Release 5.1 User's Guide

https://bridgmodel.nci.nih.gov

29 March 2018

prepared by

Wendy Ver Hoef, Smita Hastak & Julie Evans
BRIDG Modeling Team
on behalf of the
HL7 Biomedical Research and Regulation (BR&R) Work Group

Note: The BRIDG Modeling Team members are funded by the US National Cancer Institute (NCI)

Revision History

Date	Summary of Changes
18 June 2007	Initial Project Documentation for BRIDG Release 1.0
23 October 2007	Project Documentation for BRIDG Release 1.1
30 October 2009	User's Guide for Release 3.0
2 February 2010	User's Guide for Release 3.0.1
31 August 2010	User's Guide for Release 3.0.3
15 December 2010	User's Guide for Release 3.0.3
29 February 2012	User Guide for Release 3.1 (Removed much of the BRIDG background/history and glossary sections (to new upcoming website); removed redundancy within the document and among other BRIDG documents; Updated Data type section; other tasks for release 3.1)
7 September 2012	User's Guide for Release 3.2 – Re-organized Section 5.0; updated to include Statistical Analysis sub-domain and the OWL representation of BRIDG
27 March 2015	User's Guide for Release 4.0 – updated in support of the extended translational research scope of the BRIDG model. Updated content is primarily in sections 2, 3 and 4
18 March 2016	User's Guide for Release 4.1 – updated in support of HL7 DSTU ballot comment reconciliation; updated to include semantics from a software vendor for study tracking and management use case
27 July 2016	User's Guide for Release 4.1.1 – updated per the HL7 ballot comment resolutions from the informative balloting of BRIDG 4.1.
31 January 2017	User's Guide for Release 5.0 – updated in support of additional semantics in Imaging to support BRIDG/DICOM harmonization, oncology concepts from NCI's oncology program, and study management semantics from a vendor.

31 August 2017	User's Guide for Release 5.0.1 – updated per the HL7 ballot comment resolutions from the informative balloting of BRIDG 5.0.
28 March 2018	User's Guide for Release 5.1 – updated in support of additional semantics from the US FDA's Common Data Model Harmonization (CDMH) project.

Table of Contents

Revision History **2**
Table of Contents **4**
1 Guide to the Reader **7**
1.1 Overview 7
2 Executive Summary **8**
2.1 BRIDG Project Stakeholders 8
2.2 BRIDG Project Goals 8
2.3 Definition of the BRIDG Model Domain of Interest 9
2.4 Target Audience of the BRIDG Model 9
2.5 Caveats 11
2.6 BRIDG 5.1 Summary 11
2.6.1 Harmonization of the CDMH Project: 11
2.6.2 Harmonization of Anatomic Pathology Structure Report (APSR) – Work in Progress 11
2.6.3 CDISC Views Package 12
2.6.4 Deprecated Items Removed from Model: 12
2.7 Differences in Meaning vs. Differences in Representation: Normalization, Harmonization, Localization and a Note on the Challenges of Building a Domain Information Model 13
2.8 Contributing to the BRIDG Model: How Does a Project Team Get Its Content Into the BRIDG Model? 15
2.9 The BRIDG Model website 15
2.10 The BRIDG Model as a "Standard" 15
2.11 BRIDG Implementation Approaches 16
2.11.1 Physical Database Design 16
2.11.2 Data Exchange Mechanism 16
2.11.3 Service Interfaces 16
2.11.4 Meta Data Repository (MDR) 16
3 The BRIDG Model – General Considerations and Representational Conventions **17**
3.1 Computable Semantic Interoperability (CSI) 17
3.2 The BRIDG Model: Instance of a Domain Information Model 17
3.2.1 The Domain-of-interest of the BRIDG Model 17
3.2.2 The Scope of a DIM 17
3.3 UML: the BRIDG Model's Representational Language 18
3.4 Disease-Specific Model Strategy 18
3.5 BRIDG Representational Conventions 18
3.5.1 Study Lifecycle and the Grouping of Activities into "Pillars" 18
3.5.2 Identifier Type 20
3.5.3 Constraints 20
3.5.4 Relative Timing of Activities 21
3.5.5 Deprecated Model Elements 22
3.5.6 Key Distinctions Between Purpose, Objective and Reason 22
3.5.7 Color Coding of Classes in the UML-based Diagrams 22
3.5.8 An Important Addition to the Model: The Study Class 23

3.5.9 Disambiguating the Term "Protocol" 23
3.5.10 The Versioning of Study Protocol 24
4 Model Content 26
4.1 Overview 26
4.2 Canonical UML Representation 26
4.2.1 Comprehensive UML Model 27
4.2.2 Sub-Domains UML Views 27
4.2.3 Adverse Event 27
4.2.4 BioSpecimen 27
4.2.5 Common 28
4.2.6 Experiment 28
4.2.7 Imaging 28
4.2.8 Molecular Biology (out-of-scope in BRIDG 5.1 HL7 ballot) 28
4.2.9 Protocol Representation 29
4.2.10 Regulatory – Change in Deprecation Plans 29
4.2.11 Statistical Analysis 30
4.2.12 Study Conduct 30
4.3 HL7 Representation – RIM-based 30
4.4 Ontological Representation 32
4.5 Relationship to Other Models 32
4.6 Source Model Mappings 33
4.7 Views (Diagrams) in BRIDG File 35
4.7.1 BRIDG – Start Here: Overview 35
4.7.2 BRIDG Domain Analysis Model: BRIDG Sub-Domain Packages Diagram 35
4.7.3 Additional Focused Views 35
4.7.4 BRIDG Domain Analysis Model: UML-Based Comprehensive BRIDG Model 36
4.8 Class Diagrams 36
4.9 Instance Diagrams 36
4.10 State Diagrams 36
5 Appendix 37
5.1 Abstract and Complex Data Types 37
5.1.1 Temporal Grammar(s): HL7 Abstract Data Type TS Data Type 40
5.1.2 Coded Concepts: HL7 Abstract Data Type CD Data Type 40
5.1.3 Collections: BAG and DSET 40
5.2 Instances as "KindOf" vs. "InstanceOf" – holding a place vs. filling it 41
5.3 A Brief Explanation of the Concept of "Mood" 41
5.4 BRIDG and the Four Pillars of Computable Semantic Interoperability 42
5.4.1 CSI Pillar #1: A Common Reference Model 42
5.4.2 CSI Pillar #2: Complex Data Types 43
5.4.3 CSI Pillar #3: Binding to Terminologies 43
5.4.4 CSI Pillar #4: Derivation of Data Interchange Structures from Pillars #1 - #3 43
5.5 The Concepts of Visit and Event 44
5.6 Introduction to Basic UML Class Diagram Concepts and Terms 44

5.6.1 What is UML? 44
5.6.2 What is a class diagram? 45
5.6.3 Attribute names and data types 46
5.6.4 Example of an attribute list of a class from the BRIDG 46
5.6.5 Different types of associations between classes in a UML diagram 47
5.6.6 Association 48
5.6.7 Specialization 49
5.7 Aggregation 50
5.7.1 Composition 51
5.8 BRIDG Deprecation Policy 52

1 Guide to the Reader

1.1 Overview

The BRIDG User's Guide consists of six Sections:

Section 1.0 – Guide to the Reader (the present Section). This Section provides a brief summary of the content of each Section of this document.

Section 2.0 – Executive Summary. This Section provides a brief background and general information about the BRIDG Project, such as goal, scope, intended audience, etc.

Section 3.0 – The BRIDG Project. This section contains short discussions on the following topics:

- An overview of the process that enables a specific project working in the domain defined by the BRIDG Model to utilize the BRIDG Model and/or contribute to its overall semantic content;
- A brief discussion of the topic of "BRIDG as a standard";
- Example approaches of how the BRIDG Model is being used in the clinical research community.

Section 4.0 – The BRIDG Model – General Considerations. This section describes the BRIDG Model from a relatively content-independent perspective, i.e. from the perspective of topics such as the general modeling approaches, representational guidelines, naming conventions, as they are utilized in the BRIDG Model.

Section 5.0 – The BRIDG Model. This section describes the specific content of the BRIDG Model including discussions about particular *representational* choices made by the Semantic Coordination Committee as those choices relate to specific content.

Section 6.0 – Appendix. This section contains material that may be of interest to some readers of the BRIDG User's Guide but is not a primary part of the BRIDG Model. Examples include a basic UML primer and a more detailed discussion of Abstract/Complex Data Types.

2 Executive Summary

2.1 BRIDG Project Stakeholders

The BRIDG Project is a collaborative effort engaging stakeholders from five organizations:

- Clinical Data Interchange Standards Consortium (CDISC)
- US Food and Drug Administration (FDA)
- HL7 Biomedical Research and Regulation Work Group (HL7 BR&R WG)
- International Organization for Standardization (ISO)
- US National Cancer Institute (NCI)

2.2 BRIDG Project Goals

The goal of the BRIDG Project is to produce a *shared view of the process and data semantics* of a common domain-of-interest, specifically the domain of basic, pre-clinical, clinical, and translational research *and its associated regulatory artifacts*. In BRIDG, the term "translational research" refers to research that fosters the multi-directional and multi-disciplinary integration of basic, pre-clinical, and clinical research, that may include patient-oriented, population-based and even post-marketing aspects, with the long-term aim of improving the health of the public. It that sense then, it is fair to say that the scope of BRIDG is now translational research.

A shared semantic view is essential if the clinical research community, both for itself and also as part of the larger Healthcare and life sciences communities, is to achieve *computable semantic interoperability (CSI)*, i.e. the ability of computer systems to communicate information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system. Stated another way, in order to realize the various data interchange and application interactions that are known by members of the BRIDG stakeholder organizations to be requirements for CSI, a *shared view of process and data semantics must be established*. Please note that a shared semantic view is only one of the steps to achieving CSI. The other steps are described in Section 6.5.

This shared view can be used to:

- document information requirements before designing and building a solution,
- build the information structures in the solution, and
- trace requirements throughout the solution development process.

Through the efforts of the BRIDG Project, this shared view is expressed as a collection of visual diagrams which are, in turn, expressed using the iconography and grammar of the Unified Modeling Language (UML), the HL7 Reference Information Model (RIM), and a Web Ontology Language (OWL) representation of the semantics. This set of visual diagrams, the representations and the underlying inter-diagram relationships, definitions, explanations, and examples (and a few instance diagrams which illustrate

the semantics of the diagrams in more detail) are collectively referred to as the BRIDG Model. For a short introduction to UML, please see Section 6.7.

2.3 Definition of the BRIDG Model Domain of Interest

The BRIDG Model is an instance of a construct often referred to as Domain Information Model (DIM) or "problem space model." The model is specifically constructed to be *implementation-independent*, i.e. the semantics of the model are restricted to those that characterize the "problem domain" as described by domain experts. In particular, a DIM specifically excludes semantics that are introduced based on a particular "solution space" that can be built to solve the stated problem. Thus, *one or more* "solution space models" can be built to support all or part of a single problem space model. The term "domain" indicates that the semantics of the model-in-question are restricted to those that collectively define a clearly bounded domain-of-interest. In the case of the BRIDG Model, the domain-of-interest has broadened from protocol-driven/clinical research to translational research. The new domain-of-interest is formally defined as:

> ***Basic, pre-clinical, clinical, and translational research and associated regulatory artifacts,***
> *i.e. the data, organization, resources, rules, and processes involved in the formal assessment of the utility, impact, or other pharmacological, physiological, or psychological effects of a drug, procedure, process, subject characteristic, biologic, cosmetic, food or device on a human, animal, or other subject or substance plus all associated regulatory artifacts required for or derived from this effort, including data specifically associated with postmarket surveillance and adverse event reporting.*

As of BRIDG 4.0, the scope of the BRIDG model was changed to encompass the larger translational research domain rather than being limited to just representing the clinical research domain.

The rationale for broadening the scope of BRIDG to translational research was to support the use cases spanning the clinical and life sciences domains. Use cases spanning these domains are rapidly increasing, due to the tremendous advances being made in our ability to elucidate the genetic basis of disease. The goals of precision medicine require linking clinical and molecular semantics. The lack of common semantics between the clinical and life sciences domains has been referred to as "the chasm of semantic despair" (Chris Chute, http://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/080221p1.pdf, slide 8). Therefore, unifying the clinical research focused BRIDG model with the model(s) from the life sciences domain would provide consistent representation and patterns across the domains and another step toward the important goal of bridging this chasm.

2.4 Target Audience of the BRIDG Model

The target audiences for the BRIDG Model are the following:

- domain experts working within the domain scoped by the BRIDG Model;

- information analysts, architects, and software developers working on defining specific data interchange semantics (e.g., message specifications, application APIs, services or HL7 FHIR resources); and those interested in building logical/physical models from the BRIDG model.

- terminologists and ontologists interested in augmenting current work involving building ontologies in the BRIDG Model domain-of-interest.

The BRIDG Model is represented using the Unified Modeling Language (UML). However, the model uses only the most obvious and easily understandable constructs of UML and, as a result, the BRIDG modeling team believes that the model, in combination with the content of the User's Guide, should be understandable by anyone with experience and expertise in the BRIDG domain-of-interest.

The BRIDG modeling team members believe that instance diagrams will aid those domain experts with no prior UML experience in reading, understanding, and validating the content of the BRIDG Model, thereby adding the element of vetting by the larger community that is essential to the relevance, and therefore the success, of the BRIDG Project. (**Please note**: *Only instance diagrams that were added for BRIDG Release 3.2 are aligned to 3.2 semantics; all previous diagrams are still aligned with BRIDG Release 3.0.2. None of the instance diagrams were updated for the BRIDG 5.1 release, although a few new slides representing the SDTM PGx domains were added for BRIDG 4.0.)*

Large-scale validation/vetting of the content of the BRIDG Model is of particular importance, since the overarching motivation of the BRIDG Project is to present to the basic, pre-clinical, clinical and translational research communities a valid representation of the shared meaning of various concepts, relationships, and processes that collectively define the domain of "***basic, pre-clinical, clinical, and translational research*** and associated regulatory artifacts.' In the end, the value of the BRIDG Project and the BRIDG Model produced by that Project are directly related to the degree to which that representation is both valid and relevant to the community. If the BRIDG Model provides a correct and complete view of shared meanings (semantics) of the domain-of-interest in a form that is both understandable to domain experts and robust enough to be utilized by various technology teams charged with building solutions required by those domain experts, the BRIDG Project will be a success. It is towards that goal that all those involved in the BRIDG Project are continually striving.

It should be noted that the 'correctness' and 'completeness' of the view of the BRIDG Domain's shared semantics as represented in the BRIDG Model is exactly that: a representation. There are certainly other representations. The goal is to produce a single representation that is of some use to all of the BRIDG Model's stakeholders. Individual groups of stakeholders with particular, focused interests in the BRIDG Model's domain-of-interest may find alternative representations more useful for their specific purposes. The critical success metric in these situations is whether or not the shared semantics of

any alternative representation can be completely and unambiguously mapped to the BRIDG Model.

2.5 Caveats

It is not the purpose of the User's Guide to give detailed instructions as to how a given application development or message specification team can contribute to and/or work with the BRIDG Model. A subgroup of the BR&R WG is looking into developing documentation on various approaches to implementing BRIDG and providing guidance in the near future regarding BRIDG conformance and compliance.

2.6 BRIDG 5.1 Summary

BRIDG Release 5.1 is a minor release that adds semantics from the US Food and Drug Administration's Common Data Model Harmonization (CDMH) project; a new CDISC View Package, and also, several deprecated items were removed from the model as described in the BRIDG Deprecation Policy (included in the BRIDG User's Guide Section 5.8). Each of these change topics are described below.

2.6.1 Harmonization of the CDMH Project:

The goal of the CDMH project is to build a data infrastructure for conducting research using Real World Data derived from the delivery of health care in routine clinical settings. This data infrastructure will allow researchers to simply ask research questions on much larger amounts of Real World Data than currently possible, leveraging open standards and controlled terminologies to advance Patient Centered Outcomes Research. The CDMH Project will have a BRIDG-based repository that will capture query results for de-identified aggregate and patient level data from four networks: Sentinel, PCORNet, i2b2, and OMOP.

The majority of the CDMH semantics were already present in the BRIDG Model. This harmonization added 2 new classes, 16 new attributes, and 7 new associations.

2.6.2 Harmonization of Anatomic Pathology Structure Report (APSR) – Work in Progress

The BRIDG modeling team is currently working with subject matter experts (SMEs) from the HL7 Orders and Observations (O&O) Work Group and the ISO DICOM WG 26 members to harmonize the IHE's Anatomic Pathology Structured Report 2.0 (a supplement to IHE Pathology and Laboratory Medicine Technical Framework V8.0, APSR 2.0 is officially Draft for Trial Implementation IHE PaLM TF Suppl. APSR 2.0 TI (091)). The SMEs participating in this effort are some of the authors of this standard, which is based on the HL7 CDA standard. The harmonization work is in progress and work to date indicates that the majority of semantics will be supported with elements already in the BRIDG model. Some areas with existing compatibility are demographic information, medical history, procedure steps, some aspects of macroscopic/microscopic

observations, and diagnostic information. At this time, the BRIDG modeling team anticipates some new elements related to anatomic pathology will be added to BRIDG and balloted in September 2018 ballot cycle.

2.6.3 CDISC Views Package

This is a new package in Release 5.1. There is a new view for the CDISC Lab Model. This view was built by the members of the BRIDG modeling team to facilitate the work efforts with the TransCelerate Biopharma group. The SDTM 3.1.3 view and the three sample SDTM Domain views have been moved from the Additional Focused Views package to under the CDISC Views package.

2.6.4 Deprecated Items Removed from Model:

The deprecated items that were removed from the model are:

1. Animal.speciesCode - The Species of an Animal is now represented by BiologicEntityClassification.scientificNameCode.
2. PlannedActivity.purpose - This attribute is redundant and was deprecated and then removed in favor of the existing Activity.reasonCode.
3. Specimen.accessionNumberText - This semantic is more accurately derived from SpecimenCollectionGroup.accessionNumberText regardless of the number of specimens collected.
4. Specimen.typeCode - The semantic represented by the Specimen.typeCode attribute was clarified by the LSDAM team and actually is an identification of the activity that produced the Specimen, which means it maps to PerformedMaterialProcessStep > DefineMaterialProcessStep.nameCode.
5. StudySite.accrualStatusCode and StudySite.accuralStatusDate - The accrual status of a study site can overlap with the overall status of the study and study site status. The proposed solution was to eliminate the study site accrual status codes and adopt the CTRP values for study site status code.
6. Association: Each DefinedMaterialStorage always stores one Specimen. - This association was redundant since a Specimen can be the subject of an activity via the Subject to Specimen association.
7. Association: Each PerformedAdministrativeActivity might be performed at one StudySite. - This association was redundant since StudySite can be the subject of an administrative activity via the Subject to Organization association and can be a performer of an activity via the Performer to Organization association.

Note: The BRIDG release number on https://bridgmodel.nci.nih.gov is different from the HL7 BRIDG release number due to HL7 naming conventions. The corresponding release numbers are shown in the table below:

Release Date	BRIDG Website Release Name	Corresponding HL7 BRIDG Release Name
March 2018	BRIDG Release 5.1	BRIDG R4
August 2017	BRIDG Release 5.0.1	BRIDG R3
May 2017	BRIDG Release 5.0	BRIDG R3

2.7 Differences in Meaning vs. Differences in Representation: Normalization, Harmonization, Localization and a Note on the Challenges of Building a Domain Information Model

The process of building a DIM in a domain-of-interest is a process of *knowledge extraction*, and of establishing common, non-ambiguous *knowledge representations* across and among various stakeholder groups. The term "harmonization" was taken from the HL7 V3 framework, where it is used to refer to the process whereby multiple perspectives on a given problem are presented to an informed, neutral group who must decide if the various perspectives represent the same or different semantic content, i.e. the process of harmonization is focused on distinguishing true differences in *meaning* from differences of representation or synonymy, an exercise that is, at times, straightforward, while at other times quite complicated, multi-factorial, and time-consuming. When a difference in meaning, naming (i.e. an instance of synonymy) or representation (synonymy could be considered a trivial case of a representational difference but as used here, the notion of representational differences refers to multi-factorial differences between two structures which have the same *meaning/semantics* but different *structure/syntax)* is encountered, the choices are, at least in theory, fairly limited:

- on the one hand, the difference is real so that the two concepts need to be appropriately defined, separated, and granulated; or, alternatively

- the apparent differences are instances of synonymy or representation (syntax) and appropriate naming decisions and definitions (including examples of naming choices) must be made.

The operational difficulties in building a DIM arise when seemingly "simple" differences (or agreements) are examined more closely and reveal intricacies in meaning that may be rooted in deeper differences in context, culture, or deep-layer semantics.

Technically speaking, the process of disambiguation and concept granulation known as "harmonization" can be distinguished from the process of "normalization," where one concept name is assumed to be semantically equivalent to another concept's name.

Likewise, harmonization and normalization are distinguished from "localization", a process whereby a concept that is known to be non-interoperable is allowed to persist in a model, often because "that's the way things are done" at a local level. Because the goal of the BRIDG Model is to define a set of *shared* semantics, the BRIDG modeling team does not allow for differences based on localization, choosing instead to work to establish consensus across the BRIDG Model's stakeholders and project teams for semantics that must be shared and allowing 'localized' semantics to exist, but outside of a representation in the BRIDG Model itself.

Based on the above discussion, the BRIDG modeling team's job as it sorts through a new set of harmonization artifacts presented for inclusion in the BRIDG Model is to determine whether conflicts in concepts represent one of the following situations:

- naming/synonymy or representational issues, which are solved through normalization with the choice of a single name/structure assigned to the two competing concepts/structures (note that this often means that identical concepts/structures with different names end up either being named using one of the common names, or re-named using a more neutral name equally unfamiliar – but hopefully not opaque – to all stakeholders in the semantics); Note that all BRIDG model element descriptions contain an "OTHER NAME(S)" section where synonyms are captured when they are identified.

- definitional issues, which are solved by appropriate harmonization techniques such as concept splitting (same term used for two meanings necessitates the creation of two distinct terms); or

- localizations, which need to be either excluded from the model or included as appropriately tagged exemplar content (see note above on the modeling team's rationale for excluding localized content from the BRIDG Model).

Thus, an expected outcome of a BRIDG modeling team harmonization meeting is that new (or agreed-upon) names are selected for overloaded terms (i.e. situations where the same term is used to mean several things), ambiguous/overlapping concepts are granulated to express their meaning with appropriate semantic clarity, and, if necessary, a certain amount of the content is deemed as localizable or outside the scope of interoperability concerns and therefore excluded from the BRIDG Model.

It should be noted that one of the realities of a shared DIM such as BRIDG is the phenomena of "my words aren't in the model" or "that term is used more broadly than I use it." The pragmatics of building a shared model mean that not everyone's particular labels (words) for modeling team to make sure that words with conflicting or overlapping semantics are disambiguated. Sometimes, this is accomplished by using new words to point to specific, rigorously defined concepts. For example, the symbol "Protocol" is used in the domain of clinical trials and "protocol-driven research" to mean *both* "a particular kind of documentation, i.e. one that specifies all of the rules, regulations, assumptions, etc. that define a clinical study set up to answer a specific

question/address a specific hypothesis," *as well as* "the collection of the defined set of activities one plans or, in fact, accomplished, in the course of a given study." Pursuing the word "Protocol" a bit further, one realizes that there are actually two distinct concepts lurking behind a single term and that using the symbols "Study Protocol" and "Process Protocol" resolves the ambiguity by restricting the latter term to refer to the definition and specification of the steps to accomplish a goal, while the former includes not only the "rules," but also expands the meaning of the original overloaded symbol to include the overarching semantics of hypothesis evaluation. In such cases, inclusion of a concrete example is often deemed necessary to fully define the new terms: *For example, a Study Protocol would be "Protocol XYZ to evaluate the therapeutic effectiveness of drug ABC in treating disease JHK," while an example of a Process Protocol would be "To draw a CBC, execute the following sequence of steps and rules."*

In the case of a word/symbol being used more broadly than a particular local usage of the word, the guidance of the modeling team is for the local application to be aware of the restricted usage and the possible larger interoperability issues that may result from systems with the restricted usage trying to interoperate with systems with the broader usage. If significant interoperability issues arise and find their way back to the modeling team, new symbols may have to be defined for a future version of the BRIDG Model and appropriate transformations applied to the involved applications based on earlier versions of the BRIDG Model. In other words, the BRIDG Model and the underlying modeling team processes are expected to play a critical role in the overall process of message specification/application development and as such, specification/application implementations must be designed with the appropriate processes/procedures/infrastructure in place to allow smooth migration and version management going forward. *Change is a given. Architecting and planning for change is essential.*

2.8 Contributing to the BRIDG Model: How Does a Project Team Get Its Content Into the BRIDG Model?

The BRIDG modeling team members have published a set of documents to formalize the BRIDG harmonization process and to help project teams understand how to contribute to the BRIDG Model. These documents are available in the BRIDG Harmonization Package (a zipped file) that is downloadable from the BRIDG website (https://bridgmodel.nci.nih.gov/). It is recommended that project teams download this zipped package and review the content.

2.9 The BRIDG Model website

The BRIDG Model and its associated documentation can be found at: https://bridgmodel.nci.nih.gov/.

2.10 The BRIDG Model as a "Standard"

Several years ago the BRIDG Board of Directors decided to advance BRIDG to the status of an international standard through the International Organization for Standardization (ISO) process under ISO Technical Committee 215 for Healthcare Informatics, which included joint SDO balloting through the Joint Initiative Council

(JIC). This collaboration provided the opportunity to gather feedback from a variety of key member countries and stakeholders, applying recommendations to improve the BRIDG Model. In May 2010, BRIDG R3.0.1 passed the first round of joint balloting in ISO TC215, HL7 and CDISC. All comments were addressed and published as part of BRIDG 3.2 September 2012. Most recently in 2015 BRIDG R3.2 unanimously passed the ISO DIS ballot as a Draft International Standard which allows BRIDG to advance directly to publication.

2.11 BRIDG Implementation Approaches

Over the last few years, the BRIDG model has become a source of reference for clinical research semantics. The BRIDG user community has started to leverage the model to support their clinical research related systems development efforts. Following are the various BRIDG implementation approaches that have been identified by the BRIDG user community.

2.11.1 Physical Database Design

This approach involves leveraging BRIDG as a reference model or a conceptual model and then constraining it to build downstream models such as logical models and then generating the physical database model. This approach has been implemented by multiple stakeholders at varying degrees and the methodology pattern appears to be quite similar where design choices and strategies were defined regarding how to build the logical model(s) and physical database model(s). Most of these types of implementations of BRIDG have been done for a sub-set of the BRIDG model with the focus on one or more use cases of clinical research.

2.11.2 Data Exchange Mechanism

This approach involves building BRIDG-based schemas as a data exchange specification. This implementation approach would also leverage the BRIDG semantics and build use case specific schemas for exchange of a particular set of information. For example, registering a subject to a study, exchanging adverse event data, etc.

2.11.3 Service Interfaces

This approach involves leveraging the BRIDG structures and semantics to build services that can be exposed to the relevant user community.

2.11.4 Meta Data Repository (MDR)

CDISC, one of the founding stakeholders of BRIDG, has built a metadata repository (MDR) to house the CDISC standards. One of the underlying framework requirements for this project, named the CDISC Shared Health and Research Electronic Library (SHARE), is a foundation domain model. BRIDG serves as that domain model and provides the semantics of the domain. This metadata repository is intended to hold standards for individual "biomedical concepts," instances of objects which combine attributes of defined, planned, and performed BRIDG classes. To learn more about the SHARE project, please go to http://www.cdisc.org/cdisc-share.

3 The BRIDG Model – General Considerations and Representational Conventions

This section begins with a general discussion of computable semantic interoperability and then addresses BRIDG scope and approach to disease-specific content.

3.1 Computable Semantic Interoperability (CSI)

It is important to note that the overarching, primary use case for the BRIDG Model is the need to achieve computable semantic interoperability (CSI) both *within* the domain of ***basic, pre-clinical, clinical, and translational research*** as well as *between* this domain and others that may intersect with it at the interoperability level. For example, the domains of protocol-driven research and public health both share the concept of Adverse Events. It is beyond the scope of this BRIDG User's Guide to discuss the details of CSI. However, a summary of this topic, including the Pillars of Computable Semantic Interoperability and how BRIDG addresses them, is in Appendix 6.4.

3.2 The BRIDG Model: Instance of a Domain Information Model

The BRIDG Model is the most obvious and visible artifact produced by the BRIDG Project. More formally, the BRIDG Model is a domain-specific instance of a Domain Information Model (DIM), a further discussion of which can be found in the Glossary of this document.

3.2.1 The Domain-of-interest of the BRIDG Model

The BRIDG Model defines the static and dynamic semantics of a domain-of-interest. The formal definition of the domain for the BRIDG Model can be found in section 2.3.

3.2.2 The Scope of a DIM

When working with DIMs, one assumes that at some point and in some context, a given application, service, or data interchange specification ("message," "procedure call," etc.) may involve multiple overlapping/intersecting domains, i.e. domains that share common dynamic or static semantics. The occurrence of overlapping scope between two or more DIMs is *not* necessarily an indication that a given DIM's scope is too broad, but rather simply an indication that the two domains have some common semantics. In such cases, for example, there are overlapping *data* semantics around the concept "adverse event" between the BRIDG domain and the Public Health domain – the most expeditious approach often is *not* to expand a particular DIM, but rather to *harmonize* the overlapping semantics of the two domains.

Semantics from intersecting domains-of-interest (e.g. 'adverse event' in the Public Health and Protocol-driven research domains; or 'observation,' 'person,' or 'organization' concepts common to Public Health, BRIDG, and Specimen concepts common across BRIDG and Clinical genomics) are "harmonized" so that CSI is possible between the various domains.

Note that the notion of harmonization at domain-to-domain "touch points" ("intersections") may not be limited to data concepts, but may also involve interoperability/shared semantics regarding process or behavior.

As mentioned in the discussion of the genesis of the BRIDG Model (Section 3), a DIM is useful in separating domain-specific (implementation-independent) semantics from implementation-dependent semantics and is therefore helpful in identifying true cross-domain intersection points that may need to be harmonized.

3.3 UML: the BRIDG Model's Representational Language

The graphical *lingua franca* of the BRIDG Model is the software industry standard Unified Modeling Language (UML). To date, the iconography and underlying language grammar of UML V2.1 have been sufficient to express all of the semantics of both the process and data views of the domain of the BRIDG Model. However, the modeling team has occasionally chosen to not slavishly follow UML conventions when an intentional deviation supports a domain-friendly way of presenting or making visible domain semantics, such as using class level constraints to expose business rules pertaining to the class, its attributes and/or associations.

3.4 Disease-Specific Model Strategy

Early in the history of BRIDG, the team considered including disease-specific classes, such as for cancer staging and identification of metastasis sites, but decided that the number of disease-specific classes would significantly expand the model until it was utterly unmanageable. Also, since most diseases require understanding of semantics that are disease-specific and not common to protocol-driven research in general, the modeling team should be comprised of modelers familiar with those diseases. Thus it was determined early on that the BRIDG modeling team would harmonize the common semantics used in protocol-driven research across many different areas of study and leave the disease-specific semantics to teams who would create models based on existing BRIDG classes and patterns. The BRIDG modeling team welcomes input from such teams who would like to propose semantics they need that they believe are common to translational research and two or more diseases.

3.5 BRIDG Representational Conventions

3.5.1 Study Lifecycle and the Grouping of Activities into "Pillars"

The BRIDG Model spans the lifecycle of a study from the planning of a study through the implementation, execution and evaluation of the study or experiment. Consequently, the model represents several stages of activities. Since BRIDG is a domain information model, the BRIDG modeling team's intention is to represent each concept in a domain-friendly way while being analytically rigorous. Thus the team has attempted to represent each concept once, in the context in which it originates and link to it in other contexts as needed. This allows the BRIDG Model to define only the attributes and relationships required for each context, thus eliminating much of the perceived redundancy in the model.

Defined Context (Global)
Defined activities are the characterization of a kind of activity, i.e. they define "what" an activity is. Most activities included in a study are not completely brand new, rather they are usually common tests or procedures, or they may be composite activities that are composed of several component activities that form a standard treatment strategy or a new treatment strategy used in several different studies. These activities are reusable concepts that essentially form a global library of activities that can be referenced in studies being planned, implemented, executed and evaluated. These activities can be defined once and referenced in many different studies to save the time and effort of re-entering data and, more importantly, to make the semantic connection between an activity being used in two different studies or at two different points in the same study. This notion of activities being defined once and referenced in many studies is the core idea of the defined activity class and its subclasses. This part of the model is what the BRIDG modeling team calls the "Defined Pillar". For those familiar with HL7 moodCodes, this is somewhat similar to the defined mood. Other "pillars" are also somewhat similar to moodCodes though not exact.

Planned Context (Study Specific)
Planned activities are the association of defined activities to a particular study. This association also includes the characterization of the sequencing (or timing) of these activities, also referred to as the study calendar or study design. For example, a defined activity is created for the notion of allocating subjects to arms on a study; however, the notion of when it occurs in the context of a particular study and the method to be used is characterized in a planned activity. This part of the model is what the BRIDG modeling team calls the "Planned Pillar".

Scheduled Context (Subject Specific, future)
Scheduled activities are the instantiation of a planned activity for each subject on a study; this is sometimes referred to as the subject calendar. The scheduled activities also identify the intended timing of the activity as well as any anticipated resources (participants, locations, etc.) required for the activity. This part of the model is what the BRIDG modeling team calls the "Scheduled Pillar".

Performed Context (Subject Specific, past)
Performed activities represent the execution of activities for actual subjects on a study and the results that come out of those activities. This context was developed because domain experts have identified a need to capture what was actually done to a subject, as opposed to what was intended to be done. This makes it appear like there is redundant data in the BRIDG Model, e.g., the attributes in DefinedSubstanceAdministration vs. PerformedSubstanceAdministration. But in reality the attributes in the defined activity characterize what was intended and the attributes in the performed activity characterize what was actually done. Performed activities may reference a scheduled activity that they fulfill. Or, a performed activity may reference a planned activity in cases of a contingent activity, e.g., in the case of an adverse event a contingent activity may be planned but it is not scheduled. Or, a performed activity may be related directly to a defined activity in case of a totally unplanned activity such as when a subject breaks a

leg and has emergency surgery that is considered study-relevant due to possible drug interactions. This part of the model is what the BRIDG modeling team calls the "Performed Pillar".

From the above descriptions, it should be clear that the "pillars" are essentially related areas of the model and that naming them helps provide language to discuss issues that occur during the lifecycle of the study.

3.5.2 Identifier Type

Some concepts represented in the BRIDG Model use an identifier to distinguish between different instances of the same concept. When a given model concept has the possibility of being assigned more than one identifier from different sources and/or purposes, there is a need to distinguish one identifier from another by type. For example, a person may be assigned a medical record number (MRN) as well as a social security number (SSN), driver's license number, and insurance number. To distinguish one number from another, a type needs to be associated with each identifier. However, the HL7 data type for Instance Identifier (II) is intended to be a universal identifier only – it intentionally does not identify the kind of identifier. The II.identifierName data type attribute is not to be used to computably determine the type, source, or any kind of meaning for the identifier. Until BRIDG R5.0, the pattern for identifiers that require identification of a type is to move the identifier attribute into its own class and add a typeCode attribute to capture the type of that identifier with a one-to-many association between the original class and the identifier class. With BRIDG R5.0, a new datatype extension named "ID" is used so that additional identifier semantics such as type code can be included with the data type. See Section 2.6.2 Vendor Semantics for more information and examples of the new ID datatype.

3.5.3 Constraints

Since release 3.0, the modeling team has converted all business rules into constraints. To ensure they are visible in diagrams and portable when the same class is included on multiple diagrams, the modeling team has decided to define all constraints at the class level (attribute level constraints are not visually represented). Additionally, the naming convention is to include a descriptor in front of one of the types below where the descriptor is patterned after a model element (class, attribute, association). We have the following types of constraints to date:

- **Exclusive Or** – used to specify that only one of several similar associations can be used for any given instance of the class. For example, the StudyLegalSponsor class has an Exclusive Or constraint called "be a function performed by Exclusive Or" which says "A StudyLegalSponsor is a function performed by one and only one of the following: HealthcareProvider, Organization."
- **Not Applicable** – used to indicate that an attribute or association is not used in a certain context such as on a subclass that inherits an association that applies to the superclass but not this subclass. For example, the PerformedStudyAdministrativeActivity class has a Not Applicable constraint called "be participated in by Not Applicable" which says "Associations from Subject (including StudySubject) and ExperimentalUnit are not valid."

- **Qualifier** – used to restrict a value on an attribute, to require or disallow a certain association in a specific situation, and generally capture other kinds of business rules. For example, the Place class has a Qualifier constraint called "physicalAddress Qualifier" which says "physicalAddress should not contain PO Box address parts."
- **Unique Qualifier** – used to indicate that the value of a single attribute should be unique within a given context. For example, the Arm class has a Unique Qualifier constraint called "name Unique Qualifier" which says "An Arm name must be unique within the context of the StudyProtocolVersion to which it is associated."
- **Attribute Set Qualifier** – used to restrict the values of a set of 2 or more attributes within a given context. For example, the DocumentIdentifier class has a constraint called "Attribute Set Qualifier" which says 'For a given Document and a given DocumentIdentifier.typeCode, only one DocumentIdentifier can have primaryIndicator = "true".'
- **Declaration -** used to highlight the fact that for any subclasses, inherited properties or associations are valid unless explicitly constrained or excluded at the subclass level. For example, the Activity class has a constraint called "be participated in by Declaration" which says "Unless constrained at any subclass level, all associations from Subject (including StudySubject) and ExperimentalUnit are valid for all activities."
- **<association label> actualIndicator Qualifier** – Used to specify that an entity on the other end of the association must be a "kind of" or "instance of" entity. For example, the DefinedActivity class has an <association label> actualIndicator Qualifier constraint called "be participated in by actualIndicator Qualifier" which says "Only BiologicEntity, BiologicEntityGroup, Material, ProductGroup or Organization (via Subject or ExperimentalUnit) with actualIndicator = 'false' is valid."
- **Attribute Set actualIndicator Qualifier** – Used to specify when a set of attributes is valid only for an "instance of" entity. For example, the Person class has an Attribute Set actualIndicator Qualifier constraint which says 'name, birthDate, deathDate, initials, postalAddress, telecomAddress are valid only when actualIndicator = '"True".'

Note that the modeling team may add other types of constraints in the future as a pattern for handling other business rules.

3.5.4 Relative Timing of Activities

Study Protocols often require an activity to occur some amount of time before or after another activity, i.e. the timing of such activities is relative to other activities. BRIDG uses a pattern for modeling such relative timing that is drawn from HL7. The association between the relatively timed activity and the activity that it references (the anchor activity) is represented with a relationship class that carries an attribute called pauseQuantity. The pauseQuantity attribute has the data type PQ.TIME, meaning it represents a quantity of time. The relatively timed activity occurs "X" number of minutes, hours, days, etc. after the anchor activity. An example of this in the BRIDG

Model is PlannedContingentOnRelationship.pauseQuantity which defines how long after the prerequisite (anchor) activity occurs that the contingent (relatively-timed) activity occurs.

3.5.5 Deprecated Model Elements

The BRIDG Model uses the convention of the "DEPRECATED" stereotype on any elements that, after due diligence per the BRIDG Deprecation Policy, the modeling team has determined are obsolete. These items will remain in the model for a period of time but will eventually be dropped from the model if no use cases for retaining them are brought to the modeling team. Please see Appendix A7 BRIDG Deprecation Policy for more information.

3.5.6 Key Distinctions Between Purpose, Objective and Reason

In common language the term "purpose" is sometimes used as a synonym for "objective" and sometimes for "reason". Likewise, in the domain of protocol-driven research the domain also sometimes uses these terms interchangeably. The use of and examples for the terms in the BRIDG model elements may not be consistent or clear currently but nevertheless, given their placement in the model in relation to the StudyProtocolVersion class or in other classes, there seems to be a progression of specificity in the concepts: "purpose" is broad and is talked about at a high level, in fact, at the study level; objective is more specific but also defined at the study level; reason is usually more detailed and defined at an activity or low level. The key concepts can be distilled this way:

- Key concept of "purpose": A broad, high-level intent
- Key concept of "objective": A specific and measurable, but high-level goal
- Key concept of "reason": A specific and detailed rationale, i.e. the why

3.5.7 Color Coding of Classes in the UML-based Diagrams

Since BRIDG Release 3.0, the class color coding convention has been changed to reflect the sub-domain to which a class belongs. One of the guiding principles since Release 3.0 UML-based model was to remove the RIM aspects so that these diagrams would be easier to understand for domain experts. In previous releases, the classes had been color coded to RIM conventions, but since 3.0 the BRIDG Model uses color to group classes into sub-domains. With addition of three new sub-domains in BRIDG 4.0 and one more in BRIDG 5.0, the legend now shows ten different colors. The color coding legend is on each of the BRIDG UML diagrams and is also included below in **Figure 3**.

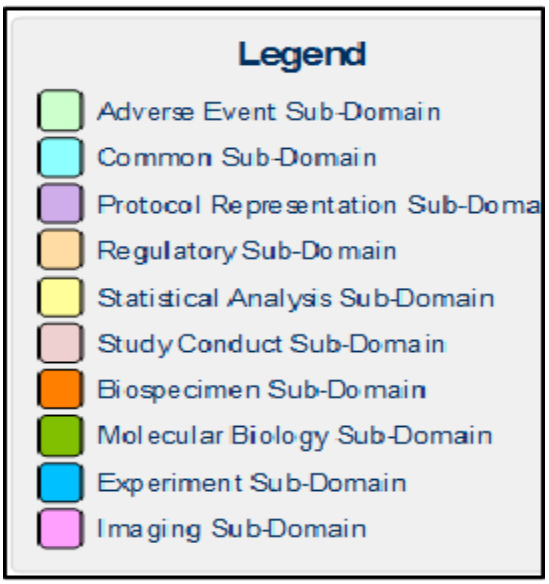

Figure 3: UML-Based Diagram Class Color Coding Legend

The modeling team does not intend to continue propagating this color coding pattern as new sub-domains of translational research are identified. At present, the color coding allows a way to group the classes and make it visually easy to align with the sub-domains. As the BRIDG framework evolves, the modeling team has considered providing more granular groupings in support of business use cases of the domain. If that does become a way of grouping the semantics, then this color-based grouping may go away and be replaced by some other categorization mechanism. The modeling team would welcome your comments on the usefulness of the color coding of sub-domains.

3.5.8 An Important Addition to the Model: The Study Class

As a result of the May 2015 HL7 DSTU ballot and the harmonization of a new study tracking and management use case, BRIDG now includes a new class: Study. After much discussion with ballot commenters and software vendor representatives, the Study class has been added to give a concrete starting point for the suite of relevant classes and to refine the relationships between them. One of the primary distinctions made is the notion of a StudyProtocol representing the plan for a Study, and that a Study exists prior to the development of the plan. Further changes may be seen in this area as new use cases are more fully understood.

3.5.9 Disambiguating the Term "Protocol"

In modeling, often the same term is used to mean different things and a single concept can have more than one name. In the healthcare arena, the term "protocol" is somewhat overloaded and must be qualified to provide semantic context. Therefore during the early years of the BRIDG project, the term "study protocol" was chosen to disambiguate the concept of the detailed plan for a clinical study (the scope of BRIDG at that time) from other kinds of protocols such as are common in life sciences. In BRIDG, the notion of a study protocol is very specific in purpose and includes (but is not limited to) the

design, statistical considerations, activities to test a particular hypothesis or answer a particular question that is the basis of the study, characteristics, specifications, objective(s), background, pre-study/study/post-study portions of the plan (including the design, methodology, statistical considerations, organization). For a more complete discussion of the notion of the study protocol see the classes StudyProtocol, StudyProtocolVersion, StudyProtocolDocument, StudyProtocolDocumentVersion, StudyConduct and all their associations.

With the addition of life sciences to the scope of the BRIDG model, there came along (with that scope) the need to identify the kind of protocol that represents a more simple or atomic concept, that of "a composite activity that serves as a rule that guides how activities should be performed." This concept, represented by the Protocol class, has a more limited size than the concept of a study protocol does and represents a standardized approach to doing tasks or activities that are not as big as the plan for a whole study.

The BRIDG modeling team acknowledges that overloaded terms are problematic. The modeling team recognizes that many different users within the BRIDG community will have differing opinions on what the meaning of a term is, which term is the best to use for each concept, and how to define them most effectively. Given that the real "meat" of a concept is in the definition, the BRIDG modeling team aims to choose the most unambiguous term to use as the class name, to make the class definition as explicit and clear as possible, to provide sufficient examples and other names to illustrate the range of possible instances that could be represented by the class. The modeling team would like to solicit feedback from the community on representational choices that have been made as well as the class name and other aspects of the model.

3.5.10 The Versioning of Study Protocol

Versioning is the process of storing the history of changes for an item, usually a document. Versioning allows viewing previous editions of an item, comparing them to the current live version, and, if desired, reverting to a previous edition. Within BRIDG the primary concept that requires versioning is the Study Protocol. The term "study protocol" has multiple meanings within the domain of protocol driven research, each used in a different context. "Study Protocol" is sometimes used to refer to the physical document that is reviewed. It is sometimes used to refer to the detailed plan for the study. And, it is sometimes used to refer to the activities of conducting the study. The BRIDG Model has a class for each of these concepts; StudyProtocolDocument, StudyProtocol, and StudyConduct, respectively. The StudyProtocolDocument class is a sub-class of the Document class. Because a study protocol can change over time (both the document and the plan) it became necessary to define versioned classes for Document, StudyProtocolDocument, and StudyProtocol. There are currently no use cases for versioning of StudyConduct.

When the BRIDG modeling team evaluated study protocol attributes and associations, it was necessary to determine to which class they belonged. The criteria the modeling team used to determine if an attribute or relationship should be assigned to the versioned class was if a change in that attribute or relationship might require an amendment to the

protocol (and subsequent approval) it belongs to the versioned class. It quickly became apparent that almost anything relating to a study protocol could change and could require an amendment. This is evident in the BRIDG Model as most of the attributes and relationships ended up in the versioned classes.

There are many things related to study protocol that were evaluated as part of this exercise; personnel, resource, sites, agents, and activities. Most of these relate to both the versioned and non-versioned classes of study protocol. Study sites were one of the exceptions. This class was split into StudySite and PlannedStudySite; where PlannedStudySite relates to StudyProtocolVersion (as part of the study plan), and StudySite relates to StudyConduct (as part of the conduct of the study). The other exception is the activity classes. The DefinedActivity and PlannedActivity classes are related to StudyProtocolVersion (through StudyActivity) because they are part of the definition of the study plan that could change with each version. The PerformedActivity class is also related to StudyProtocolVersion; not because it is part of the plan, but because it is important to know under which version of the protocol a given activity was performed.

4 Model Content

4.1 Overview

As the BRIDG Model grew in size and complexity it became increasingly difficult to satisfy the diverse audiences of BRIDG in a single model. Beginning with Release 3.0 the BRIDG Model has contained multiple representations (see **Figure 4** below). The Canonical Representation is a set of UML models (class diagram) of all the harmonized semantics. There is one large comprehensive UML model and six sub-domain specific model views. The HL7 Representation is comprised of several HL7 models representing the harmonized semantics using unambiguous RIM constructs. The Ontological Representation is comprised of a single OWL file and is primarily intended to be used for semantic validation and inferencing, but it may also be used for providing an alternate mechanism to share information (RDF), link to other models (ontologies), etc.

Note: The RIM and OWL representations of BRIDG have not yet been developed for Release 5.1

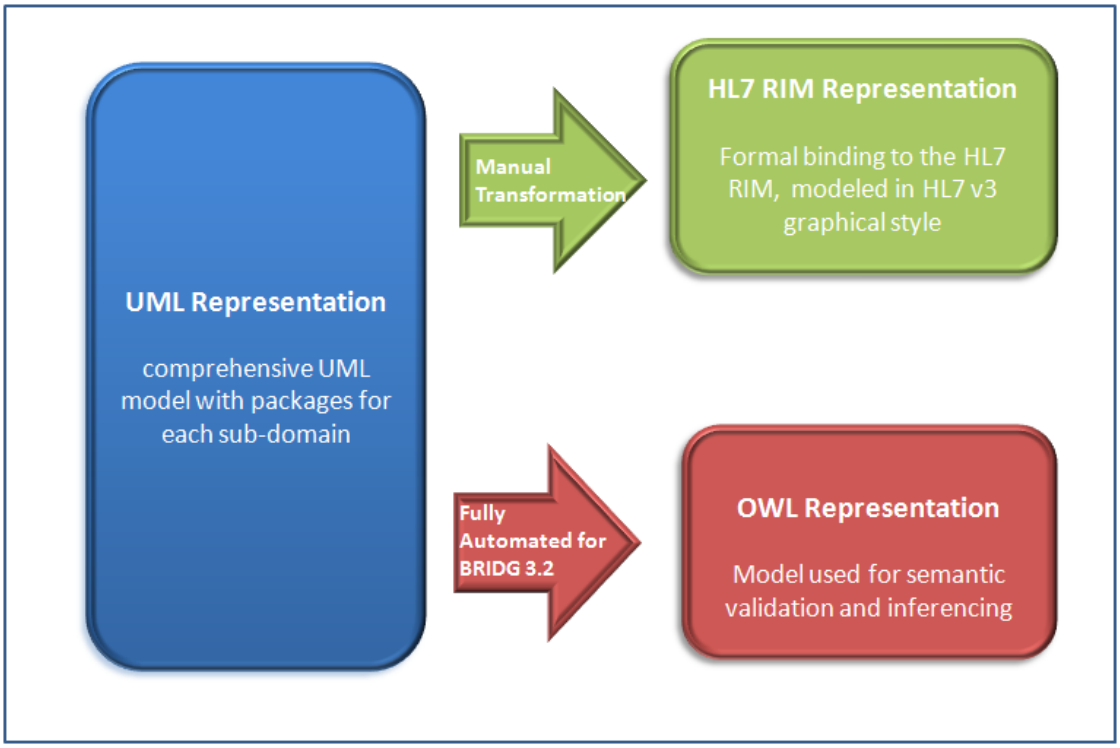

Figure 4: Diagram depicting the BRIDG multi-perspective approach

4.2 Canonical UML Representation

The intended audience of the UML class diagram is the domain experts. This audience should be technically savvy and have a good understanding of semantic modeling. The intent of the model in this representation is to allow the domain experts to visually see their domain semantics represented in an intuitive manner, such as a class diagram. Some minimal knowledge of UML class diagram constructs is needed to get

comfortable in reading the UML model. UML is a standardized general-purpose modeling language used to specify semantic requirements for a particular domain. UML offers a variety of diagrams to visually represent the semantics; BRIDG uses primarily class and instance diagrams. There is a brief introduction to UML class diagrams in the Appendix section of this document.

4.2.1 Comprehensive UML Model

The master BRIDG semantics are maintained in a UML model and the BRIDG modeling team maintains and uses the Comprehensive UML diagram for managing the BRIDG model. BRIDG modeling team members are the main audience of this comprehensive UML view of the entire BRIDG model and contains every class, attribute, association and constraint in the model.

NOTE: This comprehensive view is not intended for users to learn the model. It is created by the BRIDG modelers to ensure model integrity. It is highly recommended that you look at the sub-domain views and other additional views to learn and review the concepts and relationships in the BRIDG model.

4.2.2 Sub-Domains UML Views

The sub-domain views are intended to provide a more focused or controlled view of sub-set of the domain of BRIDG. The subject matter experts, analysts and architects can focus on these smaller models/views to verify the accuracy of the semantics captured in the BRIDG Model. The intent of the models in this perspective is to allow these domain experts to view the models and "see" their business. The sub-domain classification/grouping presents the data semantics based on a set of logical groupings as defined by the BRIDG modeling team and the SMEs from the community. These groupings also reflect a modeling team need in managing the model as well as a high level division in protocol life cycle. There has been an interest from the BRIDG user community to provide more granular groupings and views that reflect the use cases of the domain. The BRIDG team is reviewing that feedback. Three new sub-domains were defined to support the broader life sciences scope of BRIDG 4.0. No new sub-domains were added in BRIDG 4.1, but one more was added for BRIDG 5.0, so the current groups are as follows. Other sub-domains may be defined as needs arise.

4.2.3 Adverse Event

The Adverse Event sub-domain is intended for those involved in safety related activities; such as detection, evaluation, follow-up and reporting. This includes safety issues involving people or products. It also includes safety-related activities during or after a research protocol, such as post-market adverse event reporting.

4.2.4 BioSpecimen

The Biospecimen sub-domain includes concepts related to a biologic specimen, including collection and processing.

Collaboration with other HL7 Work Groups: With the broader scope of BRIDG 4.0 covering translational research and the harmonization of Life Sciences Model and

CDISC Pharmagenomic & Pharmacogenetics domains, it is likely that there is some overlap of concepts between BRIDG Biospecimen subdomain concepts and the models being developed in the HL7 Specimen (Orders & Observation) and Anatomic Pathology work groups. BRIDG recognizes these work group models as peer or sibling models and is committed to working with those teams to align on common semantics. The BRIDG team has started the conversation with the Orders & Observations work group members to this effect and plans to continue the dialogue on how to operationalize the collaboration and leverage the subject matter expertise of these HL7 work group members.

4.2.5 Common

The Common sub-domain is not intended for any one specific audience. It represents the semantics that are common to all (or most) of the other sub-domains. Most of the content is not even specific only to the BRIDG domain but might be common to any healthcare-related domain analysis model, including semantics for such things as people, organizations, places and materials.

NOTE: The larger scope of BRIDG in support of translational research has resulted in many new higher level concepts in the Common sub-domain. Review the BRIDG Backbone view in the Common sub-domain to understand the new Project related concepts. There are many other new classes that were added to the Common Sub-domain in Release 4.0.

4.2.6 Experiment

The Experiment sub-domain includes concepts related to the design, planning, resourcing and execution of experiments, which are intended to test hypotheses or lead to discoveries.

NOTE: The majority of the concepts in this new sub-domain are from the Life Sciences model (LS DAM) that was harmonized in the BRIDG 4.0 release.

4.2.7 Imaging

The Imaging sub-domain represents the core concepts related to imaging studies, images, annotations and other related concepts. This includes harmonization of semantics from parts of the DICOM standard and an NCI Imaging project (Annotation and Imaging Markup (AIM)). It doesn't intend to replicate all the semantics of imaging studies, series, images, annotations and reports, but rather contains summary-level key concepts which could serve as search criteria for interfacing between a BRIDG-based CTMS and a DICOM-based imaging system.

4.2.8 Molecular Biology (out-of-scope in BRIDG 5.1 HL7 ballot)

The Molecular Biology sub-domain represents the core concepts related to genetics and genomics, including gene, protein, molecular sequence, chromosome, genome, and numerous other related concepts. It also includes an initial set of links between these

concepts and previously existing concepts related to clinical studies. It is expected that these will get further fleshed out as this part of the BRIDG model matures in these new areas.

NOTE: The Molecular Biology sub-domain was started initially by harmonizing concepts from the LS DAM project and then was extended by harmonizing CDISC's Pharmacogenetics and Pharmacogenomics domains (PGx). LS DAM was substantially developed in a "top-down" kind of approach which means analysts and subject matter experts modeled concepts related to molecular biology from scratch essentially, by discussing terms, meanings, characteristics and relationships between concepts. However, the PGx domains were developed in a more "bottom-up" kind of approach which means that contributing projects identified specific semantics from applications that need to exchange data with one another and the concepts are grouped and organized accordingly. The resulting differences are that top-down modeling often identifies a wide range of high level information about which domain experts are concerned, but doesn't provide the detailed model elements or semantic rigor that a bottom-up approach typically provides. It's hoped that more detailed and use case specific models will be harmonized with BRIDG to help flesh out the high level concepts that were identified by the LS DAM project.

Collaboration with other HL7 Work Groups: With the broadening of the scope with BRIDG 4.0 to cover translational research and the harmonization of Life Sciences Model and CDISC Pharmacogenomic & Pharmacogenetics domains, it is likely that there is some overlap of concepts between BRIDG Molecular Biology subdomain concepts and the models being developed in the HL7 Clinical Genomics (CG) work group. BRIDG recognizes the Clinical Genomics work group models as peer or sibling models and is committed to working with the CG team to align on common semantics. The BRIDG team has started the conversation with the CG work group to this effect and plans to continue the dialogue regarding how to operationalize the collaboration and leverage the subject matter expertise of the HL7 CG group members. The CG WG has now started a sub group that is focused on building the Clinical Genomics Information Model.

4.2.9 Protocol Representation

The Protocol Representation sub-domain is intended for those involved in the planning and design of a research protocol. The majority of business requirements have come from those involved in clinical trial protocols. It focuses on the characteristics of a study and the definition and association of activities within the protocols, including "arms" and "epochs". It also includes the definitions of the roles that participate in those activities.

4.2.10 Regulatory – Change in Deprecation Plans

NOTE: NOTE: By agreement within the HL7 BR&R Work Group, the Regulatory sub-domain and its classes were marked as deprecated. The reason was that the classes have not substantially changed since BRIDG release 1.0 and are thought to be inadequately fleshed out. However, it has come to the attention of the BRIDG modeling team that a

use case for retaining some of the regulatory data elements may be presented by a BRIDG user. Consequently, the modeling team is deferring the deletion of the elements even though the sub-domain and classes have been deprecated for longer than the required minimum of year.

4.2.11 Statistical Analysis

The Statistical Analysis Domain Analysis Model includes concepts describing the planning and performance of the statistical analysis of data collected during clinical trial research and their relationships. This sub-domain currently represents the Statistical Analysis Plan semantics.

4.2.12 Study Conduct

The Study Conduct sub-domain is intended for those involved in the execution of a research study. The majority of business requirements have come from those involved in clinical trials. It focuses on the activities of conducting the study as well as the results from those activities.

4.3 HL7 Representation – RIM-based

Note: The RIM representation has NOT been developed for BRIDG 5.1.

The HL7 Reference Information Model (RIM) is a highly abstract comprehensive information model for the healthcare domain.

Because of its breadth, the RIM tends to use generic class, attribute and association names that are not necessarily domain-friendly. In addition, the RIM is completely free of the constraints and business rules that apply to domain-specific models. Its purpose is to provide a single set of reference semantics that can be leveraged across all healthcare domains. Additional, more specific, models are then created with strict derivation relationships to the RIM to support the implementation of communication interfaces.

Although the RIM is, for the most part, relatively free of implementation details, it is not a Domain Analysis Model (DAM) or a Domain Information Model (DIM) because, as just mentioned, it is not readily understandable by domain experts in any one of the listed domains (e.g., "Where are vaccinations in the RIM?", "How do I represent a provider credential?", "Where is a SNP found?", etc.)., This arises from the requirement that the RIM be an abstraction of cross-domain semantics. However, because it is important to the BRIDG stakeholders that BRIDG semantics be expressible in HL7 v3 XML, the modeling team is responsible for ensuring that BRIDG semantics can be represented in RIM structures. If this is not the case, the modeling team works with project teams to bring their specific semantics to HL7 for harmonization of the RIM, i.e. expansion of RIM semantics. To date, only a handful of such instances have been identified; some of these have already been submitted and been incorporated into the RIM. Others may need to be submitted in the future.

> NOTE: Saying that BRIDG semantics are mappable to the HL7 RIM does not mean that the mappings are one-to-one, e.g. attribute to attribute. In fact, they are most often **not** of that nature., i.e. a BRIDG class doesn't often map to a single RIM class (exceptions being concepts like "Person") and, likewise, a single attribute in the BRIDG Model may map to a combination of RIM attributes or a collection of RIM data type properties. The details of the mapping are not important. Semantic equivalence is the critical issue.

In some cases, RIM patterns have influenced BRIDG Model design where they have accurately reflected requirements and made sense from a domain perspective; for example, separating entities from the roles that played them, using 'typed' relationships between kinds of Acts, etc. In addition, some BRIDG attributes share the same name as the 'equivalent' attribute from the RIM. However, BRIDG is **not** truly a RIM-based model. It contains numerous classes, attributes and associations that are not directly derived from RIM classes, attributes and associations. This was necessary to ensure BRIDG served its primary purpose as a domain-friendly model.

Custom UML leveraging domain-friendly business names and putting attributes where-ever domain experts feel they make the most sense works extremely well for a domain analysis model. However, it does not work as well when the time comes to exchange data in a standardized manner or to map BRIDG concepts to HL7 communication specifications. For HL7, this communication needs to occur using instances based on static models that are strictly derived from the RIM. Because there is no strict derivation of the BRIDG Model itself, an additional set of models was needed. These RIM-based BRIDG Models mirror the content of the BRIDG but are created using HL7 tooling to ensure strict derivation from the RIM.

The HL7 perspective presents the same semantics as in the UML model but using HL7 RIM classes grouped into a central DMIM (Domain Message Information Model) diagram and a number of supporting diagrams. The HL7 models are constructed using HL7's Visio RMIM Designer tool and thus have a different graphical representation than a typical UML class diagram. However, the underlying constructs are all UML. The HL7 models have been carefully crafted to represent each of the associations and attributes present in the BRIDG domain analysis model. In some cases, a BRIDG attribute may be represented as the presence or absence of an HL7 association. In addition, HL7 requires many attributes that are not found in BRIDG. These are assigned fixed values and help to define the semantics of the HL7 model.

To allow cross-referencing between the BRIDG Model and the HL7 models, mapping annotations are captured within the HL7 classes and attributes identifying the BRIDG class, attribute or association end that corresponds to each element. Cross-reference reports are then generated to ensure a consistent representation between the UML and HL7 representations. In a few cases, mappings will not be one-to-one. This happens most frequently on status codes and similar coded attributes where the semantics found within a single BRIDG attribute must be spread out across multiple RIM attributes to adequately represent the semantics. In these circumstances, the mapping includes a

conditional clause that identifies what "subset" of the BRIDG attribute semantics map to a particular RIM attribute.

While the BRIDG HL7 models are "standard" HL7 models, they are not intended to be used directly as the foundation for exchanging messages. Instead, they serve as a basis of discussion with other HL7 groups who are modeling content relevant to BRIDG. By having the BRIDG semantics clearly expressed in HL7 terms, it will be easier to ensure that BRIDG requirements are incorporated in the various HL7 standards specifications related to the domains covered by BRIDG.

4.4 Ontological Representation

Note: The OWL representation has not yet been developed for BRIDG 5.1

The Ontological Perspective – OWL (Web Ontology Language) is a knowledge representation language for authoring ontologies. OWL stands for Web Ontology Language. It is a W3C-defined language for representing knowledge in a web-friendly way. Further information on OWL can be found here: http://www.w3.org/TR/owl2-overview/.

One of BRIDG's primary purposes is to act as a focus for mapping of data elements from different standards, specifications and implementations. OWL provides a number of capabilities that are useful in this process including helping to verify that mappings are not contradictory. In addition, some BRIDG stakeholders are leveraging OWL for their own processes and wanted a standardized representation of BRIDG in OWL to support that work. The BRIDG modeling team has now committed to publishing a representation of BRIDG semantics in OWL. The scope of the published OWL content is limited to the information found in the BRIDG UML model (as opposed to the RIM-based representation of the semantics).

BRIDG's OWL representation is expressed using one of the OWL syntaxes called OWL XML. It is not intended for direct human consumption. It can be navigated using web ontology tools such as the freely available Protege Ontology Editor: http://protege.stanford.edu/.

4.5 Relationship to Other Models

The domain for other models (such as the Tuberculosis DAM, or the Cardiology DAM) may intersect or overlap with the domain of BRIDG. (See section 2.3 for the formal definition of the domain for the BRIDG Model.) In these cases, the common semantics must be reconciled; BRIDG, the other model, or both should be changed to bring the models into alignment. BRIDG recognizes these other models as peers with a different scope and focus and plans to align on the common semantics in the future.

As explained briefly in section 2.6.1 above, the BRIDG model is moving towards the principle of "modeling-by-reference". "Modeling-by-reference" essentially means that when an established standard/model exists in a particular domain then not all the

semantics of the referenced standard will be harmonized with BRIDG. Instead the harmonization effort will focus on aligning the common semantics between the two domains to support implementable interoperability use cases.

4.6 Source Model Mappings

While there are mappings between various perspectives in the BRIDG Model, there is also a need for mappings between the canonical BRIDG UML Model and the source models and projects whose semantics have been harmonized into BRIDG. Such information is useful to the BRIDG modeling team so that they know the impact of any model changes under consideration, but also to the users of the BRIDG Model so that they know where their harmonized semantics "landed" in the BRIDG Model.

With BRIDG 3.0, the modeling team introduced a new convention for Mapping Tags in the EA file. Each UML class and attribute has one or more tags indicating the source model element from which the concept was derived or to which the element maps. The tags are comprised of a tag name indicating the type of tag and the source model (e.g. Map:AE) and a tag value indicating the name of a source model element (usually a class and attribute name).

The BRIDG modeling team also builds and maintains a document, an Excel mapping spreadsheet, that includes additional information, such as the full path of mappings for when a source concept spans more than one BRIDG class, "where clause" criteria or conditions that may apply to mappings, source model elements deemed implementation-specific, source model elements that remain a gap in BRIDG, etc. This mapping spreadsheet is published with every BRIDG release and explicitly shows the mapping of BRIDG model elements to the project semantics that were harmonized for that particular release.

It should be noted that the relationships shown in a given entry in a mapping document are often not one-to-one and often not at the same level of abstraction between the two models. The mapping documents form an essential part of the BRIDG knowledge base that project teams will use as they both begin new projects and extend existing ones.

It should also be noted that, as the BRIDG model evolves, the mapping tags in the UML model are maintained from version to version, however the various tabs in the Excel mapping spreadsheet are not necessarily updated with each new release of BRIDG.

BRIDG Mapping Tag Name	Harmonized Project's Full Name
Map:AE	CDISC, FDA, NCI Adverse Events Model
Map:AIM v4 rv48	NCI Annotation and Image Markup model
Map:BRIDGSCC	BRIDG Work Group Model Integrity Changes (added when the BRIDG team was called the Semantic Coordination Committee, SCC)
Map:BRIDGv2.2	BRIDG version 2.2 concept (otherwise unattributed)
Map:C3PR	NCI Central Clinical Participant Registry version 2

BRIDG Mapping Tag Name	Harmonized Project's Full Name
Map:C3PRv2.9	NCI Central Clinical Participant Registry version 2.9
Map:caAERSv2.2	NCI Cancer Adverse Event Reporting System version 2.2
Map:CDASHv1.1	CDISC Clinical Data Acquisition Standards Harmonization
Map:CDISCLabv1.0.1	CDISC Lab Model
Map:CDMHv1.0	FDA Common Data Model Harmonization
Map:CoopGrp	NCI Cooperative Group organizational semantics
Map:CTGOV	ClinicalTrials.gov
Map:CTOM	NCI Clinical Trial Object Model version 0.53
Map:CTR&Rr2	HL7 Clinical Trials Registration and Results Model release 2
Map:CTRPv1.0	NCI Clinical Trials Reporting Program version 1.0
Map:CTRPv3.8	NCI Clinical Trials Reporting Program version 3.8
Map:CTRR	HL7 Clinical Trials Registration and Results Model release 1
Map:CTRRr3	HL7 Clinical Trials Registration and Results Model release 3
Map:CTRv1.0	FDA Clinical Trial Repository version 1.0
Map:DICOM	ISO Digital Imaging and Communications in Medicine
Map:FDA HL7 SD SD DSTU2012	HL7 Study Design Structured Document Draft Standard for Trial Use (DSTU) R1 May 2012
Map:GSK MDRv1.0	Glaxo Smith Kline Meta Data Repository version 1.0
Map:HCTv1.0	Hematopoietic Cell Transplant (HCT) Common Data Elements (CDEs) from the Center for International Blood and Marrow Transplant Research (CIBMTR)
Map:HL7SD	HL7 Study Design RMIM
Map:HL7SDr1	HL7 Study Design RMIM release 1
Map:HL7SP	HL7 Subject Participation RMIM
Map:HSDBv1.0	Clinical and Translational Science Award (CTSA) Human Studies Database version 1.0
Map:ICSRr2	HL7 Individual Case Safety Report RMIM release 2
Map:Lab	NCI Lab Hub Model
Map:LabViewer2.2	NCI LabViewer Model version 2.2
Map:LSDAMv2.2.3Plus	NCI Life Science Domain Analysis Model version 2.2.3Plus
Map:NCI CRF Standard	NCI Standard Case Report Form (CRF) Common Data Elements (CDEs)
Map:PRM	CDISC Protocol Representation Model
Map:PSC	NCI Patient Study Calendar version 1
Map:PSCv2.6	NCI Patient Study Calendar version 2.6
Map:RPS1	HL7 Regulated Product Submission RMIM version 1
Map:RPS2	HL7 Regulate2 Product Submission RMIM version 1
Map:SDTM IGv3.1.1	CDISC Study Data Tabulation Model Implementation Guide version 3.1.1
Map:SDTM IGv3.1.2	CDISC Study Data Tabulation Model Implementation Guide version 3.1.2
Map:SDTM IGv3.1.3	CDISC Study Data Tabulation Model Implementation Guide version 3.1.3
Map:SEER 2015	NCI Surveillance, Epidemiology and End Results Program

BRIDG Mapping Tag Name	Harmonized Project's Full Name
Map:Statistics v1.0	CDISC Statistical Analysis Model version 1.0
Map:TDM	CDISC Trial Design Model version 1
Map:TDMv2	CDISC Trial Design Model version 2
Map:Vendor1v1.0	HL7 Clinical Trial Management Application Vendor
Map:Vendor1v1.1	HL7 Clinical Trial Management Application Vendor
Map:WHO	World Health Organization (clinical trials registry concepts)

4.7 Views (Diagrams) in BRIDG File

4.7.1 BRIDG – Start Here: Overview

Users unfamiliar with BRIDG should read this view. It provides an overview of the BRIDG project.

4.7.2 BRIDG Domain Analysis Model: BRIDG Sub-Domain Packages Diagram

This diagram shows the ten sub-domains and the classes in each sub-domain. There are two additional packages that have been added to provide a more focused look into a small sub-set of BRIDG semantics. The packages are called "CDISC View" and "Additional Focused Views". These smaller set of semantics are based on either a logical grouping or a disease area. The CDISC package is new in BRIDG 5.1 and shows BRIDG representation of two CDISC standards - Lab Model and SDTM. The complete list of these additional views is as follows:

4.7.3 Additional Focused Views

- Activities
- Adverse Events
- Biologic Entities and Related Activities
- CDISC Lab Model v1.0.1
- Observation and Results
- Oncology
- Organization
- Organization-Related Classes
- Participant Registration
- Performer
- Product
- SDTM 3.1.3
- SDTM Exposure (EX) Domain
- SDTM Disease Response (RS) Domain
- SDTM Vital Sign (VS) Domain
- Surveillance, Epidemiology, and End Results (SEER)
- Study Site and Study Subject

- Subject and ExperimentalUnit Comparison

4.7.4 BRIDG Domain Analysis Model: UML-Based Comprehensive BRIDG Model

This view shows the complete BRIDG Model (current release) and specifically shows, for each class where it's applicable, the complete set of attributes for the class, partitioning the attributes as to whether they are "local" to the class or inherited from the class' super-type hierarchy. It also shows all associations and class-level constraints.

Please Note: This comprehensive view of the BRIDG model is not intended for users to learn the model. It is created by the BRIDG modelers to ensure model integrity. It is highly recommended that you look at the sub-domain views to learn and review the concepts and relationships in the BRIDG model.

4.8 Class Diagrams

Class Diagrams are the *lingua franca* for representing the concepts, attributes, and relationships that define the static semantics of the domain of the BRIDG Model.

4.9 Instance Diagrams

Instance diagrams aid domain experts with no prior UML experience in reading, understanding, and validating, the content of the BRIDG Model. They are useful for exploring "real world" examples of objects and the relationships between them. Instance diagrams look similar to class diagrams, except the classes and attributes contain real world examples of data values so that one can see what types of data are expected. (**Please note**: *Only instance diagrams that were added for BRIDG Release 3.2 are aligned to 3.2 semantics; all previous diagrams are still aligned with BRIDG Release 3.0.2. The modeling team members may update these diagrams in a future release of the BRIDG model).*

New instance diagrams were added to BRIDG 4.0 for the CDISC SDTM PGx domains.

4.10 State Diagrams

State Diagrams depict the possible life cycle stages and legal transitions that an instance of a given concept may undergo during its existence and are associated with a small percentage of BRIDG static concepts. In common modeling practice, approximately 10% of classes in a given class diagram have sufficiently interesting life cycles to warrant an associated State Diagram. All classes for whom state is of interest carry an obligatory "statusCode" attribute.

The essential semantics of a UML *State Diagram* are:

- State (including "initiate" and "final status")
- Transition
- Guard (optional and not currently used in BRIDG)
- Intra-state Activities (not currently used in BRIDG)

The BRIDG Model includes several state transition diagrams representing several status code attributes which needed to be elaborated for the purposes of accurately representing the UML-based semantics in the RIM-based model. Here are some sample State Diagrams in the BRIDG file:

- Review Board Approval Status
- Study Overall Status
- Study Site Accrual Status
- Study Subject Status
- Submission Status

5 Appendix

5.1 Abstract and Complex Data Types

Definition of abstract data type (from Wikipedia): *"In computing, an **abstract data type (ADT)** is a specification of a set of data and the set of operations that can be performed on the data. Such a data type is abstract in the sense that it is independent of various concrete implementations. The definition can be mathematical, or it can be programmed as an interface. The interface provides a constructor, which returns an abstract handle to new data, and several operations, which are functions accepting the abstract handle as an argument."*

Virtually everyone is familiar with the concepts of integer and character. Most people have little trouble moving from a single character to the concept of a string of characters. Integer, character, and string are often referred to a simple or primitive data types. In contrast, complex data types combine primitives (and other complex data types) to express increasingly complex semantics, e.g. notions of date and time, physical quantities, names, etc. In a computationally tractable model, each attribute in a Class Diagram must be of a single data type, be it complex or primitive.

Analysis models, being implementation-independent, are often not concerned with specifying a particular data type for a given attribute of an analysis-level class. Reasons for not specifying data types vary from not knowing enough about the implementation specifics of a given attribute to reliably specify a data type with any real accuracy, fear of confusing domain experts with notions of computational concerns that are outside of their realm-of-interest, and/or lack of a common data type standard that can be reliably applied to a given domain. However, there are certain times when it becomes necessary to define analysis-level complex data types to capture the semantics of a particular domain concept. The two most common reasons for defining a complex data type in the context of a given analysis model are:

- a set of attributes that have inter-relationships (e.g. a Value and a Unit for a given measurement such as 8 inches vs. 8 days)
- a set of attributes that is used as an attribute in multiple classes (e.g. a DateTime stamp)

In the case of the BRIDG Model, the modeling team has decided that referencing complex data types was necessary in order to support the semantic interoperability requirements around which the BRIDG Project was initiated. Rather than define its own set of data types, the modeling team chose to leverage the HL7 and ISO data types specification specifically developed for the domains of Healthcare, clinical research, and life sciences. Each attribute in the static representation (Class Diagram) will be assigned to a specific HL7 Abstract Data Types R2 data type. The modeling team has made this commitment for several reasons:

1. The need for complex data types in the BRIDG Model (e.g. for coded concepts and various complex timing specifications) is becoming ever more clear;
2. The increasing emphasis of BRIDG as a key component in application development (where data type definition is critical to final interoperability requirements);
3. The viability and world-wide vetting of the HL7 data type specification has made it a natural choice for a DIM that lives within the composite domains-of-interest of HL7 (which BRIDG clearly does);
4. The modeling team's belief that, to domain experts, the concept of an Abstract Data Type is either interesting or understandable or uninteresting and unimportant, but in neither case distracting to the overarching problem of robustly defining domain knowledge using UML/BRIDG constructs.

Please Note: Since BRIDG R3.1, the BRIDG model is bound to HL7 Abstract Data Type R2. The HL7 Abstract Data Types Release 2 Specification is fully compliant with the ISO 21090 Data Type Specification, but provides a greater level of precision and expressivity over the semantics supported by an attribute and is more appropriate for the conceptual/logical level for which BRIDG is designed.

As of the publication date of this User's Guide, the HL7 Abstract Data Type R2 specification is freely available to the public on the HL7 website. A description of this specification and download instructions are here: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=264.

Following are the HL7 Abstract Data types that are being used in the BRIDG Model:

BRIDG Data Types Cheat-sheet	
Data Type	**Description**
AD	Address
ANY	Abstract type – can use any legal BRIDG type
BAG<AD>	Collection of addresses. Duplicates allowed
BAG<TEL>	Collections of phone, fax, e-mail, etc. Duplicates allowed
BL	Boolean
CD	Coded value
DSET<CD>	Collection of coded values
DSET<EN>	Collection of entity names
DSET<II>	Collection of identifiers
DSET<ON>	Collection of organization names
DSET<SC>	Collection of strings with optional code
DSET<ST>	Collection of strings
ED	Encapsulated data (text, images, video, etc.)
EXPR<PQ>	Physical quantity, possibly expressed as a formula
ID	Identifier (BRIDG-specific data type)
II	Instance identifier
INT.NONNEG	Integer of 0 or more
INT.POS	Integer of 1 or more
IVL<EXPR<TS.DATETIME>>	Interval of time, possibly expressed as a formula. Approximate values allowed
IVL<INT>	Collection of integers between two bounds. Negatives allowed
IVL<PQ>	Interval of physical quantities
IVL<TS.DATE>	All dates within a specified range, approximate dates allowed
IVL<TS.DATETIME>	All times within a specified range, approximate dates and times allowed
IVL<TS.DATE.FULL>	All dates within a specified range – full dates only
OID	Object identifier
PQ	Physical quantity
PQ.TIME	Physical quantity of time
REAL	Unitless real number
RTO<INT.NONNEG,INT.POS>	Ratio between two non-negative integers
RTO<INT.NONNEG,PQ.TIME>	Ratio of an integer per unit of time
RTO<PQ,PQ.TIME>	Ratio of a physical quantity per unit of time
RTO<PQ,PQ>	Ratio of two physical quantities
SC	String with Optional Code
ST	String value
ST.SIMPLE	String value where language is not relevant
TEL	Phone, fax, e-mail, etc.
TEL.URL	Universal Resource Locator (website, ftp, etc.)
TN	Trivial name
TS.DATE.FULL	Fully-specified date
TS.DATETIME	Point-in time. Partial values allowed
URG<INT.NONNEG>	Integer value greater than or equal to 1 between two specified bounds
URG<INT.POS>	Integer value greater than or equal to 0 between two specified

	bounds
URG<PQ>	Physical quantity that falls between two specified bounds
URG<PQ.TIME>	Duration value that falls between two specified bounds

5.1.1 Temporal Grammar(s): HL7 Abstract Data Type TS Data Type

The inherent temporal complexity of a clinical trial – that is, the importance of expressing complex timing relationships between planned, scheduled, and performed activities as well as activities and events that occurred as either outside the scope of the trial or as unplanned or out of sequence within the context of the trial – requires that complex temporal semantics be captured from domain experts. From a computational perspective, this requirement is best stated by saying that the model requires a temporal grammar. To date, the HL7/ISO data type Timing Specification (TS) meets most (if not all) of the requirements for such a grammar. As such, the modeling team uses that data type's semantics in the UML-based portion of the BRIDG Model.

5.1.2 Coded Concepts: HL7 Abstract Data Type CD Data Type

The modeling team has liberally used the CD complex data type through the UML-based part of the model to emphasize the consistent message from domain experts of the importance of the use of coded concepts rather than free text in capturing critical clinical trial information for computational analysis (e.g. reporting, cross-trial data pooling, etc.). It should be emphasized that the assignment of a given BRIDG attribute to the CD complex data types does not mean that the modeling team believes that a standardized set of coded concepts exist. Rather, the use of coded concepts most often begins with a locally-defined set of concept codes which, over time, is harmonized with other locally-defined sets of concept codes for similar purposes to eventually produce a standardized set of codes that can be used by all systems collecting data in a given context. At the point where a single set of coded concepts for a given concept exists, true computational semantic interoperability is realized.

5.1.3 Collections: BAG and DSET

Computational representations of collections of things recognize three types of collections with differing semantics:

- BAG – a collection of elements in which redundancy/non-uniqueness of elements may occur (for example, the set of phone numbers for a study subject – each phone number has a date range associated with it which is useful if the subject lives in Florida in spring and fall and in Maine in summer and winter)
- DSET – a collection of elements without order but with a guarantee of element uniqueness (e.g. the set of all names for a given organization)

The modeling team has utilized these collection semantics as needed throughout the BRIDG Model.

5.2 Instances as "KindOf" vs. "InstanceOf" – holding a place vs. filling it

The separation of the Planning phase of a study from the Scheduling or Performing phases raises an issue of how to represent the notion of a 'generic' occurrence for a particular instance of a class vs. representing an 'actual' occurrence of that instance. For example, when in the Planning phase of a study, one speaks of 'potential research subjects,' counting their numbers and assigning them to activities as if they were 'real,' knowing that they are simply placeholders for 'real' research subjects that will be recruited once the study officially begins, and knowing that these various 'real' research subjects will then 'replace' the 'generic' instances that are referred to in the planning phase.

The BRIDG Model must, of course, deal with these concepts of 'generic instances' vs. 'real instances' of a given class (the best – but not only – example being that of BiologicEntity). This is done through an attribute called actualIndicator, a Boolean that will have the value false if the instance is 'generic' and true if the instance is, in fact, real. Thus, a PlannedActivity will be associated with a BiologicEntity instance whose actualIndicator attribute = false, whereas a ScheduledActivity or PerformedActivity will be associated with a BiologicEntity instance whose actualIndicator attribute = true.

5.3 A Brief Explanation of the Concept of "Mood"

One of the most significant knowledge representation decisions made so far in the course of the BRIDG Project has been around the representational choices used to capture the well-known but not standardized representation of a somewhat obtuse concept that HL7 has named *mood* (evident in the value of the *moodCode* attribute of the RIM class *Act and its various subtypes.*) The name of the concept comes from the world of grammar:

> In linguistics, many grammars have the concept of **grammatical mood** (or **mode**), which describes the relationship of a verb with reality and intent. (Wikipedia)

The term is used in HL7 to distinguish "phases of a business process through which *multiple instances* of a concept can pass" from "state: the named phases of the life cycle of an *instance* of a concept." If one studies this definition carefully, one realizes that *the concepts of state and mood* each describe different perspectives on complex business processes and workflow, *state* describing phases of the lifecycle of a single *instance* of the concept, while *mood* describes the phases of the business process itself, through which *multiple instances – each in a different mood* -- may pass. An example of a state is that a PerformedActivity can be "aborted", "cancelled", "completed", or "on-going". All states in BRIDG are captured in the statusCode attributes of various classes. All of the statusCodes are then further specified in state diagrams, which can be found in the BRIDG State Transitions Diagrams section of the model. Examples for mood are further described below. For more details on the notions of mood and state, please see the HL7 website for the list of valid values for moodCode.

In the domain of BRIDG, there are several well-known phases of the clinical trial business process, e.g. *Plan, Schedule, Perform, Report.* In HL7 terms, these would be values of the *moodCode* of the various 'Act instances' that occurred during this 'business lifecycle' of a clinical trial that were collected within an instance of a particular *Study).* As is the case in many domains besides BRIDG, one finds that at the analysis model level, there are many common structures (i.e. classes and attributes) shared across business process phases. However, one also quickly discovers that there are small but essential differences, e.g. a *Performed Activity* is associated with a specific *Study Subject* whereas a *Planned Activity* is not, even though the two activities contain virtually identical information *with the exception that the Performed Activity uses actual dates and the Planned Calendar uses relative dates.* (A similar difference in "mood-based attributes" is found in clinical care when an order for a specific test does not contain a result value and contains somewhat different details of time specification, in many ways similar to the timing differences between a Planned Activity and a Scheduled Activity.)

Early versions of the BRIDG Model attempted to use a single attribute called *businessProcessMode* which was modeled directly off of the HL7 attribute *Act.moodCode* to simplify the model by allowing reuse of classes (with the caveat that there needed to be mood-specific restrictions on attribute usage). This approach was found to be too confusing and obfuscating to domain experts. As a result, the composite BRIDG Model contains explicit structural manifestations of each 'mood' of a clinical trial. Specifically BRIDG contains the 'moods' of "Defined", "Planned," "Scheduled," and "Performed."

5.4 BRIDG and the Four Pillars of Computable Semantic Interoperability

Computable Semantic Interoperability (CSI) rests on the four pillars identified in this section. This section also identifies how BRIDG addresses each of the pillars.

5.4.1 CSI Pillar #1: A Common Reference Model

The BRIDG Model is an instance of a common reference model. It should be noted that it is the express purpose of the BRIDG Model to inform data interchange, message specification and application and service development. Thus, the BRIDG Model defines -- and graphically represents -- both dynamic information (business process activity flows and interactions) and static information (the concepts, attributes, and relationships) that together define the semantics of the BRIDG domain. Both aspects of semantic clarity are essential if one is to achieve CSI to the degree deemed necessary by the BRIDG stakeholders.

It is beyond the scope of this document to describe the similarities and differences between the BRIDG Model and traditional purely static structures such as ontologies. Members of the healthcare informatics and knowledge representation communities are actively involved in developing the necessary transformations to move between BRIDG and OWL representations of BRIDG Model static content.

5.4.2 CSI Pillar #2: Complex Data Types

As of Release 1.1, each attribute in the static class diagram of the BRIDG Model has been bound to an appropriate complex data type. With Release 3.0, the BRIDG modeling team moved to using the HL7 and ISO data type standard and starting with Release 3.1, the modeling team is using the HL7 Abstract Data Type Release 2 specification *(please see note regarding this change in Release 3.1 release notes)*. This strategy is a reflection of the modeling team's assessment of the maturity and degree of adoption of this specification in combination with the modeling team's view that all of the semantics of interest to the modeling team and the BRIDG Model are currently represented in HL7Abstract data types release 2.

5.4.3 CSI Pillar #3: Binding to Terminologies

The binding of attributes in the BRIDG Model to specific terms within locally or globally specified terminologies – a process which occurs in the context of a run-time instance of BRIDG Model concepts – has not historically been part of the BRIDG Model's collective content. However, the binding of well-defined standardized vocabulary/terminology concepts to agreed-upon semantics (e.g. BRIDG Model attributes) is fundamental to achieving CSI and must therefore be addressed at some point in the evolution and use of the BRIDG Model. NCI has an effort underway to address this need, but the process is in the early stages and is expected to be a significant effort, not just because of the size and scope of the model but also because of reasons such as the lack of agreement in the extremely diverse community on what code systems to use. Additionally, terminology bindings may be provided by CDISC which has a team working on standardizing various coded concepts. However, at this time, it is up to the users of the BRIDG Model to determine at implementation time what terminology bindings will suffice for the data exchanges and interoperability they are targeting.

Thus, on the one hand, the BRIDG Model, is an implementation-independent representation of domain semantics applicable to a number of design/implementation solutions. However, each of these designs/implementations must, at some point in their lifecycle, specify appropriate terminology bindings as part of the computational-level semantic definitions of attributes. In order to achieve CSI, the BRIDG Model can – and must – specify these bindings in the future if the semantics of the BRIDG Model are truly to be "shared." The specification of terminology bindings will not make the BRIDG Model implementation-dependent. Rather, it specifies the shared semantics at a deeper/more comprehensive level than is possible without these bindings. In other words, the BRIDG Model can be "implementation independent" in the sense that it should not specify structures like primary or foreign database keys, but it cannot remain implementation independent in the sense of ignoring run-time terminology bindings.

5.4.4 CSI Pillar #4: Derivation of Data Interchange Structures from Pillars #1 - #3

With BRIDG Release 3.0, the modeling team introduced an HL7 RIM-based model of the BRIDG static semantics, akin to HL7's Domain Message Information Model (DMIM). At this time, the purpose of the RIM-based model however is more about ensuring the understanding of the semantics in the HL7 community than about creating a

model from which messages can be derived. In the future, it is hoped that the BRIDG RIM-based model may form the basis of a DMIM for the HL7 RCRIM Work Group. It is possible that NCI-defined service specifications based on the BRIDG Model may also serve as a means of achieving the goals of pillar #4.

5.5 The Concepts of Visit and Event

While the terms *visit* and *event* are heard quite often in the discussion of clinical trials/protocol-driven research, a careful examination of the meaning behind the concepts reveals that the terms, although usually understood in a specific contextual usage, can have a number of meanings which range from *a point in time where something is planned to happen, scheduled to happen, or actually happens* (visit or event) to *an activity of interest* (event) to *a collection of activities* (visit). When faced with terms as heavily overloaded (i.e. multiple meanings or interpretations for the same term), the modeling team tries to either establish new but related names or to define all of the various concepts described by the overloaded term with careful definitions and neutral, basically unused terms.

In the case of both *visit* and *event,* the modeling team decided that because both terms can cross the boundary between being *timing information* and *activity information* – not to mention the boundaries between the process steps of *planning, scheduling,* and *performing* – that neither term would explicitly be used in the BRIDG Model. Instead, the modeling team has been – and will continue to be – extremely explicit (including providing multiple example instance diagrams) in presenting the various core concept building blocks that collectively can be used to construct all of various complex meanings of terms like *visit* and *event* as they are used by domain experts throughout the BRIDG domain. The present version of the User's Guide contains several such examples. In addition, the definitions of the core concepts (e.g. *Arm, Epoch, etc.)* contain clear definitions and practical, real-world examples of the concepts. In addition, the modeling team has included a class named DefinedSubjectActivityGroup to include the semantics of this concept. Finally, the modeling team looks forward to additional use cases and clarifying dialogue with project teams and domain experts at any time.

5.6 Introduction to Basic UML Class Diagram Concepts and Terms

For readers who are unfamiliar with UML, a basic introduction to UML concepts is provided in this section.

5.6.1 What is UML?

The Unified Modeling Language (UML) is a software industry-standard language for specifying, visualizing, constructing, and documenting the requirements of software systems. Through UML, developers can visually describe and represent the components and activities of a system they are creating. This is done by using standard lines, arrows, connectors, shapes, and colors to draw diagrams. Generally speaking, there are two major categories of UML diagrams: static diagrams and dynamic diagrams. Within these categories, there are a variety of diagram types developers can choose from in order to represent the activities of a system.

5.6.2 What is a class diagram?

A class diagram describes the structure of a system by depicting classes, class attributes, and relationships. A "class" is usually an entity that represents a person, place, or thing. **Please Note**: To help understand the various structures/concepts of Unified Modeling Language (UML) Class Diagram, the authors of this document have often, though not always, used the BRIDG Model as an example UML Class Diagram. The various BRIDG examples provided below are not updated with every BRIDG release. Therefore, do not use these examples as reference to any given BRIDG Model release.

See **Figure A1** below for a non-BRIDG example.

The diagram shows a class called Employee. An Employee has certain attributes (e.g., a "name") and can perform certain functions. Classes are, moreover, represented as boxes or rectangles, as depicted by the boxes in the figure below. Each class has two compartments: the top compartment contains the class name and the bottom compartment contains one or more attributes. Also, classes are often related to each other in some way. Such relationships are depicted by different types of lines connecting the classes (i.e., "relationships" or "associations"), and multiplicity values tell you about the numeric aspects of that relationship.

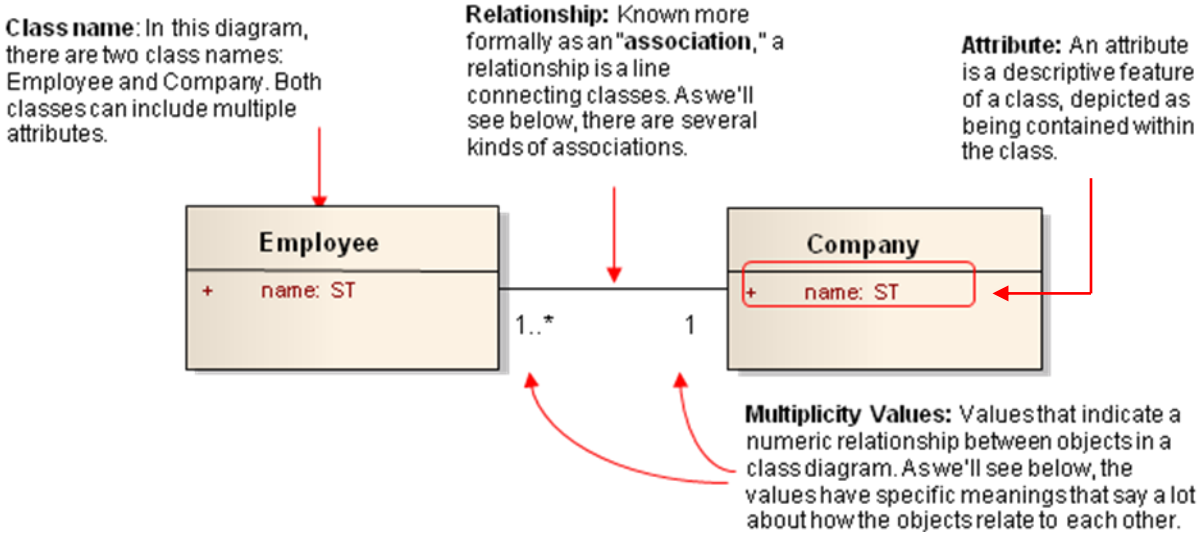

Figure A1: High level overview of the parts of a class diagram

This diagram is communicating several things. First, it's saying that both instances of the classes called Employee and Company have the attribute "name." Furthermore, the diagram says there is certain kind of relationship between an Employee and a Company. You'll notice a few other things like the numbers of the multiplicity values and the letters "ST" but don't worry about them now. We'll address them below.

5.6.3 Attribute names and data types

Let's look more specifically at the chief characteristics of a class's attributes, since this will help immensely in being able to read the BRIDG Model. As explained above, every class has a class name, and each class has one or more attributes. Most attributes have at least three parts: one, an **attribute name**; two, an **attribute type**; and three, a **visibility** mark, represented by a sign such as the plus (+) or minus (-) sign in front of the attribute name. The plus sign indicates that the attribute is "public," and the minus sign indicates that it is "private." The plus and minus sign are values for the scope (also called visibility) property. This property allows classes or attributes to selectively be displayed by class or diagram. Other marks can be used to indicate whether the attribute is "protected" or "package," but the BRIDG UML model does not use them.

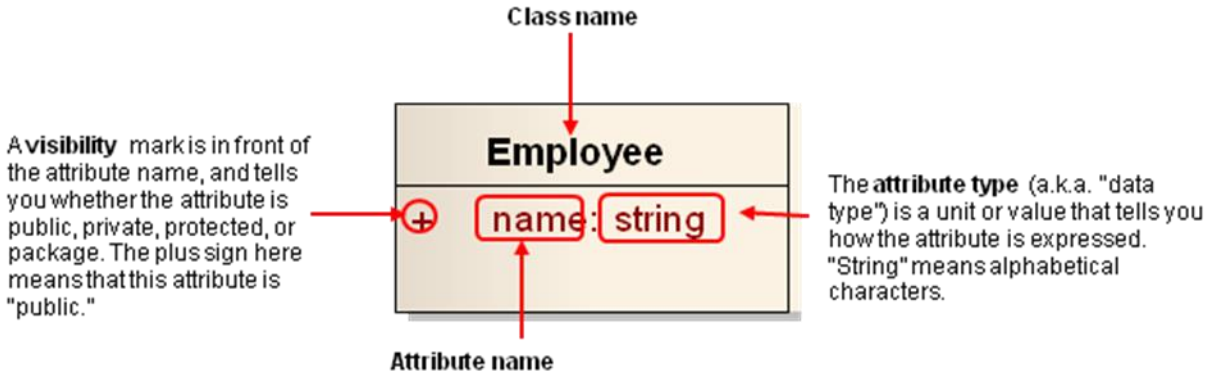

Figure A2: Anatomy of a UML class

What does this particular class tell us? First, the plus sign tells us the "name" attribute is public. In fact, in the BRIDG, all elements are public, so the only visibility indicator you will encounter in the BRIDG is the plus (+) sign.

What does the attribute type "ST" mean, though? "ST", which stands for "string", in particular means that the name attribute must be represented as a set of alphanumeric characters. For example, an employee's name could be "Julia Roberts". Of course, there can be other attribute types such as "integer" (i.e., numbers only) and "Boolean" (i.e., true or false), as shown in the next example. Together, this group of attributes makes up what is called an "attribute list" for a class.

To apply these concepts, let's look at a specific example of a class from the BRIDG Model. The class we will examine is the StudyProtocolVersion class.

5.6.4 Example of an attribute list of a class from the BRIDG

The StudyProtocolVersion class, shown in Figure A3, is one of the most important classes in the BRIDG Model. Let's use our newly-gained knowledge of attributes to interpret what it says.

First, notice that each attribute has a plus sign next to it, meaning that all the attributes are "public." Second, take a look at the attribute names. Many of them, such as

accrualReportingMethodCode and studySubjectTypeCode, are pretty self-explanatory to people working in clinical research. The former is a coded value used to represent the format for how subject accrual data should be reported back to the study sponsor (e.g., as "complete" or "abbreviated"). The latter is a coded value used to represent the target entity of the study of investigation. For example, in a clinical study, the subject type would be "human," but in other studies it could be animals such as "rats" or "mice."

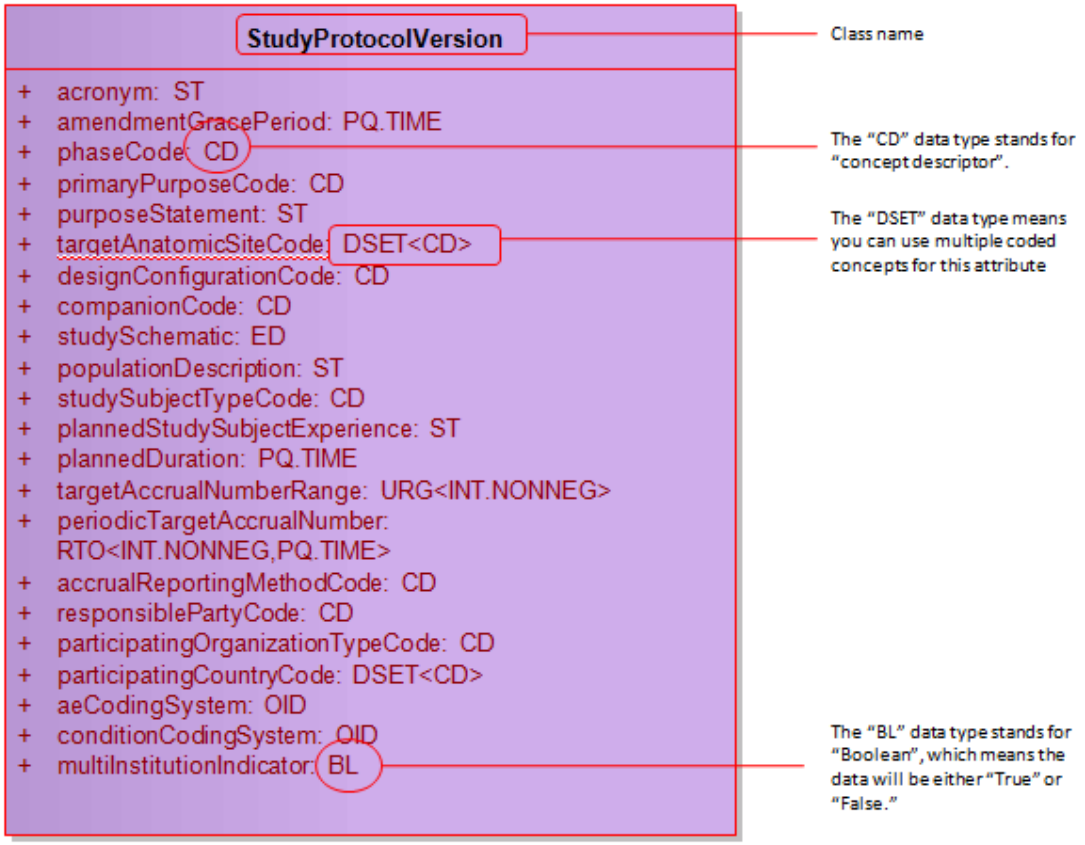

Figure A3: Attributes of the StudyProtocolVersion class

Next, notice that each attribute has a corresponding attribute type, or data type. These types are all based on HL7 abstract data type R2 specification. Each type indicates what kind of data is used to represent the attribute. For example, look at the last attribute in the list, *multiInstitutionIndicator*. You can imagine that this attribute is designed to answer the question, "Is this study conducted at multiple institutions?" The answer must be either yes or no. Hence, it makes sense that its data type is BL, or Boolean, which means the data can be expressed as either "true" or "false."

5.6.5 Different types of associations between classes in a UML diagram

The next important UML concept to know in order to read and understand BRIDG is the concept of relationships or associations. There are multiple types of relationships between classes: simple association, specialization, aggregation, and composition. While

the names sound pretty technical, these are relatively simple ideas, each of which we'll cover below.

5.6.6 Association

An association means that both classes are aware of each other and their relationship. That is, both classes are "known" to each other. An association is indicated by a solid line between the classes. At either end of the class is a multiplicity value. For example, in the diagram below, the multiplicity value of 0..* next to the StudySite class means that when an instance of HealthcareFacility exists, it can have *zero or more* instances of a StudySite associated with it. Taken in the context of clinical research, this makes sense. After all, even when a HealthcareFacility exists, it could take time for the first study site to be identified. Moreover, the number of study sites associated with a healthcare facility should be left open-ended in order for a variable number of study sites to be identified.

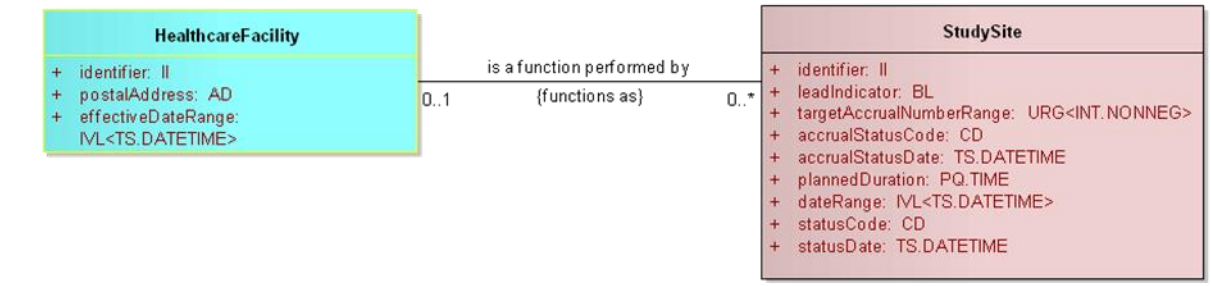

Figure A4: Example of an association between two classes

Conversely, the multiplicity value of 0..1 next to the HealthcareFacility class means that when an instance of StudySite exists, it can have zero or one *HealthcareFacilties* associated with it. This also makes sense in context, because in a particular study, a study site is not assigned to more than one healthcare facility.

There are, of course, a variety of ways you can designate multiplicity values. A few examples are shown in the table below.

Multiplicity Value	Meaning
0..1	Zero or one
1	One only
0..*	Zero or more
*	Zero or more
1..*	One or more
2	Two only
0..6	Zero to six
5..15	Five to fifteen

5.6.7 Specialization

Specialization is the ability of a class to receive or acquire the same exact attributes and functionality of another class, in addition to its own set of unique attributes and functionalities. This relationship is expressed in terms of "parent" and "child." Take an example from real life: a child may inherit a parent's brown hair. However, that same child may be the only one in the family to have blue eyes. Likewise in UML, a class may inherit attributes from a parent class but still retain unique attributes of its own.

In the context of UML, specialization is shown by drawing a solid line from the child to the parent class. At the end of the solid line is a closed, unfilled triangle (or arrowhead) pointing to the parent class. The child class also includes a listing of the parent class's elements under an italicized heading.

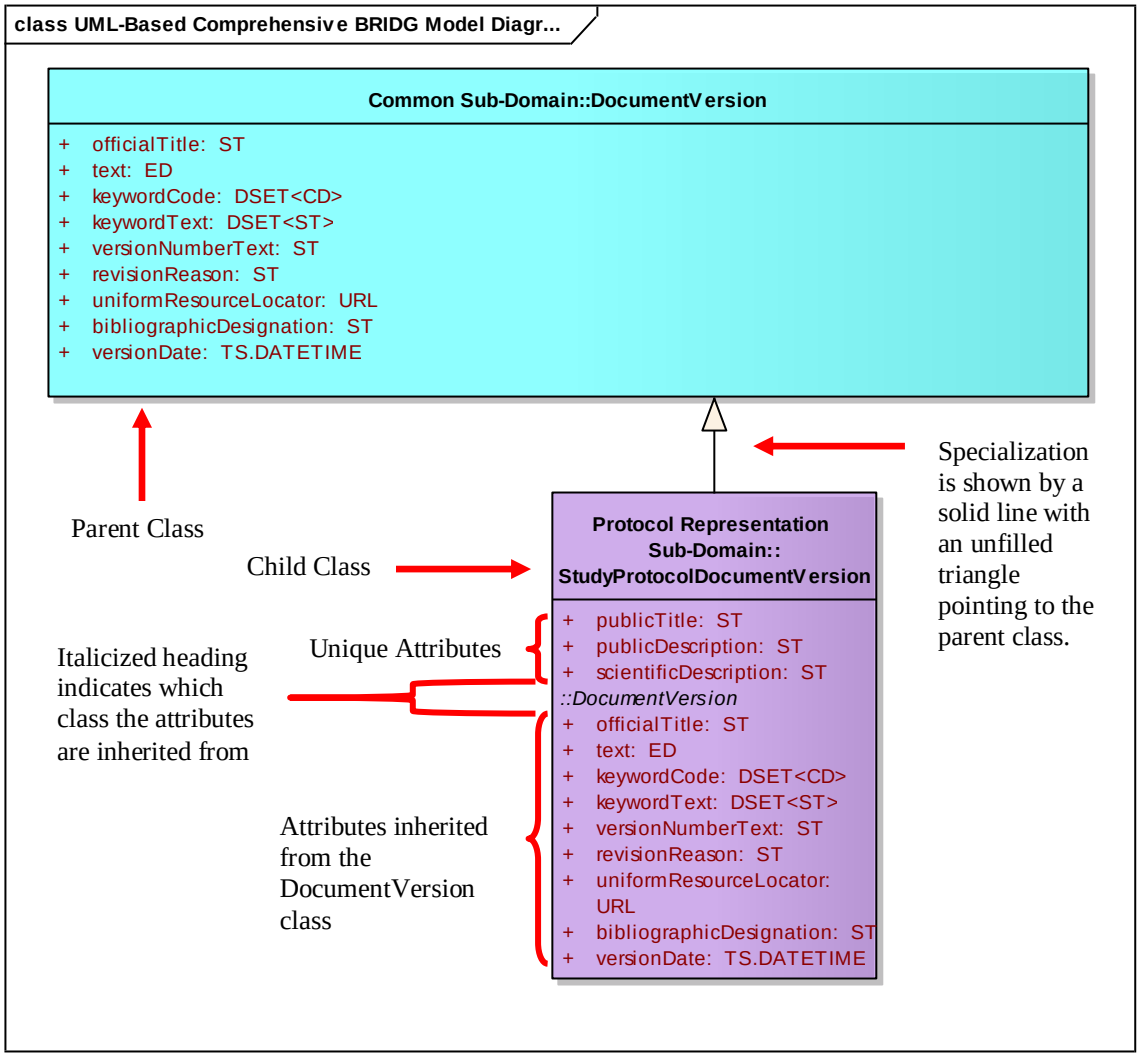

Figure A5: Example of specialization within BRIDG

As an example from the BRIDG Model, the StudyProtocolDocumentVersion class is a "child" to the DocumentVersion class. This means that, in addition to its unique

attributes, the StudyProtocolDocumentVersion class receives or acquires the full set of attributes from the DocumentVersion class.

Now consider this relationship in context. "Document" is a generic term that could encompass many kinds of documents. Indeed, all clinical research involves regulatory processes that require the submission of multiple documents from the Applicant to the Regulatory Authority. The DocumentVersion class is, therefore, an abstract concept that contains attributes common to all types of documents, such as officialTitle, versionNumberText, etc., which should be present no matter what type of document is being developed, whether an adverse event report or autopsy report. A StudyProtocolDocumentVersion is a specific type of document that falls under the broader category of document. It thus makes sense for the StudyProtocolDocumentVersion class to inherit attributes from the DocumentVersion class.

5.7 Aggregation

The association known as aggregation is a type of association that shows that a parent class is composed of or contains child classes. It shows how more complex parent classes (aggregates) are built from a collection of simpler child classes (component parts; e.g. a car from wheels, tires, motor and so on). If the parent class is removed, the child class is not. This relationship is represented by a solid line with an empty diamond shape on the parent class's association end.

In the diagram below (Figure A6), one or more BiologicEntityGroups is composed of zero or more BiologicEntityGroups. If BiologicEntityGroup is removed, BiologicEntity remains in the model.

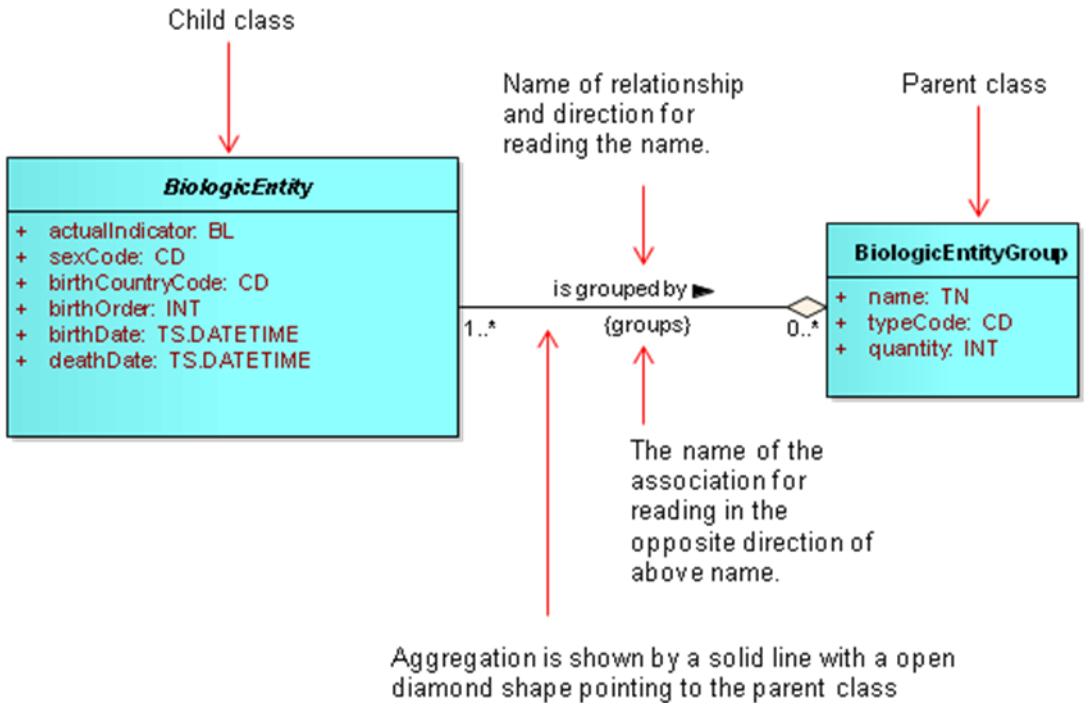

Figure A6: Example of an Aggregation relationship in BRIDG

5.7.1 Composition

The association known as composition or composite aggregation shows a "parts to the whole" relationship between a parent and a child class. In this relationship, the child class is *part of* the parent class, which is a stronger relationship than the aggregation relationship discussed above. Additionally, the child class's existence depends on the parent class's existence. If the parent class is removed, the child class is automatically removed as well. This relationship is represented by a solid line with a filled diamond shape on the parent class's association end.

In the diagram below, one or more Submission classes are part of the RegulatoryApplication class. There is only one RegulatoryApplication, and if it is removed, then Submission is removed as well.

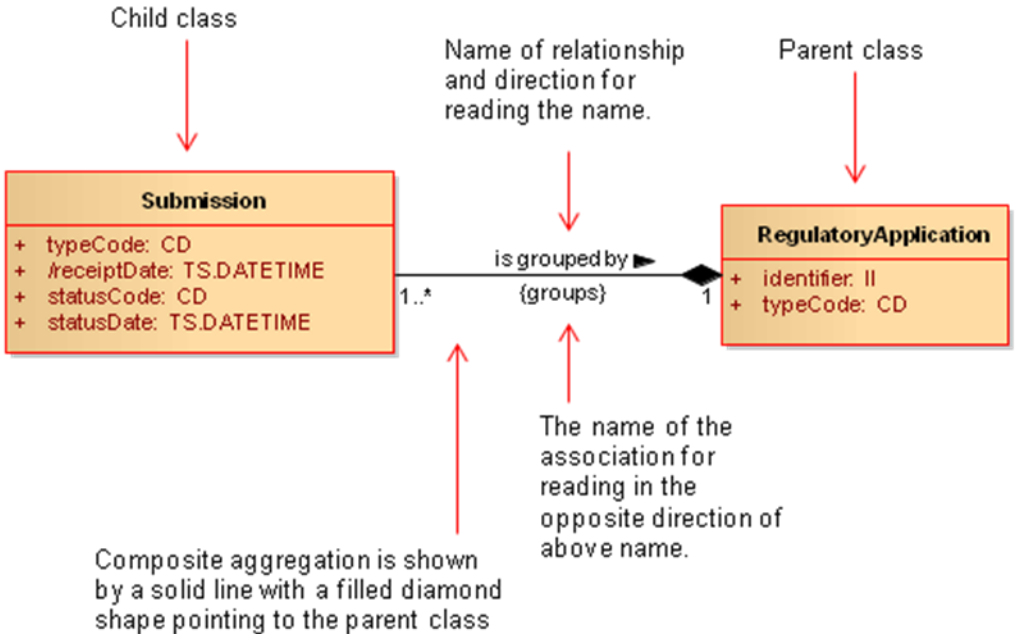

Figure A6: Example of an Aggregation relationship in BRIDG

As the name implies, the RegulatoryApplication class refers to a collection of submissions that are grouped together for regulatory purposes, and are usually specific to a particular device, food, feed additive or biopharmaceutical substance. The Submission class refers to any compilation of the contents of one or more submission units that supports a specific regulatory purpose or decision. It therefore makes sense that there is a composition or composite aggregation relationship between these two classes, since a regulatory application in the world of clinical research may consist of multiple submissions.

5.8 BRIDG Deprecation Policy

During the process of harmonization or model maintenance, the BRIDG modeling team periodically makes representational changes to the model. Most of these changes involve additions to the model, but some involve moving a concept or otherwise altering existing semantics to make the model more robust or more accurate. The BRIDG modeling team recognizes that occasionally BRIDG model elements are identified that are no longer needed, were erroneously added or came from source models that have since deprecated the source model element. The following process describes how such situations will be handled starting with BRIDG release 3.0.3.

When an item is identified as no longer needed, as erroneously added, or as coming from a source model that has since deprecated the source model element, the BRIDG modeling team will perform the due diligence required to determine whether or not the item is truly obsolete. This effort may include, but is not limited to, tracing the mapping tags to the source model and the appropriate mapping spreadsheet, researching related changes in the model change list, and discussing the item with a point of contact for the originating project if one can be identified. If this research effort does not bring to light a current reason for retaining the model element, the modeling team will mark the attribute, association or class in the BRIDG UML model using the "DEPRECATED" stereotype. This visually sets an item apart from the other items in the model and alerts potential users that the item will be dropped in a future release.

The modeling team will also move any tags still needed to an appropriate location in the model (if they are not already there). The corresponding elements in the RIM-based BRIDG representation might also be updated to include an implementation note indicating that the element is considered deprecated. The element will be published in the next release of BRIDG in this deprecated status – still available if needed but clearly indicating the modeling team's intent to drop the element at a later date.

The model element will then be retained in the BRIDG model for 1 year after its first release in this deprecated status. Any project team or BRIDG model user who has a use case for retaining the model element may contact the BRIDG modeling team (bridgTHC-L@list.nih.gov) to schedule a meeting to discuss their requirement. Any deprecated items that are subsequently deemed useful after all will be restored to non-deprecated status and published as such in the next release of the BRIDG model. Any items that have spent one year in the deprecated status with no requests to restore them will be dropped from the model all together in the first release of BRIDG after the one year mark.

A diagram of this process is shown in Figure A7 below.

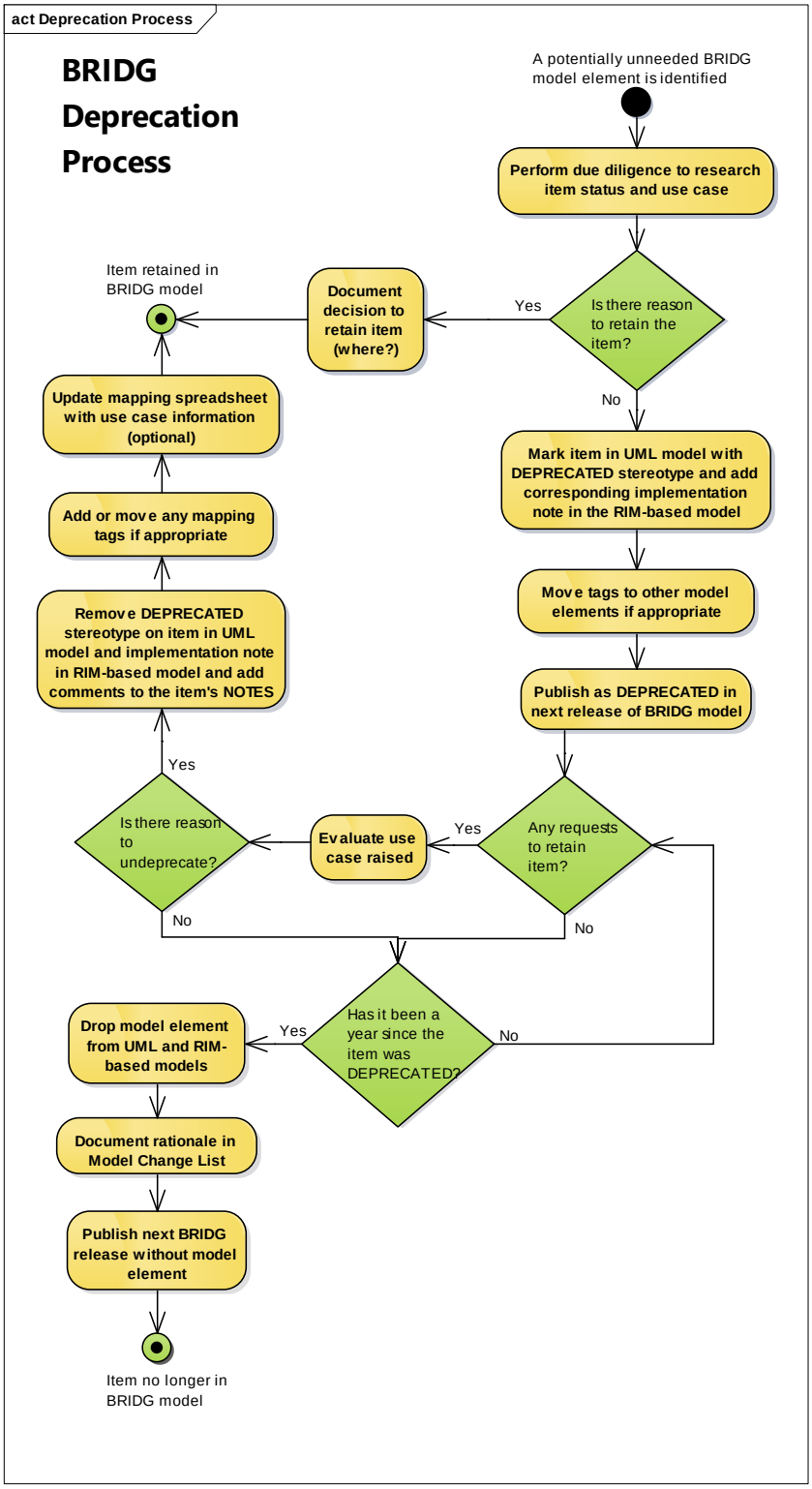

Figure A7: BRIDG Deprecation Policy Process Diagram

End of Document