

The BRIDG Project: A Technical Report

DOUGLAS B. FRIDSMA, MD, PhD, JULIE EVANS, SMITA HASTAK, CHARLES N. MEAD, MD, MS

Abstract **Objectives:** The Biomedical Research Integrated Domain Group (BRIDG) project is a collaborative initiative between the National Cancer Institute (NCI), the Clinical Data Interchange Standards Consortium (CDISC), the Regulated Clinical Research Information Management Technical Committee (RCRIM TC) of Health Level 7 (HL7), and the Food and Drug Administration (FDA) to develop a model of the shared understanding of the semantics of clinical research.

Design: The BRIDG project is based on open-source collaborative principles and an implementation-independent, use-case driven approach to model development. In the BRIDG model, declarative and procedural knowledge are represented using the Unified Modeling Language (UML) class, activity and state diagrams.

Measurements: The BRIDG model currently contains harmonized semantics from four project use cases: the caXchange project and the patient study calendar project from caBIG™; the standard data tabular model (SDTM) from CDISC; and the regulated products submission model (RPS) from HL7. Scalable harmonization processes have been developed to expand the model with content from additional use cases.

Results: The first official release of the BRIDG model was published in June 2007. Use of the BRIDG model by the NCI has supported the rapid development of semantic interoperability across applications within the caBIG™ program.

Conclusions: The BRIDG project has brought together different standards communities to clarify the semantics of clinical research across pharmaceutical, regulatory, and research organizations. Currently, the NCI uses the BRIDG model to support interoperable application development in the caBIG™, and CDISC and HL7 are using the BRIDG model to support standards development.

■ J Am Med Inform Assoc. 2008;15:130–137. DOI 10.1197/jamia.M2556.

Introduction

Clinical trials are an important method of scientific discovery. Every year, vast amounts of resources are invested into clinical trials to discover new mechanisms of disease, evaluate the effectiveness of new therapies, and assess the safety of therapeutic interventions.¹ As scientific discovery becomes more complex, it becomes increasingly important to systematically analyze—and integrate across different trials—data collected as part of clinical trials. Unfortunately, the valuable data gained from these studies are often collected in disparate databases that limit the ability to exchange,

share, and systematically analyze and integrate clinical trials data.

There have been a number of recent developments that have addressed this problem. Healthcare level 7 (HL7) was established in 1985 to develop system-independent data interchange specifications and messaging standards to support clinical applications in healthcare.² In 2000, the Clinical Data Interchange Standards Consortium (CDISC) was established to develop platform-independent standards to support the acquisition, exchange, submission, and archiving of clinical trials data between pharmaceutical companies and the FDA.³ More recently, HL7 (in collaboration with CDISC and the FDA), established the Regulated Clinical Research Information Management (RCRIM) Technical Committee (TC) to develop information exchange specifications relevant to clinical research. Additional initiatives to support global trial banks^{4,5} other standards development projects^{6–11} have tried to address the problem of sharing clinical trials data among organizations and researchers.

Although these standards development initiatives have been a step forward in standardizing and exchanging clinical trials data, there has been little coordinated effort across the standards development initiatives. The current challenge in exchanging clinical trials data has not been the lack of standards, but rather the existence of many divergent local

Affiliations of the authors: Department of Biomedical Informatics, Arizona State University (DBF), Phoenix, AZ; Clinical Data Interchange Standards Consortium (JE), Washington DC; ScenPro Inc. (SH), Washington DC; Booz Allen Hamilton/National Cancer Institute (CNM), Washington DC.

The authors thank Rebecca Kush, Ken Buetow, and Randy Levin for their leadership, and the hundreds of domain experts and modelers from caBIG, CDISC, and HL7 who have made the model possible.

Correspondence: Douglas B Fridsma, MD, PhD, Department of Biomedical Informatics, Arizona State University, Arizona Biomedical Collaborative, 425 N. 5th Street, Phoenix, AZ 85004-2157; e-mail: <fridsma@asu.edu>.

Received for review: 07/13/07; accepted for publication: 12/10/07

standards. Different standards in basic science or clinical research, as opposed to clinical care, limit translational research.^{6,7} The Biomedical Research Integrated Domain Group (BRIDG) project is an effort to coordinate the standards development efforts in RCRIM, CDISC, and the NCI, by both harmonizing existing standards, and developing the specifications for new standards to support clinical research.

Background

The BRIDG project developed from two complementary streams of research, one within CDISC and the HL7 RCRIM TC, and the other within the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG™) project. The goal of these projects has been (and continues to be) to support practical application and data interchange message development based on a shared understanding of the semantics of clinical research.

History of the BRIDG Project

In 2004, the National Cancer Institute initiated a research project called the cancer Biomedical Informatics Grid (caBIG™)—a large-scale research initiative aimed at creating an interconnected grid of researchers, patients, and data to support cancer research.¹¹ One of the early projects supported by the caBIG™ project was the development of a structured protocol representation that could be used to exchange clinical trial protocol information among caBIG™ participants. Rather than create “yet another standard”, the caBIG™ structured protocol representation project joined forces with CDISC and HL7 to understand the underlying semantics of the data collected as part of clinical research and to develop a shared model that could be used by all the stakeholders. Using the initial model developed by CDISC as part of the harmonization efforts between CDISC and the HL7 RCRIM TC as the basis for the initial modeling effort, the BRIDG project was initiated.

Organization of the BRIDG Project

The organization of the BRIDG project is based on other open source collaborative projects that have demonstrated success, as well as the HL7 Version 3 ‘harmonization’ process for managing the evolution of analysis artifacts developed in a disconnected, distributed environment.^{12,13} To support the collaboration, the BRIDG project uses GForge, an open-source collaborative software development environment,¹⁴ Subversion to manage code changes,¹⁵ standards-based modeling platforms (UML)¹⁶ and a transparent collaborative process to engage and coordinate a wide variety of participants.

Organizationally, the BRIDG project is divided into two areas. First, a BRIDG Advisory Board (BAB) was established with representation from each of the stakeholders (NCI, FDA, HL7, CDISC). The charge of the BAB is to identify the harmonization priorities for the BRIDG model, to coordinate standards development efforts between the various constituencies, and to provide a strategic direction to the project. It is the responsibility of BAB to facilitate the dissemination of the model and to solicit feedback from their stakeholder communities, regarding the adequacy of the model for their particular use.

In addition to the BRIDG Advisory Board, a Technical Harmonization Committee (THC) was established to help

with the operational details of managing, supporting, and interrelating the technical infrastructure of the BRIDG model. In keeping with other open source collaborative projects, the THC was purposely kept small so that representational choices and harmonization activities could proceed rapidly. The work was done with as much transparency is possible—all working copies of the model were uploaded to a public website and could be reviewed at any time in the development process. This transparency allowed the BRIDG project to solicit additional feedback from a larger community so if there was disagreement in some of the representational choices made within the model, the recommendations of the larger community could be included quickly as the model developed. The mantra of “build in small groups, test in large groups” was a fundamental organizational principle that successful open source collaborative projects have used and served as the foundation for the BRIDG project.

Design Objectives

In developing the BRIDG project and the BRIDG model, we followed specific design objects: A clearly defined scope; use-case driven priorities; an open-source approach to collaboration, harmonization, and consensus development; and implementation-independent representation of domain semantics. Each of these objectives is described more completely below.

Project Scope

Early on in the BRIDG project, we formalized the scope of the information to be contained within the model. Much of this was driven by the requirements of both CDISC and the HL7 RCRIM community, and the need for these two groups to effectively collaborate in development of standards to support a broad range of data in clinical trials and in the submission of regulatory information. Thus, the scope of the BRIDG model is defined as “protocol driven research and its 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, or device on a human, animal, or other biologic substance or substrate plus all associated regulatory artifacts required for our derived from this effort.”¹⁷

Use-case Driven Development

The BRIDG model was not developed in an abstract top-down method, but rather used use cases to define smaller projects and then developed organizational processes for harmonization to support integrating these use-case driven models into a larger implementation-independent model of the semantics of those use cases. For example, as part of the caBIG™ project, Northwestern University, was awarded a contract to develop a Patient Study Calendar application to support managing patients who were enrolled in protocol driven clinical research. As part of their development efforts, they constructed an analysis model that captured the requirements, the business processes, and the data structures necessary for their application. These artifacts were submitted to the BRIDG project technical harmonization committee and the semantics were incorporated within the BRIDG model. Overlapping concepts were harmonized with those defined within the patients at the calendar and new concepts for which there were no current existing concepts within the

BRIDG model were added to the model. Thus the BRIDG model is being built from the bottom up with each use case adding new concepts and definitions to the larger model.

Ultimately, as the BRIDG model grows, it will be possible for many application and message development use cases to be supported in a top-down fashion in which projects can use the BRIDG model as a starting point and add only those concepts and relationships that are necessary to support their new applications.

Semantic Harmonization

Although we use real-world use cases and application development to drive much of the model development, it is important to note that the BRIDG model is not just the aggregation of existing standards, or the construction of new content in the abstract. Instead, the model is intended to harmonize and unify existing resources (scoped to the BRIDG domain-of-interest) into one semantically coherent conceptual model. For example, concepts that are similar (such as *ScheduleOfActivities* or *ScheduleOfEvents*) are given a definition derived from consensus-driven synthesis of the two concept definitions. In other cases, subclass relationships are created when one concept is a more specific concept of another concept. In creating one semantically coherent conceptual model, all definitions for classes or attributes in the BRIDG model are based on consensus of all participants, and on feedback from community reviewers and domain experts.

Achieving Consensus

There are many different ways that teams can achieve consensus. In the following section we describe three alternative methods of achieving consensus, and describe the approach chosen within the BRIDG project.

One method of achieving consensus is *consensus through ambiguity*. This means creating definitions and processes that are sufficiently vague so that each of the stakeholders can interpret statements as they see fit. Often consensus statements developed in the UN or in other large organizational bodies use definitions and statements that all participants can agree upon, but which lack semantic clarity required for computable semantic interoperability. In achieving consensus through ambiguity, definitions are left purposefully vague, so that individual participants are free to interpret the consensus statement in a way that they see fit.

Achieving *consensus through abstraction* is another method of achieving consensus. Using abstraction allows individuals to gain consensus, but these concepts often lack domain specificity. The HL7 V3 Reference Information Model (RIM) has been extremely successful in achieving cross-domain consensus because the concepts represented within the RIM are abstract enough to be interpreted or modified for the local needs of different technical committees. For example, rather than speaking of a medical record number, the HL7 RIM uses an "instance identifier" as a more abstract concept that can be generalized to a medical record number for patients in an electronic medical record system, a accession number to specimens in a tissue bank repository, or to an invoice number in a financial billing system. Whereas cross-domain consensus is an appropriate and powerful way to view cross-domain semantics, it is a relatively ineffectual approach to representing the semantics of a single domain,

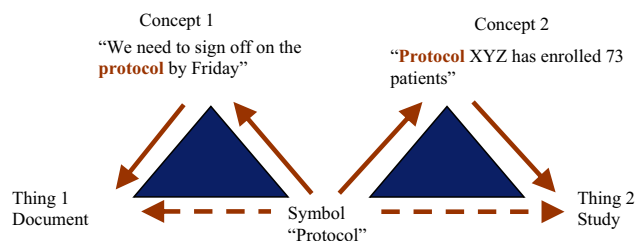


Figure 1. Type I semantic mismatch and the semiotic triangle. The notion of a semiotic triangle is a concept from linguistics. A symbol or name is associated with an underlying concept that individuals have within their minds. Confusion and difficulty with communication typically occur when the same symbol is applied by different people to different concepts. In this diagram, the same symbol "protocol" is used by two different individuals to refer to two different concepts. Concept 1 refers to a protocol as a policy document, while concept 2 describes the protocol as a study. Conceptually, it is not possible to enroll a patient into a document. The process of harmonization attempts to make explicit the semiotic triangles and forces domain experts to be explicit about the definitions that they apply to names within the clinical trials domain.

where the emphasis is on concept and process disambiguation realized through explicit presentation using domain-friendly terms.

In the BRIDG project, we have chosen to use *consensus through harmonization* rather than *ambiguity* or *abstraction* as a means of achieving consensus about the underlying semantics that exist between different models. When a disagreement or confusion arises over the semantics of a particular concept, we consider three different kinds of "semantic mismatches" that must be overcome to harmonize the semantics. Type I semantic mismatches occur when the same symbol or term are associated with multiple concepts. Type I semantic mismatches are resolved by creating two new concepts and asking domain experts to articulate the differences between those two concepts. New terms or symbols are applied to each of these concepts so that the "overloaded" term is not used. This is described in Figure 1.

Type II semantic mismatches occur when many different symbols or terms are applied to the same underlying concept. Type II semantic errors are much easier to resolve, particularly when the terms are explicitly defined and have illustrative examples associated with them. For example, the terms "protocol" and the term "clinical trial" may both refer to the concept of a "study." In harmonizing these terms, we chose a term to use in the BRIDG model, but provide explicit definitions (and references to the other terms) in the definitions of the concept. A Type II semantic mismatch is described in Figure 2. When there is a combination of Type I and Type II semantic mismatches we must construct both new concepts, new labels, and provide explicit definitions and examples.

Implementation-independent Semantics

Throughout the development of the BRIDG model, we have chosen to represent the semantics of clinical research in an implementation-independent way. In particular, the BRIDG model is not a database specification or design model that describes a particular architecture or implementation. It is

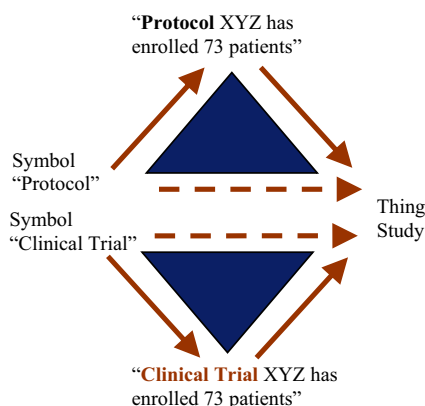


Figure 2. Type II Semantics mismatches. In this diagram two different symbols ("protocol" and "clinical trial") both point to the same concept but use different terms to describe that concept.

best characterized as an implementation-independent "information model" expressed in UML.¹⁸ The BRIDG model only captures the semantics and as much as possible, it excludes the implementation details of the clinical trial management application. For example, in a model drive architecture, BRIDG would be considered a Platform Independent Model (PIM) that could be used to create many implementation-specific models which share the same semantics and ensure the interoperability among applications based on the BRIDG model. This means that the semantics captured within the BRIDG model can be mapped to the HL7 V3 RIM and the message development processes of HL7RIM, the CDISC defined XML standard, and other implementation-specific representations of clinical research.

Technical Description

The BRIDG model exists as a collection of UML views (diagrams) that describe the declarative and procedural semantics of clinical research. It was built using a UML modeling tool called Enterprise Architect (SPARX systems), and can be read by using a free downloadable viewer (in the same way that .pdf files can be read by a free viewer, like Adobe Acrobat.) In addition to the Enterprise Architect version of the BRIDG model, the model is documented in RTF, and available for download in a standardized Unisys XMI file.

Declarative and Procedural Semantics

The semantics of a particular domain comes in two forms, declarative semantics (descriptions of objects, classes, and relationships between classes), and procedural semantics (descriptions of behaviors, work processes, and organizational processes).¹⁹⁻²¹ In the BRIDG model we represent both kinds of semantics. We represent the declarative semantics of clinical research in UML class diagrams that describe concepts and the relationships between those concepts, and represent the business processes or the procedural semantics of clinical research in UML activity and state diagrams.

This approach has been beneficial in a number of ways. First, when there is disagreement among domain experts about definitions or the semantics of declarative structures within the BRIDG model, focusing on the procedural aspects of the

semantics and the context in which the data is used helps to clarify the underlying definitions. For example, achieving consensus around definition of "adverse events" was made easier when domain experts examined the business processes that must be followed from the point in time in which a patient or subject experiences a symptom, its interpretation within the context of a clinical trial as being adverse, and the subsequent analysis required to determine causal and temporal relationships between the observed change in the patient's clinical state and its association the clinical trial activities. In understanding the definition of adverse events, the business processes clarified the context of the declarative semantics.

Second, within the BRIDG model, the activity diagrams provide an orthogonal view of the declarative semantics of the class diagrams and provided a double check of the declarative representations in the model. For example, if there are activities that require data that is not represented in the declarative structures, this suggests that there are missing concepts in the class diagrams. Likewise, if there are data elements that are referenced within the class diagrams but not used in any activity diagrams, these may be concepts that are unnecessary or redundant within the model. This validation helps to reduce redundancy and maintain consistency between the declarative and the procedural semantics.

Declarative Semantics

There are currently 10 classes that have been identified as comprising the "backbone" classes at the BRIDG model. These include:

- Person
- Organization
- Material
- Study Protocol
- Documentation
- Activity
- Activity Relationship
- Observation Result
- Observation Result Relationship
- Assessment

In representing the declarative semantics of clinical research, we have made a number of representational choices. Because the model has been adopted by HL7 as the domain analysis model for the RCRIM TC of HL7, there are similarities between the high-level classes of the HL7 V3 RIM, and the backbone classes of the BRIDG model. However, when choosing between more abstract RIM structures and explicit domain structures, we always chose to represent the explicit, detailed semantics of the domain to provide clarity. This means that the BRIDG model is informed by, but not an extension of, the HL7 V3 RIM.

For example the BRIDG model differs from the RIM is that we explicitly represent different phases of the clinical trial life cycle as subclasses of a more general class. The HL7 RIM uses "mood codes" designate similar concepts in different phases of the business process. Although this is an elegant modeling solution to a complex problem, it can be difficult for domain experts to understand. To make it easier for domain experts to define concepts across moods, we have descriptions of the clinical trial protocol in a "scheduled" phase as well as in a "performed" phase. This differentiation

helps with clarifying the semantics for domain experts. The backbone of the declarative semantics of the BRIDG model is shown in Figure 3, available as a JAMIA online document at www.jamia.org.

Procedural Semantics

One of the valuable features of the BRIDG model is the emphasis on capturing the procedural semantics of clinical research. This procedural knowledge is captured in activity diagrams that represent the business processes supported within clinical trials such as protocol authoring, protocol set up, protocol execution, and protocol analysis. These high-level activities are decomposed into finer-grained activities until they can be associated with the declarative information (data) associated with the clinical protocol. An example of the procedural semantics represented in an activity diagram is shown in Figure 4.

The activity diagrams clarify the semantics of the data that is used in clinical research by providing the context of how

(and in what processes) the data is used. These process descriptions provide a critical foundation for standardized services running in a service-oriented architecture and help articulate how different applications or messages to support clinical trials activities can be constructed. Within the caBIG community for example, these process descriptions can serve as the basis for developing clinical research services (such as “eligibility determination”) that will be available on the caBIG grid.

Modeling Approach

Developing Source Models

Many of the individuals in the BRIDG modeling community came from different organizations, and widely distributed geographically. The teams constructing source models used biweekly teleconferences, email listservs, shared project workspaces, and regular face-to-face meetings with facilitated modeling to develop their models, based on either

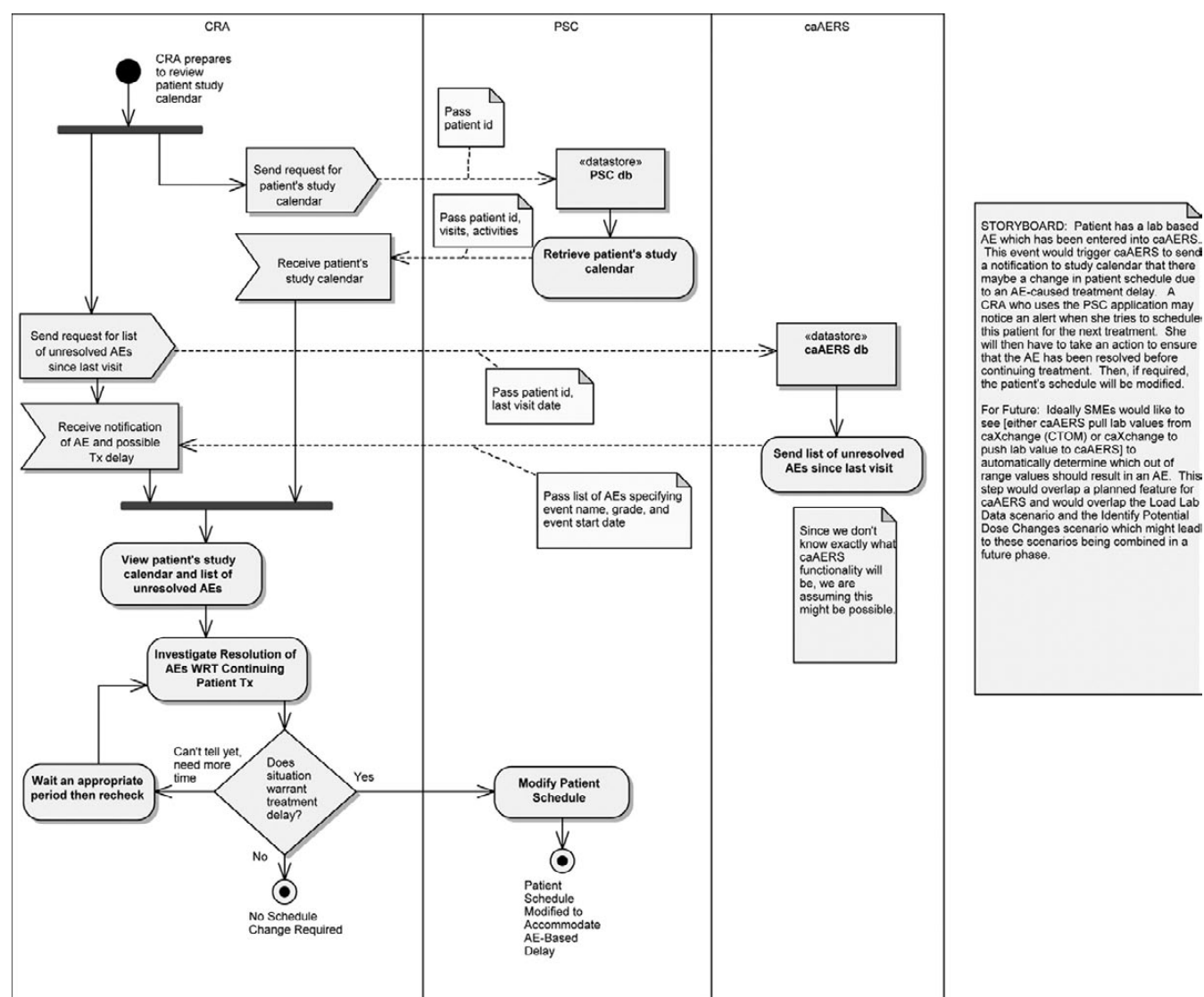


Figure 4. Capturing the procedural semantics using activity diagrams. This is an implementation specific diagram of the organizational processes for managing adverse events, and illustrates how data (declarative semantics) can be associated with particular activities (procedural semantics) used in the conduct of a clinical trial. This particular diagram describes how adverse events can trigger schedule changes for a patient in the data that is required to evaluate those changes.

specific use-cases or existing standards. Similar to the Delphi technique, facilitated modeling helped domain experts to articulate the semantics of the particular domain and to translate them into formal UML representations. Participants included a modeling facilitator, domain experts, and the UML expert to construct the model using a UML modeling tool. Using a spreadsheet to organize the work, and the facilitator to help the discussion, every class and attribute in the model was given a semantically rich, expressive definition and included examples of that concept in use. These definitions are based on not only the opinions of the domain experts, but on other standards efforts such as the CDISC glossary.³ The outcome of these modeling sessions was a set of “harmonizable artifacts” that included a mapping spreadsheet that described proposed relationships between the BRIDG model and the source model being constructed, and a UML model of the semantics of the use cases that included both declarative (data) and procedural (activity and state diagram) representations of the semantics.

Harmonization of the Source Models

Once the source models were completed and vetted within the team, the harmonizable artifacts were submitted to the THC. Representatives of the THC and the source model teams met face-to-face at a BRIDG harmonization meeting, during which the THC and all stakeholders reviewed the concepts in the mapping spreadsheet and developed a definition that satisfied all the participants. A model of 50–80 concepts (classes and attributes) could generally be modeled and harmonized over a 2½ day period.

Once the model was harmonized, the model was reviewed by the teams that submitted the model for harmonization to evaluate the model for errors or omissions. Thus, the harmonization process involved an iterative and cumulative process of knowledge assimilation and unification based on existing knowledge resources.

Representational Language

To represent the semantics of clinical research, we used the unified modeling language (UML).¹⁶ This decision followed from two important considerations. First, the primary use case for the BRIDG model was to support computable semantic interoperability. As a result, we wished to use a representational language that had strong industrial backing for application development and interchange specifications. Although other representational languages in formal ontology languages²² were considered, we found that UML was sufficiently expressive to capture the semantics of clinical trial concepts that have been articulated by domain experts and to support application and message specification development.

Second, we wished to provide a link between the domain expertise as captured in a domain analysis model and the technical expertise of software developers. By representing the domain experts semantics in a UML diagram, the BRIDG model was able to capitalize on the extensive application development environments constructed as part of model driven architectures.²³

Current Adoption and Utilization Status

There has been active use of the BRIDG model by standards, regulatory, and research organizations. Currently, the model has been adopted by HL7 as the official domain

analysis model for the regulated clinical research information management technical committee (RCRIM) and will serve as the semantic foundation for all HL7 messages related to regulated clinical research. CDISC has committed to harmonizing all of their existing standards to the BRIDG model and continues to support CDISC/BRIDG as the domain analysis model for their standards development. The National Cancer Institute (NCI) is actively using the BRIDG model to support application development within the caBIGTM program as part of the clinical trial management workspace. Finally, through the HL7 RCRIM technical committee, the FDA is developing four HL7 messages based on the BRIDG model to support electronic submission of Study Design, Study Participation, Subject Data and Adverse Event reporting.²⁴

The first official release of the BRIDG model was published in June 2007 and is available at <https://cabig.nci.nih.gov/inventory/infrastructure/bridg/>. This model includes harmonized semantics from four important projects: the caXchange project, (based on the CDISC lab message and the H7 periodic trial update message), and used within the caBIGTM program to support the exchange of laboratory information; the patient study calendar model, used by caBIGTM to support an application to manage patients on clinical trials; the standard data tabular model (SDTM, a CDISC initiative to support the submission of clinical trials data to the FDA); and the regulated products submission model, an HL7 RCRIM specification representing those artifacts necessary to support documentation and submission of data to the FDA for review. In addition, the BRIDG model release 1.0 has been informed by ongoing work with the CDISC trial design model (TDM), work on modeling adverse events, and other harmonization tasks.

Pursuant to the formal release of the BRIDG model 1.0, the model has been officially housed at the NCI GForge site with all the supporting documentation. The full model and release notes can be found at <https://cabig.nci.nih.gov/inventory/infrastructure/bridg/>.

Discussion

In the past three years, the BRIDG project has demonstrated a collaborative, open-source approach to the development of a model of the shared semantics of clinical research. In doing so, the BRIDG project has developed processes based upon open source, collaborative projects and have used these processes to provide an environment that is open source, collaborative and scalable. Many would argue that it is impossible to achieve global consistency across all of the semantics within the clinical trials lifecycle.²⁵ However, by separating the representations of the semantics from the implementation of those semantics, focusing on specific use-cases derived from real applications, and recognizing the need to harmonize within a wide community of stakeholders, we have achieved considerable success.

Limitations

We note that the overarching goal of the BRIDG project—and the BRIDG model that it has been building—has been to develop computable semantic interoperability. In juxtaposition to many other formal ontology development efforts, the BRIDG project was not intended to support inference or annotation of clinical data, nor was the BRIDG model *per se*

meant to serve as input into an inference engine. We also note that the BRIDG project presented one piece of the puzzle that must be solved to support computable semantic interoperability. For example, the BRIDG project currently does not include domain specific vocabularies or terminologies that are ultimately required for semantic interoperability. However, we recognized the importance of specific vocabularies and terminologies to support semantic interoperability. Similar to the HL7 version 3.0 abstract datatype Concept Descriptor (datatype CD),² many of the attribute concepts in the BRIDG model are modeled with a datatype of type “CD” that will bind to specific terminologies (e.g., a SNOMED-CT conceptID) as part of developing implementation specific solutions. Many of the stakeholders involved within the BRIDG project are developing processes and applications to support the full spectrum of models, common data elements, and terminologies necessary for achieving computable semantic interoperability.

For example, the NCI uses the cancer Data Standards Repository (caDSR)^{9,10,26} to represent and reuse Common Data Elements (CDE) across applications that use the BRIDG model. In addition, the Enterprise Vocabulary Service (EVS)²⁶ is a source to manage the vocabulary and terminologies necessary for cancer research. Using a unified process approach to application development, the NCI integrates a shared analysis model across to application development, integrates these models into a repository of common data elements, and binds these common data elements with vocabulary and terminology, drawn from its vocabulary servers.

An Example of Success: The CTMSi Project

Although a formal evaluation of the BRIDG model has not yet been done, we have demonstrated practical success in using the BRIDG model to achieve semantic interoperability within the caBIGTM project. The Clinical Trials Management Systems interoperability (CTMSi) project was an initiative sponsored by the caBIGTM program to demonstrate semantic interoperability across five applications in the clinical trials management systems workspace that would support the registration of a patient, the import of laboratory data related to that patient, and the assessment of that laboratory data to report a potential adverse event. Three of the five applications (patient study calendar, caXchange, and adverse event reporting) had harmonized their data definitions with the BRIDG model prior to the initiation of the CTMSi project.

The CTMSi project started with the BRIDG model as the basis for the analysis model and in collaboration with domain experts, developed story boards and activity diagrams for each of the use cases and mapped these activity diagrams and data elements to the BRIDG model. In the course of nine weeks (including holidays), the teams developed implementation-level exchange specifications between the applications and were able to demonstrate registering a patient, collecting laboratory information on the patient, and evaluating the patient based on those labs for an adverse event. The conclusion was that the project could not have succeeded in that time frame without the use of the BRIDG model as the semantic foundation for interoperability.

Next Steps

There is still much that needs to be done. The BRIDG project is a complex large-scale distributed collaborative informatics research project, with different stakeholders, who may have competing priorities and different user groups that will contribute to and draw from the BRIDG project. Success in the BRIDG project will require addressing educational, technical, and organizational challenges that collaborative research demands.

Education and Best Practices

The BRIDG project requires input from both domain experts who can articulate the key semantics of a domain, and technical experts who can translate that information into interoperable software and interchange messages. Domain experts must understand how the BRIDG project can help solve their problems, and technical experts must understand how to use the BRIDG model to design interoperable solutions. Educational materials that provide a non-technical view of the BRIDG project for domain experts, as well as detailed descriptions of modeling and harmonization best practices for technical experts will be needed. As the project matures and more individuals use the BRIDG model, an ongoing evaluation, modification, and refinement of these processes will be needed.

Analysis of Other Clinical Research Models

The BRIDG project will also benefit from a more critical analysis of how the legacy models have served as the basis for the model (such as HL7 and CDISC), and more detailed analyses of how the BRIDG project model compares to other initiatives that are developing models for clinical research. Although the CTMSi project demonstrated practical success in using the BRIDG model for interoperability, a more detailed and formal analysis of how the BRIDG model relates to other clinical trial models would improve the quality of the BRIDG model, and articulate points of collaboration across other related initiatives.

Collaboration and Tool Support

As the BRIDG project matures, we anticipate additional tools will be needed to support the model development, validation, and maintenance. For example, the process of semantic harmonization in the BRIDG project can be a time consuming manual process. Additional tools to support the harmonization process will help to resolve bottlenecks in model development, and speed the construction of the BRIDG model. Other tools are needed to support the technical experts developing interoperable applications. For example, in much the same way that the unified process requires additional tools to help with “requirements traceability”, the BRIDG project will require additional tools to support “tracing” the semantics from the BRIDG model to design and implementation models. Maintaining the provenance of the semantics ensures that semantics used within a particular application matches the semantics in the BRIDG model.

Finally, the development of collaboration tools to support the community-driven modeling of the BRIDG project across groups of distributed domain experts will be an important challenge to overcome. By developing semi-automated and automatic methods of model construction, vali-

dition, and maintenance, the BRIDG project can continue to support interoperability in clinical research.

Conclusions

The BRIDG project has achieved some important milestones. First, the BRIDG project has brought together different standards communities to work together in constructing a shared semantics of clinical research. CDISC is an important standards organization in the clinical trials community and HL7 is an important standards organization in clinical care. With the increased emphasis on translational research, the BRIDG project could serve as a “semantic bridge” between clinical research and clinical care.

Second, by using UML as a modeling formalism, the BRIDG model can serve as the initial analysis model for model-driven application development as well as the analysis model for the RCRIM driven HL7 V3 message specifications. Tools that support unified process and model-driven approaches to software development use UML, and should speed the development of interoperable applications and messages based on BRIDG.

Finally, the BRIDG project has begun to clarify the semantics of clinical research across pharmaceutical, regulatory, standards, and research organizations. As one of the important pillars of semantic interoperability, domain models such as BRIDG provide a context in which terminology and common data elements can be integrated. The earlier in the process of application development in which the semantics can be harmonized, the easier it will be for different implementations to share common semantics. Ultimately, the success of the BRIDG project will come in the adoption of the model by increasing numbers of stakeholders, and the demonstration of computable semantic interoperability between clinical trials information systems.

References ■

- CenterWatch. Grant Market to Exceed \$4 Billion in 2000. Centerwatch Newsletter 2000;17(11):193.
- HL7. Health Level 7. 2007. Available at: www.HL7.org. Accessed Jan 6, 2008.
- CDISC. Clinical Data Interchange Standards Consortium. 2007. Available at: www.CDISC.org. Accessed Jan 6, 2008.
- Sim I, Detmer DE. Beyond trial registration: a global trial bank for clinical trial reporting. *PLoS Medicine/Public Library of Science* 2005;2(11):e365.
- Sim I, Rennels G. A trial bank model for the publication of clinical trials. *Proc Annu Symp Comp Appl Med Care* 1995: 863–7.
- Wang X, Gorlitsky R, Almeida JS. From XML to RDF: how semantic web technologies will change the design of ‘omic’ standards. *Nature Biotechnology* 2005;23(9):1099–103.
- Nardon FB, Moura LA. Knowledge sharing and information integration in healthcare using ontologies and deductive databases. *Medinfo* 2004;11(Pt 1):62–6.
- Sim I, Rennels G. Standardized Reporting of Clinical Trials into Electronic Trial Banks: In Support of Computer-assisted Evidence-based Medicine; 1996. Report No.: SMI-96-0630.
- Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P. Common data element (CDE) management and deployment in clinical trials. *AMIA 2003; Annu Symp Proc AMIA* 2003:1048.
- Phillips J, Chilukuri R, Fragoso G, Warzel D, Covitz PA. The caCORE Software Development Kit: streamlining construction of interoperable biomedical information services. *BMC Med Inform Dec Making* 2006;6:2.
- von Eschenbach AC, Buetow K. Cancer Informatics Vision: caBIG. *Cancer Informatics* 2006;2:22–4.
- Mockus A, Fielding RT, Herbersleb JD. Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*; 11(3):309–46.
- Fielding RT. Shared leadership in the Apache project. *Communications of the ACM* 1999;42(4):42–3.
- Copeland T, Mas R, McCullagh K, Perdue T, Smet G, Spisser R. GForge Manual. 2006. Available at: http://gforge.org/docman/view.php/1/34/gforge_manual.pdf. Accessed Jan 6, 2006.
- Pilato CM, Collins-Sussman B, Fitzpatrick BW. Version Control with Subversion: O'Reilly; 2004.
- OMG. Unified Modeling Language. UML Resource Page. Available at: <http://www.uml.org>. Accessed Jan 6, 2008.
- Evans J, Fridsma DB, Hastak S, Mead CN. BRIDG V 1.0 Release notes. 2007. Available at: www.BRIDGproject.org. Accessed June 15, 2007.
- Mead CN. Data interchange standards in healthcare IT—computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? *J Healthc Inform Manag* 2006;20(1):71–8.
- Anderson JR. Rules of the Mind. Hillsdale, NJ: Lawrence Erlbaum; 1993.
- Friedland P. Acquisition of Procedural Knowledge From Domain Experts. In: Hayes P, editor. *International Joint Conference on Artificial Intelligence*; 1981; Vancouver, BC: William Kauffman; 1981. p. 856–61.
- Georgeff M, Bonollo U. Procedural Expert Systems. In: *International Joint Conference on Artificial Intelligence*; 1983; Karlsruhe, Germany; 1983.
- Musen MA. Domain Ontologies in Software Engineering: Use of Protege with the EON Architecture. *Methods Inform Med* 1998;37:540–50.
- Ambler S. The Object Primer: Agile Model-Driven Development with UML 2.0. 3rd ed. New York: Cambridge University Press; 2004.
- FDA. Food and Drug Administration. CDISC Data Standards Resources. 2007. Available at: <http://www.fda.gov/oc/dataacouncil/cdisc.html>. Accessed Jan 6, 2008.
- Wiederhold G. The impossibility of global consistency. *Omics J Integ Biol* 2003;7(1):17–20.
- Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, et al. caCORE: a common infrastructure for cancer informatics. *Bioinformatics* 2003;19(18):2404.