

Local Cancer Center's Legacy Adoption Strategy For CaBIG Projects

Fangxia (Scott) Li, M.D.M.S., Jianguo Yang, B.S., Karl J. Smalley, B.B.A.,

Denise Warzel*, Jack W. London, Ph.D.

Kimmel Cancer Center, Thomas Jefferson University (TJU), Philadelphia, PA

*Associate Director, National Cancer Institute, Center for Bioinformatics (NCICB), Rockville, MD

Abstract:

To ensure caBIG silver-level compatibility, an application created at a developer's site must iterate through the semantic integration process for the classes and attributes of all its domain objects. However, it is much more difficult to modify a legacy system to be caBIG compliant when adopting a silver compatible project. EVS level concept mappings for legacy domain objects must be taken into consideration at local cancer centers before, during, and after the caBIG application is released to adopters. Kimmel Cancer Center have learned lessons in the process of adopting several caBIG silver compatible projects.

In this poster, we intend to document the caTISSUE CORE adoption work at Jefferson, to illustrate Jefferson's view and our proposed solutions for integrating local legacy systems. We describe paths for handling UML models using tools such as EA, Middlegen, and SIW, for a legacy breast cancer tissue database. We also describe a local knowledge database, caMetaMix, we created to maintain the concept mappings between the legacy and standard models, and how to review, analyze, and merge the models using caMetaMix. We also indicate the potential for the application of MBean technology, a powerful integration technology in software industry, and the MBean Service Toolkit (MST) being created at Jefferson.

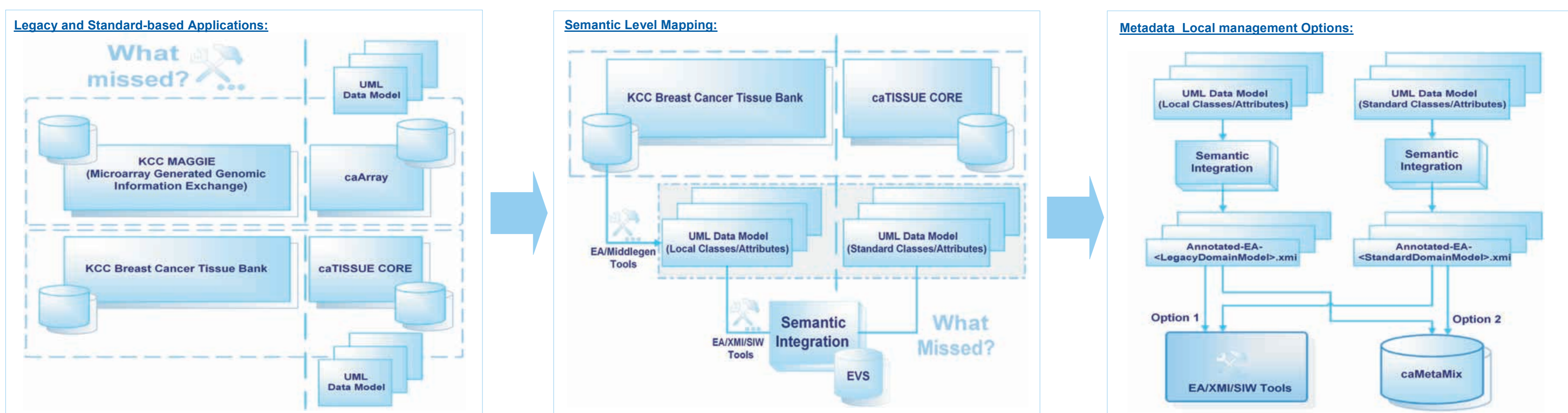
Challenges:

1. Mapping between legacy and standard systems is not an easy task at the semantic level.
2. Lack of local data mapping support.
3. Lack of tools and middleware to integrate legacy and standard-based applications.

Work Involved:

1. EA (Version 6.1), Semantic Integration Workbench, Version 3.0 (SIW)
2. CaMetaMix schema (Visible Analyst, version 7.6), GUI tool for managing mapping.
3. Tools for exporting mapping. MST (MBean Service Toolkit) to conduct integration.

Semantic Data Mapping Path :



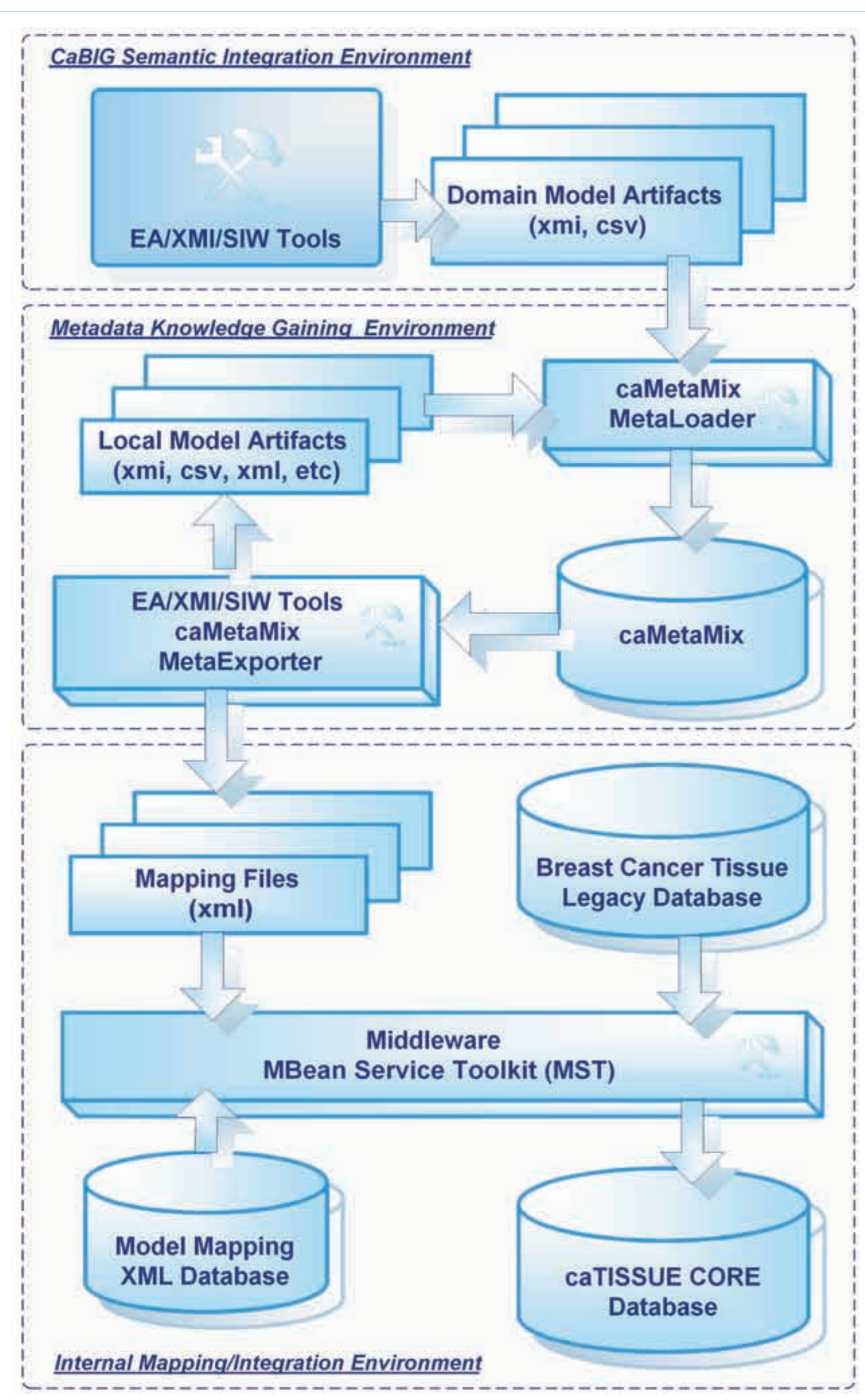
Discussions:

There are several options for managing metadata for legacy models and standard caBIG models. Annotated models must be properly managed because of:

1. Unsynchronized timing to reconcile legacy data model and standard data model. Local cancer centers often have maintained legacy systems for long time.
2. Lack of tools for properly handling mapping concerns, such as reviewing, merging, and analyzing differences between silver and non-silver compatible models.
3. Dynamically changing data standard which may require domain inputs to harmonize the models rooted from many institutions in a collaboration project.

Therefore, local metadata management leveraging the EVS API and/or the functions of Semantic Integration Workbench (SIW) must be provided to support local data porting and integration, preferably in a transparent fashion. Currently there still are missing tools and easy-to-use middleware to fully support local metadata management.

Metadata Knowledge Gaining Process:



For integration of a legacy application in caBIG, cancer centers wish to either port legacy data to a standard repository or maintain the current system with adoption strategies that allow transparent data access. Two key resources that make this possible are:

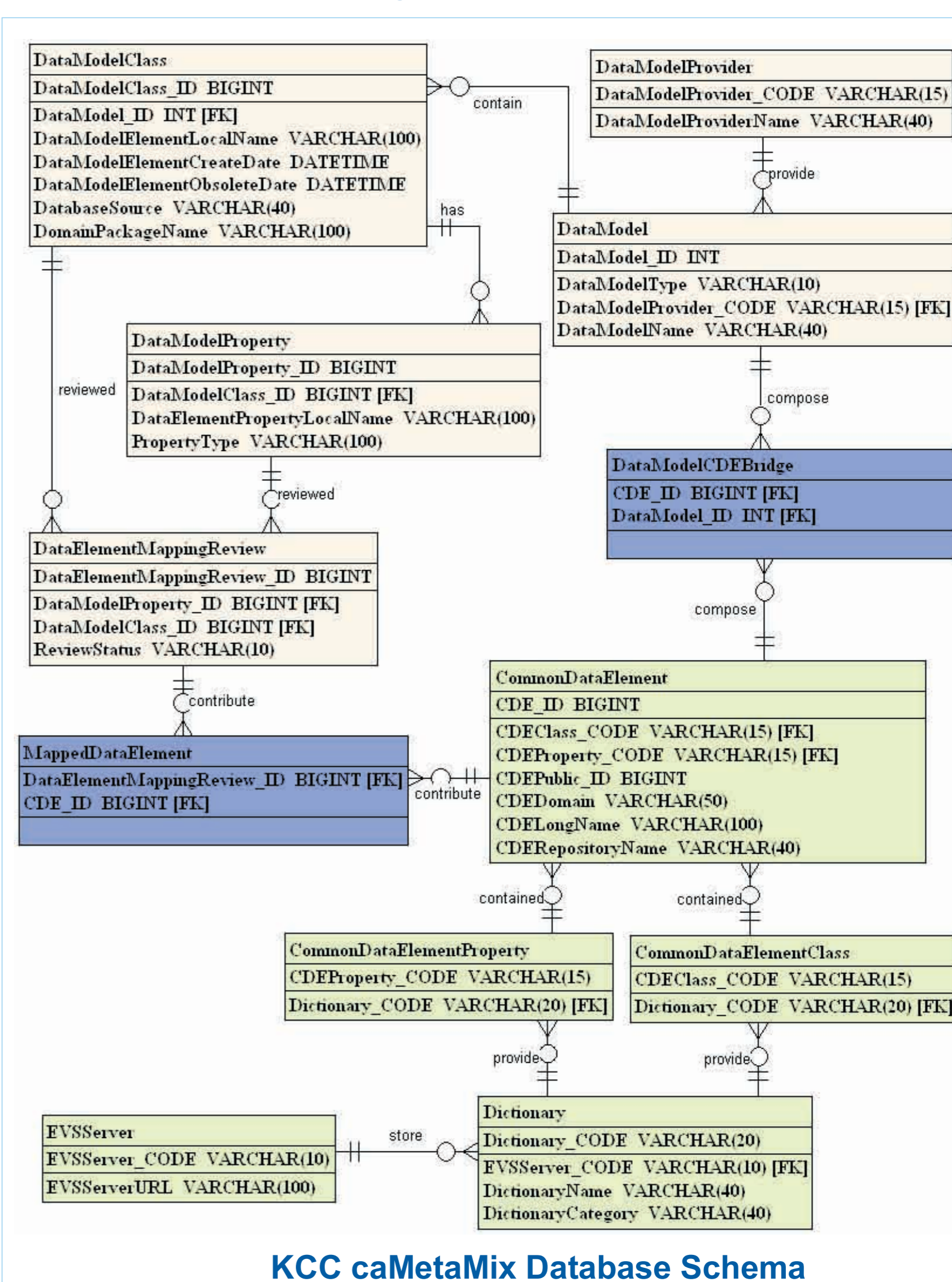
1. Tools that load, export, extract, review, and manage concept level metadata.
2. Middleware that enables transparent data access in the caBIG and caGRID environment.

Tagged values in XMI files before caCORE code generation provides EVS concept codes for all domain objects after running the SIW toolkit. Metadata artifacts generated by developers are then available as a reference set for future mapping of a given legacy system by adopters. The same semantic integration process is necessary for a legacy system to achieve silver compatibility.

However, for adopters, the semantics integration process can be intractable, without clearly addressing the final integration goal. This dilemma is faced by many adopters today. We propose that maintaining local metadata in caMetaMix can facilitate maintaining and supporting further adopter mapping needs.

Use of tools and integration middleware can help review, manage, monitor, extract, and drive data mappings in existing local applications. When mature, a local model can be the basis for porting and integrating data between a legacy system and the adopted caBIG project. This can be achieved with a middleware toolkit, such as the MBean Service Toolkit (MST) created at Jefferson.

Metadata Local Management:



KCC caMetaMix Database Schema