied ways in which similar or identical concepts have been collected and stored in databases. This inconsistency in data representation makes it nearly impossible to aggregate and manage even modest-sized data sets in order to ask basic questions and obtain meaningful answers. Common information building blocks or "common data elements," used for capture of data and for reporting, facilitate understanding and sharing of cancer research information. Using registered metadata facilitates interoperability of the data collected by disparate applications developed in research centers and deployed on [caGrid](#).

For in depth information about caDSR, including links to documentation and contacts as well as technical background and product status, visit the [caDSR wiki](#) .

Enter an Approach called Specific Code Lists



CDISC Share - Conclusion

- Precise definitions
- Rich metadata
- 24/7 access
- Linked to NCIt
- Links Clinical Research to Healthcare.

All-new design

Less in your hands.
More at your fingertips.



Enter an Approach called



UMLS®

The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records. [More information...](#)

[Metathesaurus License](#)

[UTS](#)

[Downloads](#)

[Source Documentation](#)

[UMLS® Reference Manual](#)

Requires login.

Quick Links:

New Users

- [UMLS Quick Start Guide](#)
- [Licensing Information](#)
- [Basics Tutorial](#)
- [More...](#)

UMLS Knowledge Sources

Documentation for:

- [Metathesaurus](#)
- [Semantic Network](#)
- [SPECIALIST Lexicon and Lexical Tools](#)
- [More...](#)

UMLS News and Announcements

RxNorm Overview updated...

[Subscribe to the UMLS News RSS Feed.](#)

User Education

- [UMLS Video Learning Resources](#)
- [Glossary](#)
- [Presentations](#)
- [More...](#)

Implementation Resources

For advanced users:

- [MetamorphoSys](#)
- [Database Query Diagrams](#)
- [Load Scripts](#)
- [More...](#)

Related Resources

- [MeSH®](#)
- [RxNorm](#)
- [SNOMED CT®](#)
- [SNOMED CT CORE Subset](#)

Enter an Approach called

VTSL

SNOMED CT Core Browser

Concepts

Parent(s):

(Select a parent to make it the "Current Concept")
[Allergy to biocide \(disorder\)](#)

Current Concept
[Pesticide allergy \(disorder\)](#)

Child(ren):

(N=7) (Select a child to make it the "Current Concept")
[Benzyl benzoate allergy \(disorder\)](#)
[Carbamate pesticide allergy \(disorder\)](#)
[Chlorinated pesticide allergy \(disorder\)](#)
[Monosulfiram allergy \(disorder\)](#)
[Organophosphate pesticide allergy \(disorder\)](#)
[Pyrethroid pesticide allergy \(disorder\)](#)
[Warfarin allergy \(disorder\)](#)

This Browser is built and maintained by:



Virginia-Maryland Regional College of Veterinary Medicine

This browser is designed to search and browse SNOMED CT (January 2011 release).

This material includes SNOMED Clinical Terms® (SNOMED CT®) which is used by permission of the International Health Terminology Standards Development Organisation (IHTSDO). All rights reserved. SNOMED CT®, was originally created by The College of American Pathologists. "SNOMED" and "SNOMED CT" are registered trademarks of the IHTSDO

[- All Is a antecedents -](#)

[- All descendents and related subtypes -](#)

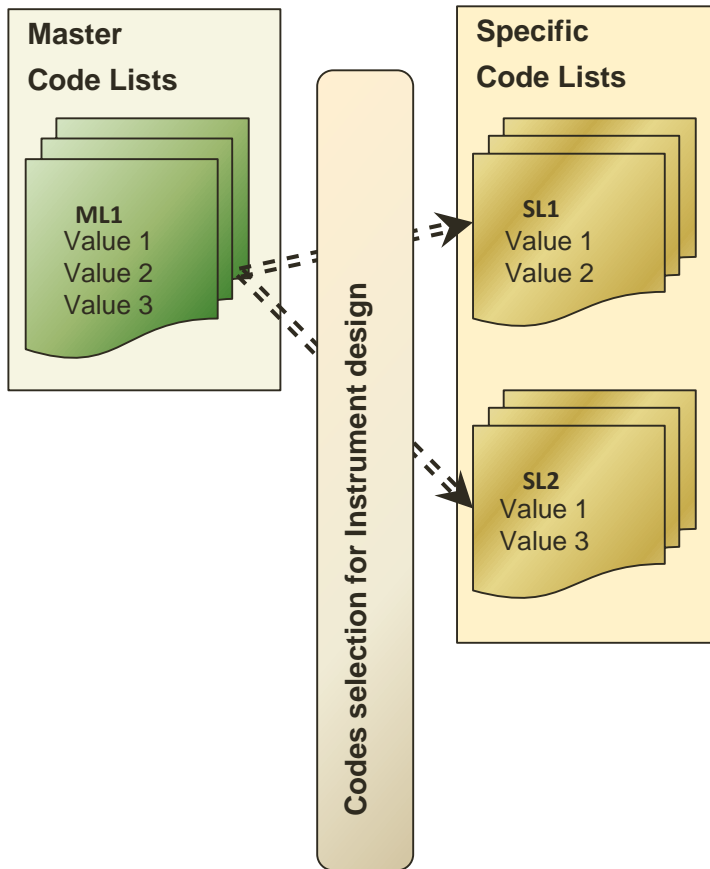
[Pesticide allergy \(disorder\)](#)
4619002

[Allergy to biocide \(disorder\)](#)
[Benzyl benzoate \(substance\)](#)

cts

[Pesticide allergy \(disorder\)](#)
[Pesticide allergy](#)

Master and Specific Code Lists in Action



PA004/(WHO_APPLY). Who applied this product? Was it....

You	1
A friend or family member	2
Building maintenance, or	3
A professional exterminator?	4
REFUSED	-1
DON'T KNOW	-2

PA004/(WHO_APPLY). Who applied this product? Was it....

You	1
{C_FNAME}	2
A friend or family member	3
Building maintenance, or	4
A professional exterminator?	5
REFUSED	-1
DON'T KNOW	-2

Harmonization...

Internal

External

PA004/(WHO_APPLY). Who applied this product? Was it...

You	1
A friend or family member	2
Building maintenance, or	3
A professional exterminator?	4
REFUSED	-1
DON'T KNOW	-2

PA004/(WHO_APPLY). Who applied this product? Was it...

You	1
{C_FNAME}	2
A friend or family member	3
Building maintenance, or	4
A professional exterminator?	5
REFUSED	-1
DON'T KNOW	-2

PESTICIDE_APPLICATOR_ML0901

Master Category	Globally Universal Identifier (GUID)	Concept ID (NCI Meta Thesaurus)
Respondent	e562d7ff..	C0282122
Friend or family member	b0e18823..	C1709486
Maintenance	7e852cab..	C0335350
Exterminator	f9c090e5..	****
Subject (Child)	5be1616a..	C0681850

NCI metathesaurus

Exact Match
 Begins With
 Contains

Name/Code
 Property
 Relationship

Source

[Home](#) | [NCIt Hierarchy](#) | [Sources](#) | [Help](#)
[Visited Concepts](#)

Quick Links

Respondent (CUI C0282122)

[Suggest changes to this concept](#)
[Add to Cart](#)

- Terms & Properties
- Synonym Details
- Relationships
- By Source
- View All

Terms & Properties

Concept Unique Identifier (CUI): C0282122

Semantic Type: Population Group

Synonyms & Abbreviations: [\(see Synonym Details\)](#)

Respondent
Respondents

External Source Codes: *(none)*

Other Properties:

TERMUI	T010562	MSH
TH	NLM (1996)	MSH

Additional Concept Data: *(none)*

URL to Bookmark: <http://ncim.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI%20MetaThesaurus&code=C0282122>

Section II – Sophia Kuan

DDI SUPPORT FOR THE PROPOSAL

Category and Code Schemes

- In DDI code lists are represented using Category and Code Schemes
- A 'scheme' is a bank /collection/list of similar items
- A Category represents a question response domain that does not have code representations assigned to it, e.g.

Question: What is your marital status

Response Domain:

- **Single**
- **Married**

- Code schemes organize categories from one or more Category Schemes and provide the code representation for the category as it is found in the question or variable:

- 1** **Single**
- 2** **Married**

Category Schemes

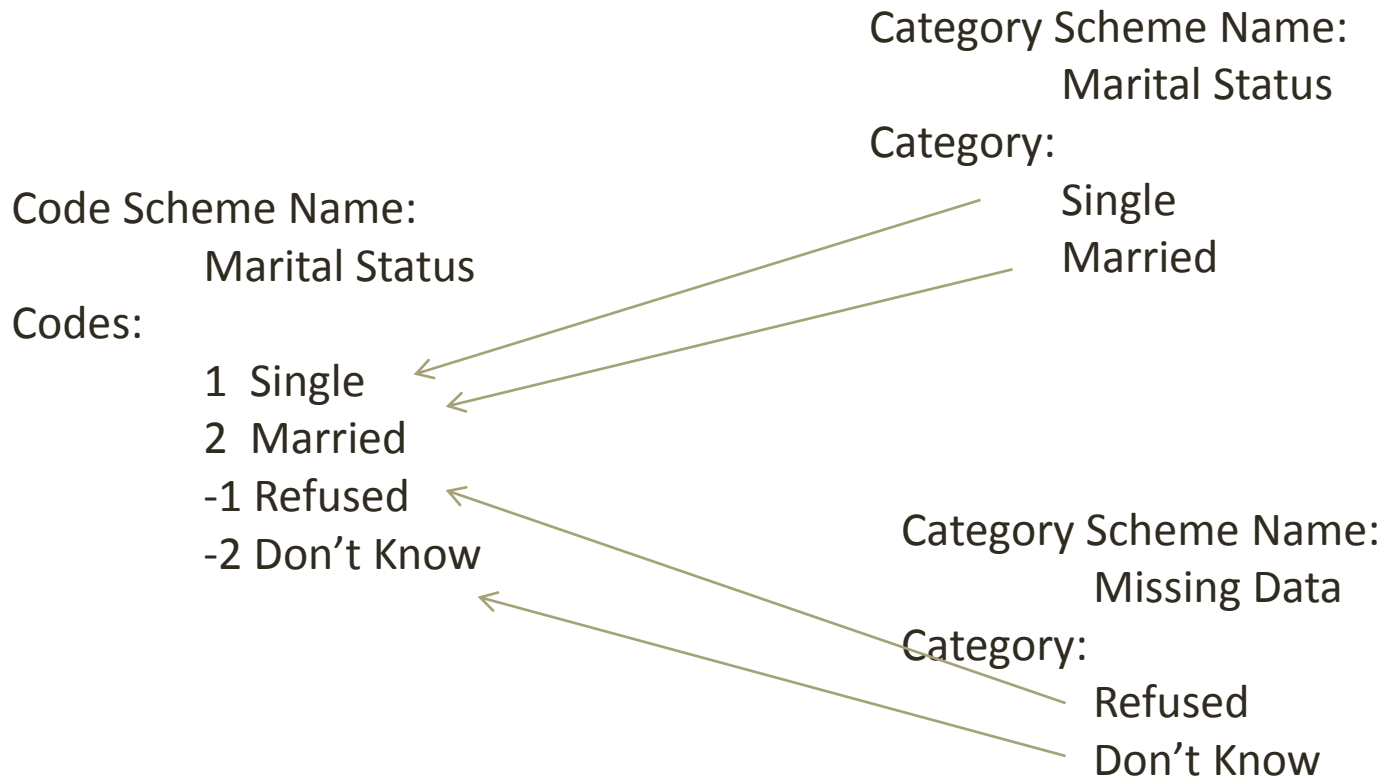
Category Scheme Name: Marital Status
Category Scheme ID: 1
Category 1 (id=1): Single
Category 2 (id=2): Married

Category Scheme Name: Language
Category Scheme ID: 2
Category 1 (id=1): English
Category 2 (id=2): Russian
Category 3 (id=3): French

Category Scheme Name: Missing Data
Category Scheme ID: 3
Category 1 (id=-1): Refused
Category 2 (id=-2): Don't Know

Code Schemes

Code Schemes can be assembled from different Category Schemes



DDI promotes reuse

- Happily, Master Code List and Specific Code List can be mapped to Category and Code Scheme, respectively
- Category Schemes can be declared once and used in many Code Schemes by reference

```
<|:CategoryScheme id="111">
  |:CategorySchemeName>MARITAL_STATUS</|:CategorySchemeName>
  |:Description>Marital status</|:Description>
  |:Category id="1">
  | |:Label>Single</|:Label>
  |</|:Category>
  |:Category id="2">
  | |:Label>Married</|:Label>
  |</|:Category>
</|:CategoryScheme>
```

```
<|:CodeScheme id="222">
  |:CodeSchemeName>MARITAL_STATUS</|:CodeSchemeName>
  |:Description>Person's Marital Status</|:Description>
  |:Code>
  |:CategoryReference>
  | |:ID>1</|:ID>
  |</|:CategoryReference>
  |:Value>1</|:Value><!-- This is the local Value-->
</|:Code>
```

Section III – Alexandra Shlionskaya

BENEFITS AND CHALLENGES

Benefits of two-level approach

- Master Lists can be connected to external sources, but a system that uses Master lists will keep relative independence from a structure, formats, access rules and change management procedures of external sources.
- Master Lists may be structured in a more sophisticated way, for example, as taxonomies for terms grouping or ontologies for various relationships to support data analysis, browsing, and querying, while related Specific Lists could have simple, flat structure for practicality and ease of coding and data collection procedures.
- **Code Lists can be used consistently for all data related processes throughout the data life cycle, including various types of data collection, consistency checks, analysis and reporting.**

CHALLENGES OF CODE LISTS IMPLEMENTATION

Why it is challenging to

- **Achieve maximum reusability of Specific Code Lists?**
 - **Synonyms:**
 - Frequency -
 - Once a day
 - One time per day
 - Daily
 - **Context-specific set of answers :**
 - In guided interview –
 - Yes
 - No
 - **Refused**
 - Don't know
 - In self-administered interview –
 - Yes
 - No
 - **Prefer not to answer**
 - Don't know

Why it is challenging to

- **Define Master lists in the most logical and usable way?**
 - Multiple meanings depending on context
 - Tablespoon – medication unit measure
 - Tablespoon - silverware
 - Uncertainty about appropriate granularity level
 - Childcare center – Organization
 - Childcare center –Child care organization
 - Values that are difficult to categorize
 - Broken vacuum cleaner

Why it is challenging to

- **Keep values in just one Master list?**
 - Responses to different questions –same spelling/different meaning - homonyms
 - Outreach target culture – Russian
 - What is the language you speak at home? – Russian

Why it is challenging to

- **Link to external sources?**
 - There is no single authoritative source of relevant terminology
 - Multiple external vocabularies that are pairwise inconsistent
 - Most vocabularies contain only a subset of terms that a study needs
 - Search for terms is often challenging process that is labor intensive and time consuming

TERMINOLOGY HARMONIZATION AS A SOLUTION

Terminology Harmonization

Harmonization: Enables interoperability and data exchange which then allows aggregation and comparison of data collected at different times and using different protocols resulting in richer data analysis.

Internal Code Lists and Data Elements harmonization process - ensure consistency of data across a study – define mapping and relationships between two internal code lists

External Code List and Data Element harmonization enable data interoperability across multiple data sources – define mapping and relationships between local code list values and external vocabulary

DDI Comparison as harmonization mechanism

Comparisons are used to compare two DDI entities of the same type

DDI entity	Study entity	Mapping type	Order
Concept	Instrument	ConceptMap	1
Variable	Data Element	VariableMap	2
Question	Question	QuestionMap	3
Category	Master Code List	CategoryMap	4
Code	Specific Code List	CodeMap	5
Universe	Population	UniverseMap	6

Code lists Harmonization

Local

DE11:	Person Language
CL11:	LANGUAGE_CL1
Values:	1-English
	2-Spanish
	3-French
	4-Farsi
	5-Chinese

ISO 639

DE11:	Spoken Language
CL11:	LANGUAGE_CL2
Values:	1-English
	2-Spanish
	3-French
	4-Persian
	5-Mandarin

Same as

Broader than

```

<cm:CategoryMap id="1000000010">
  <cm:SourceSchemeReference>
    <r:URN>Chinese</r:URN>
  </cm:SourceSchemeReference>
  <cm:TargetSchemeReference>
    <r:URN>Mandarin</r:URN>
  </cm:TargetSchemeReference>
  <cm:Correspondence>
    <cm:Commonality>Same group</cm:Commonality>
    <cm:Difference>Broader than</cm:Difference>
  </cm:Correspondence>
</cm:CategoryMap>

```

Finally - it is essential to....

- Convey the idea to everyone in the pipeline to increase overall benefit
 - Communication through Governance is a key
 - If there are no Conventions, Rules and Responsibilities in place that are clearly communicated to everyone involved - the message is going nowhere

That's all folks...

shlionskaya_alexandra@bah.com

kuan_sophia@bah.com

greenfield_jay@bah.com

