



The Value of Value Sets

Journal:	<i>2011 AMIA Clinical Research Informatics Summit</i>
Manuscript ID:	AMIA-042-C2011.R1
Manuscript Type:	Panel
Date Submitted by the Author:	29-Jan-2011
Complete List of Authors:	Wynden, Rob; UCSF, Academic Research Systems Solbrig, Harold; Mayo Clinic Tu, Samson; Stanford University, Center for Biomedical Informatics Research Brinkley, James; University of Washington, Biological Structure
Classification for Review Purposes:	Data Management, Sharing, and Representation

SCHOLARONE™
Manuscripts

The Value of Values Sets

Rob Wynden, BSCS¹; Harold Solbrig, BS²; Samson Tu, MS³; Jim Brinkley⁴, PhD

¹University of California San Francisco

²Mayo Clinic

³Stanford University

⁴University of Washington

Abstract

A common definition of value set will be provided and fully characterized relative to its proposed uses. We will describe, compare, and contrast several approaches to specifying and referencing value sets in a stable manner over time. The term “value set”, although ubiquitous within biomedical informatics has no common definition and has yet to be fully described in a formal manner. It is essential for the design and launch of new ontologies, biomedical informatics applications and data sharing environments that a common and well-understood definition of “value set” is provided. It is also essential that options and trade-offs be understood for what type of technology is appropriate for the implementation and usage of particular types of value set for particular use cases.

Introduction

We will discuss several potential solutions and important topics related to value sets. 1) A motivating example of why value sets need to be fully defined will be provided based on the Human Studies Database Project, using a trial on cardiovascular interventions (Aversano et al, JAMA 2002) as an example. 2) A review and theory of value sets will be provided including a summary of the technologies for defining and using value sets. 3) In order to use value sets with current technologies gaps in the transformation of value sets must be addressed. Those challenges will be identified.

4) We will fully characterize the various forms of terminological value sets as an enumeration of allowed terms that may reference each other and require cached instantiation. 5) We will define the use cases for which value sets are required. This will include lookup and constraints and we will define how value sets are used within EDC (Electronic Data Capture) platforms and how value sets are mapped from local codes to standard values. 6) We will discuss how value sets are best used for queries including the expressiveness of possible query-languages, ease of use and dynamic execution. 7) We will discuss what is and should be the format of value sets. This will include a recommendation of what formal flat file representations should be supported, how source OWL representations are used, and how an RDF graph may be generated.

The Need for a Robust Definition of Value Sets

Rob Wynden

University of California San Francisco

The practical utility of terminologies to biomedical informatics is dependent on the proper definition and usage of value sets. Without first addressing that issue, the transformation of ontologies into useful sets of terminologies and the effective usage of those terminologies within biomedical informatics applications may be difficult to achieve. As one example; ontologies within biomedical informatics may be modeled within OWL where

value constraints are expressed as axioms; whereas much of the current infrastructure that exists for EDC, biomedical application development, grid based query, common data element definition and value domain definition are all based on UML (Unified Modeling Language) and object relational technology. However, a robust and isomorphic means of transformation between OWL and UML does not currently exist. Without either solving this issue, or describing how to carefully avoid it, a seamless transition between these two sets of technologies will remain elusive.

The ISO 11179 Model of Value Domains

Harold Solbrig

Mayo Clinic, Rochester, MN

Harold Solbrig will discuss ISO 11179 Edition3 and its model of value domains, permissible values and value meanings. He will then summarize how this model has been translated into the HL7 domain, the grammar that was developed at HL7 to define value sets, and some of the issues that arise when permissible values and the identifiers for value meanings are not kept separate. He will also touch briefly on the CTS2 model of value set definitions and their resolution.

Competing Methods of Referencing Value Sets

Samson Tu

Stanford University

Central to the use of biomedical ontologies as terminologies for information models used in applications is a common and well-understood definition of value set. What are the means by which and ontology (a formalized conceptualization of a domain) provides value sets that can be referenced within biomedical applications? Should value sets be maintained within the ontology itself or should coding classes be created that houses instances of terms to form a dynamic value set? What forms of value set best support the later query

of data once collected? Should the query of value set information be based on the ontology or should queries run against the instance data from which value sets might be derived? A discussion of these topics and a comparison of possible approaches to these problems will be discussed.

SparQL-based Views and Value Sets

Jim Brinkley

University of Washington

A value set can be described as a fixed list of coded terms but also as a derived list based on rules. But these rules for derived value set lists can become quite complex and computationally expensive to maintain. Also, the set of permissible values can be quite large and may require that the values they contain be cached on the terminology server; otherwise their reference from within biomedical applications would take too long to be useful. Dr. Brinkley will show an approach whereby value sets can be generated from RDF sources (including OWL sources such as OCRE) using a query-based approach that is highly expressive albeit potentially complex for an end-user. The queries can be saved in a View Query Manager, then re-run at any time via a REST service, which means that a given value set can be accessed from a data acquisition or data integration application via a URI. He will show example queries and result sets for the specific use cases discussed in the panel, and will participate in discussions comparing this approach presented by other panelists.

Participation of Panelists

All panelists have agreed at the time of submission to participate in this panel, and have prepared and approved the included topic summary and presentation overviews.