SAND2016-0450PE

Exceptional service in the national interest

Sandia National Laboratories

IDC Reengineering Phase 2

Data Model Overview

Chris Young

January 2016

U.S. DEPARTMENT OF ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Why a Data Model?

Sandia National Laboratories

- The System is fundamentally concerned with acquiring and processing waveform data for the purpose of detecting events of monitoring interest.
- To properly design the System it is necessary to model the types of data used by the system and the relationships between those types of data, i.e. to formulate a data model.
- Data modeling can represent a high, medium, or low level of detail. Our work so far focuses on the high to medium level of detail, establishing the most significant classes and their inter-relationships.

Data Model vs. Schema

Sandia National Laboratories

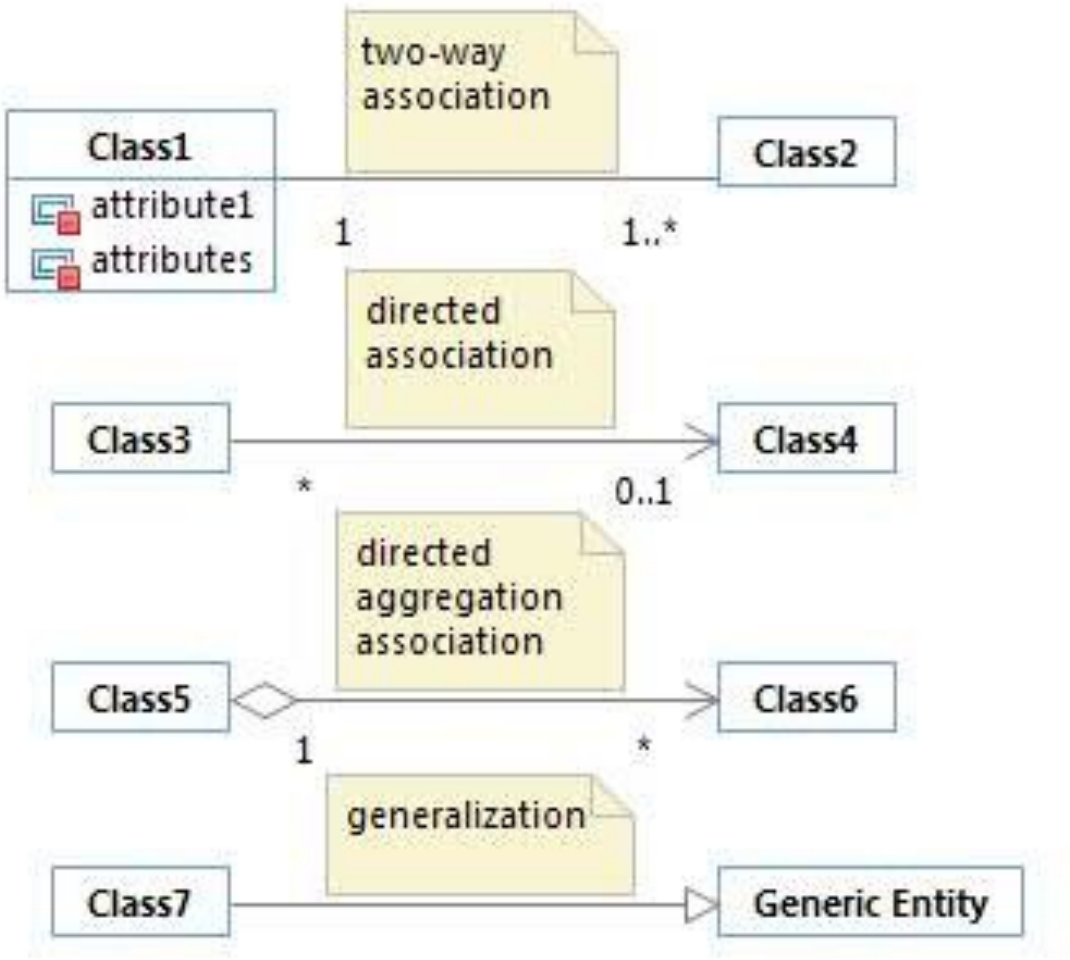
- The data model is how the information is represented within the System, i.e. how system developers will interact with it
- Ultimately this information must be persisted, likely in a relational database schema, but design of that schema will come later, after the data model has been designed
- It is a major goal of the Re-engineered IDC System to hide the storage details from users of the System; they will access data through a Common Object Interface (COI)
- Regardless of the way the information is represented within the System, there will be capability to export information in a variety of standard formats such as CSS3.0 or QuakeML. Thus customers tied to legacy formats will be supported.

Phase I Data Model Goal: Model Information in CSS3.0 (Core Tables)

Sandia National Laboratories

- The original CSS3.0 schema modeled the basic information needed for monitoring: stations, waveforms, detections, events
- This provided a good first phase for our data modeling work
- We have attempted to address problems/shortcomings of CSS3.0, both major and minor
- All useful information in CSS3.0 should be in the data model, but the converse is definitely NOT true: exporting from the data model to CSS3.0 WILL result in loss of important information

Class Diagram Conventions

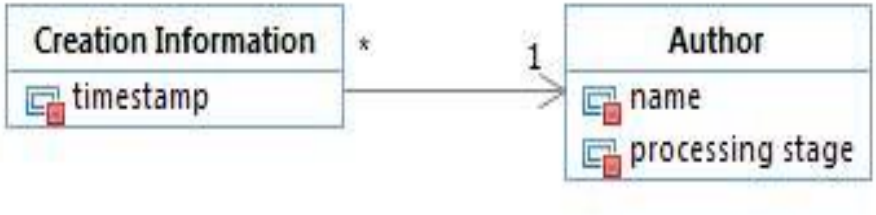


WARNING: Not All Class Attributes Included!

Sandia National Laboratories

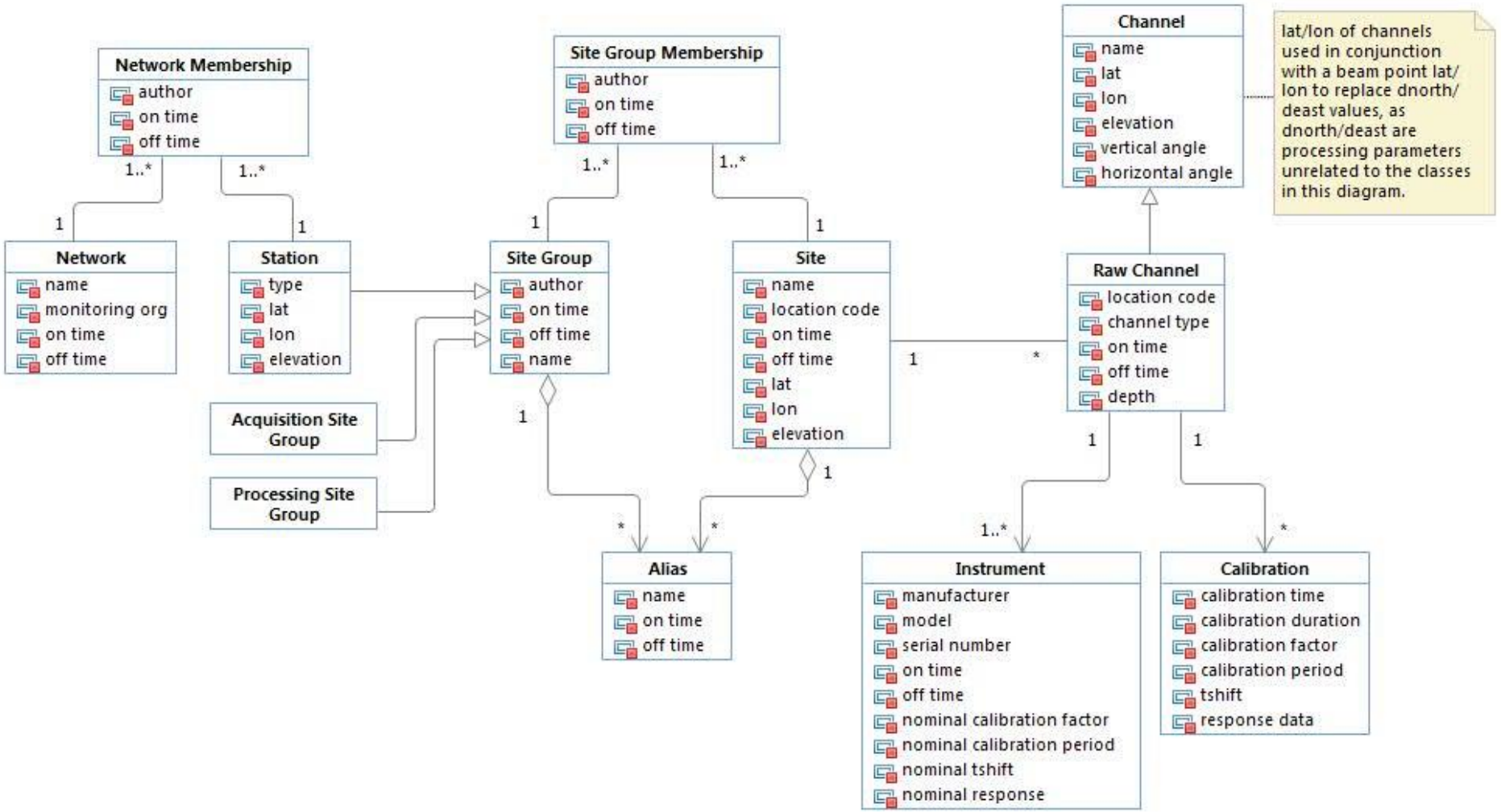
- Our work so far has been at a fairly high level, focusing on trying to identify all the necessary classes to capture the same types of information that are included in CSS3.0
- We included attributes for each class as needed to clarify what each class represents and how the classes relate to each other
- We have NOT checked carefully to make sure that each class has all of the necessary attributes, but this will eventually have to be done.
- The full set of class attributes will be included in the final, complete data model.

Creation Information



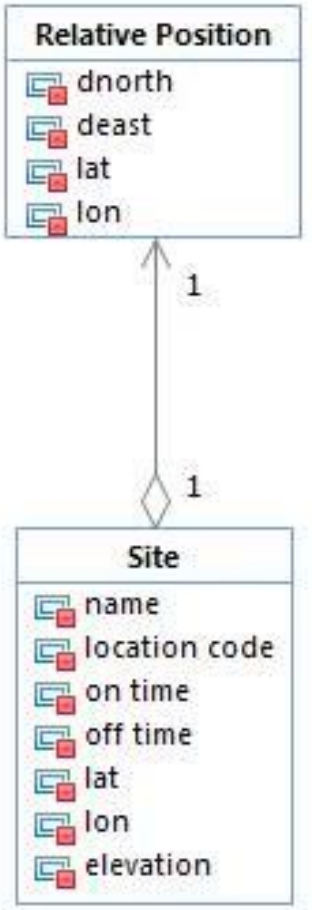
- When objects are created and changed within the System, the history of those objects are just as important to capture as the values they currently contain.
- The Creation Information class is one step towards this goal of object provenance. It contains:
  - a creation **timestamp**, representing the date and time that this object was created.
  - a reference to the **Author** that created it.
  - the **processing stage** the object was created in.

Station Information

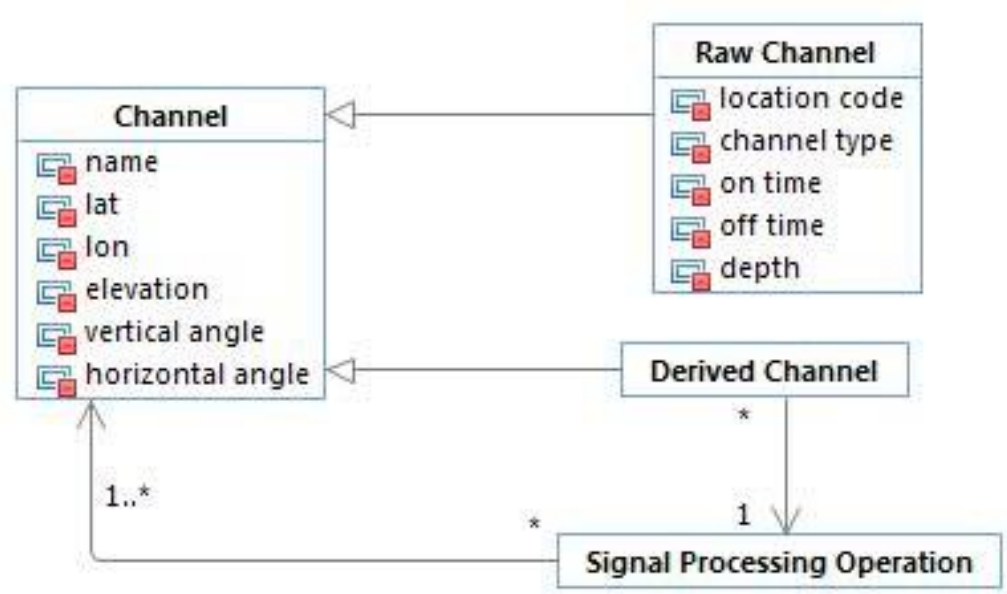


Relative Site Information (Array Processing)

- Introduced in CSS3.0 because of issues with capturing sufficient precision in site positions within small arrays
- Used ONLY to capture information for legacy stations
- Precision no longer a problem, so grouping of sites for array processing will be handled differently within the data model
- Array processing was out of scope for this initial effort so that information is not shown in this version of the data model.

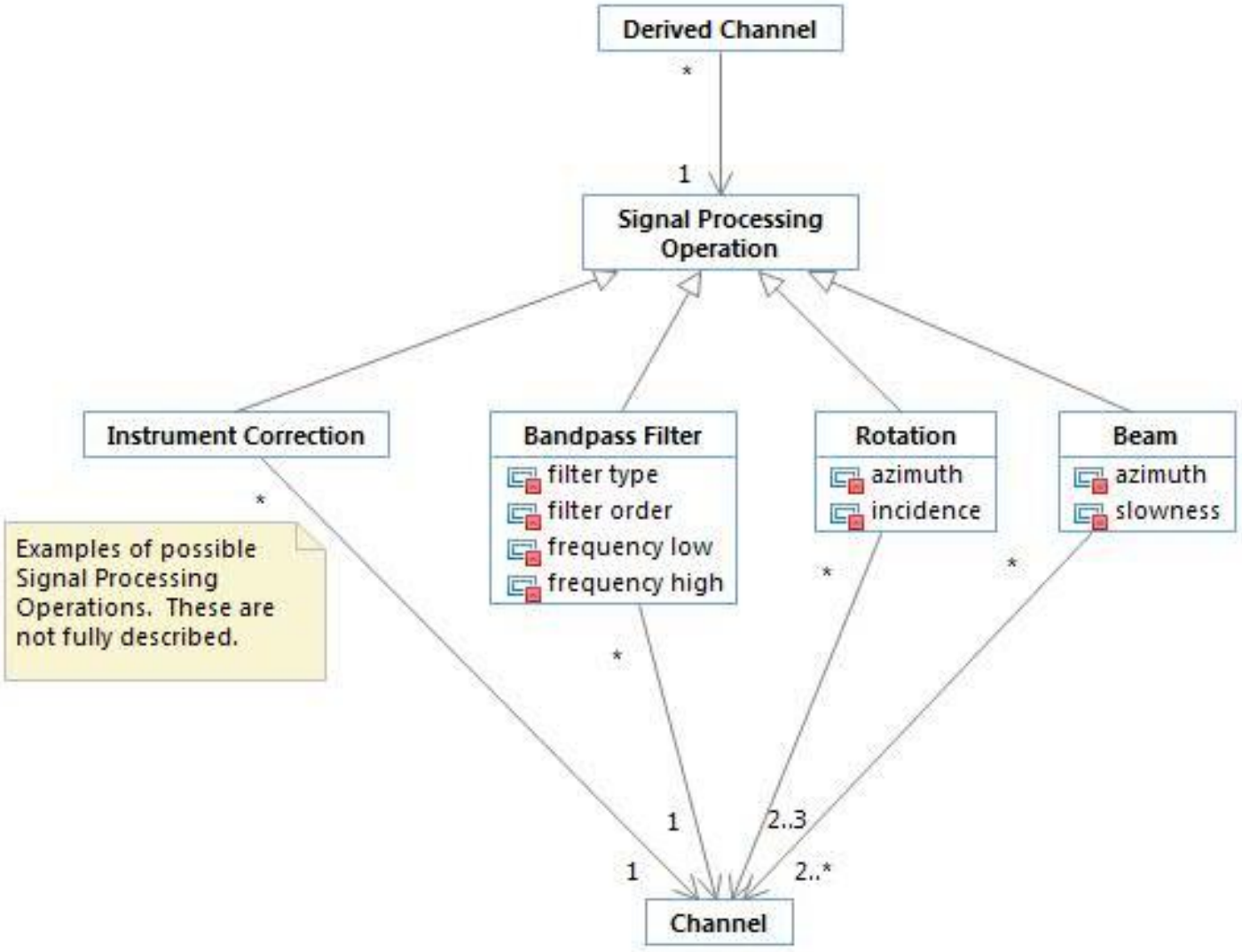


Channel Information

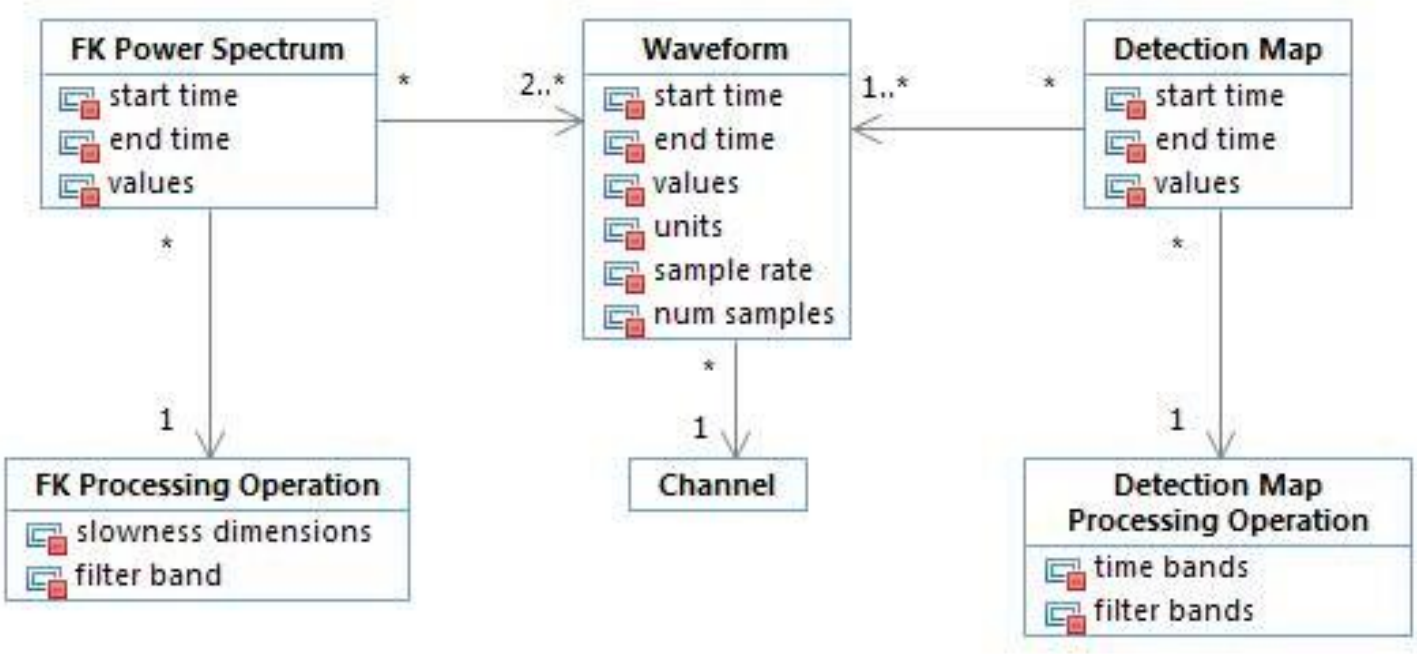


- Raw Channel and Derived Channel (e.g. filtered, beamed, etc.) are generalization of the Channel base calss.
- A Derived Channel is a nested collection of Channels
- Signal Processing Operations describing the processing history of that Channel's data, all the way back to the Raw Channel(s).

Signal Processing Information

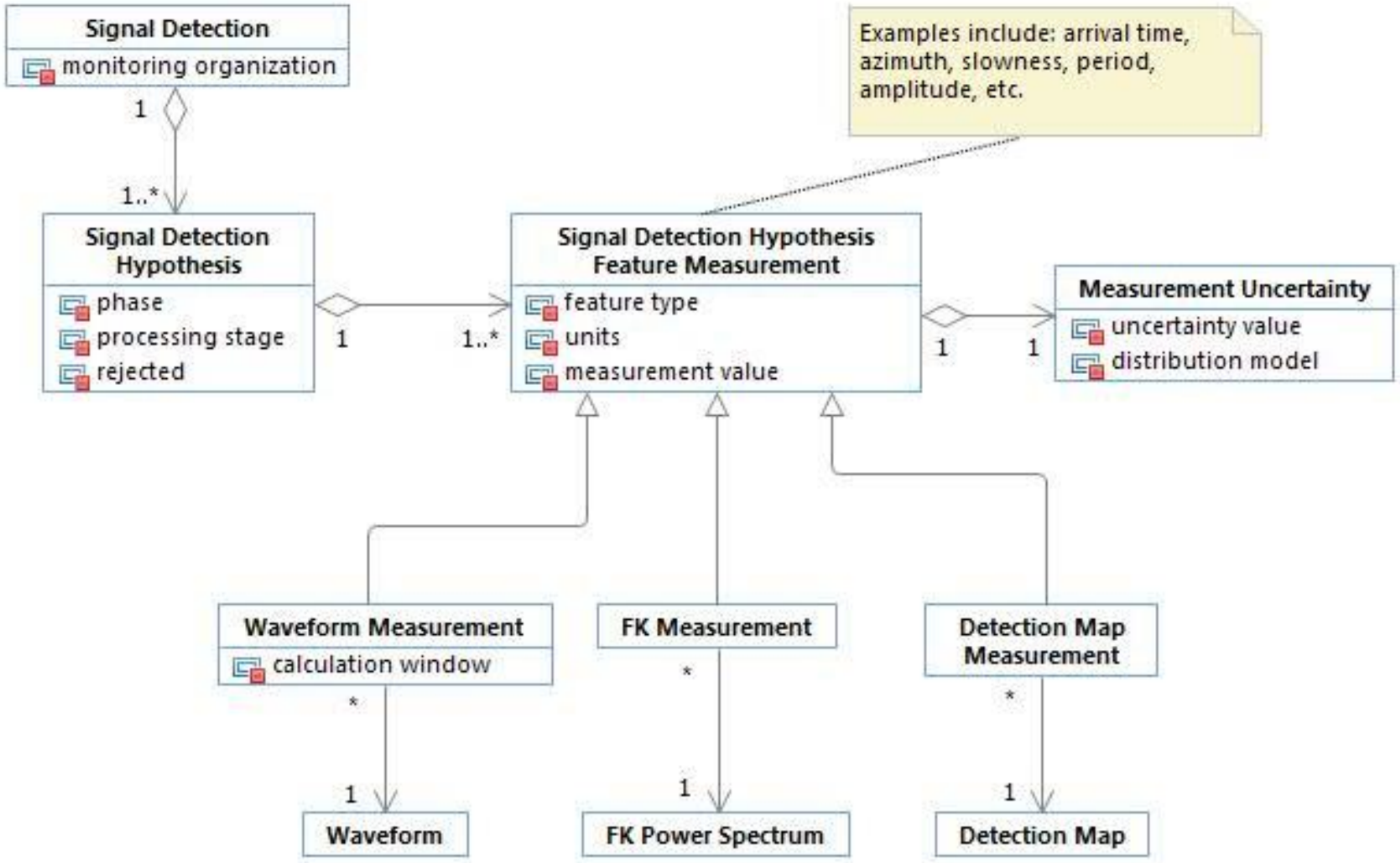


Signal Processing Information

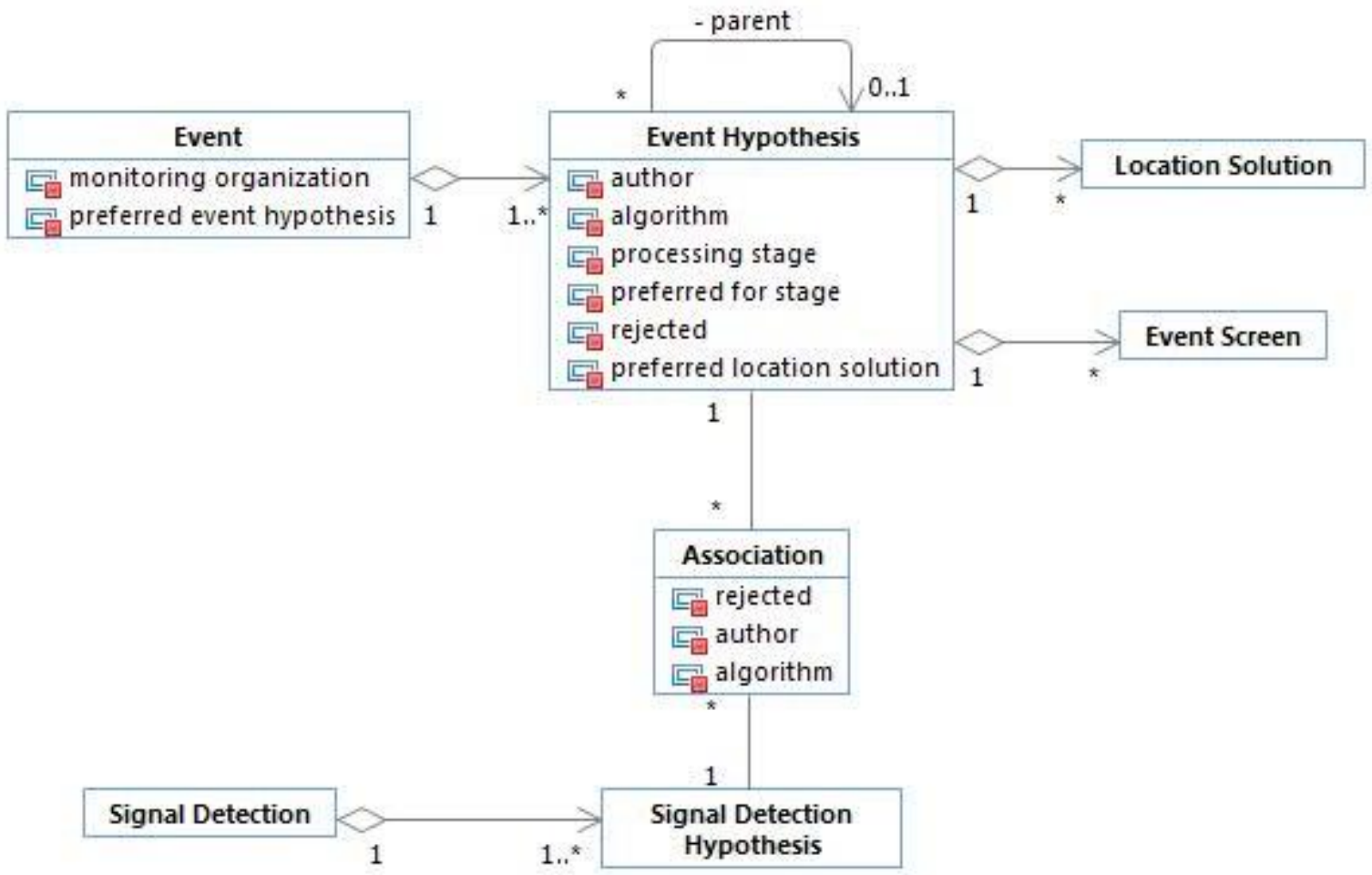


- Waveform data from Channels can be processed to produce a variety of object types (e.g. , FK, PMCC pixel map) that are used to detect and characterize signals.
- All of these object types will be calculated across a particular time window, but the data values they hold can be vastly different.

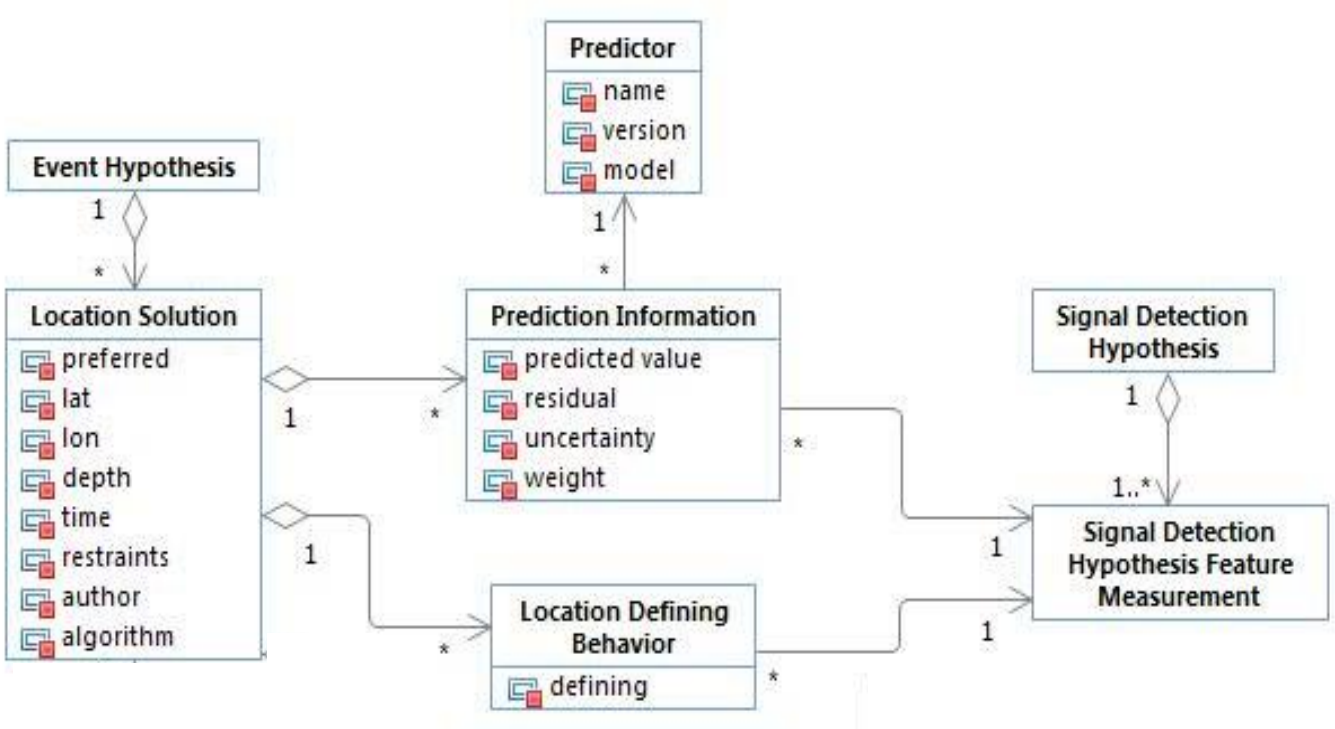
Signal Detection Information



Event Information

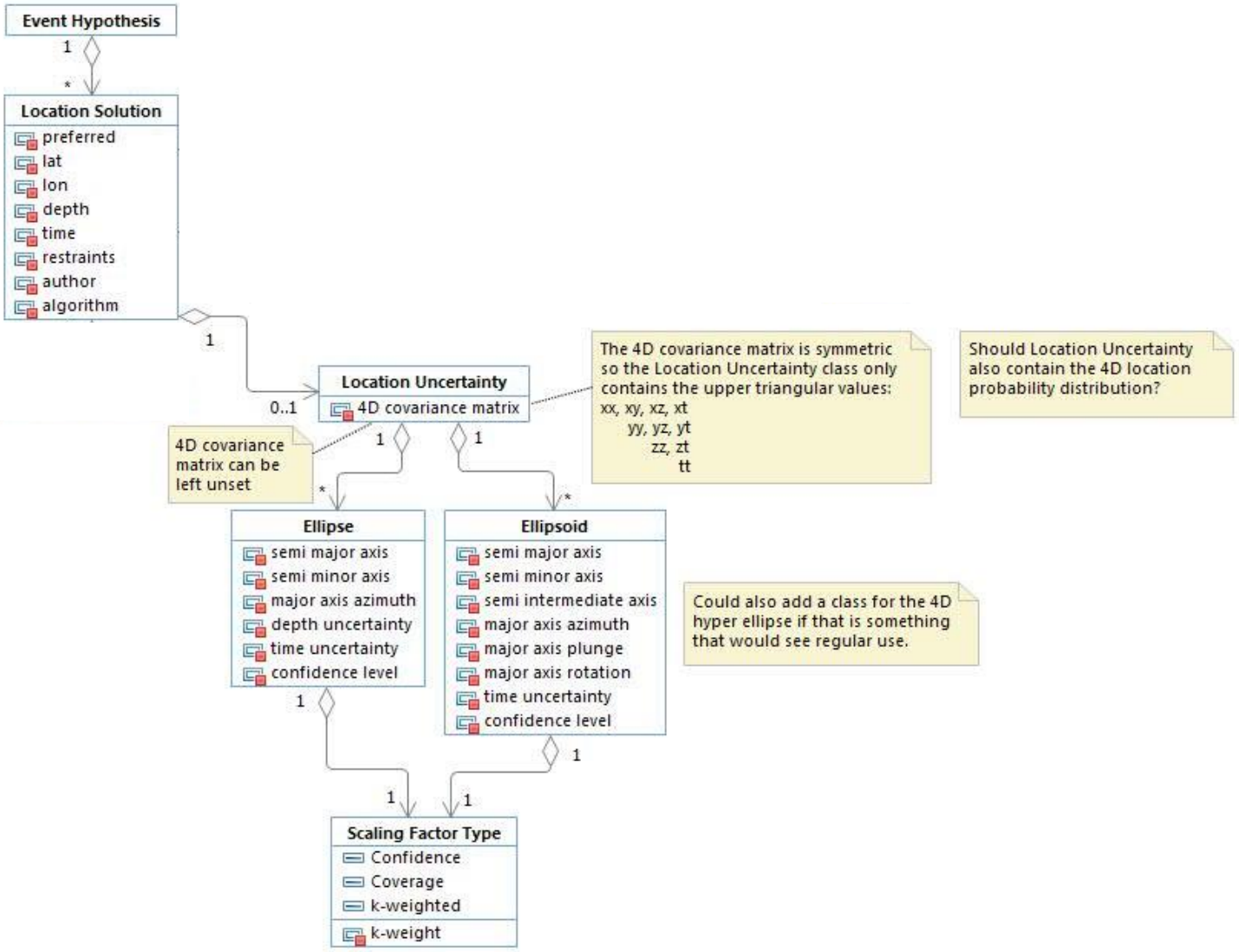


Event Location Information



- An Event Hypothesis (group of Signal Detection Hypotheses) can be located in multiple ways (free depth, fixed depth, etc.), hence there can be multiple Location Solutions

Event Location Uncertainty Information



Event Magnitude Information