|  |
| --- |
|  |
| BRIDG-LS DAM Integration Proposal |
| Version 1.4 |

|  |
| --- |
| Smita Hastak & Lisa Schick, ScenPro, Inc.  6/19/2012 |

Table of Contents

[Document History 3](#_Toc327874137)

[Overview 4](#_Toc327874138)

[1. Rationale for Integrating BRIDG and LS DAM 4](#_Toc327874139)

[2. LS DAM Background 5](#_Toc327874140)

[a. LS DAM Project Goals 5](#_Toc327874141)

[b. LS DAM Project Scope Statement 6](#_Toc327874142)

[c. LS DAM Project Stakeholders 6](#_Toc327874143)

[d. The LS DAM Project Team 7](#_Toc327874144)

[e. Project Release History 8](#_Toc327874145)

[f. LS DAM Core Areas 9](#_Toc327874146)

[3. LS DAM’s BRIDG-based Approach 14](#_Toc327874147)

[a. Modeling Approach 14](#_Toc327874148)

[b. Modeling and Naming Conventions 15](#_Toc327874149)

[c. Tooling 15](#_Toc327874150)

[d. Release Artifacts 16](#_Toc327874151)

[4. Approach to Integration 18](#_Toc327874152)

[a. Mechanics of the Integration 18](#_Toc327874153)

[b. Resolving Semantic Differences 20](#_Toc327874154)

[c. Maintenance and Model Management 21](#_Toc327874155)

[Effects on BRIDG and Open Questions 22](#_Toc327874156)

[References 22](#_Toc327874157)

[Glossary 23](#_Toc327874158)

# Document History

|  |  |  |
| --- | --- | --- |
| **Version Number** | **Description of Changes** | **Date** |
| Version 1.0 | Initial document | March 27th, 2012 |
| Version 1.1 | Addressed feedback from Juli Klemm | April 5th, 2012 |
| Version 1.2 | Addressed feedback from BRIDG SCC members | April 9th, 2012 |
| Version 1.3 | Final edits and clarifications after review by NCI leadership – submitted to BRIDG Board Members | April 30th, 2012 |
| Version 1.4 | Added Rationale from Juli Klemm ; Added data type clarification from Charlie Mead ; Since a new version of LS DAM (2.2.3) has been published since version 1.3 of this document was published– updated the related content in Approach to Integration section and removed the steps that have already been completed. Other clarification edits from Baris Suzek; Updated the responses to Open Questions based on discussions with the Task Force members; formatting changes | June 19th, 2012 |

Overview

This document contains a proposal for integrating the Biomedical Research Integrated Domain Group (BRIDG) and the Life Science (LS) Domain Analysis Models (DAMs). The proposal summarizes the rationale for integrating the models and details the process by which integration may be achieved. With the support of the BRIDG Board of Directors (BoD), the leadership of the National Cancer Institute (NCI) assigned the task of developing this proposal to the authors identified herein.

The document is organized into four sections:

**Rationale for Integrating BRIDG and LS DAM:** This section summarizes the rationale for integrating BRIDG and LS DAM

**LS DAM Background**: This provides an overview of the LS DAM project and current model sub-domains.

**LS DAM’s BRIDG-based Approach**: This explains how the LS DAM was modeled in accordance with the BRIDG project framework and highlights the similarities and differences between the two.

**Approach to Integration**: This section contains the actual proposal and highlights the steps needed to accomplish the integration.

While developing thisdocument, the authors decided that it was important to provide historical background explaining what the LS DAM is and how it was developed as a peer to the BRIDG model from its inception. Although the document is longer than we originally planned, we believe that the additional information provides the context readers need to better understand the integration effort.

This proposal assumes that the current publicly available version of LS DAM (LS DAM 2.2.3) is to be merged with the BRIDG 3.2 model which is scheduled to be published in August 2012. The actual versions of the models to be integrated will depend on when the BRIDG BoD approves the effort.

When integration has been completed, the LS DAM will cease to exist as a stand-alone entity. The semantics currently represented in the LS DAM will be represented within the BRIDG model and will be managed by the BRIDG Semantic Coordination Committee (SCC).

# Rationale for Integrating BRIDG and LS DAM

There are both scientific and technical goals for integrating the BRDG and LS DAM models. From a scientific perspective, unifying the models will more readily support use cases spanning the clinical and life sciences domains and will eliminate the need for users determine which model to go to for needed semantics. 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. Indeed, the goals of precision medicine require linking clinical and molecular data, as described in a recent article from The New England Journal of Medicine:

*“…the fundamental idea behind personalized medicine: coupling established clinical–pathological indexes with state-of-the-art molecular profiling to create diagnostic, prognostic, and therapeutic strategies precisely tailored to each patient's requirements — hence the term “precision medicine.” Recent biotechnological advances have led to an explosion of disease-relevant molecular information, with the potential for greatly advancing patient care. However, progress brings new challenges, and the success of precision medicine will depend on establishing frameworks for regulating, compiling, and interpreting the influx of information that can keep pace with rapid scientific developments.” (*[*http://www.nejm.org/doi/full/10.1056/NEJMp1114866*](http://www.nejm.org/doi/full/10.1056/NEJMp1114866)*)*

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/080221p1.pdf> ). Unifying BRIDG and LS DAM into a single model with shared classes is another step toward the important goal of bridging this chasm.

From a technical perspective, unification of the models will significantly facilitate model management.  Approximately one third of the LS DAM is comprised of BRIDG-derived classes. Maintaining this level of harmonization as separate models requires frequent model import and export, which is labor intensive and error prone. Combining the models into a single Enterprise Architect file will eliminate the need to maintain shared classes across two separate models. Additionally, an integrated model will ensure more consistency in the way semantics are represented, patterns are applied, and processes are followed. The “BRIDG-LS DAM Integration Proposal” provides details for how this unification will be achieved from a technical perspective.

# LS DAM Background

In February 2009, the NCI Center for Biomedical Informatics and Information Technology (CBIIT) initiated the Life Sciences Domain Analysis Model (LS DAM) project to support interoperability across several NCI cancer Biomedical Informatics Grid® (caBIG®) Life Sciences applications. The successful implementation of the BRIDG model as a basis for achieving interoperability across the caBIG Clinical Trials Suite applications was a driving factor in determining the approach to be used in developing the LS DAM.

## LS DAM Project Goals

The primary goal of the LS DAM effort was to produce a shared view of the “static,” or data, semantics of the life sciences domain in order to facilitate and support semantic interoperability in that domain. A secondary, but still important, goal was to build representations of the Life Sciences semantics in an approach similar to that of the BRIDG model’s representations of protocol-driven clinical research while re-using the existing BRIDG model semantics where applicable. LS DAM has always been envisioned as a “sibling” to the BRIDG model, and therefore the requirement to share and re-use common concepts between the two models has always been an NCI objective. These shared concepts are areas where the semantics of the two domains overlap. The “sibling” approach has resulted in a set of models that can be used together to describe the semantics across the translational research continuum, which includes clinical research, the life sciences, and electronic health care records (EHRs), as well as other sub-domains.

## LS DAM Project Scope Statement

*Life Sciences research includes (i) in vivo experiments (ii) ex vivo, in vitro, or in situ experiments, and (iii) in-silico experiments modeling and analyzing processes or other biological phenomena.*

*The Life Sciences Domain Analysis Model (LS DAM) focuses on concepts important for conducting hypothesis-driven and discovery science at the organismal, cellular and molecular level. The LS DAM includes and defines relationships between concepts that are central to specimen collection, processing and banking (human, model organism, cell lines, etc.), in vitro imaging, and molecular biology.*

This scope statement was formally documented in order to provide a guideline for assessing whether concepts that came to harmonization are appropriate for inclusion in the model. It was developed early in the life of the model when the LS DAM Project Team was focused on providing support for basic science use cases. As the model has gained visibility in the broader life science community, the LS DAM Project Team has realized that the model more globally represents complex, molecularly based assays that are just as valid in a clinical research setting as they are in a basic research setting. The Project Team has discussed revising the scope statement, but has not yet formally made any changes.

## LS DAM Project Stakeholders

As the LS DAM has evolved and gained some visibility in the broader life sciences community, the project stakeholders have expanded beyond the original NCI caBIG Life Sciences community.

### NCI caBIG Life Sciences Community

The caBIG life sciences community is the original stakeholder involved in the LS DAM effort.  It is represented by the Information Representation Working Group (IRWG), which is part of the caBIG Integrative Cancer Research (ICR) Workspace.   The IRWG has functioned as the LS DAM Project Team since the development of LS DAM 2.0 was initiated, and the group has been a primary source of feedback on earlier iterations of the DAM.  It comprises subject-matter experts and bioinformaticists associated with various cancer centers and laboratories that participate in NCI’s caBIG Integrative Cancer Research (ICR) WS, the Tissue Banking and Pathology Tools (TBPT) WS, and other NCI caBIG communities**.**

### HL7 Clinical Genomics WG

In June 2010, the HL7 Clinical Genomics Work Group (HL7 CG WG) began participating in the LS DAM Project. Since the LS DAM provides a common representation from which the ever-expanding array of omics technologies can be supported, the HL7 CG WG is using the LS DAM as a base model on which the Omics DAM[[1]](#footnote-1) will be developed. This development effort is ongoing and will be posted for Informative Ballot as part of the May 2012 HL7 ballot cycle. The Omics DAM includes portions of the LS DAM Experiment and Molecular Biology Cores (sub-domains) and several concepts from the Common area.

### HL7 Orders and Observations WG

As follow-up to a joint meeting between the HL7 Orders &Observations (O&O) Working Group and the CG WG at the HL7 Working Group Meeting in January 2012, the Specimen Project within the HL7 O&O WG was begun to build an HL7 Specimen DAM. The project team agreed to use the LS DAM as a starting point and is in the process of reviewing the LS DAM and mapping use cases to it. This exercise may identify new requirements for the LS DAM.

## The LS DAM Project Team

The LS DAM Team is composed of a single modeler/analyst and a core set of subject-matter experts (SMEs) who have experience in biological research, are familiar with NCI caBIG projects, and have expertise in industry-standard reference models and industry-accepted data-exchange formats. Some of the SMEs are also bioinformaticians who have the expertise necessary to provide direction on modeling as well. This team is similar in nature to the BRIDG SCC and has the same range of responsibilities. Like the BRIDG SCC, HHthe size of the team has been kept small to ensure effectiveness when making representational choices, but the composition and expertise of LS DAM Project Team members are quite different from those of the BRIDG SCC team. The BRIDG SCC is composed of modelers/analysts who are quite familiar with the domain, but are not SMEs. Subject-matter expertise in BRIDG is brought in through the SMEs of the project teams that come forward to harmonize with BRIDG and the large BRIDG SME volunteer group whose members represent pharmaceutical companies, Cancer Centers, regulators, and clinical researchers, among others.

The composition of the current LS DAM Project Team is more similar to the composition of BRIDG Technical Harmonization Committee (THC; the predecessor to the SCC) in the early days of the BRIDG Project. The BRIDG modeling experience confirmed that in order to lay out the fundamental framework of LS DAM properly, it would be best to have SMEs from across the life sciences domain participating directly in model development. Like the BRIDG SCC, the LS DAM Project Team is responsible for four sets of tasks that must be performed to support the continued evolution, management, and maintenance of the model. These tasks are

1) Supporting the operational management of the LS DAM;

2) Addressing semantic content in the LS DAM;

3) Maintaining the UML model; and

4) Producing a robust set of documentation to support each release and advocating the LS DAM within the life sciences community.

## Project Release History

Below is a table that shows the release history of the LS DAM. The LS DAM release cycle has been based on the needs of the life sciences community.

|  |  |  |
| --- | --- | --- |
| **LS DAM Release** | **Release Date** | **Comments** |
| R 2.2.3 | May 2012 | Addition of semantics to support Specimen Processing Protocol from the tissue banking domain and binding to the HL7 Abstract Data Type Release 2 |
| R 2.2.2 | January 2012 | Re-alignment of the Experiment Core with ISA-TAB based on feedback from community reviewers, restructuring the Enterprise Architect project to leverage the defined core areas, and introduction of the LS DAM User Guide |
| R 2.2.1 | May 2011 | Minor technical (non-domain semantics) updates to resolve some comments/concerns raised by community members |
| R 2.2 | April 2011 | Additions to the Experiment Core and Molecular Biology Core and introduction of the Experiment Implementation Guide |
| R 2.1 | February 2011 | Resolution of concerns within the Molecular Biology Core and the Specimen Core |
| R 2.0 | October 2010 | Introduction of the Experiment Core and alignment with BRIDG 3.02 |
| R 1.2 | February 2010 | Addition of several key semantics from the Nano WG within the ICR WG |
| R 1.1 | October 2009 | Further alignment with BRIDG 3.0.1 |
| R 1.0 | July 2009 | Initial release |

Table 1: Release history of the LS DAM Model

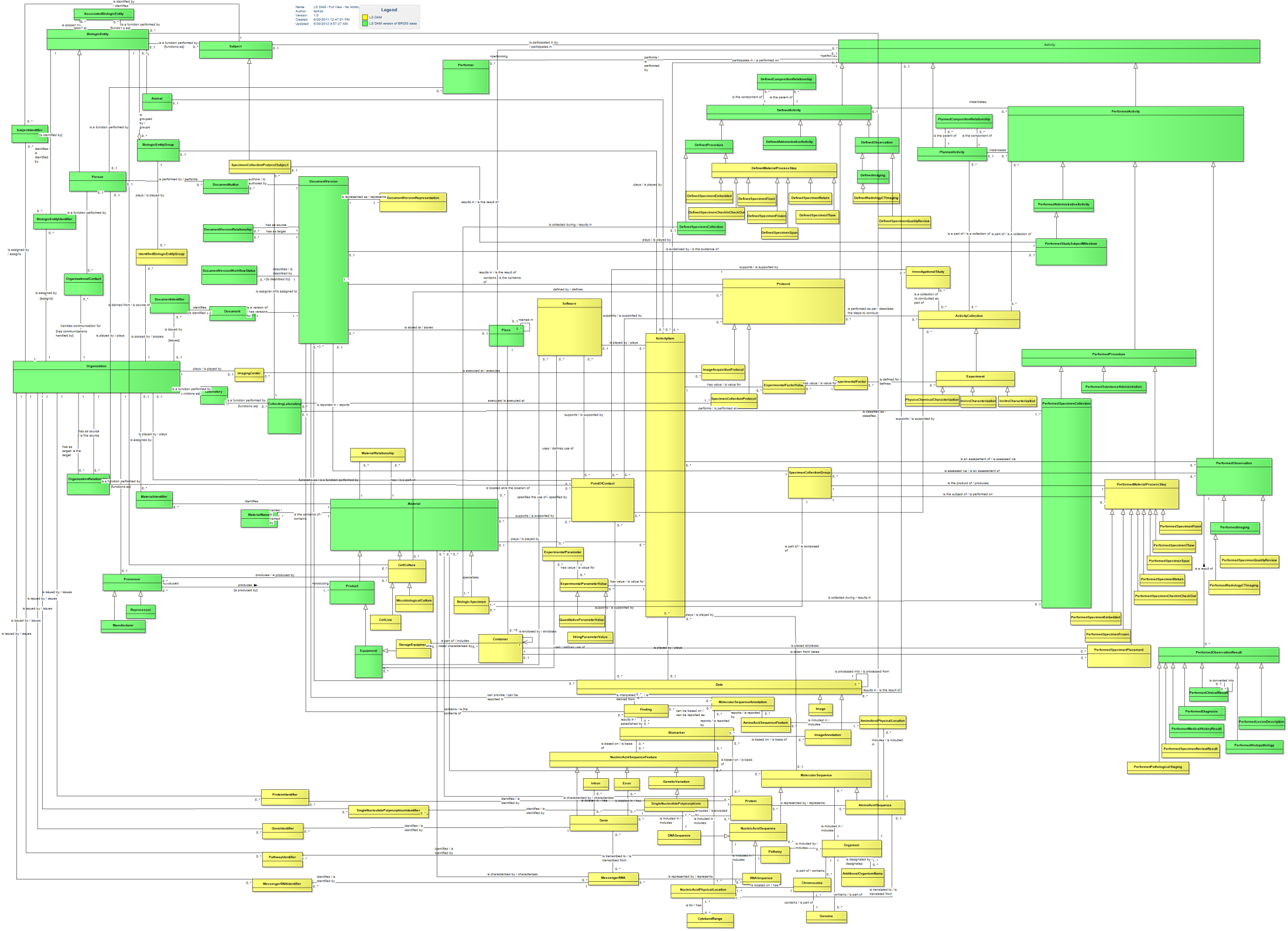
### LS DAM 2.2.3 (Current Release)

The LS DAM 2.2.3 was published at the end of May 2012. The diagram below shows the major subject areas of the LS DAM. The model and all supporting documentation are available on the LS DAM wiki: <https://wiki.nci.nih.gov/x/cxRlAQ>.

As part of this release, the data type binding for the model was changed to the HL7 Abstract Data Type Release 2 (HL7 ADT R2) specification from the ISO 21090 specification. This was done in the spirit of maintaining alignment with BRIDG, which had moved to the HL7 ADT R2 as part of BRIDG 3.1. The BRIDG SCC came to understand that the ISO 21090 data type specification is geared more towards implementation than towards conceptual/logical information models. The HL7 ADT R2 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 therefore the more appropriate choice for the conceptual/logical level for which the BRIDG and the LS DAM are designed.

It is important to note that the "binding" of the BRIDG and LS DAM models to the HL7 Abstract Data Type (R2) specification does NOT represent a "formal binding" of a particular set of data types. Rather, it represents a statement of semantic requirements of the domain(s) explicated by the two models. In particular, users of the BRIDG and/or LS DAM models are not required to develop down-stream Logical and/or Implementable (or sub-set "sibling" Conceptual domain models) using the ADT(R2) data types. Rather, they are expected to support the stated semantics of those data types irrespective of any particular data type specification.

As part of this release, an analysis of the differences in the BRIDG classes included in LS DAM 2.2.3 and BRIDG 3.1 was published. These differences are due to the fact that LS DAM 2.2.3 is aligned to BRIDG 3.0.2. As part of the integration of the LS DAM semantics into BRIDG, these differences will need to be resolved. Because the concepts that span the models tend to be foundational in nature (*e.g*., Person, Organization, Material), it is expected that the differences between any future version of BRIDG and the BRIDG 3.1 representations will be relatively minor and should not have a significant effect on the LS DAM.



**Experiment Core**

**Specimen Core**

**Molecular Biology Core**

**Molecular Databases Core**

**LS DAM 2.2.3**

**Common**

Figure 1: LS DAM 2.2.3 (full view, attributes hidden) – green classes are BRIDG classes brought into LS DAM; yellow classes are LS DAM native

## LS DAM Core Areas

As the LS DAM has evolved over time, the Project Team has found that for the purpose of managing the static semantics described in the model, it is helpful to break the model into a set of four logical groups of related classes, which are referred to as “core areas.” These core areas are supported by a set of commonly used classes. The core areas are similar in purpose to BRIDG sub-domains. Each of the core areas as currently defined is briefly explained below, and a class-level image of the core model has been added to provide an idea of what kind of semantics are present in that particular area. It is suggested that the reader download the model from the NCI wiki (link above), rather than study the diagrams in this document, in order to navigate through the LS DAM and learn the details.

### Experiment Core

The Experiment Core is the conceptual representation of the process of experimentation in hypothesis-driven and discovery research in the life sciences domain. It was developed in collaboration with the HL7 Clinical Genomics WG as an attempt to model a common representation that can provide a framework of support for interoperability in the ever-expanding array of omics technologies. At present, the HL7 CG WG is basing the framework of the Omics DAM on the LS DAM Experiment Core and will bring the gap semantics forward for harmonization into the LS DAM. This area was heavily informed by the ISA-TAB and MAGE models.

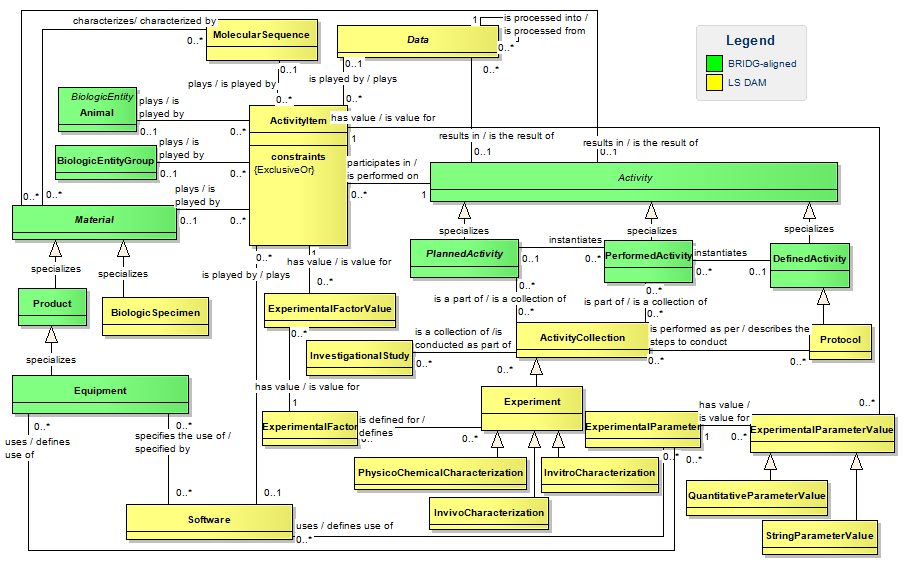


Figure 2: LS DAM 2.2.3 Experiment Core (attributes hidden) – green classes are BRIDG classes brought into LS DAM; yellow classes are LS DAM native

### Molecular Biology Core

The Molecular Biology Core is the representation of components of cells and organisms that will enable linking to genomic and proteomic annotations and experimental findings. It is expected that the HL7 CG WG will use this area of the model in the development of the Omics DAM, and that they will bring gap semantics forward for harmonization into the LS DAM. This core area was heavily informed by ISA-TAB, MAGE-TAB, FuGE, dbSNP, and various public molecular databases.

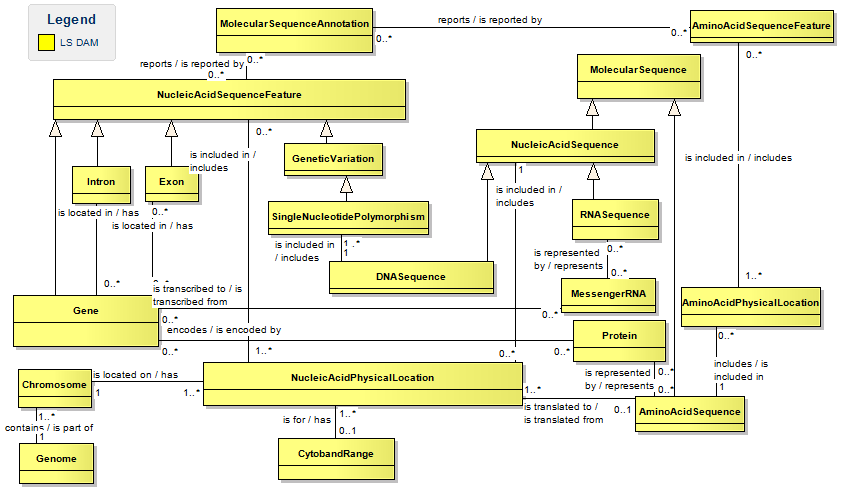


Figure 3: LS DAM 2.2.3 Molecular Biology Core - yellow classes are LS DAM native

### Molecular Databases Core

The Molecular Databases Core supports linking identifying information held in public databases such as Entrez, Ensembl, UniProt, and GenBank to the molecular components defined in the Molecular Biology Core. For example, the classes in this core will provide the ability to support linking the BRCA1 gene to the code it is assigned in Entrez (672), as well as to the code it is assigned in Ensembl (ENSG00000012048).

A separate “Identifier” class is required in the Molecular Databases Core (*i.e*., the identifier is not captured as an attribute in the class that represents the molecular component) because there is the need to support the distinction of “types” of identifiers for each molecular component.  A given gene will have multiple identifiers, each coming from a different public database – one from Ensembl and one from Entrez, for example – and there is a need to be able to distinguish which identifier comes from which data base.  The representational choice to have four distinct identifier classes, one for each molecular component, was a consensus decision driven by the desire to create an easily understandable model for SMEs, who are one of the primary consumers of this domain analysis model.

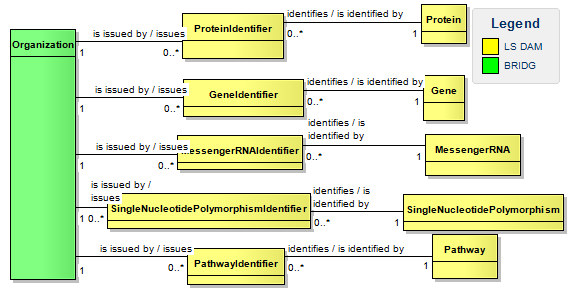


Figure 4: LS DAM 2.2.3 Molecular Databases Core (attributes hidden) – green classes are BRIDG classes brought into LS DAM; yellow classes are LS DAM native

### Specimen Core

The Specimen Core is focused on the representation of the static semantics supporting the collection, processing, and storage of physical substances obtained from biological entities. This core area was heavily informed by the NCI’s caBIG caTissue project model, the caNanoLab project model, and the HL7 Specimen Common Message Element Type (CMET). The HL7 O&O Specimen Project team has elected to use the LS DAM Specimen Core as the basis on which its Specimen DAM will be developed.

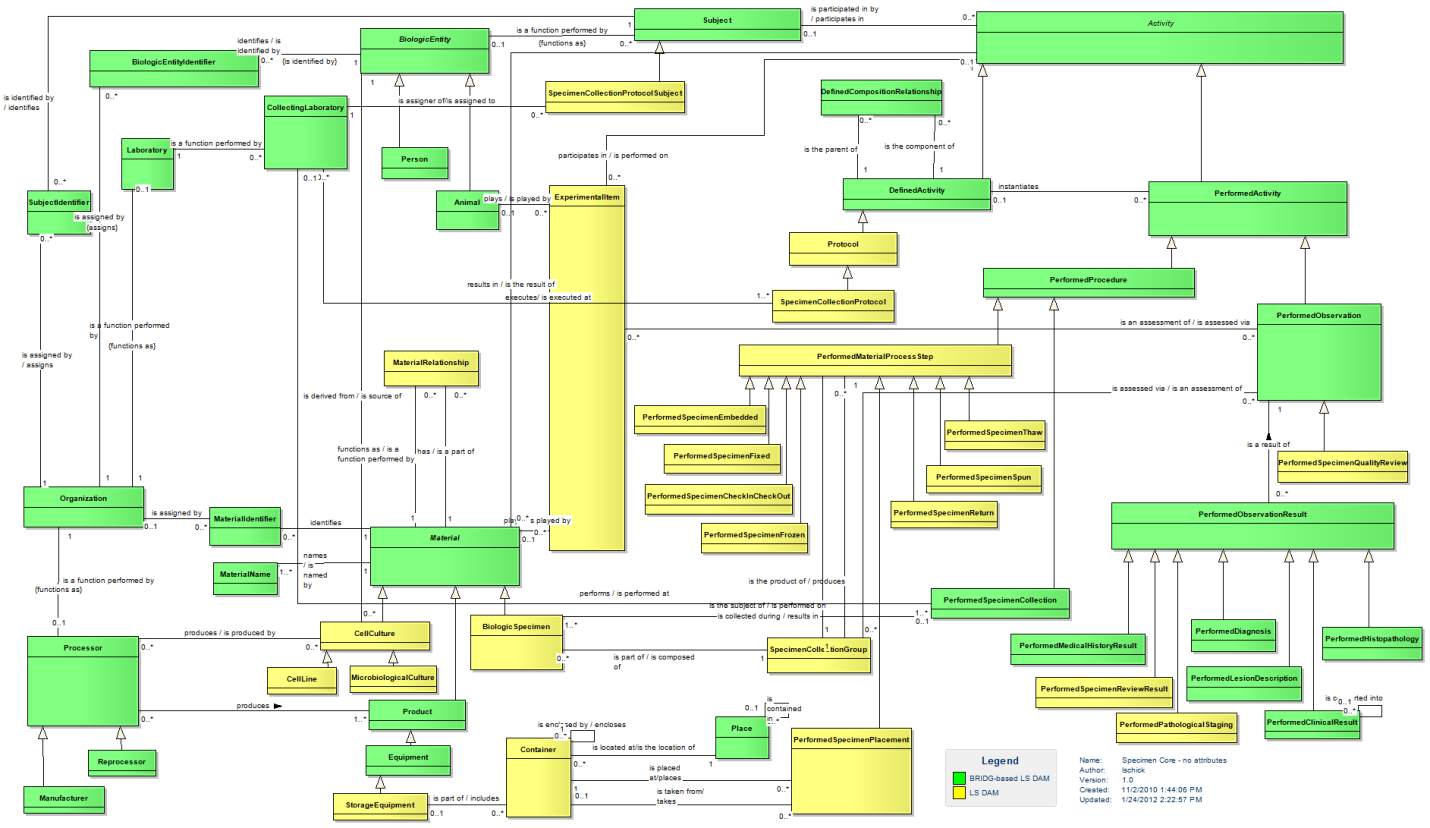


Figure 5: LS DAM 2.2.3 Specimen Core (attributes hidden) – green classes are BRIDG classes brought into LS DAM; yellow classes are LS DAM native

### Common Core

The Common area of the LS DAM represents the modeling of concepts that support several, or in many cases all, of the core components. It is similar in intent to the BRIDG Common sub-domain, and is the area in which the majority of the BRIDG classes/semantics reside within the LS DAM (*e.g*., BiologicEntity, Person, Animal, Organization, appropriate portions of the Activity hierarchy).

List of Classes in LS DAM Common Core

        \* indicates the class comes from BRIDG

|  |  |  |
| --- | --- | --- |
| Activity\* | DocumentVersionRelationship\* | PerformedHistopathology\* |
| AdditionalOrganismName | DocumentVersionRepresentation | PerformedImaging\* |
| Animal\* | DocumentVersionWorkflowStatus\* | PerformedLesionDescription\* |
| AssociatedBiologicEntity | Equipment\* | PerformedMedicalHistoryResult\* |
| BiologicEntity\* | IdentifiedBiologicEntityGroup | PerformedObservation\* |
| BiologicEntityGroup\* | Image | PerformedProcedure\* |
| BiologicEntityIdentifier\* | ImageAcquisitionProtocol | PerformedRadiologyCTImaging |
| CollectingLaboratory\* | ImageAnnotation | PerformedStudySubjectMilestone\* |
| DefinedActivity\* | ImagingCenter | PerformedSubstanceAdministration\* |
| DefinedCompositionRelationship\* | Laboratory\* | Performer\* |
| DefinedImaging\* | Manufacturer\* | Person\* |
| DefinedObservation\* | Organism | PlannedActivity\* |
| DefinedProcedure\* | Organization\* | PlannedCompositionRelationship\* |
| DefinedRadiologyCTImaging | OrganizationRelationship\* | Processor\* |
| DefinedSpecimenCollection\* | OrganizationalContact\* | Product\* |
| Document\* | PerformedActivity\* | Reprocessor\* |
| DocumentAuthor\* | PerformedAdministrativeActivity\* | Subject\* |
| DocumentIdentifier\* | PerformedClinicalResult\* | SubjectIdentifier\* |
| DocumentVersion\* | PerformedDiagnosis\* |  |

# LS DAM’s BRIDG-based Approach

## Modeling Approach

The LS DAM Team continues to use the following approach to build the model:

1. Based on the scope of the effort, the team identified the initial static semantics to be represented in the LS DAM on the foundation provided by the intersection of several well-established, community-vetted caBIG life sciences applications and several use cases or scenarios written by working groups across caBIG that were focused on fostering collaboration and defining interoperability requirements.
2. Once these data semantics were identified, the BRIDG model and several other standard reference models and data-exchange formats (*e.g*., ISA-TAB, FuGE, MAGE, and HL7 Specimen CMET) that are commonly used in the life sciences domain were reviewed and used to inform modeling and representational decisions. This same approach is followed now when new semantics are brought forward for harmonization.
3. To ensure alignment and re-use of BRIDG concepts when adding new semantics to the LS model
   1. The LS DAM Project Team looks to see if it, or something similar, exists in the BRIDG model.
   2. If something similar exists in BRIDG, the LS DAM team brings the relevant BRIDG class(es) and attribute(s) into the LS DAM to review these BRIDG semantics in the context of the life science use case.
      1. A thorough review of the BRIDG concept by the LS SMEs can result in many different conclusions. Here are examples from the recent past of the LS model:
         1. Renaming of the BRIDG concept: Sometimes the LS DAM Project Team has found the required semantic value in BRIDG, but the name is not acceptable to the life sciences SMEs. In this case, the BRIDG concept is brought into the LS DAM (one or more classes are exported from BRIDG and imported into LS DAM) and renamed. This re-naming, in effect, provides the life sciences domain alias for that BRIDG concept/semantic value. For example, the BRIDG.Device class is the same as LS DAM.Equipment class. In this case, the LS DAM Team asked the BRIDG SCC to consider re-naming the class in the BRIDG model itself (call that class Equipment rather than Device in BRIDG). Since the Device concept is fundamental to clinical research and identified by the FDA as one of the kinds of Products in their rulings, the BRIDG SCC decided that they needed to keep the name “Device” and would recognize that LS DAM is using a domain-friendly alias for it.

NOTE: *The BRIDG SCC was very supportive of the LS DAM need and did expand the definition of the BRIDG.Device class and added new examples to better reflect LS DAM requirements.*

* + - 1. Extending the BRIDG concept: The LS DAM Project Team finds that the BRIDG semantics need to be extended in order to support the life sciences domain’s needs. In this scenario, the BRIDG concept is brought into the LS DAM and a specialized class is added to carry the semantic distinction required in the life sciences domain. For example, LSDAM.StorageEquipment is a specialization of the BRIDG.Device (LS DAM.Equipment) class; LS DAM.SpecimenCollectionProtocolSubject is a specialization of BRIDG.Subject.
      2. Re-using the BRIDG concept as is: To date, the LS DAM Project Team has been able to re-use several BRIDG concepts that are foundational to the model, including for example, Person, Organization, Material, and much of the Activity class hierarchy.
  1. If a similar concept/semantic does not exist in the BRIDG model, the LS DAM Project Team looks to other standard reference models and data-exchange formats commonly used in the life sciences domain for insights into how to represent the semantics.
     1. Once the concept is understood well by the Team members, they model this new concept in the LS DAM in a manner consistent with the BRIDG framework. For example, LS DAM.Experiment and LS DAM.ExperimentalStudy were informed by the ISA-TAB data exchange format.

1. The LS DAM analyst continues to work with a BRIDG SCC point of contact in order to gain a better understanding of how a specific semantic value might be represented in BRIDG and how to resolve questions. The LS DAM analyst also works with the BRIDG SCC point of contact to bring proposals forward to the full SCC to modify some of the BRIDG semantics so they will support the LS DAM in addition to BRIDG (*e.g*., BRIDG.Device).

## **Modeling and Naming Conventions**

The LS DAM Project Team has always looked for ways to re-use the work done by the BRIDG effort and not re-create or duplicate previous work. The section above described how the LS DAM is much like an extension of BRIDG. From an operational perspective, this has been relatively easy to do because the LS DAM Project has adopted and re-used BRIDG modeling conventions, as applicable, from the BRIDG Modeling and Naming Conventions document (authored by SCC members). This document is available as part of the BRIDG Harmonization Package from the BRIDG website: <http://www.bridgmodel.org/>.

## Tooling

Both the BRIDG model and the LS DAM are modeled and managed with the Enterprise Architect (EA) tool[[2]](#footnote-2). This enables easy export of information from the BRIDG EA file into the LS DAM EA file. The LS DAM EA project package organization has been structured in a manner similar to the BRIDG model EA project. The LS DAM also leverages the tagging capability of EA in the same manner as the BRIDG model. Tagging is done at class and attribute level to capture the provenance of the semantic element.

## Release Artifacts

The set of supporting artifacts published with each release of the LS DAM is based on the set of artifacts that is published with each release of the BRIDG model. Each LS DAM release includes the publication of several supporting artifacts, including a User Guide, a Change List, a UML model, and an RTF report of the UML model.

Some of the LS DAM artifacts may not be as refined or detailed as the corresponding BRIDG model artifacts (the Change List for example), but the intent has been to move in that direction as the LS DAM model evolves. The process the LS DAM Project Team has defined to harmonize project semantics into the LS DAM is based on the process the BRIDG SCC has defined for harmonization of semantics into BRIDG. The differences between the two processes are due to the differences between the structure and operation of the two teams (BRIDG SCC vs. LS DAM Project Team).

Below are a table and a diagram that illustrate the similarities and differences between various aspects of the BRIDG model and the LS DAM.

|  |  |  |
| --- | --- | --- |
| **Type** | **BRIDG Model** | **Life Sciences (LS) Model** |
| UML Representation | UML Class Diagram | UML Class Diagram |
| RIM Representation | RIM representation | No RIM representation |
| OWL Representation | OWL version published for BRIDG 3.0.1 | A pilot effort explored the OWL representation of a small set of LS classes. Results were documented in a paper – [click here](https://wiki.nci.nih.gov/download/attachments/59212462/LS+OPS+Final+Report+1.0.doc?version=1&modificationDate=1326123198000) |
| Data Types | HL7 ADT R2 (since Feb 2012) | HL7 ADT R2 (since May 2012) |
| Modeling Conventions | BRIDG naming conventions | BRIDG naming conventions |
| Modeling Tool | Sparx Systems - Enterprise Architect | Sparx Systems - Enterprise Architect |
| Release Cycle | Annual/Bi-annual | As needed by the community, generally bi-annually with exceptions driven by contractual obligations. |
| Stakeholder(s) | CDISC, NCI, FDA, HL7 (RCRIM) | NCI and HL7 (CG & O&O WGs) |
| Model Publishing site | [www.bridgmodel.org](http://www.bridgmodel.org) | [NCI wiki](https://wiki.nci.nih.gov/x/cxRlAQ) |
| Infrastructure Needs | Supported by NCI | Supported by NCI |

Table 2: Similarities and differences between the BRIDG model and the LS DAM

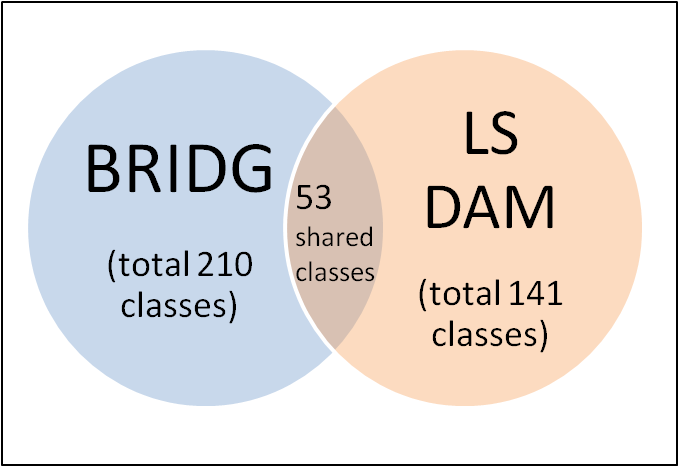


Figure 6: Overlap of UML classes between the two analysis models

As can be seen in Figure 6, there is substantial re-use of classes between the latest releases of BRIDG (3.1) and the LS DAM (2.2.3). Overall, as this section has described, the LS DAM was built with an eye towards creating a peer to the BRIDG model that incorporates the lessons learned, and best practices developed by the BRIDG SCC members. A *de facto* responsibility of the LS DAM analyst is to keep up with BRIDG model development and understand the SCC’s operations and deliverables. Some of the NCI BRIDG SCC members have also stayed informed about and followed the evolution of the LS DAM. The LS DAM analyst/modeler has been asked to join the BRIDG SCC F2F meetings at times to observe the harmonization decision-making process and gain a deeper understanding of the BRIDG model.

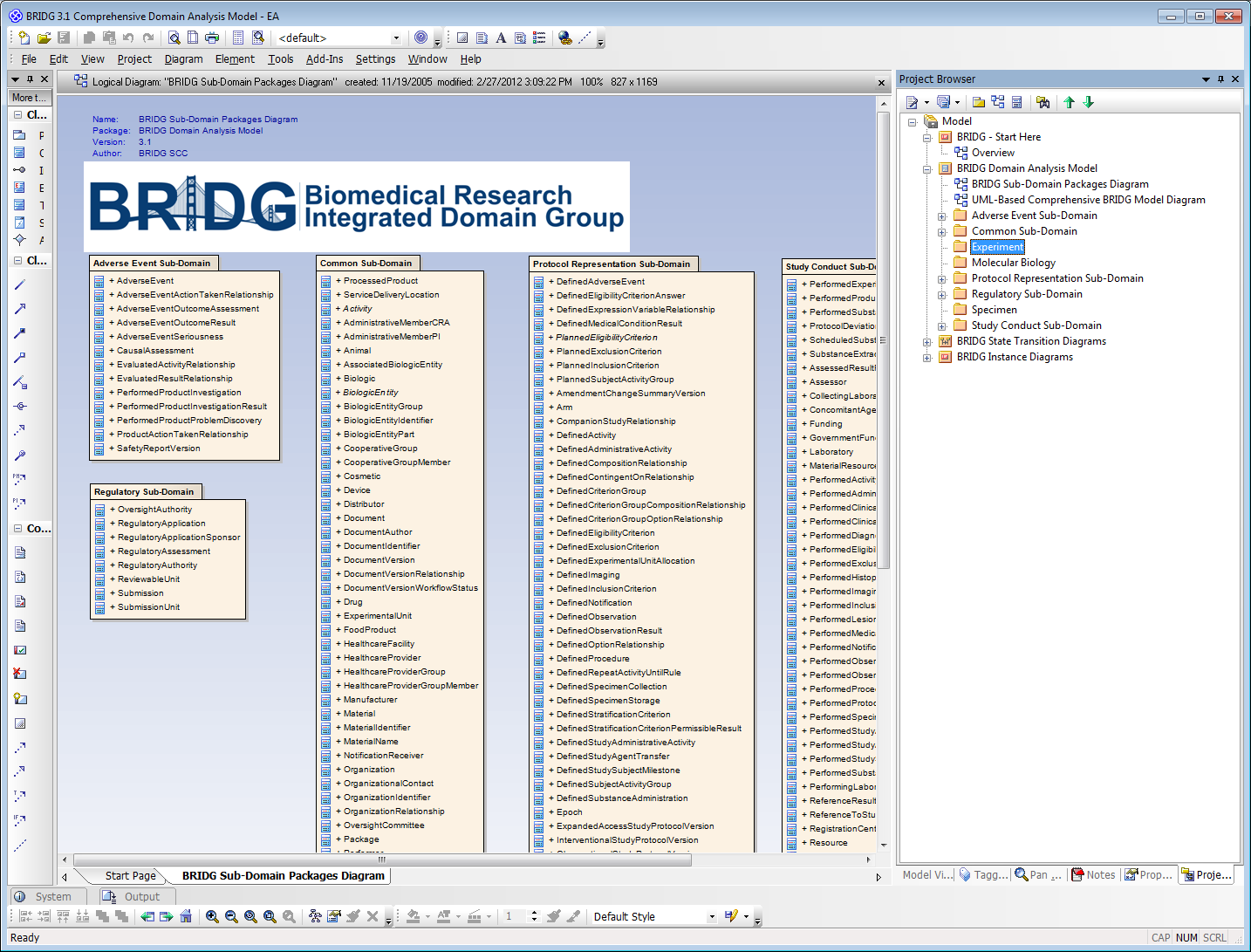
# Approach to Integration

The LS DAM was built as a BRIDG-aware model and can be considered an extension of the BRIDG model that is focused on representing the data semantics of the Life Sciences domain. Therefore, from an operational viewpoint, the LS DAM as a BRIDG extension provides a framework for how to approach this unification in support of the larger scope of translational research. This integration process will need to be a combined effort involving the BRIDG SCC members, the LS DAM Team members, and appropriate subject-matter experts. The integrated model will likely include modifications to the current representations of semantics in both the LS DAM and BRIDG, as is expected with any major harmonization effort. The authors of this document believe that this unification/integration can be separated into three distinct activities described below (note, however, that the order in which these activities are discussed is not meant to dictate the actual sequence of events):

## Mechanics of the Integration

1. One of the reasons this integration can be accomplished is due to the similarities between the models. If BRIDG is intended to support the larger scope of translational research, then we propose that the Core areas of LS DAM be considered as new sub-domains of the BRIDG model. Create a new sub-domain package in the BRIDG model project file for each of the new life-sciences-based sub-domains:
   1. Experiment Sub-Domain (LS DAM Experiment Core),
   2. Molecular Biology Sub-Domain (we propose combining LS DAM Molecular Biology Core and the Molecular Databases Core), and
   3. Specimen Sub-Domain (LS DAM Specimen Core).

NOTE: The BRIDG model already contains a package for the Common Core, and LS DAM has leveraged several of the BRIDG Common sub-domain classes as is.



Life Sciences sub-domains shown

here as new sub-domains of BRIDG

after the harmonization.

Figure 7: Sample of BRIDG Enterprise Architect file project structure with new LS sub-domains

1. Use the export/import capability of EA to bring the LS DAM native classes from each core area into the newly created BRIDG model sub-domain package, one core area at a time.
   1. Assign a new fill color for each new sub-domain and change the fill color for each class brought into the BRIDG model.
   2. Create associations from the newly added classes (that were previously native to the LS DAM) to the appropriate BRIDG classes as originally specified in the LS DAM.
   3. Tag the newly added classes as appropriate according to BRIDG SCC guidelines, while maintaining the information that was previously captured in the LS DAM.
2. Document ALL changes in BRIDG (new classes, attributes, associations, etc.) for the BRIDG Change List for release of integrated model.
3. Resolve semantic differences tracking where the LS DAM 2.2.3 semantics land in the integrated BRIDG model
4. Make updates to the packages and diagrams in the BRIDG file to support the new extensions, which include, for example, the diagrams in the new LS sub-domain packages, the Overview, the BRIDG Sub-Domain Packages Diagram, and the UML-Based Comprehensive BRIDG Model Diagram (as appropriate).
5. Re-create the LS DAM Instance Diagrams in the BRIDG EA File.

## Resolving Semantic Differences

As the LS DAM was being developed, the LS DAM team and the BRIDG SCC identified and acknowledged a few differences in semantics between the two domains of clinical research and life sciences. Some of these have been discussed within the LS DAM Team. Detailed harmonization discussions involving SMEs from both domains will be needed to align the models and develop consensus around these current semantic differences. Resolving these semantic differences will likely result in changes in current representations in both models. This harmonization will not be as simple or straightforward as bringing the LS DAM representation of its semantics into the BRIDG model.

### Resolve outstanding potential points of alignment

#### Protocol

As per the NCI Thesaurus (NCIt), “Protocol” is defined as “A rule which guides how an activity should be performed.” There are many types of “protocols” that fall within the biomedical umbrella: study protocols, specimen-processing protocols, hybridization protocols, and standard-care protocols, among others. It is important to note that there can be relationships among these types of protocols: for example, a given clinical trial protocol may specify the use of a certain specimen-processing protocol.

As we develop the integrated BRIDG model, we should consider this problem and determine whether, and if so how, we want to document relationships among the more abstract “Protocol” concept and all the types we currently have represented in the two models.

#### BiologicSpecimen

BRIDG’s definition of the Specimen concept and its associations does not align with life-sciences SMEs’ understanding from a tissue-banking perspective, where the specimen is considered a physical thing (that is, an entity and therefore more of a BRIDG Material). To support the tissue-banking perspective, the LS DAM has created a specialization of Material called BiologicSpecimen.

#### LS DAM.Finding and BRIDG.PerformedObservationResult (consider “Data” concept also)

The LS DAM defines two separate concepts “Data” and “Finding” that seem to be somewhat aligned with the PerformedObservationResult concept defined within BRIDG. Aligning these concepts has been discussed within the LS DAM Project Team several times since the initial release of the LS DAM.

#### Harmonization of “ActivityCollection” and “Experiment”

LS DAM defines “Experiment” as “A coordinated set of actions and observations, with the ultimate goal of discovery or hypothesis testing,” which is based on the definition in NCIt. It is one specialization of the more generic “ActivityCollection” which is defined as “A coordinated set of actions and observations.” Theoretically, the execution of a clinical trial is an “experiment,” albeit a highly regulated one. As part of this effort, the alignment of the LS DAM Experiment (or ActivityCollection) concept with the BRIDG StudyExecution concept should be explored. BRIDG defines “StudyExecution” as “An ongoing and/or past performance of a formal investigation as specified in a study protocol.”

### Re-assess the Common sub-domain of the integrated model

The existing BRIDG Common sub-domain (or package) contains classes that are integral to the support of two or more of the other defined sub-domains. As we integrate the LS DAM into BRIDG by adding additional sub-domains, we will need to evaluate whether additional classes need to be moved into the Common sub-domain (*i.e.*, a class that previously was not in Common, but is integral to the support of a newly added sub-domain), and move them into the Common package as appropriate.

## Maintenance and Model Management

The assumption is that once the Life-Sciences semantics are included in BRIDG, they will become native to the BRIDG model and the LS DAM will cease to exist as a separate model. Any activities and decisions made in relation to model maintenance and management will be for the entire BRIDG model, including the life-sciences sub-domains.

In addition to a well-defined harmonization processes and standard modeling patterns, the BRIDG SCC has a structured set of manual and automated processes to manage and maintain the model. This includes running utilities against the model files from time to time to check for model inconsistencies, like duplication and integrity violations, among others. The BRIDG SCC has documented a set of modeling and naming conventions which can help in building consistent semantics across the domain regardless of who from the SCC is working on a particular subset of the model.

# Effects on BRIDG and Open Questions

|  |  |
| --- | --- |
| **Open Questions** | **Discussion points** |
| Resources: BRIDG SCC already has a full plate; how will the additional task load be handled? | The harmonization effort will require support from SMEs across domains (both clinical and life sciences), noting that a high level of SME effort will be required during initial harmonization. We propose that LS SMEs join the BRIDG SME group. Additionally, we propose that an analyst be added to the BRIDG SCC to support both initial harmonization as well as ongoing tasks related to the LS portions of BRIDG. |
| BRIDG model size: Adding LS DAM semantics to BRIDG will make it substantially larger. As the BRIDG model grows, will it lose coherence and understandability? | It is not necessary that the full model be leveraged by the user community. Users do and will (likely) continue to use only the subsets of BRIDG semantics relevant to their own projects and it will be important that we prioritize tools and approaches to aid in use of the model. As such, we will create more views and revisit existing BRIDG views. Other mechanisms for aiding use of the model will be explored as well. |
| BRIDG scope statement: Does the current BRIDG scope statement need to be revised to accommodate LS DAM semantics? | We recommend that the BRIDG scope statement be revised to reflect the broader scope of biomedical research that can be support through unification with the LS DAM. |
| Life-Sciences expertise on the Board: The BRIDG BoD does not include a representative with expertise in the life-sciences domain. Should a Life-Sciences domain expert be added to the Board? | We recommended that the BRIDG BoD be expanded to include one or more individuals with expertise in the life sciences domain. |
| LS DAM is not mapped to RIM. | We recommend that RIM mapping be approached on a use case basis rather than attempting to comprehensively map the unified model to the RIM. |

# References

LS DAM wiki: <https://wiki.nci.nih.gov/x/cxRlAQ>

LS DAM ontological paper: <https://wiki.nci.nih.gov/download/attachments/59212462/LS+OPS+Final+Report+1.0.doc?version=1&modificationDate=1326123198000>

BRIDG Harmonization Package: <http://www.bridgmodel.org>

# Glossary

|  |  |
| --- | --- |
| ADT | Abstract Data Type |
| BRIDG | Biomedical Research Integrated Domain Group |
| caBIG | Cancer Biomedical Informatics Grid |
| CDISC | Clinical Data Interchange Standards Consortium |
| CG | Clinical Genomics – an HL7 work group |
| CMET | Common Message Element Type is an HL7 work product produced by a particular committee to express a common, useful, and reusable concept. |
| dbSNP | The Single Nucleotide Polymorphism Database is a free public archive of genetic variation within and across different species developed and hosted by the [National Center for Biotechnology Information](http://en.wikipedia.org/wiki/National_Center_for_Biotechnology_Information) (NCBI) in collaboration with the [National Human Genome Research Institute](http://en.wikipedia.org/wiki/National_Human_Genome_Research_Institute) (NHGRI). |
| FDA | US Food and Drug Administration |
| FuGE | The Functional Genomics Experiment model: <http://fuge.sourceforge.net/> |
| ISA-TAB | The ISA-TAB file format (Investigation, Study, Assay) is an extensible hierarchical structure that focuses on the description of experimental metadata (*i.e*., sample characteristics, technology and measurement types, sample-to-data relationships: <http://isatab.sourceforge.net/>). |
| ISO | International Standards Organization |
| LS DAM | Life Sciences Domain Analysis Model |
| MAGE | Microarray and Gene Expression model : <http://www.mged.org/Workgroups/MAGE/mage.html> |
| MAGE-TAB | A data exchange format that can be used for annotating and communicating microarray data in a standard format using a spreadsheet. |
| NCI | US National Cancer Institute |
| NCIt | NCI Thesaurus: <http://ncit.nci.nih.gov/> |
| omics | Informally refers to a field of study in [biology](http://en.wikipedia.org/wiki/Biology) ending in -omics, such as [genomics](http://en.wikipedia.org/wiki/Genomics), [proteomics](http://en.wikipedia.org/wiki/Proteomics), or [metabolomics](http://en.wikipedia.org/wiki/Metabolomics). |

1. The Omics DAM will be developed within the HL7 CG WG as an extension to the LS DAM. The semantics introduced in the Omics DAM will eventually be brought back into the LS DAM. [↑](#footnote-ref-1)
2. Sparx Systems Pty Ltd. [↑](#footnote-ref-2)