



Concept Sets in Secondary-Use Health Analytics

Sigfried Gold

July 17, 2018



The problem

Ad hoc definition of concept sets:

- Wastes effort and knowledge
- Doesn't encourage validation and documentation
- Can lead to large, unintended effects on study results

Concept sets link vocabulary values to everything else

*We can reuse them.
Do we?*

The screenshot displays the ATLAS web application interface. On the left is a dark navigation menu with the following items: Home, Data Sources, Vocabulary, Concept Sets, Cohort Definitions, Incidence Rates, Profiles, Estimation, and Prediction. A large blue arrow on the left points from the navigation menu towards the main content area. To the right of the menu, the 'Concept Sets' page is shown. It features a 'New Concept Set' button, a search filter, and a table of concept sets. The table has columns for Id, Title, Created, Modified, and Author. The 'Showing 1 to 10 of 2,248 entries' text at the bottom of the table has the number '2,248' circled in red. Below the table is an 'Export Mode' button.

Id	Title	Created	Modified	Author
1769872	COPY OF: Dx.T1DM.eMerge	07/16/2018 3:02:33 PM	07/16/2018 3:02:33 PM	anonymous
1769873	Surgery	07/16/2018 9:47:39 PM	07/16/2018 10:05:20 PM	anonymous
1769871	[c2q] AD	07/14/2018 5:13:21 PM	07/14/2018 5:13:21 PM	anonymous
1769870	TestAbcd	07/12/2018 6:49:29 PM	07/12/2018 6:49:29 PM	anonymous
1769868	T2DiaHCAbcd	07/12/2018 1:25:18 PM	07/12/2018 2:54:22 PM	anonymous
1769864	T2DiaConceptAbcd	07/11/2018 7:24:09 PM	07/12/2018 1:15:13 PM	anonymous
1769866	PMN_bw	07/12/2018 10:49:08 AM	07/12/2018 10:49:08 AM	anonymous
1769863	[C2Q]myocardial infarction	07/11/2018 4:01:10 PM	07/11/2018 4:01:10 PM	anonymous
1769862	[C2Q]symptomatic congestive heart failure	07/11/2018 4:01:07 PM	07/11/2018 4:01:07 PM	anonymous
1769861	[C2Q]active angina	07/11/2018 4:01:04 PM	07/11/2018 4:01:04 PM	anonymous



AMIA Symposium paper

Clinical Concept Value Sets and Interoperability in Health Data Analytics

Sigfried Gold, MFA, MA^{1,2}; Andrea Batch, MA¹; Robert McClure, MD³; Guoqian Jiang, MD, PhD^{4,2}; Hadi Kharrazi, MD, PhD⁵; Rishi Saripalle, PhD⁶; Vojtech Huser, MD, PhD^{7,2}; Chunhua Weng, PhD^{8,2}; Nancy Roderer, MLS¹; Ana Szarfman, MD⁹; Niklas Elmqvist, PhD¹; David Gotz, PhD¹⁰

¹University of Maryland, College Park; ²Observational Health Data Sciences and Informatics; ³MD Partners, Lafayette, CO; ⁴Mayo Clinic, Rochester, MN; ⁵Johns Hopkins University, Baltimore, MD; ⁶Illinois State University, Normal, IL; ⁷National Library of Medicine, Bethesda, MD; ⁸Columbia University, New York, NY; ⁹Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD; ¹⁰University of North Carolina, Chapel Hill, NC.

Abstract

This paper focuses on *value sets* as an essential component in the health analytics ecosystem. We discuss **shared repositories of reusable value sets** and offer **recommendations for their further development and adoption**. In order to motivate these contributions, we explain **how value sets fit into specific analytic tasks and the health analytics landscape** more broadly; their **growing importance and ubiquity with the advent of Common Data Models, Distributed Research Networks**, and the availability of higher order, reusable analytic resources like electronic phenotypes and electronic clinical quality measures; the **formidable barriers to value set reuse** and our introduction of **a concept-agnostic orientation to vocabulary collections**. The costs of ad hoc value set management and the benefits of value set reuse are described or implied throughout. Our standards, infrastructure, and design recommendations are not systematic or comprehensive but invite further work to support value set reuse for health analytics. *The views represented in the paper do not necessarily represent the views of the institutions or of all the co-authors.*

http://sigfried.org/writing/sgold_et_al_amia_2018.pdf. This is a pre-copy-editing, author-produced PDF of a paper accepted for presentation at the AMIA 2018 Annual Symposium following peer review. The definitive publisher-authenticated version is not yet available. (S61: Oral Presentation - Management Information Systems on November 6, 2018 from 8:30 AM to 10:00 AM.)



Value set uses

- Data element selection options — e.g., EHR dropdown
- Validation — e.g., of code systems against authorized, versioned vocabulary resources
- **Query** — codes/concepts treated as equivalent for use in a clinical query or analytic task



Characteristics of a Formal Value Set Definition

Release 1. June 2016.

[http://www.hl7.org/implement/
standards/product_brief.cfm?
product_id=437](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=437)



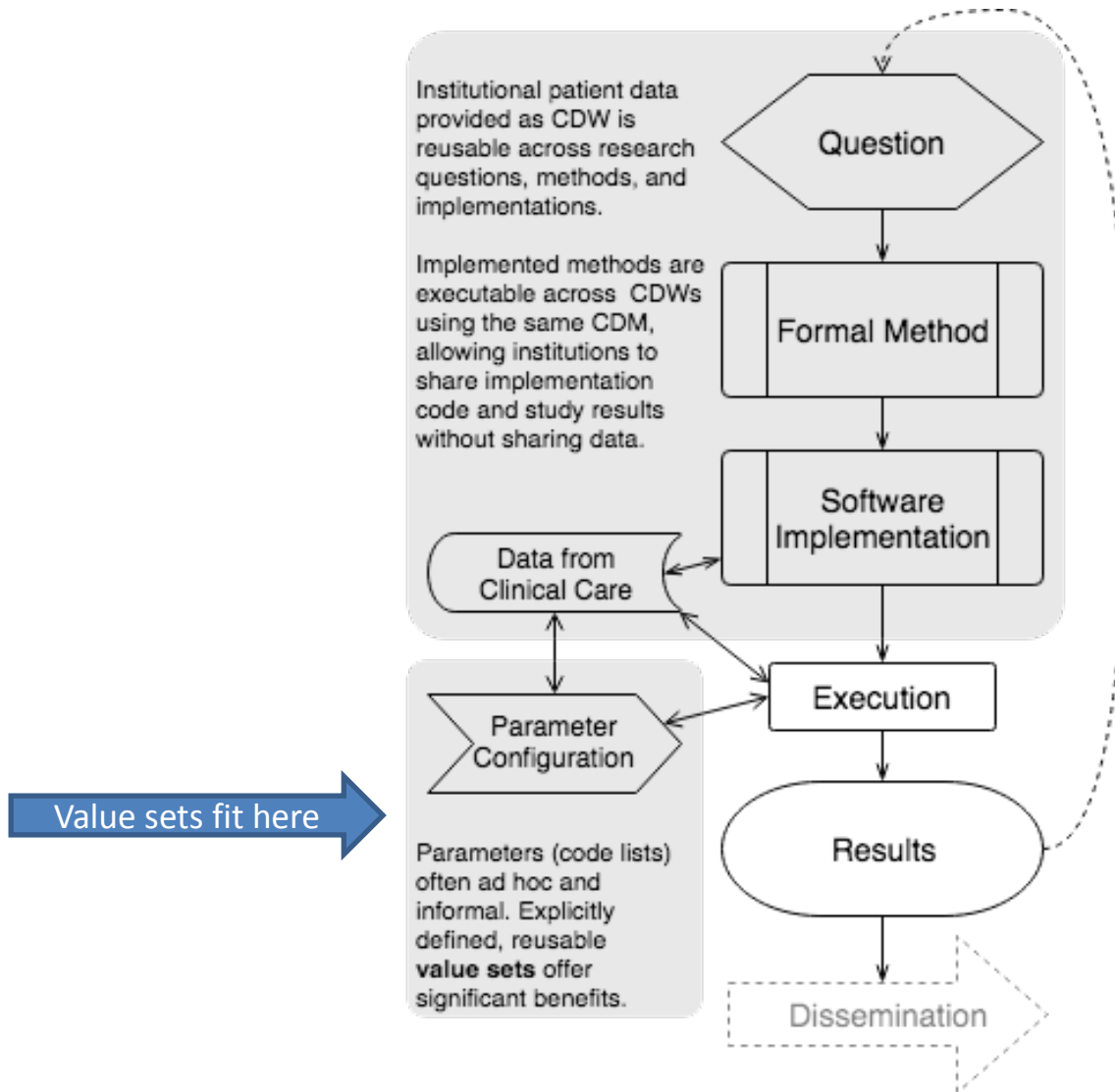
HL7 VSD relevance to OHDSI

4	CONTROLLED VOCABULARIES AND VALUE SETS	12
4.1	CONCEPTS, CODE SYSTEMS AND VALUE SETS	13
4.2	CODED DATA TYPES IN HL7	13
4.3	TERMINOLOGY BINDING	14
4.4	VALUE SET DEFINITION CONSISTENCY IN HL7 MODELS	15
4.5	VALUE SET AUTHORING AND MAINTENANCE	16
5	VALUE SET DEFINITION SPECIFICATION	16
5.1	VALUE SET ARCHITECTURE	16
5.1.1	GENERAL OVERVIEW	16
5.1.2	RELATIONS	6
5.1.3	ALIGNMENT	6
5.1.4	CLASS MODE	6
5.2	ELEMENTS	6
5.2.1	IMPLIED CC	6
5.2.2	VALUE SET	6
5.2.2.1	Value Set	6
5.2.2.2	Value Set	6
6	CONTENT LOGICAL DEFINITION	34
6.1	CONTENT LOGICAL DEFINITION GENERAL MODEL	35
6.2	CONTENT LOGICAL DEFINITION - ELEMENTS	36
6.2.1	LOCKEDDATE	36
6.2.2	ACTIVEONLY	38
6.2.3	CLDSYNTAXREFERENCE	38
6.2.4	CONTENT EXPRESSION	39
6.2.4.1	Syntax-based Content Expressions	39
6.2.4.2	Non-computable Content Expression	40
7	AN HL7 VALUE SET DEFINITION EXPRESSION SYNTAX	40
7.1.1	CONTENT DEFINING ELEMENT TYPES	42
7.1.1.1	CodeSystemElement	42
7.1.1.1.1	DrawnFromCodeSystem	42
7.1.1.1.2	CodeBasedContentSet	42
7.1.1.1.2.1	CodeBasedContent	42
7.1.1.1.2.1.1	Code	42
7.1.1.1.2.1.2	IncludeRelatedCodes	42
7.1.1.1.3	PropertyBasedContentSet	42
7.1.1.1.3.1	IncludeWithProperty	42
7.1.1.1.3.1.1	Name	42
7.1.1.1.3.1.2	Value	42
7.1.1.1.3.1.3	Expression	42
7.1.1.2	RelationshipBasedContent	42
7.1.1.2.1	RelationshipType	42
7.1.1.2.2	MinimumMultiplicity	42
7.1.1.2.3	MaximumMultiplicity	42
7.1.1.2.4	TargetConcepts	42
7.1.1.3	CodeFilterContent	42
7.1.1.3.1	ExpressionType	42
7.1.1.3.2	Expression	42
8	VALUE SET EXPANSION FILE	55
8.1	METADATA NEEDED TO DESCRIBE A VALUE SET EXPANSION	57
8.2	VALUE SET DEFINITION IDENTIFIER AND VERSION	57
8.2.1	VALUE SET IDENTIFIER	57
8.2.2	VALUE SET DEFINITION VERSION	57
8.3	DATE OF EXPANSION	57
8.4	CODE SYSTEM(S) AND VERSION	58
8.4.1	CODESYSTEMFOREXPANSION	58
8.4.1.1	CodeSystem	58
8.4.1.2	VersionDate	58
8.4.1.3	VersionString	59
8.4.1.4	DescriptiveName	59
8.4.1.5	UsesCodeSystemPartition	59
8.4.1.6	UsesCodeSystemSupplement	60
8.5	EXPANSION STEWARD	60
8.6	EXPANSION CONCEPT ATTRIBUTES	60
8.6.1	SORTKEY	61
8.7	EXPANSION IDENTIFIER	61
9	IMPLEMENTATION CONSIDERATIONS	62
9.1	IMPLEMENTATION TECHNOLOGIES	62
9.2	AUTHORING	62
9.3	REUSE OF VALUE SETS	63
9.4	DISTRIBUTION	63
9.5	IMPACT OF CODE SYSTEM EVOLUTION	63

- OHDSI Architecture Work Group addressing concept set specification.
- Is HL7 VSD overkill? “Shelf-ware”?
- Maybe, but...
- Provides foundation for FHIR’s simpler value set specs
- Years of thought and effort we could benefit from
- VSD authors essential to any discussion of widespread reuse
- Opportunities for interoperability and reaching more users



Secondary-use analysis task model





Values sets tie it all together

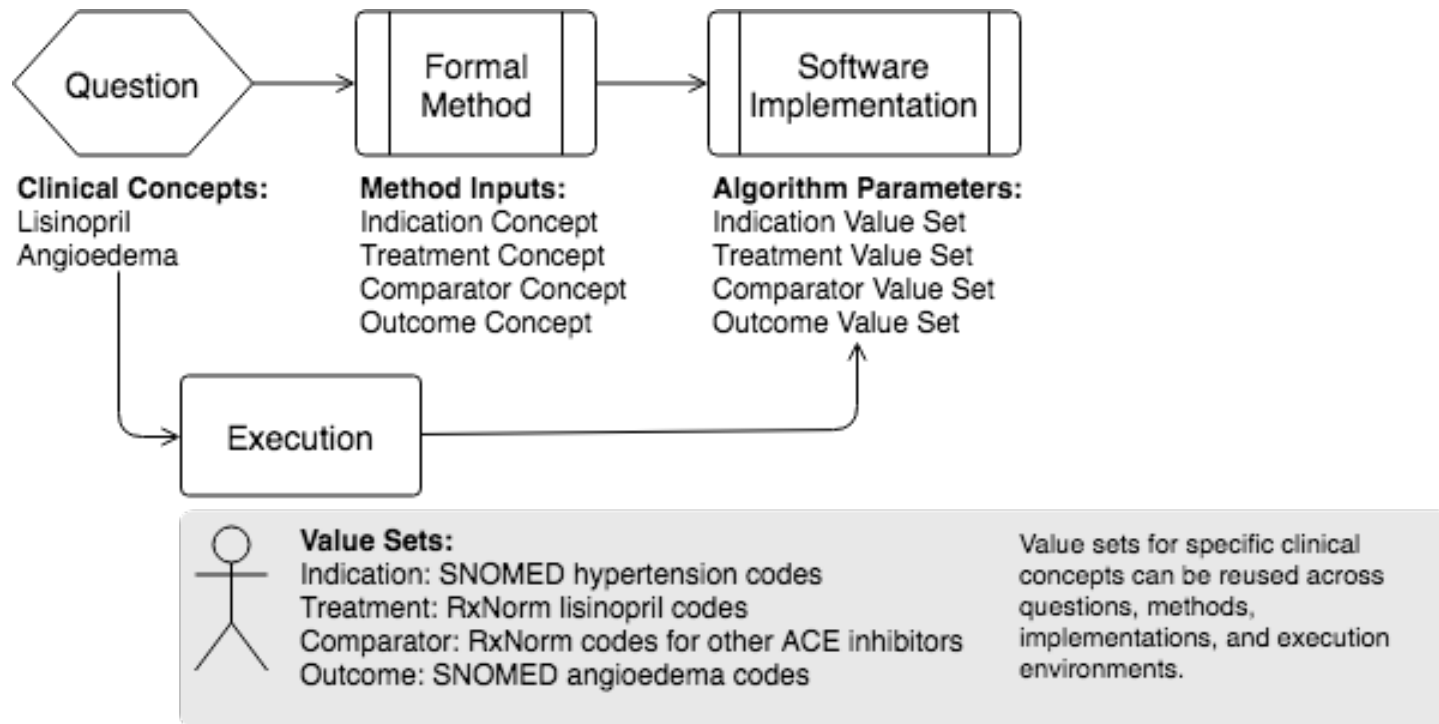


Figure 2. Value sets represent question concepts as parameters for method execution.

- Analyst, as researcher, asks, "Does lisinopril cause angioedema?"
- As biostatistician, chooses a formal effect estimation method.
- As data scientist or programmer, chooses or constructs implementation that require method input be specified as value sets.
- As informaticist or end user, chooses or constructs value sets to represent clinical concepts for execution.



Standards, Infrastructure, and Design Recommendations

- Requirements and specifications for definition syntax, metadata, traceable provenance, documentation, definition processing, and resolution
- Open standards, resources, and governance
 - Maximal openness consistent with users' licensed access
 - Affordances for both controlled and open value set management and curation
 - Certification and access control by institutional authorities (e.g., CMS, NCQA)
 - Crowdsourced curation through public commentary, usage statistics, and well-documented provenance trails
- Achieve network effects by encouraging synergistic cooperation on harmonization or consolidation projects
- Interactive, information-rich, high-performance visual interfaces
 - Modular components for integration into health analytics development environments and other analytic interfaces
 - Semantic graph visualization linked to local patient data