

DDI and GSIM – Impacts, Context, and Future Possibilities

Arofan Gregory
Metadata Technology

Overview

- The general situation for GSIM – DDI
- Implementing GSIM with DDI
- Detailed view of some GSIM areas and overlap/gaps with DDI
 - Describing data
 - Describing questionnaires
 - Describing codelists, categories, and concepts
 - Describing events and processing
- Looking forward

GSIM and DDI

- GSIM is a creation of the HLG-BAS group under UN/ECE
- DDI is a creation of the DDI Alliance
- There is no immediate formal relationship between these organizations
- However, both organizations have made statements that they will work together to make DDI a good implementation vehicle for GSIM

GSIM, DDI, and Official Statistics

- GSIM is a key standard for official statistics organizations
- Some official statistical organizations already use DDI or are planning to do so
 - IHSN Metadata tools (developing world)
 - DDI-Lifecycle (ABS, Stats NZ, INSEE, Eurostat)
- GSIM is a potential vehicle for the widespread adoption of DDI among official statistical organizations

Models at Different Levels

- GSIM is a *Conceptual Model*
 - It is technology and implementation-neutral
- DDI is an *Implementation Model*
 - It is cross-platform and application-neutral
 - It is implemented in XML (and soon, RDF), but isn't technology-neutral
- Specific applications have their own, internal models
 - These are bound to specific technologies and platforms

Implementing GSIM at a Technical Level

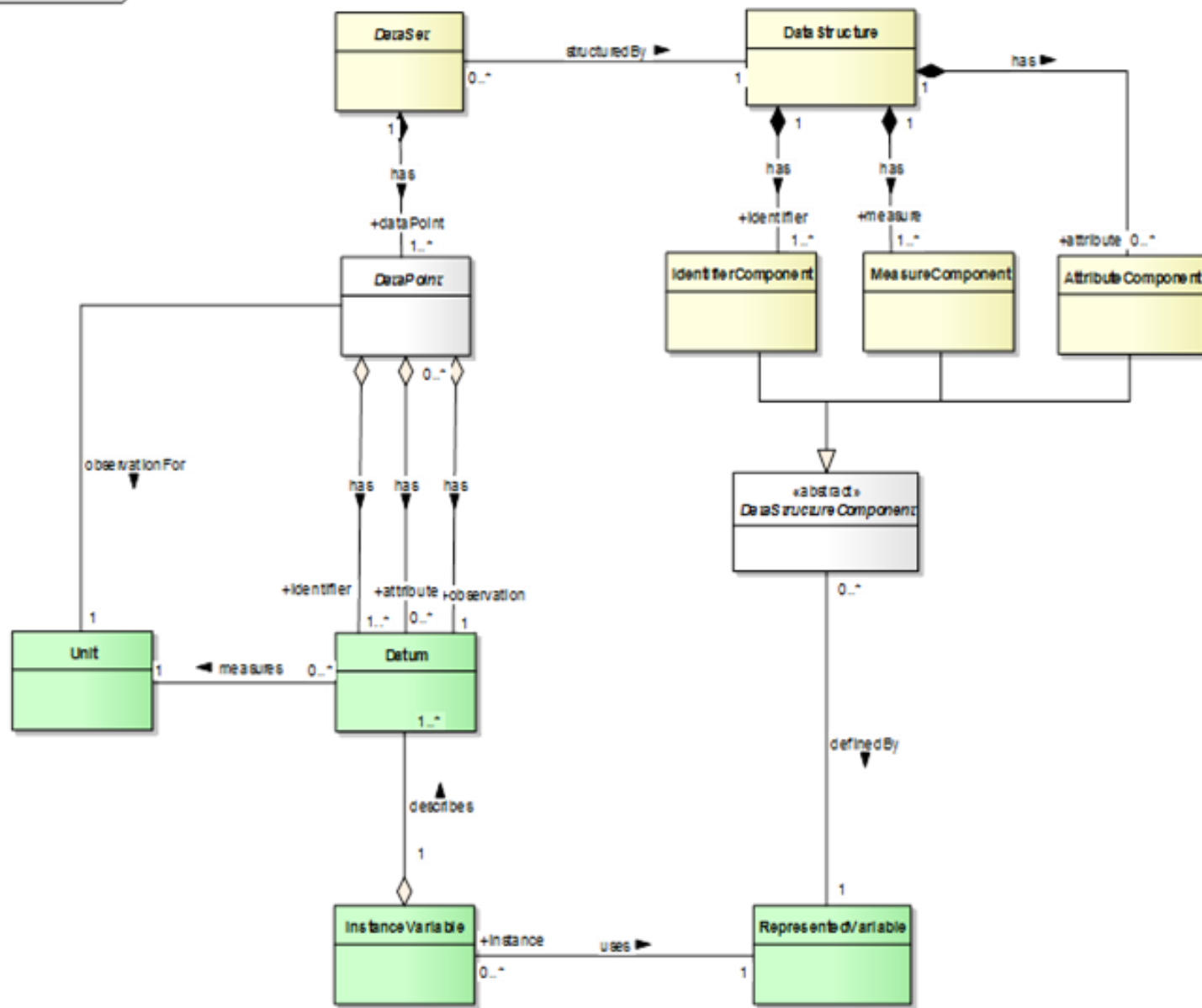
- To allow re-use of applications and services, agreements must exist on many levels
 - Conceptual models must match (GSIM)
 - Implementation models must match (DDI)
 - Application models must match (TBD – web services? Others?)
- There is still a lot of work around mapping DDI to GSIM, and then agreeing on how DDI XML will be used within applications before we have reusable, interoperable GSIM-based services and applications

What is the Usefulness of GSIM?

- To make applications work together on all levels, we will need to map existing application models to each other
 - On the basis of DDI
 - On the basis of GSIM
- From a technical perspective, this can be very difficult
 - Having an agreed base model at the conceptual and implementation level makes it easier/possible

Describing Data

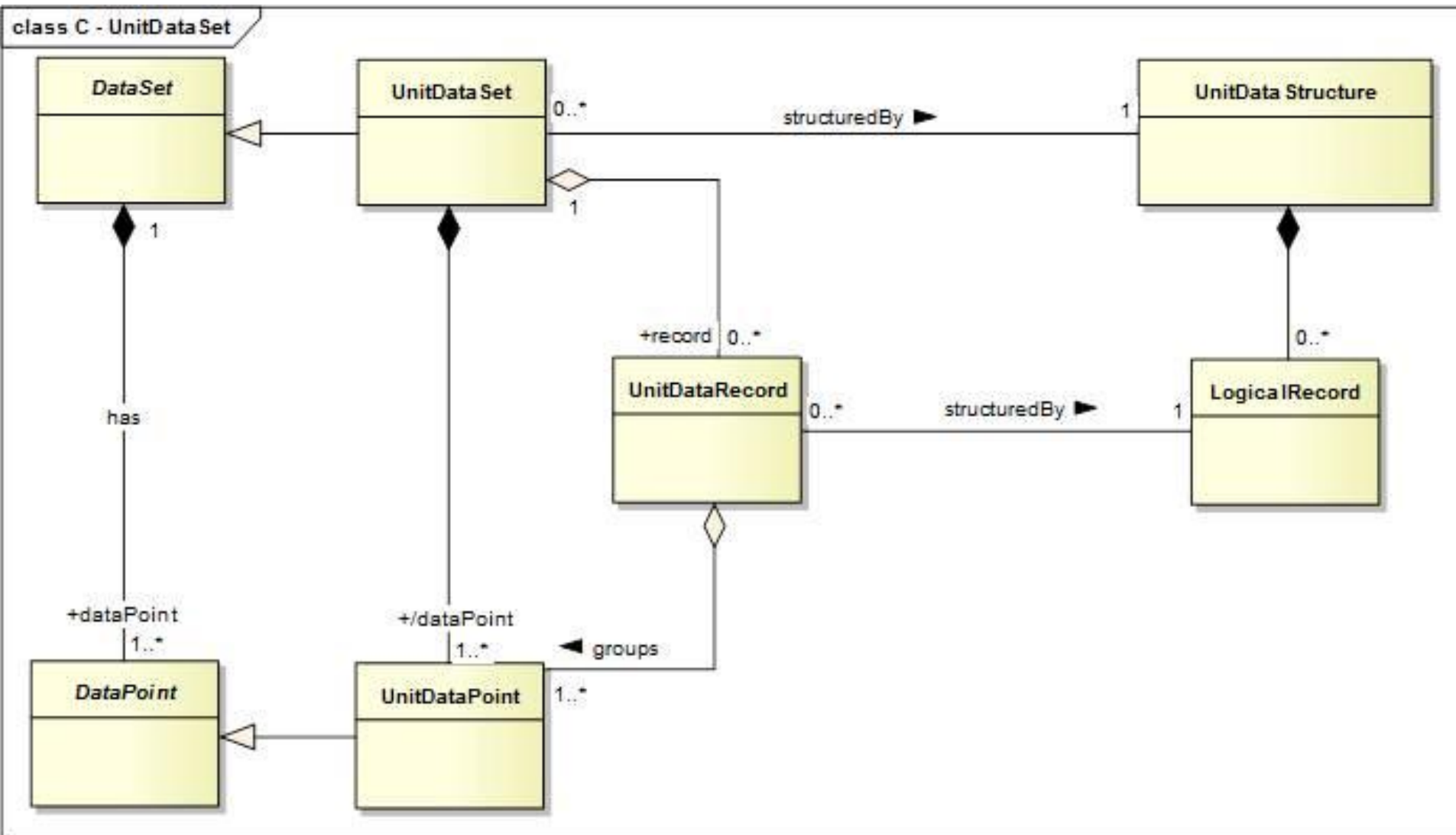
- DDI describes two kinds of data
 - Microdata sets
 - Aggregate (“dimensional”) data sets – Ncubes
- Both exist in GSIM
- In DDI microdata, each case/unit has a set of variables, at least one of which is the case identifier
 - Others hold observations or derived or supporting values (such as weights)
- Ncube structures use variables as dimensions, observations, and attributes to describe the matrix structure of tables



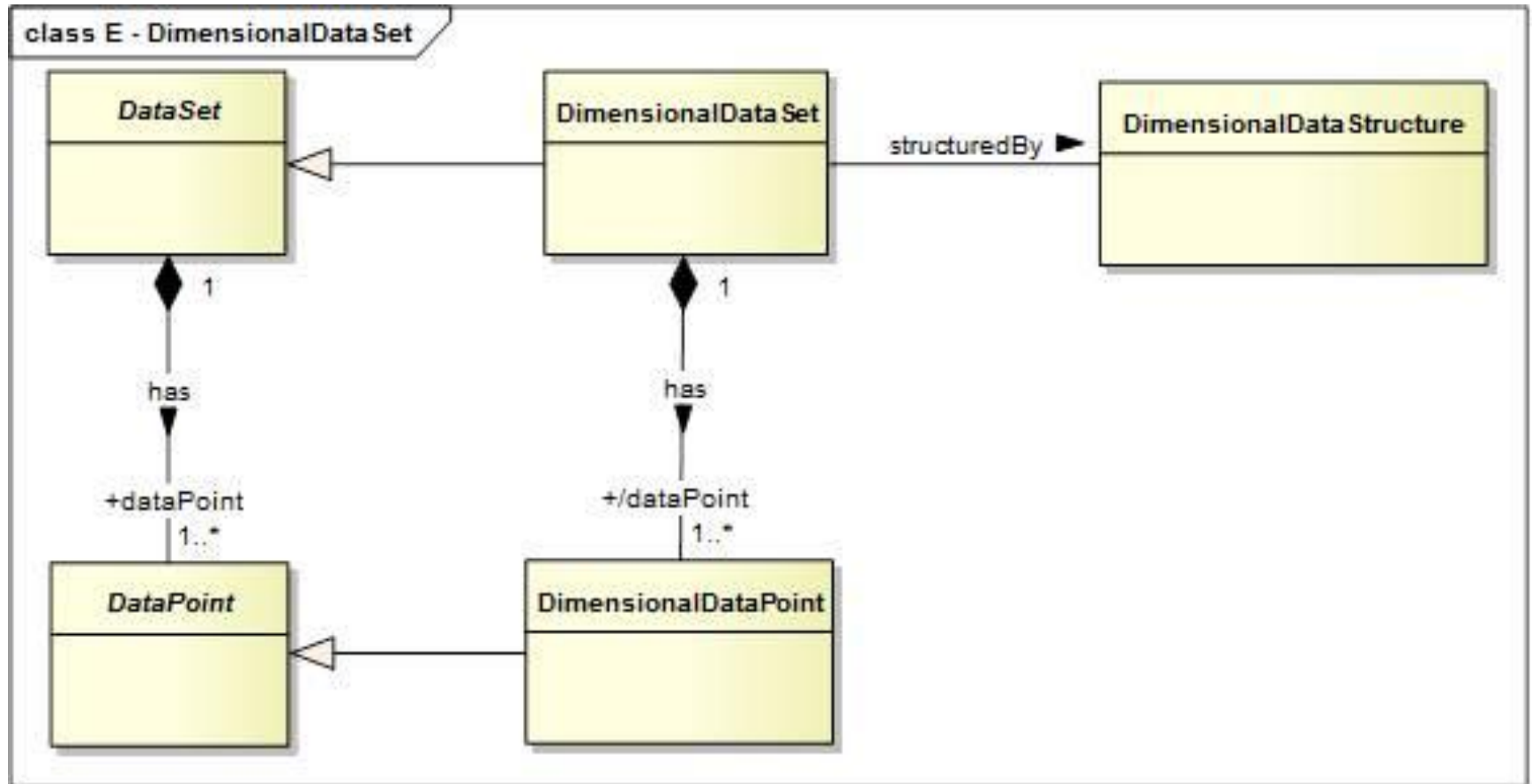
Other GSIM Constructs

- GSIM does make a distinction between “unit data structures” and “dimensional data structures”
 - GSIM supports hierarchical relationships in data sets
- Both are based on the core data model you have seen

Unit Data Set



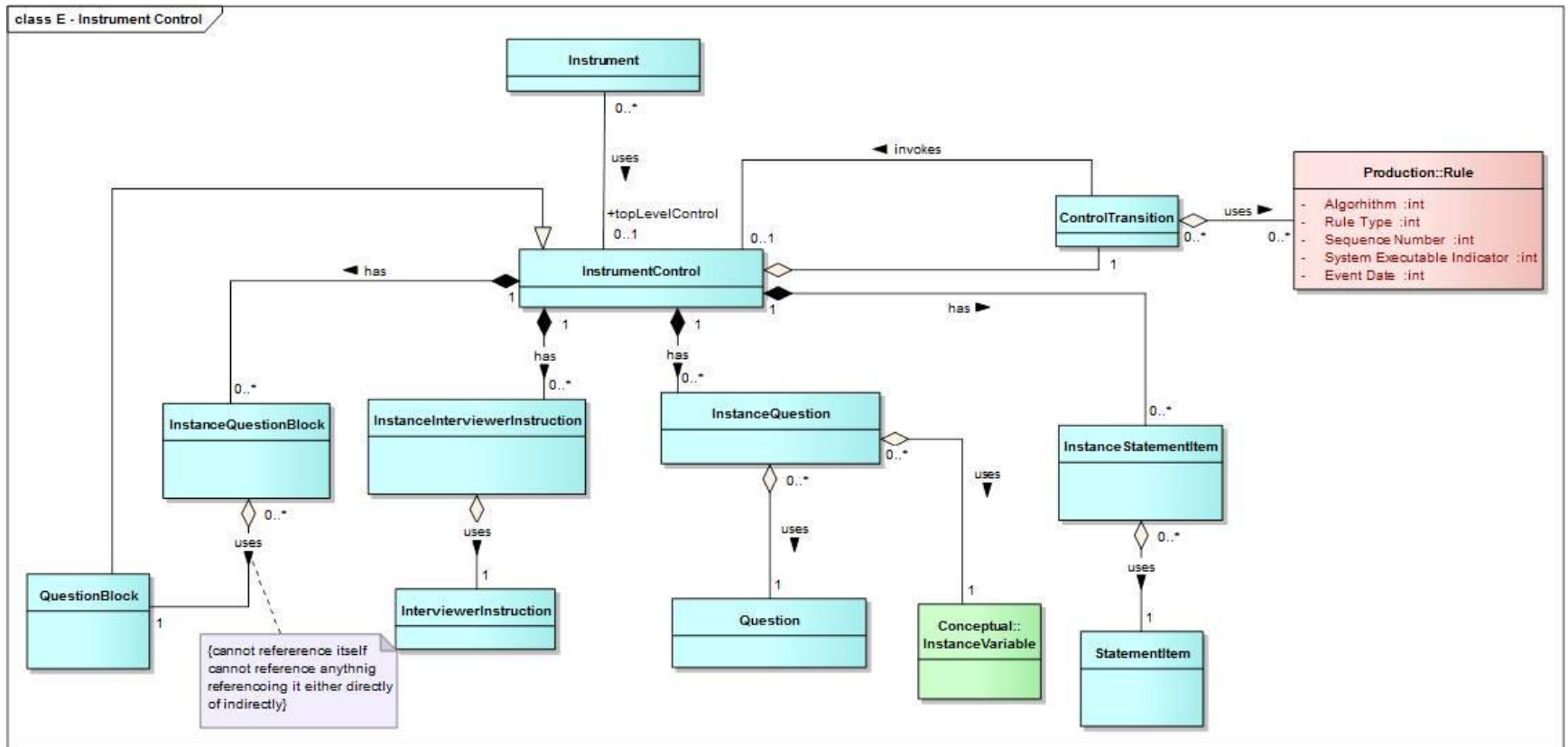
Dimensional Data Set



Describing Questionnaires

- DDI – Lifecycle has a very complete description of a questionnaire/instrument
 - Includes the mode and specifics of the instrument
 - Includes the questions, statements, and instructions used
 - Includes the flow logic of the questionnaire
 - Can have multiple-question “blocks”
- GSIM does the same
 - With less detail
 - Largely based on DDI

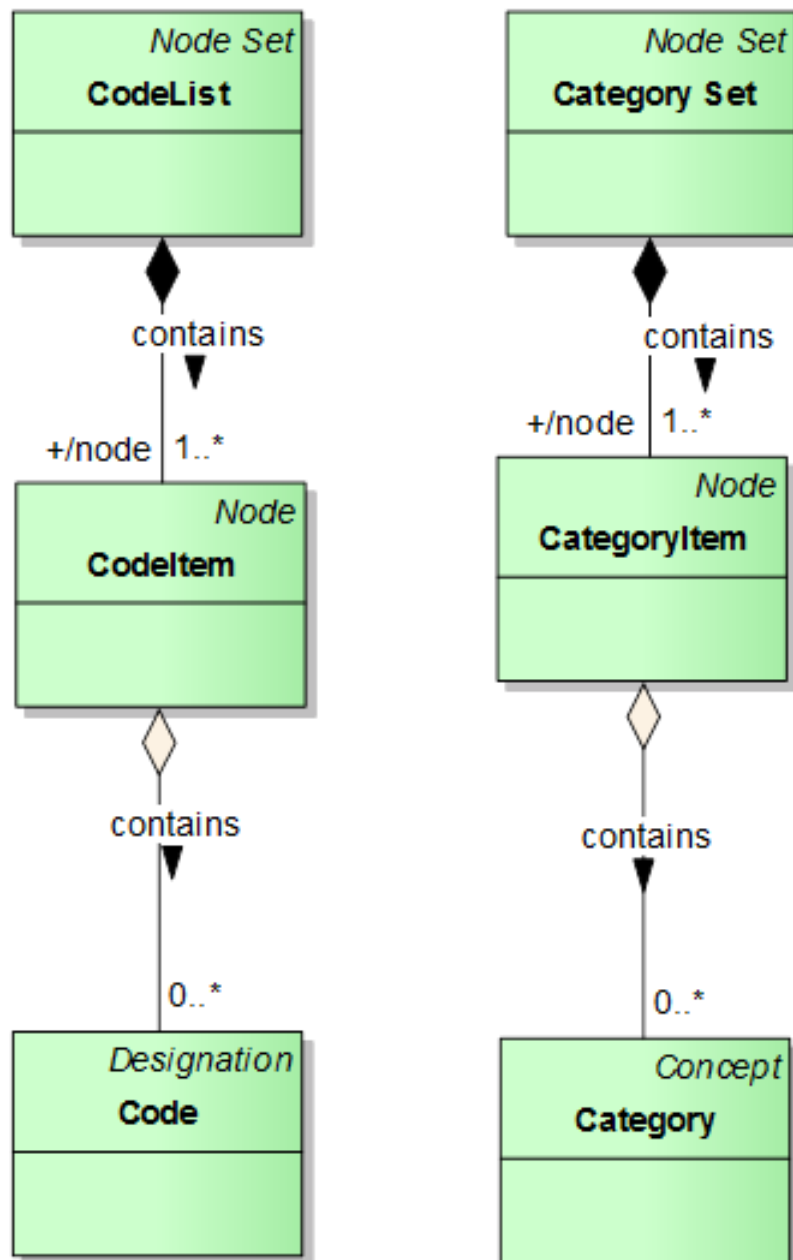
GSIM Survey Instrument



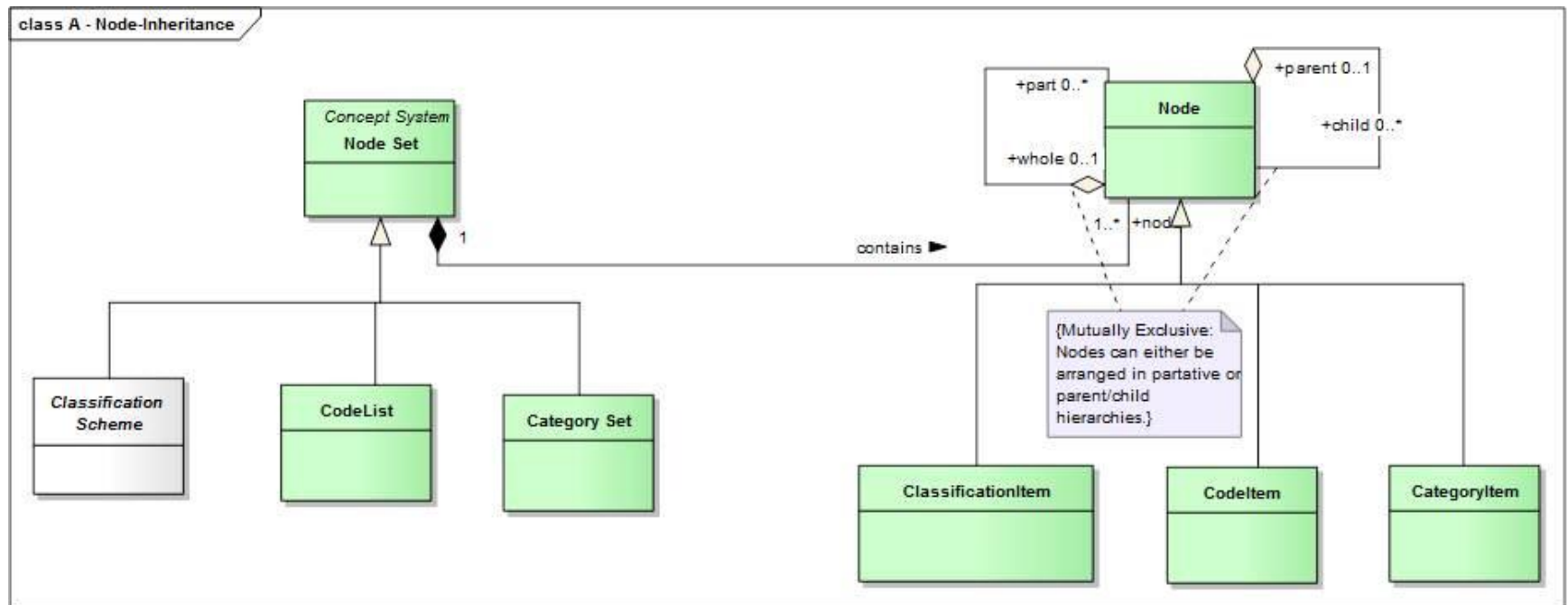
Classifications, Codelists, Categories and Concepts

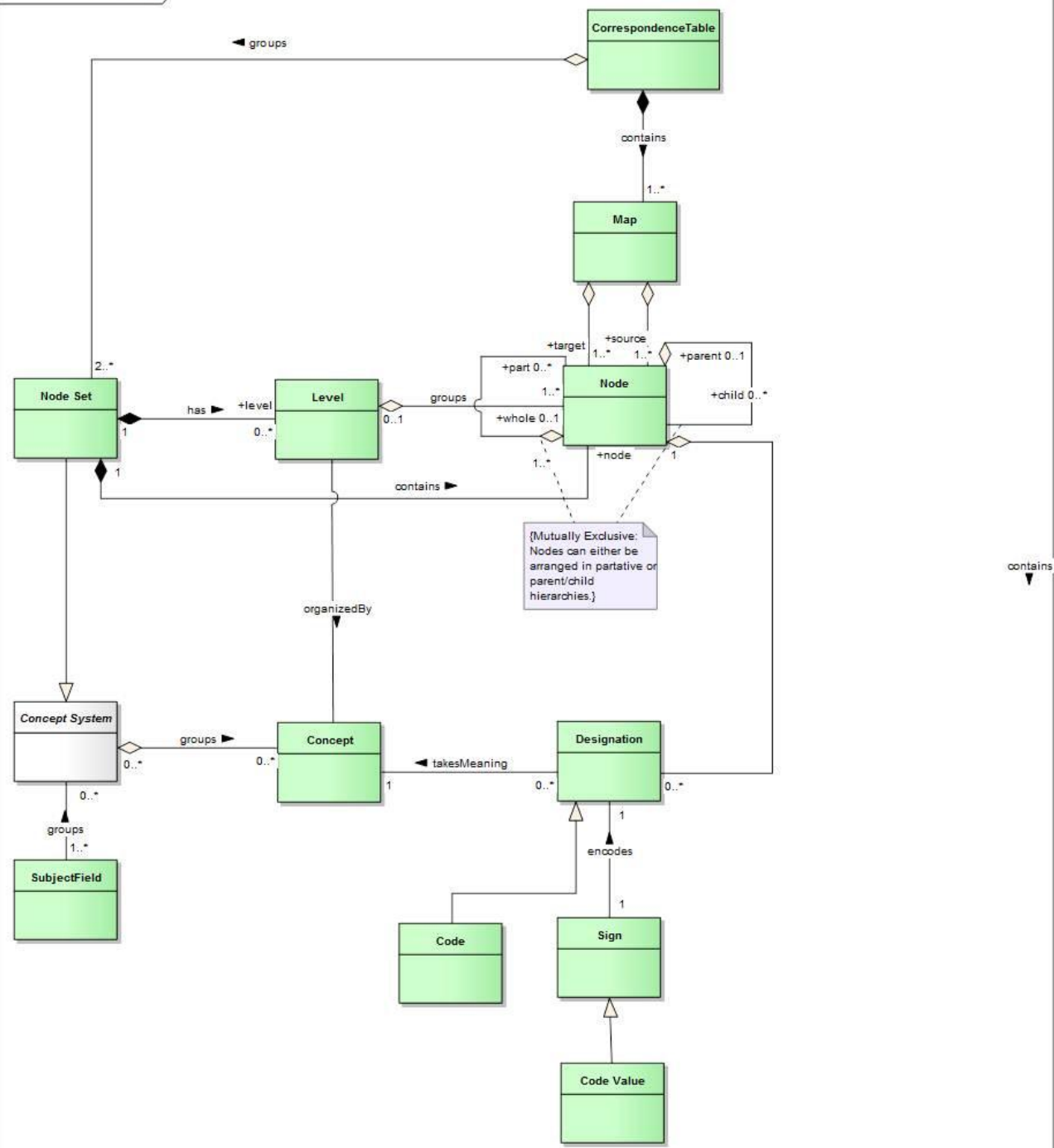
- DDI has codelists which take their meaning from categories.
- DDI has concepts associated with variables and questions.
- GSIM has all of this, and more!
 - GSIM is “concept-rich”
 - GSIM also has a pure classification model, which is not as complete in DDI-Lifecycle (a bit in 3.2)

class E - Category-Code

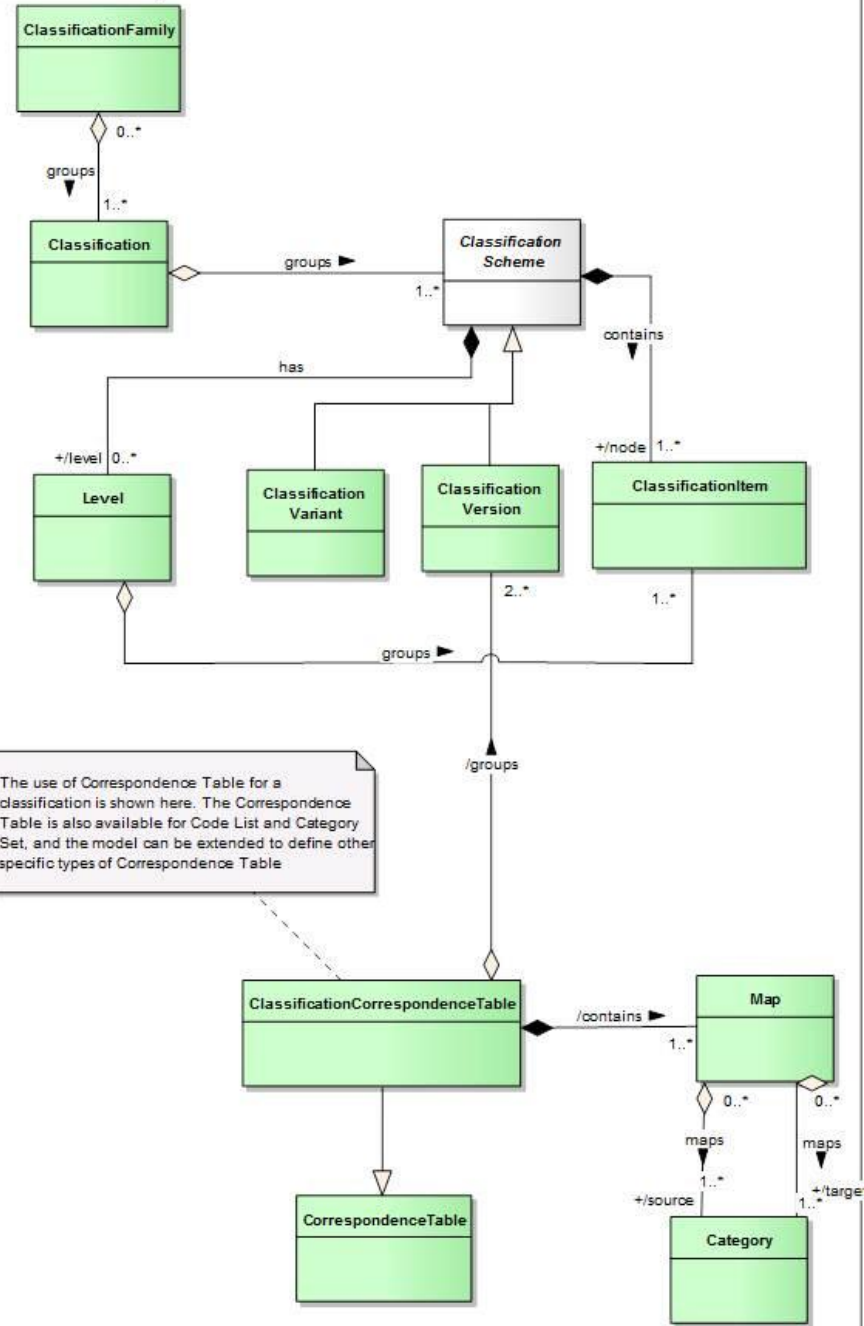


Nodes and Node-Sets





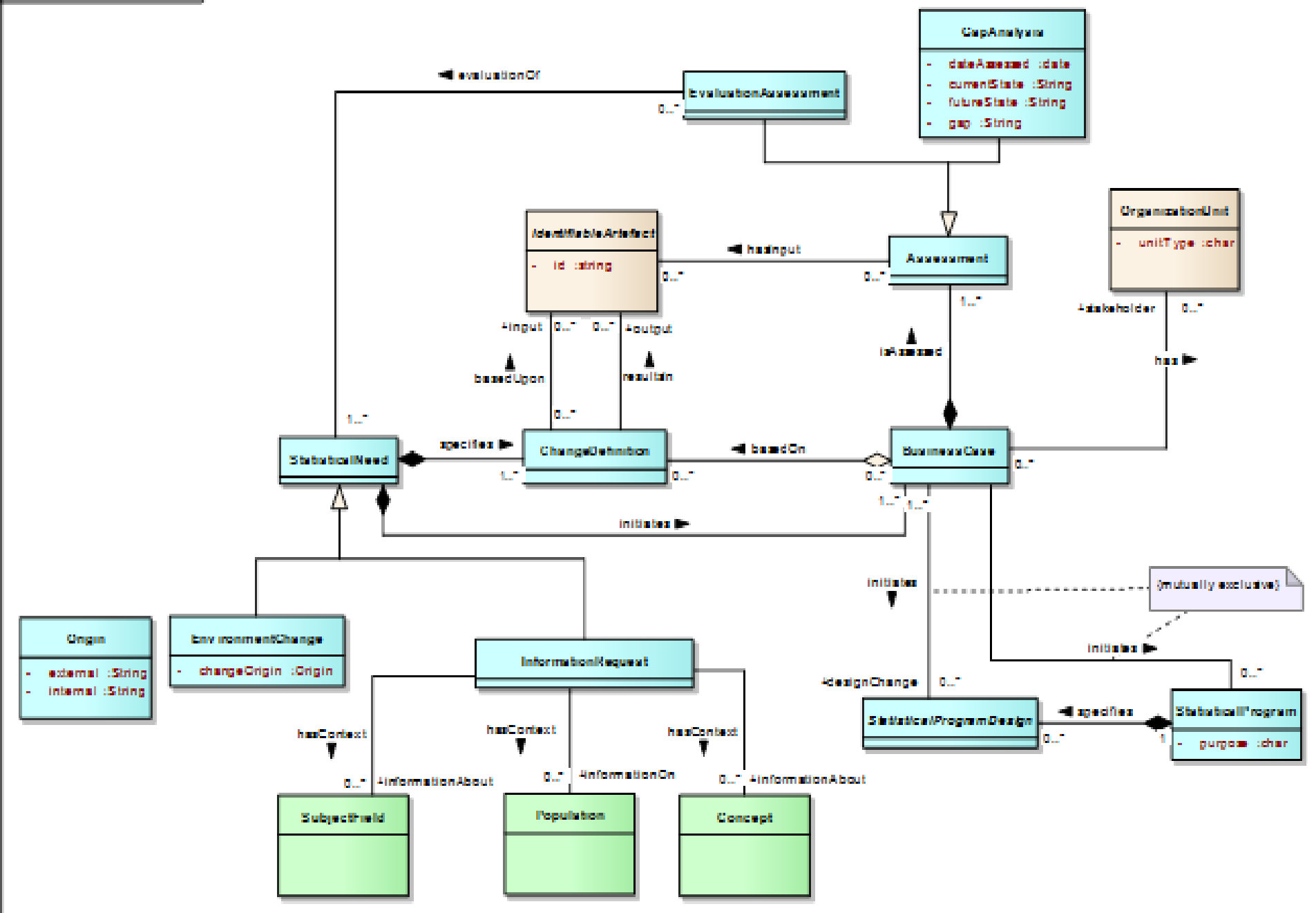
class D - Classification



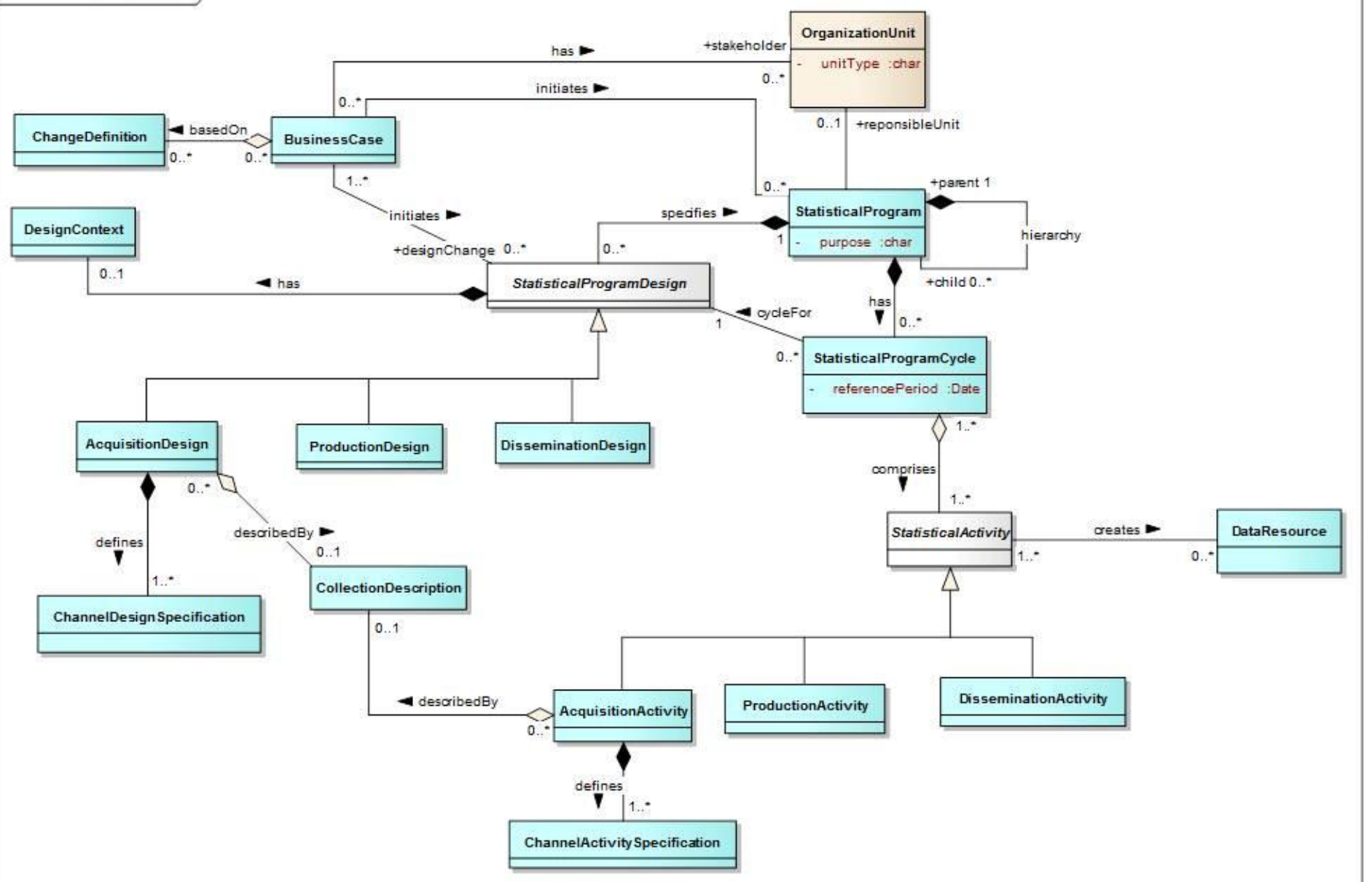
The use of Correspondence Table for a classification is shown here. The Correspondence Table is also available for Code List and Category Set, and the model can be extended to define other specific types of Correspondence Table

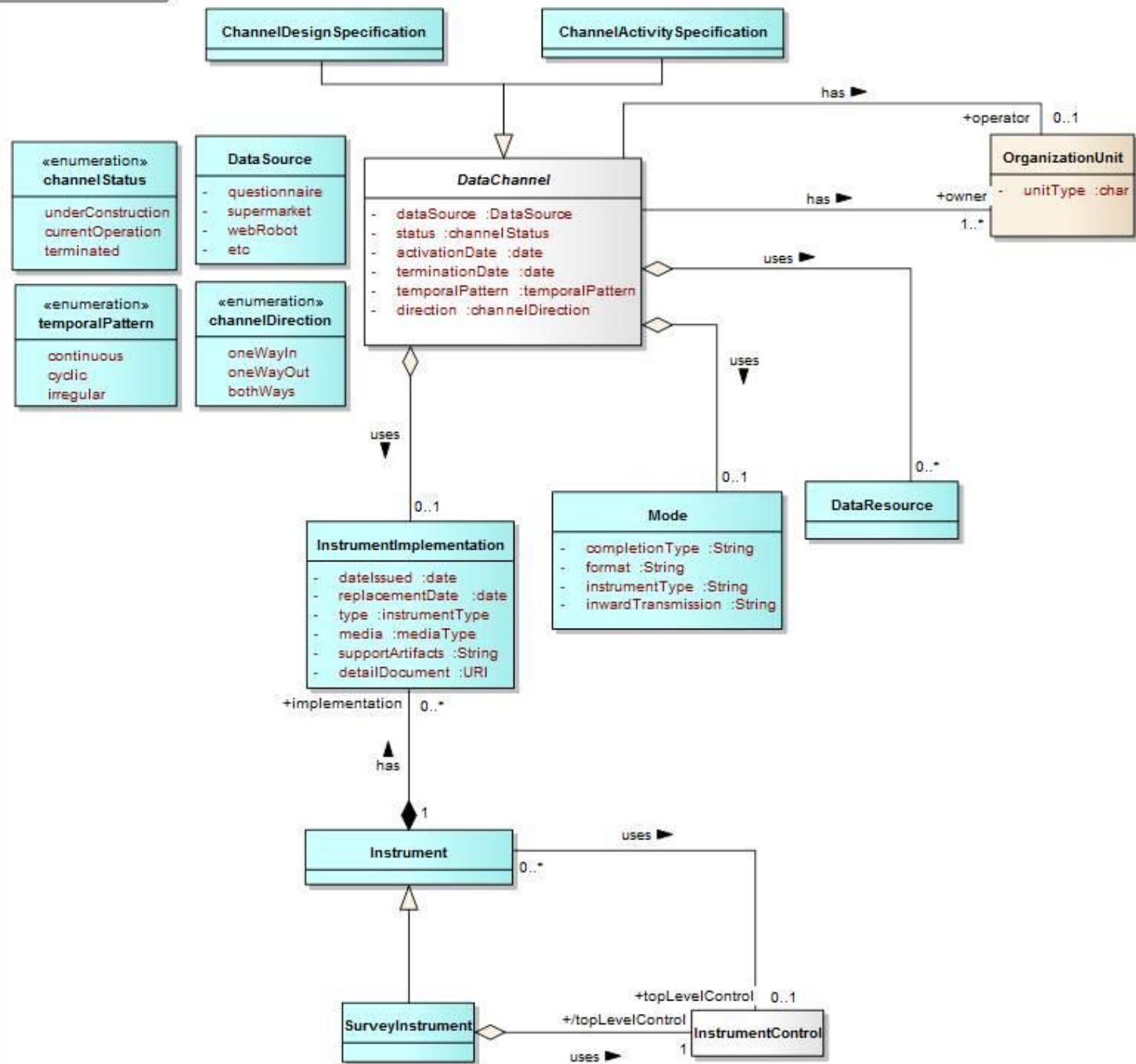
Events and Processing

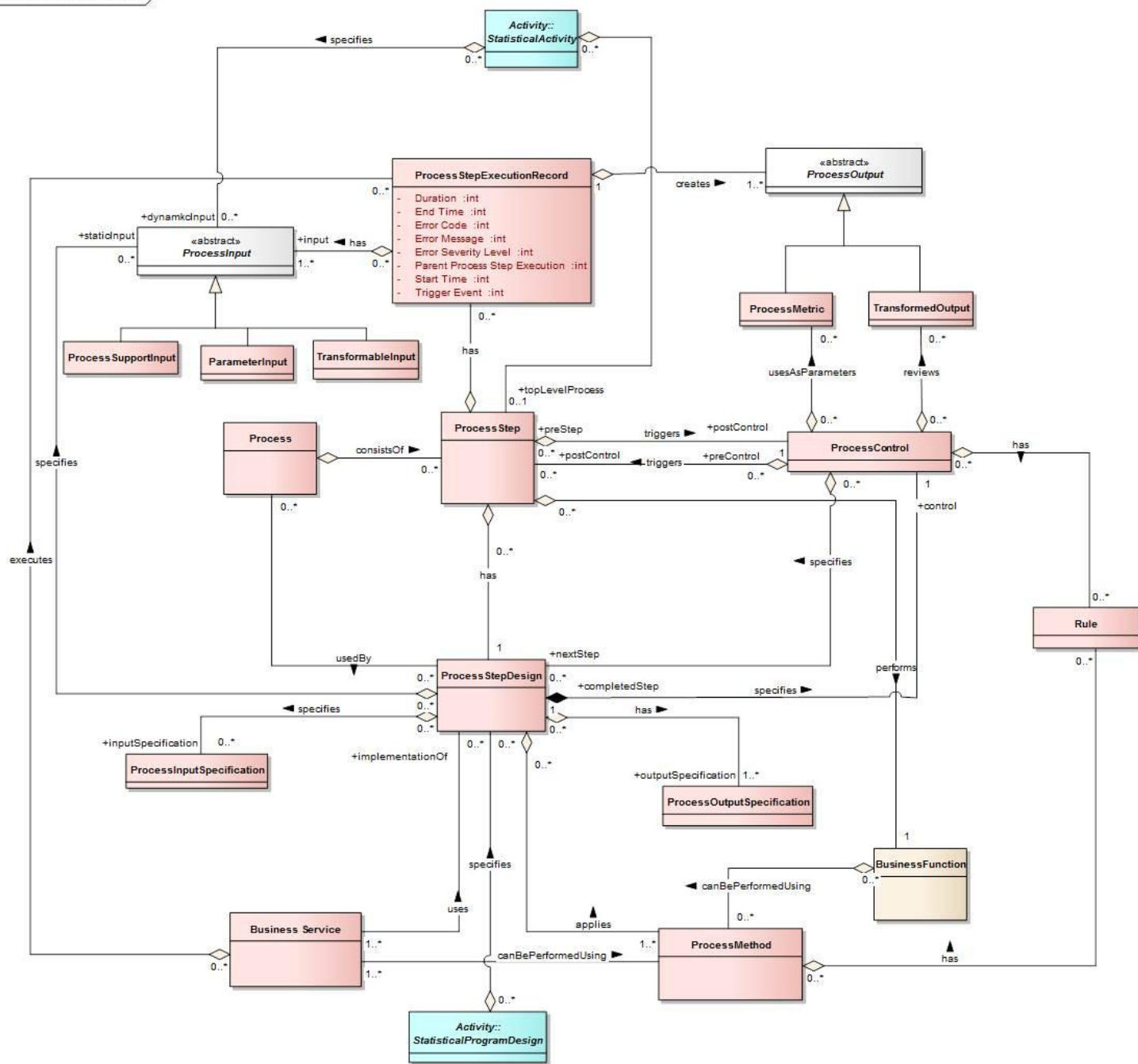
- DDI provides us with several ways to describe events and processing
 - Lifecycle Events
 - Collection Events
 - “Coding” Elements
 - Generation Instructions
 - General instructions
- GSIM gives us much, much more!
 - Some of this is very specific to statistical agencies



class B - Statistical-Program







Looking Forward

- DDI and GSIM have some very strong alignments
- There are also some gaps
- DDI may need to add support for some functionality
 - But maybe not everything – maybe SDMX can fill some gaps
- This is a two-way alignment
 - GSIM may need to adjust to better fit DDI implementation

Looking Forward (cont.)

- As we look to the next major re-design of DDI, we will be working proactively with GSIM
 - Representative from GSIM were invited to the first working session this year at Schloss Dagstuhl
- DDI will continue to attend events around GSIM sponsored by HLG-BAS
 - Like the Geneva meeting this past November
- Possibility for proactive engagement at a technical level
 - SDMX-DDI Dialogue
 - DDI Working Groups?
 - Others?
- GSIM may also provide a strong basis for other types of work within the DDI Community, less focused on official statistics
 - Like the “Generic Longitudinal Process Model”, which was based on GSBPM

Looking Forward

- Some external projects involve both archives and statistical agencies
 - Data without Boundaries (DwB) is a prime example
 - DwB is using a DDI-based metadata model
 - May lead to production implementations in future
- If archives and statistical agencies use the same metadata...
 - Archiving of official data becomes much easier
 - Both communities can leverage the same tools. Approaches, and resources (where appropriate)
 - Microdata access is an obvious point of synergy

Questions?