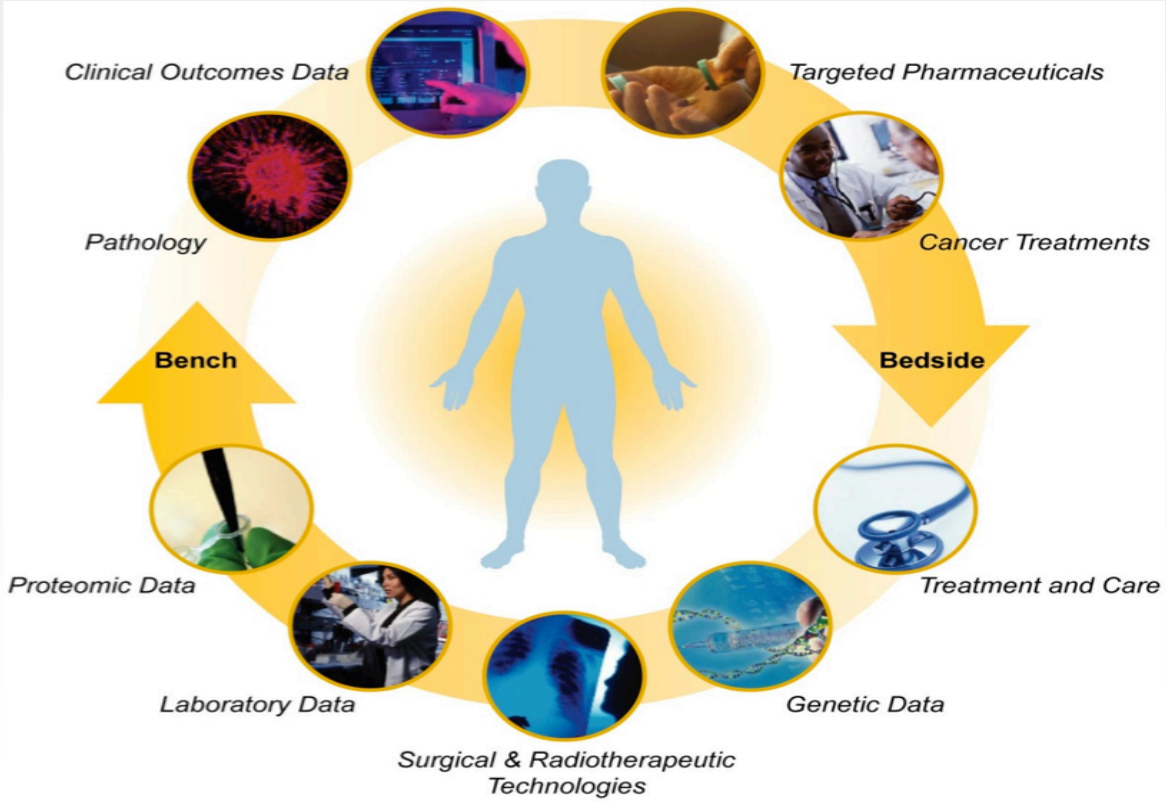
*RDF as **the** Healthcare Interchange Language*

Charles Mead, M.D., MSc., CoChair
W3C Healthcare Life Sciences Work Group

21st-Century "Personalized Medicine"



Used by permission: National Cancer Institute Cancer Bioinformatics Grid (caBIG)

Barriers to Achieving Computable Semantic Interoperability ("sharing semantics")

- Non-semantic
  - Representation technologies and traceability gaps (design-time)
  - Serialization brittleness(run-time)
  - Implementation technology diversity
- Semantic
  - Lack of "layered" standards
  - "Semantic overlap" at standards' boundaries

Non-Semantic Barriers to Achieving CSI

- Representation technologies and traceability gaps (design-time)
  - UML separation of views of instances from views of classes → modeling of domain content (instance data and schema/metadata) is often difficult for domain experts
    - Fido a dog
    - Dog rdf:subClassOf Mammal
  - UML tools non-standard mechanisms for associating exemplar but unbounded value sets with class attributes

- Requirements-to-Design-to-Implementation traceability is difficult
  - Disconnect between analysis-level definitional semantics and their computational “equivalence” when traced through design-time logical and run-time physical representations.

Non-Semantic Barriers to Achieving CSI(2)

- Serialization brittleness(run-time)
  - XML nested structure can obscure semantic equivalence
    - Simple example: Students, Classes, Teachers
    - Necessitates one-off Xpath, XSLT transforms for semantic alignment
- Implementation technology diversity
  - Different physical data models and/or access technologies for same schema
    - E.G. *raison d'être* for CDISC ODM exchange standard (XML)

Why RDF?

Semantics as a First-Class Citizen

- Representation of data (and metadata) in a RDF-based graph eliminates the problem of “semantics wrapped in syntactic structures”

- Transmission of RDF graphs utilizes existing internet standards, e.g. HTTP, REST, etc.

- Application of Linked Data Best Practices increases the value of data (and metadata) in both intranet- and internet-based communities.

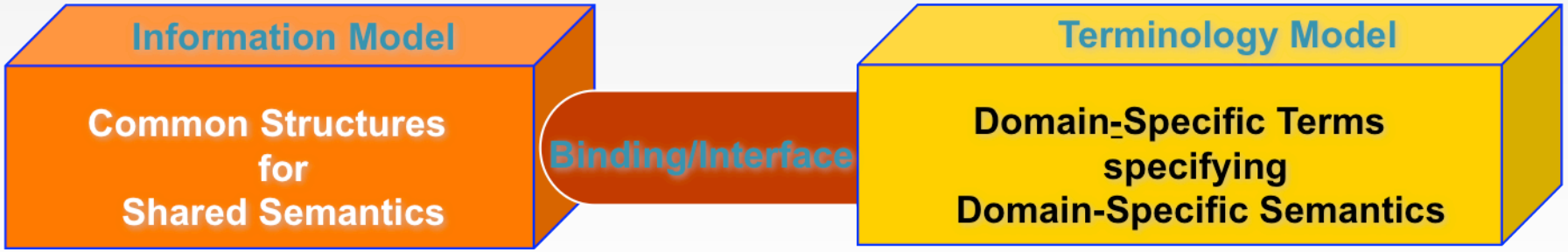
Semantics as a First-Class Citizen (2)

- Use of Semantic Web technologies ***does not*** magically solve core issues of semantic ambiguity or cross-graph disagreement.
  - Those issues must still be addressed through person-to-person/organization-to-organization dialogue.
- Use of SW technologies does, however, enable ***bottom-up*** harmonization/semantic resolution
  - Siloed ontology development if more amenable to harmonization than traditional RDBMS or XML siloed development

Information Models and Terminology Models: *The Perfect World*

"Grade IV anaphylactic reaction to Penicillin injection."
(HL7 Reference Information Model and SNOMED-CT)

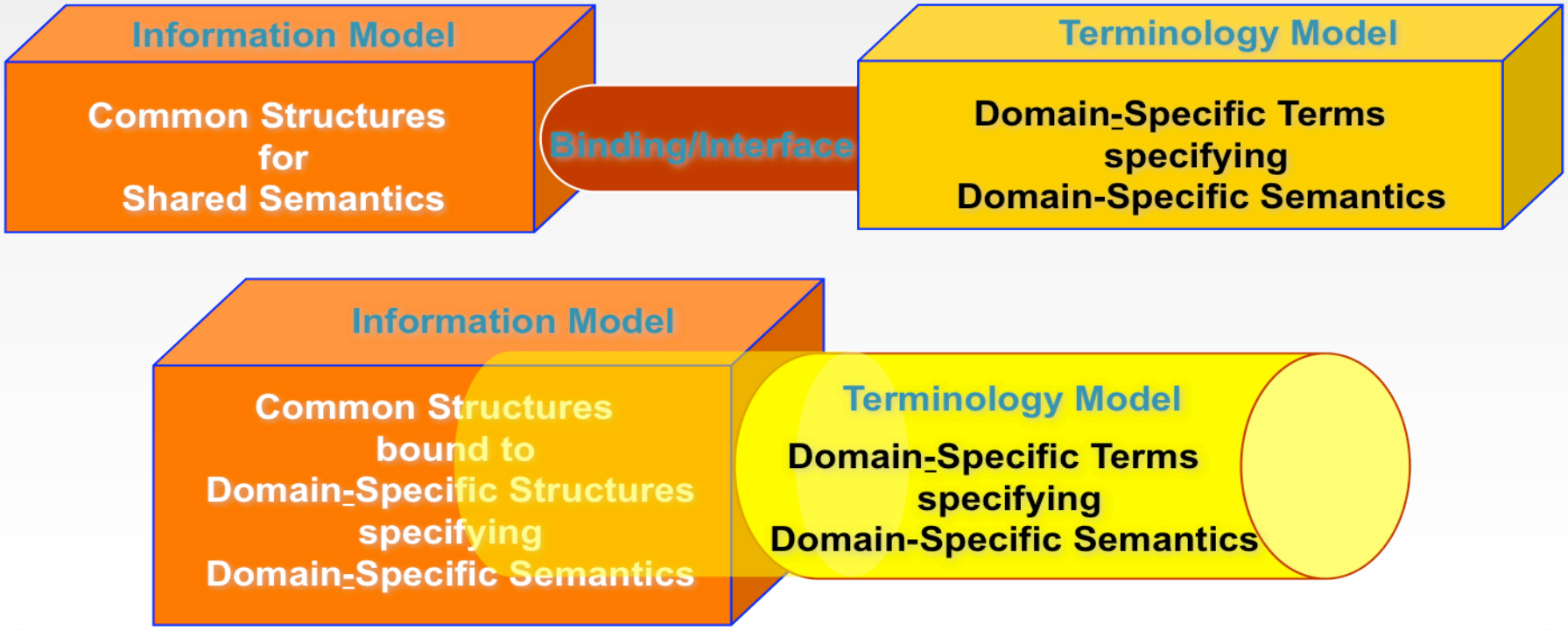
http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo



"Grade IV anaphylactic reaction to Penicillin injection."
(HL7 Reference Information Model and SNOMED-CT)

http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo

Information Models and Terminology Models:
The Real World (TermInfo problem)



TermInfo Example

http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo

- Clinical observation: *"Grade IV anaphylactic reaction to a penicillin injection."*
  - Representation using HL7 Reference Information Model (RIM) and SNOMED-CT terminology
    - RIM is an "cross-domain information model"
      - "compositional" capability based on use of instances of Act and ActRelationship
    - SNOMED-CT is a "healthcare terminology model"
      - "compositional" capability based on concatenation of multiple "atomic" terms from multiple "axes."

TermInfo Example (2)

http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo

- The semantic boundary between the RIM and SNOMED is not "clean" → multiple representations of the same semantics
  - Complex RIM model with atomic SNOMED code bindings
  - Simple RIM model with complex SNOMED code bindings
  - Continuum of possibilities between these two extremes
  - Although all representations are "semantically equivalent" from a clinical perspective, they are not *a priori* CSI because of serialization differences

- SW "solution" to the problem involves identifying and expressing the "overlapping semantic boundaries"
  - RIM ontology
  - SNOMED ontology
  - RIM-SNOMED cross ontology "boundary" mapping

The Semantic Web Value Proposition

- Common "design-time" and "run-time" representation (RDF *graph patterns*)
  - Eliminates "impedance mismatch" that often occurs between these two contexts
- Semantics as a "first-class citizen"
  - Reduces technology barriers
    - SPARQL end-points (a broadly-applicable standard) to query RDBMS (and other) persistence layers
  - Reduces syntax and lexical barriers
    - Single RDF model and syntax → technology-independent schema → *serialization brittleness*
  - Enables solutions to semantic CSI barriers
    - Layered metadata
    - Overlapping standards' boundaries

The Semantic Web Value Proposition(2)

- Tools and infrastructure designed to promote reuse, discovery, amalgamation, harmonization of data and meta-data from multiple contexts
  - "Point-wise" (bottom-up) semantic harmonization
  - Decentralized extensibility
    - Reuse made easy: "Interoperability emerges from laziness…"
- Significant amount of work already done in TMC domains

Healthcare Standards and RDF… plus a few examples

- SALUS project – https://github.com/srdc
- CDISC2RDF – http://cdisc2rdf.com
- HL7 Reference Information Model – http://www.hl7.org/special/committees/projman/searchableprojectindex.cfm?action=edit&ProjectNumber=983
- BRIDG Model – www.bridgmodel.org
- HCLS demo projects
  - http://www.w3.org//2013/02/ODM/
  - http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo
  - http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo
  - http://www.w3.org/2013/05/11179/
  - http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability -