



MAGE: a caBIG Data Standard

Dianne M. Reeves

Biomedical Data Standards

National Cancer Institute Center for Bioinformatics



What is a Standard?

- **Standard** - The set of parameters we use to determine what is acceptable. A level of expectations designed to bring out the best of us. A realistic, consistent guideline set to help us attain our goals.
 - caBIG Annual Meeting, Best Practices SIG Breakout 2006
- **Standards** can be regulatory, funding, sponsor or practice based in nature



Today in Review

- **Ken Buetow:**
- NCICB brings biomedical informatics as an enabling tool, not just a solution to these problems.
- caBIG is committed to large scale data sharing.
- caBIG does not set standards, but identifies and uses existing standards.

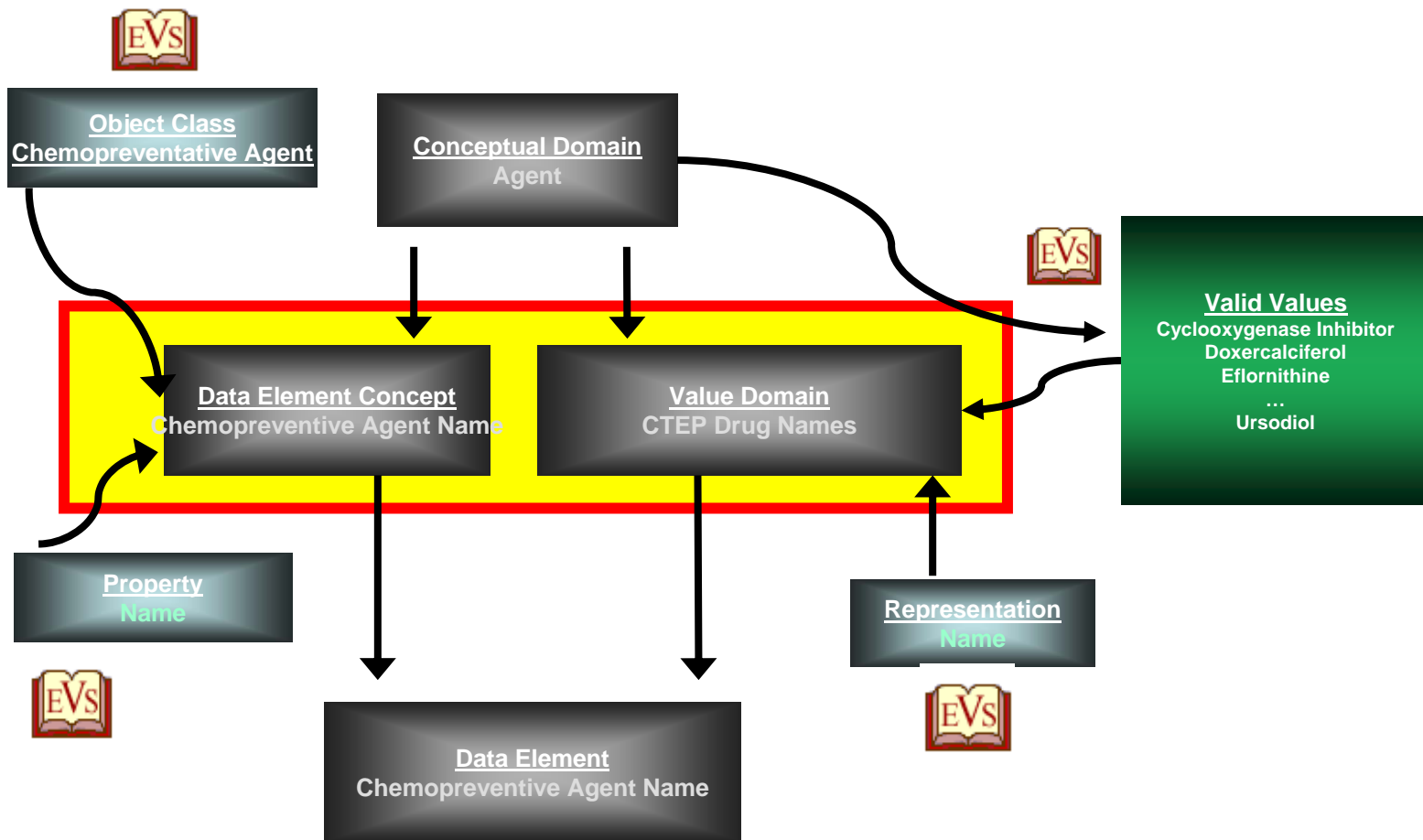


Today in Review

- **George Komatsoulis:**
- **Interoperability** is the ability of a system to access and use the parts or equipment of another system.
- We devise a set of guidelines for an interoperable data system.
- Levels of interoperability – bronze through gold
 - caBIG compatibility reviews (VCDE)
- Architecture and VCDE use Mentors to assist with compatibility guidance.



caDSR Implementation of ISO/IEC 11179 Model





Interoperability

ability of a system to
access and use the
parts or equipment of
another system

Syntactic
interoperability

Semantic
interoperability



Maturity Model	Legacy	Bronze	Silver	Gold
Interface Integration	<ul style="list-style-type: none"> - No Programming interfaces to the system are available. Only local data files in a custom format can be read - Some ad hoc data transfer mechanism such as FTP 	<ul style="list-style-type: none"> - Provide baseline* programmatic access to data. Data can be read from remote electronic sources or from commonly used file formats Data can be pushed out to from applications to other external data sources 	<ul style="list-style-type: none"> - Well-described API's that provide access to data objects. - System architecture separated into tiers and interoperable components - Data read in from standards-based electronic sources that support standard or commonly used interchange formats - Documented component description of the underlying data structures that are accessible - Standard messaging systems where appropriate 	<ul style="list-style-type: none"> - All features of Silver, plus: - Interoperable with data grid architecture to be defined by caBIG - Fully componentized provide access to individual resources in the form of grid services
Vocabularies / Terminologies & Ontologies	<ul style="list-style-type: none"> - Free text used throughout for data collection 	<ul style="list-style-type: none"> - Use of publicly accessible standardized controlled vocabularies as well as local terminologies 	<ul style="list-style-type: none"> - Standard terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace are used for all relevant data collection fields. 	<ul style="list-style-type: none"> - All features of Silver, plus: - Fully compliant with caBIG recommended standards for vocabulary terminology services and content sources
Data Elements	<ul style="list-style-type: none"> - No Structured metadata is recorded 	<ul style="list-style-type: none"> - Some type of metadata describing the information in the system is used for data collection and external reporting. Metadata is retrieved from external repository shared by multiple applications. - Common Data Elements should be built using controlled terminology 	<ul style="list-style-type: none"> - Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard - CDEs are harmonized and re-used from across the Domain Workspace - Common Data Elements are built using standard controlled terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace 	<ul style="list-style-type: none"> - All features of Silver, plus: - Programmatic access to all metadata, including data class descriptions, site and source information, and any other caBIG-defined metadata requirements and use information models - Use the caBIG standard or electronic representation of metadata and Common Data Elements
Information Models	<ul style="list-style-type: none"> - No particular information model is used to represent data 	<ul style="list-style-type: none"> - Some type of diagrammatic model describing the data relationship is available in electronic format 	<ul style="list-style-type: none"> - Information models defined in a standard modeling language such as UML 	<ul style="list-style-type: none"> - All features of Silver, plus: - Information models are harmonized with other s across the caBIG Domain Workspace



Today in Review

- **Carl Schaefer:**
- MAGE was accepted for Microarray data – via ICR
- ICR is the largest Workspace in caBIG



Today in Review

Catherine Bell:

- ? What are the next steps?

Chris Stockert:

- MGED-Ontology provides a controlled vocabulary for MIAME annotation
- Importance of standards in relation to MGED-Ontology
- MGED-Ontology is explicitly tied to MAGE



MGED Ontology in EVS

NCI Terminology Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://nciterns.nci.nih.gov/NCIBrowser/Dictionary.do>

Please select a terminology

<input checked="" type="radio"/>	NCI_Thesaurus	Published by NCI, this knowledgebase contains the working vocabulary used in NCI data systems. It covers clinical, translational and basic research as well as administrative terminology.	March 2006 (06.03d)
----------------------------------	---------------	--	---------------------

Other Terminologies:

<input type="radio"/>	GO	GO Ontology, published by the Gene Ontology Consortium.	April 1, 2006
<input type="radio"/>	SNOMED_CT	Published by the College of American Pathologists, Systematized Nomenclature of Medicine Clinical Terms. Consists of SNOMED CT Jan-05 content loaded by Apelon from files received directly from CAP. The "5AA" part of the DTS version name represents the UMLS version from which CUI and Semantic Type values are obtained (UMLS 2005AA) and inserted as SNOMED CT properties in DTS. Licensed through National Library of Medicine, National Institutes of Health, Department of Health and Human Services (NLM). See http://www.nlm.nih.gov/research/umls/Snomed/snomed_license.html	2005.07.05AB
<input type="radio"/>	MGED_Ontology	The MGED Ontology, produced by the MGED Society. Concepts, definitions, terms, and resources for standardized description of a microarray experiment in support of MAGE v.1. The MGED ontology is divided into the MGED Core ontology which is intended to be stable and in synch with MAGE v.1; and the MGED Extended ontology which adds further associations and classes not found in MAGE v.1.	1.1.9
<input type="radio"/>	MedDRA	MedDRA (Medical Dictionary for Regulatory Activities). International medical terminology developed under the auspices of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). This online version of MedDRA is licensed by NCI for use within NCI. Users who are "NOT" NCI employees or contractors of NCI must obtain permission from the MedDRA MSSO before using MedDRA codes, terms or other content in their systems. Contact: MedDRA MSSO; 12011 Sunset Hill Road; Reston, VA 20190-3285; phone: 877-258-8280; e-mail: MSSOhelp@ngc.com; http://meddramsso.com	6.0
<input type="radio"/>	VA_NDFRT	Published by the US Veterans' Administration. National Drug File RT covers clinical drugs used at the VA. Not yet approved for use in clinical systems.	October 2004
<input type="radio"/>	LOINC	Published by the Regenstrief Institute, the Logical Observation Identifier Names and Codes covers clinical laboratory terminology.	local 3MN (11.01.01)
<input type="radio"/>	PRE_NCI_Thesaurus	This is a snapshot of the development version of the NCI Thesaurus. It is ONLY intended to facilitate early review by our collaborators. It is NOT for use in any production application, and is sometimes shut down without warning. It has NOT been processed for editing history. It has NOT gone through quality control. It is NOT necessarily a release candidate for publication. In the Terminology Browser, the "Printable Page" and "History" features will NOT work with this version of the Thesaurus.	April 2006 (06.04d)

Done Internet

start 6 Internet Explorer MAGE Slides update ~... Microsoft PowerPoint ... 2:41 PM



MGED Ontology as a Standard: Next Steps

- What is the role of MGED Ontology as a Standard?
- Version 1.3 is the latest version
- Every concept needs textual definition, CUI
- Versioning of ontology

Action Item is to move MGED Ontology as standard



MAGE-OM as a Standard: Next Steps

- What is the role of MAGE-OM as a standard?
 - **135 Classes with 145 Attributes in caDSR**
- What are the next steps?
 - Version 2.0 under development
 - Match model developers and VCDE Mentor
 - MAGE v2.0 is an extension to FuGE (ready for validation by end of 2006)



Questions?

Thank you for your attention!

That was Easy!