SAND2023-04908D

SnT 2023
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
HOFBURG PALACE - Vienna and Online
19 TO 23 JUNE

Data Model and Data Access Architecture in the Geophysical Monitoring System

Ben Hamlet and Mark Harris
Sandia National Laboratories

Sandia National Laboratories

INTRODUCTION

Data model and data access architecture for GMS, a new waveform processing system for the United States National Data Center (US NDC)

METHODS/DATA

Architecture overview, COI data model and data access concepts, and the legacy data bridge to the US NDC database

START

RESULTS

Architecture context for the GMS data access components and example architecture descriptions for the Event COI class

CONCLUSION

The COI data model and data access architecture supported the system's migration to using a legacy data bridge to the US NDC system's data and file storage. In the future, it may support migration to new GMS data persistence solutions.

Please do not use this space, a QR code will be automatically overlayed

P4.3-375

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Introduction

Sandia National Laboratories is developing the Geophysical Monitoring System (GMS) to modernize the United States National Data Center (USNDC) waveform processing system. The United States is providing the common architecture and processing components of GMS as a contribution-in-kind to accelerate progress on International Data Centre (IDC) Re-engineering. A key aspect of GMS architecture is the Common Object Interface (COI) specification, which includes a Data Model and an Access Application Programming Interface (API). The Data Model is a collection of data structure definitions describing all the data stored by GMS and the Access API specifies the available query and storage operations. The COI specification is independent of any storage solution. This decoupling facilitates changes to the storage solution with minimal impact to the GMS application software; likewise, the Data Model and GMS applications can be updated with minimal changes to the storage solution or schemas. One GMS COI implementation, the data bridge, provides clients access to legacy system data through web services. This poster describes GMS COI and data bridge architecture.

GMS
Geophysical
Monitoring
System

Objectives

Our primary objective is to specify a programming language independent Common Object Interface (COI) for GMS using both a data model and an API describing the available query and storage operations (create, read, update, and delete). The API declares both service interfaces and the repository interfaces directly encapsulating the data persistence solution.

Our secondary objective is to define a Legacy Data Bridge to provide an implementation of the COI repository interfaces using the legacy US NDC system's database and waveform file storage.

Methods

The following slides describe the GMS COI architecture and how the Legacy Data Bridge is an extension of the COI architecture concept.

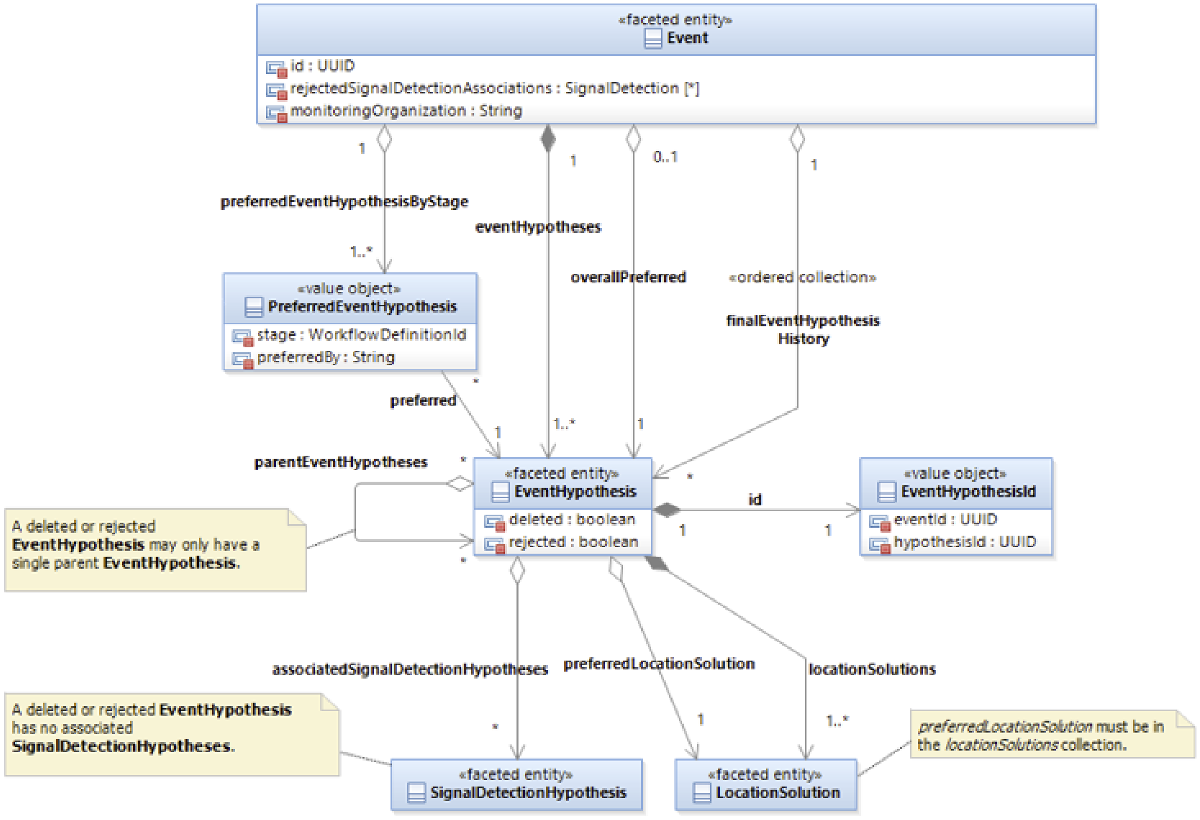
The COI consists of two primary elements:

> **COI Data Model -** a collection of data structure definitions that together describe all the data stored by the system, including acquired data, processing results, provenance, etc.
>
> **COI Access API -** consists of library and network service interfaces providing Create, Read, Update & Delete (CRUD) access to persistent COI Data Model objects.

The Legacy Data Bridge is an implementation of the COI Access API.

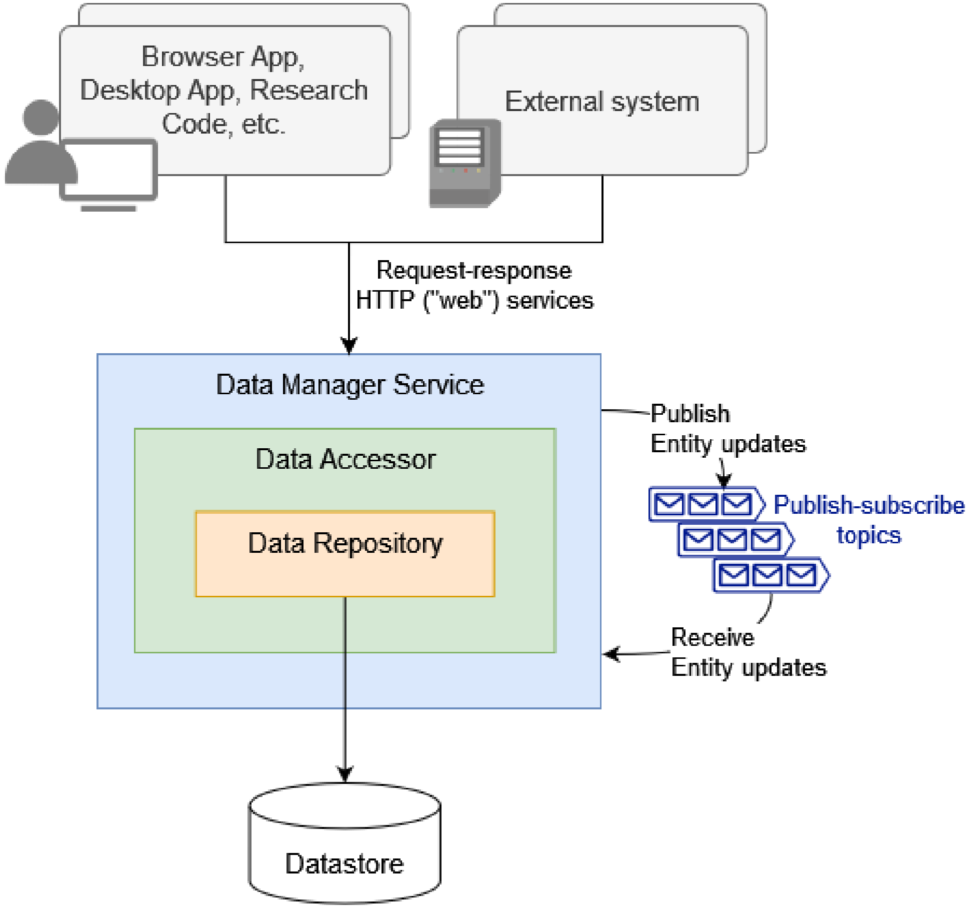
COI Data Model

- Motivation: the legacy US NDC system lacks separation between the application data model and physical database schema, a critical impediment to cost-effectively maintaining and extending the system.
- The GMS architecture avoids this problem using an abstraction layer encapsulating the underlying physical storage solution.
- The COI Data Model includes object oriented data structure definitions describing all the data stored by GMS, including acquired data, processing results, provenance, etc.
- The COI Data Model specification is independent from the physical schema of the underlying storage technology, insulating GMS applications from storage implementation details.



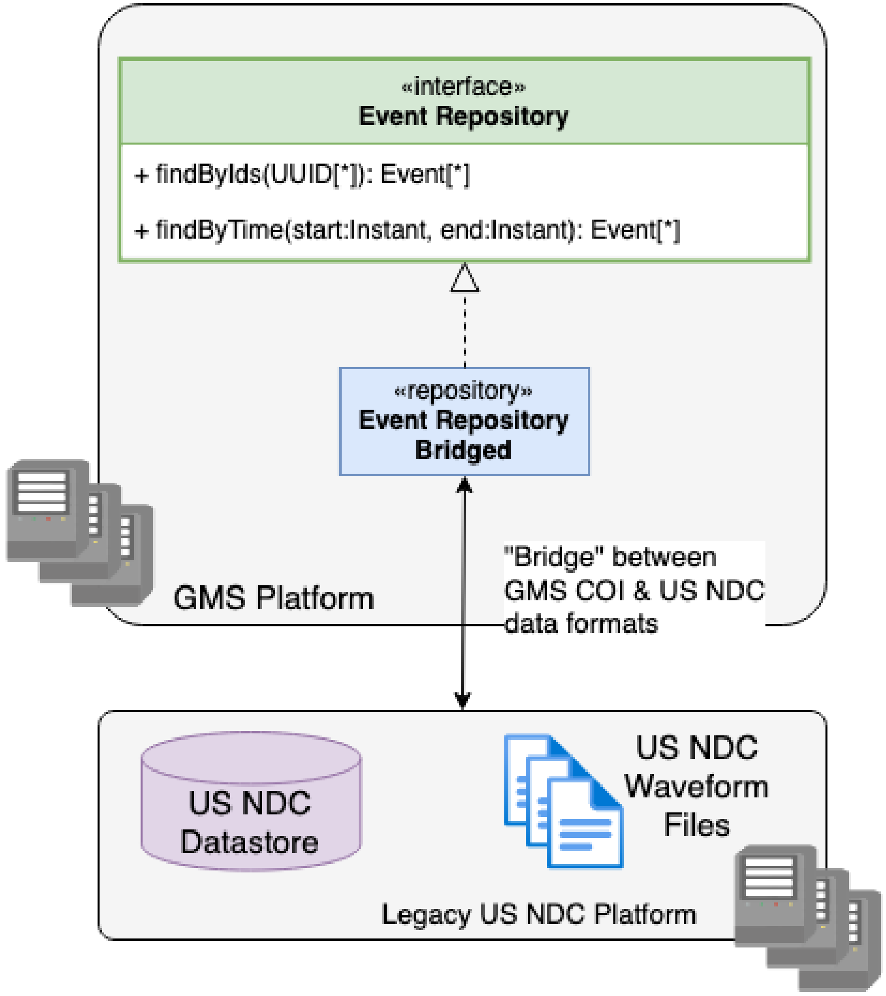
COI Access API Architecture

- Consists of library and network service interfaces providing Create, Read, Update & Delete (CRUD) access to persistent COI Data Model objects.
- The Repository/Accessor/Manager pattern defines a standard structure and allocation of responsibilities for data services that provide access to persistent COI Data Model objects.
- Data Repositories provide interfaces with replaceable implementations for access to entities stored in the underlying persistence solution. They encapsulate the details of accessing underlying storage technologies and converting between the storage schemas and the COI Data Model.
- Data Accessors provide aggregation across Data Repositories (e.g. to provide Signal Detection Hypotheses together with associated measured Waveform Channel Segments) and may provide forward caching of retrieved Entities to reduce access latency.
- Data Managers provide network service packaging for Data Accessors. Service interfaces may include HTTP request-response ("web") services and asynchronous publish-subscribe message queues.



Legacy Data Bridge

- Motivation: the GMS project is currently prioritizing development of the interactive analysis tools used by Analysts. Since the GMS data acquisition and automatic processing components are too immature to support the interactive analysis tools, GMS temporarily includes the Legacy Data Bridge to provide the interactive analysis tools access to the legacy US NDC database and waveform file storage.
- The set of Data Manager Services the interactive analysis UI depends on for data have been modified internally to provide access to data from the existing US NDC data store while preserving the COI interface.
- Specifically, the Repository implementations and related software in the Data Manager Services provide the COI Data Access operations through read/write operations against the legacy data store, converting between the legacy schema and GMS COI data model.
- This concept enables the interactive analysis UI to use the raw data and processing results created by the legacy system's acquisition and automated processing software.



Results

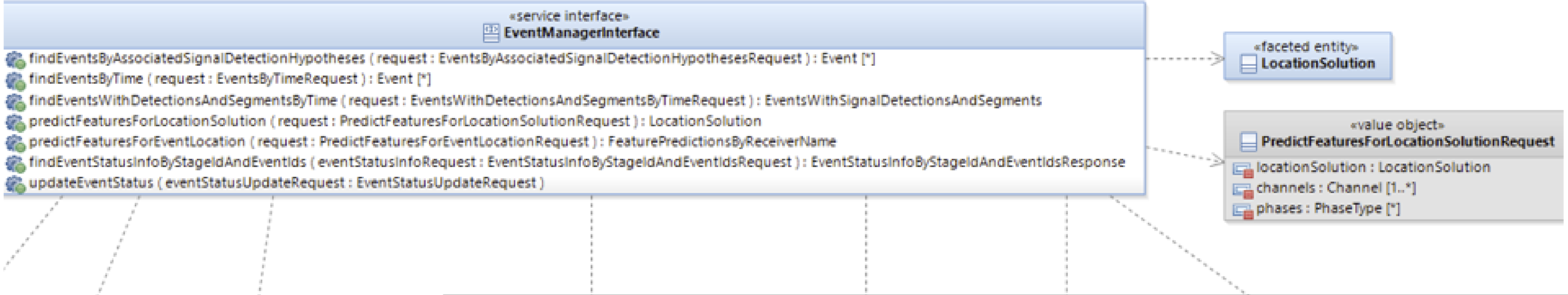
The following slides show example architecture descriptions for a portion of the GMS COI Data Model, COI Access Interface, and Legacy Data Bridge.

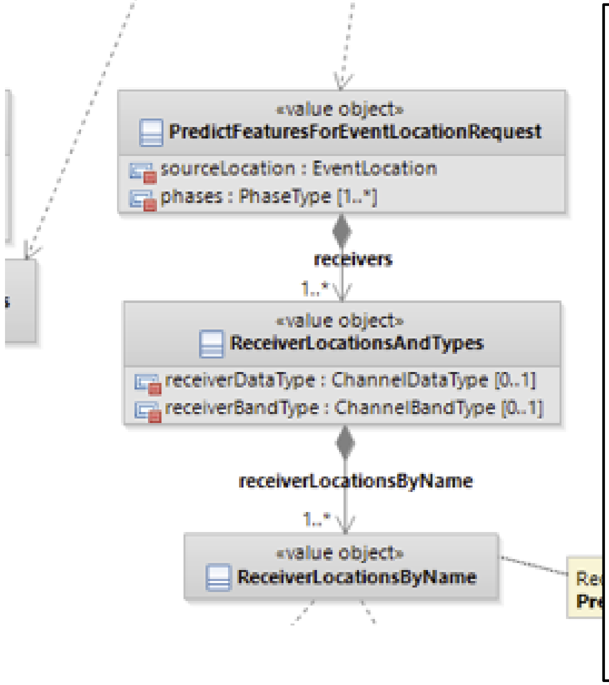
The architecture describes COI Data Model classes using diagrams along with textual descriptions discussing the purpose of each class in GMS and how it relates to a concept from the explosion monitoring domain. Each class attribute is described in detail, including its datatype, units, range, whether it is optionally populated or must have a value in each class instance, and a description of what its values represent.

The architecture describes COI Data Access interfaces using diagrams along with textual descriptions of each operation discussion the operation's behavior, constraints on its input parameters, and guarantees about its output parameters.

The Legacy Data Bridge is an implementation of the COI Data Access interfaces. The architecture descriptions for these components focus on mappings between the legacy and COI data models.

Example COI Access Interface – Event Manager





Example – Event Manager Operation Description

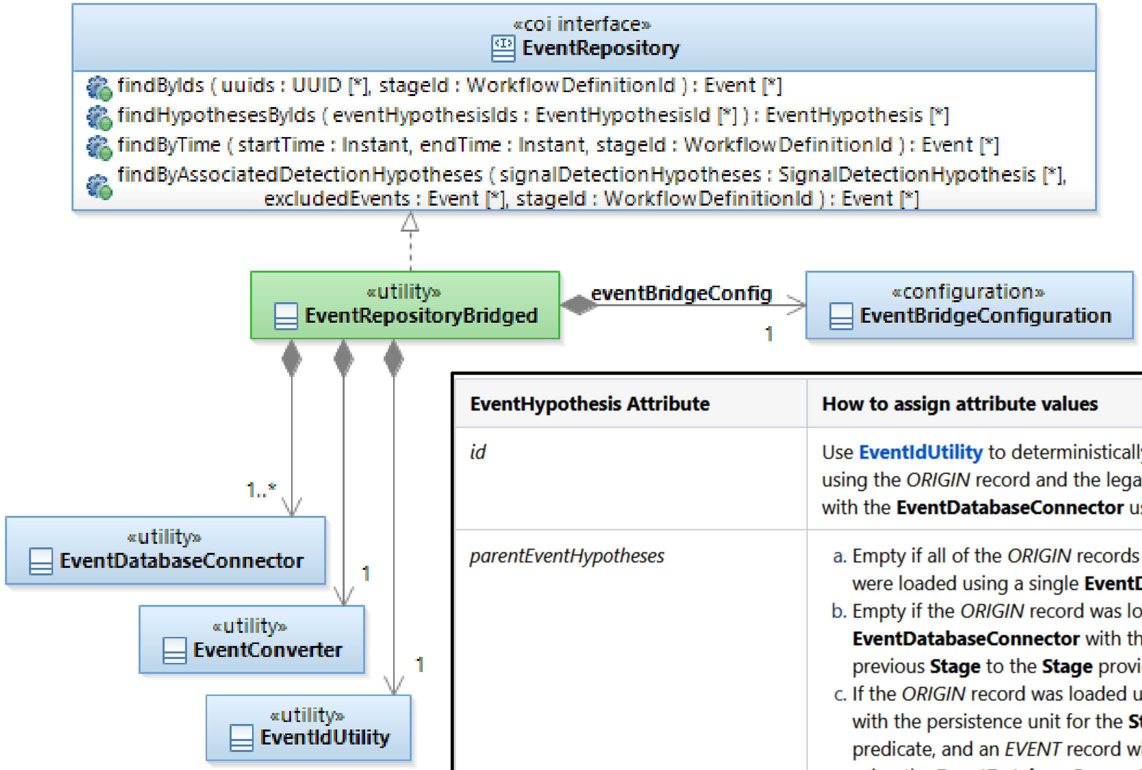
findEventsWithDetectionsAndSegmentsByTime(startTime : Instant, endTime : Instant, stageId : WorkflowDefinitionId) : EventsWithSignalDetectionsAndSegments - this operation finds the **Event** objects matching the time range and **Stage** provided in the query predicate following the same semantics as **EventRepository's** *findByTime(...)* operation. However, this operation populates the **EventHypothesis** objects in each **Event's** *eventHypotheses* collection rather than populating them as version references. Additionally, this operation returns a collection of **SignalDetection** objects (each including populated **SignalDetectionHypothesis** and **Channel** objects), for each **SignalDetection** with a **SignalDetectionHypothesis** associated to any **EventHypothesis** included within the returned **Event** collection. Further, this operation returns a collection of **ChannelSegments** containing the measured **ChannelSegment** objects referenced by the **FeatureMeasurements** of the returned **SignalDetectionHypothesis** objects. Within the returned **ChannelSegment** objects, this operation populates the *maskedBy* **ProcessingMask** objects, leaving their *maskedQcSegmentVersions* populated as identifier-only instances.

Example Legacy Data Bridge Component - Event Bridge

Database Connector components provide read/write access to the USNDC data stores, encapsulating access details, including credential management; SQL queries, inserts and updates; file access to the US NDC waveform files (for waveform-related repositories)

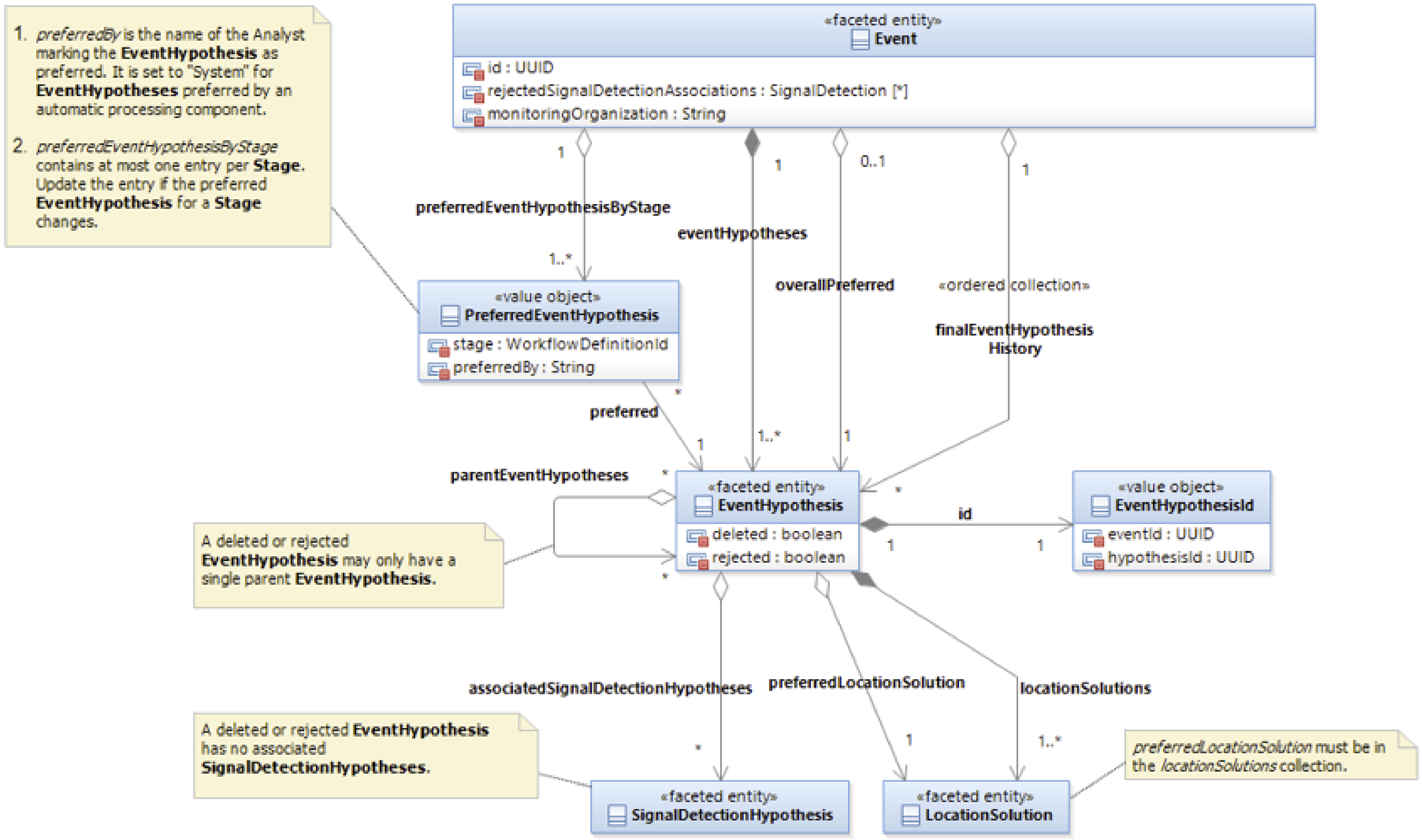
COI Converter components support conversion between the USNDC data formats exposed by the **Database Connector** and the GMS COI data model format.

ID Utility components manage mappings between COI format and legacy format object identifiers.



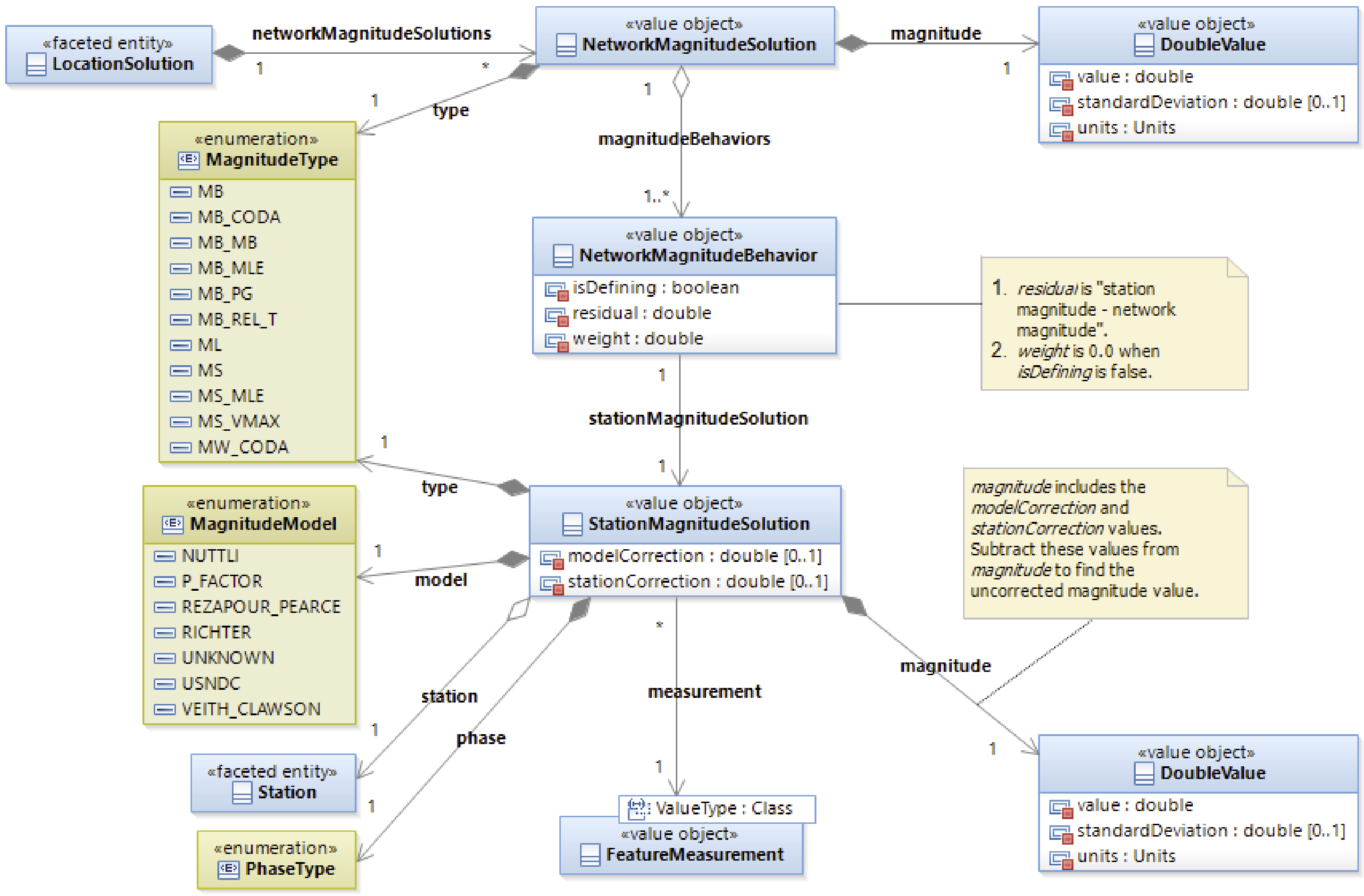
EventHypothesis Attribute	How to assign attribute values
id	Use **EventIdUtility** to deterministically generate an **EventHypothesis** id using the *ORIGIN* record and the legacy database account associated with the **EventDatabaseConnector** used to load the *ORIGIN*.
parentEventHypotheses	a. Empty if all of the *ORIGIN* records associated with a particular *evid* were loaded using a single **EventDatabaseConnector**. b. Empty if the *ORIGIN* record was loaded using the **EventDatabaseConnector** with the persistence unit for the previous **Stage** to the **Stage** provided in a query predicate. c. If the *ORIGIN* record was loaded using the **EventDatabaseConnector** with the persistence unit for the **Stage** provided in the query predicate, and an *EVENT* record with the same *evid* was also loaded using the **EventDatabaseConnector** with the persistence unit for the previous **Stage**, then use **EventIdUtility** to generate an **EventHypothesis** id for that *EVENT* record's *prefor ORIGIN* record (i.e. an *ORIGIN* record from the previous **Stage**) and add that id to the *parentEventHypotheses* collection.
associatedSignalDetectionHypotheses	a. Find the *ASSOC* records associated to the *ORIGIN* record. b. Use **SignalDetectionIdUtility** to generate the **SignalDetectionHypothesis** id corresponding to each *ASSOC*. c. Use the **SignalDetectionHypothesis** ids to populate the *associatedSignalDetectionHypotheses* collection.

Example COI Data Model – Event



Results

Example COI Data Model – Event (Magnitude)



Conclusion

Sandia National Laboratories is developing GMS for the US NDC.

The US is providing the common architecture and processing components of GMS as a contribution-in-kind to accelerate progress on IDC Re-engineering.

The GMS COI Data Model and Data Access Interface describe new data structures and access operations for GMS acquired data, processing results, provenance, etc.

The COI is agnostic to the underlying data persistence solution.

The Legacy Data Bridge is a COI implementation using the legacy US NDC data stores.

The COI concept allowed GMS to pivot from an initial implementation using a new database structure to temporarily focus on the Legacy Data Bridge implementation.

The COI abstractions support other implementations. Possible implementations include replacing the Legacy Data Bridge with new GMS data persistence solutions, or a data bridge to the IDC database and file storage solutions.