

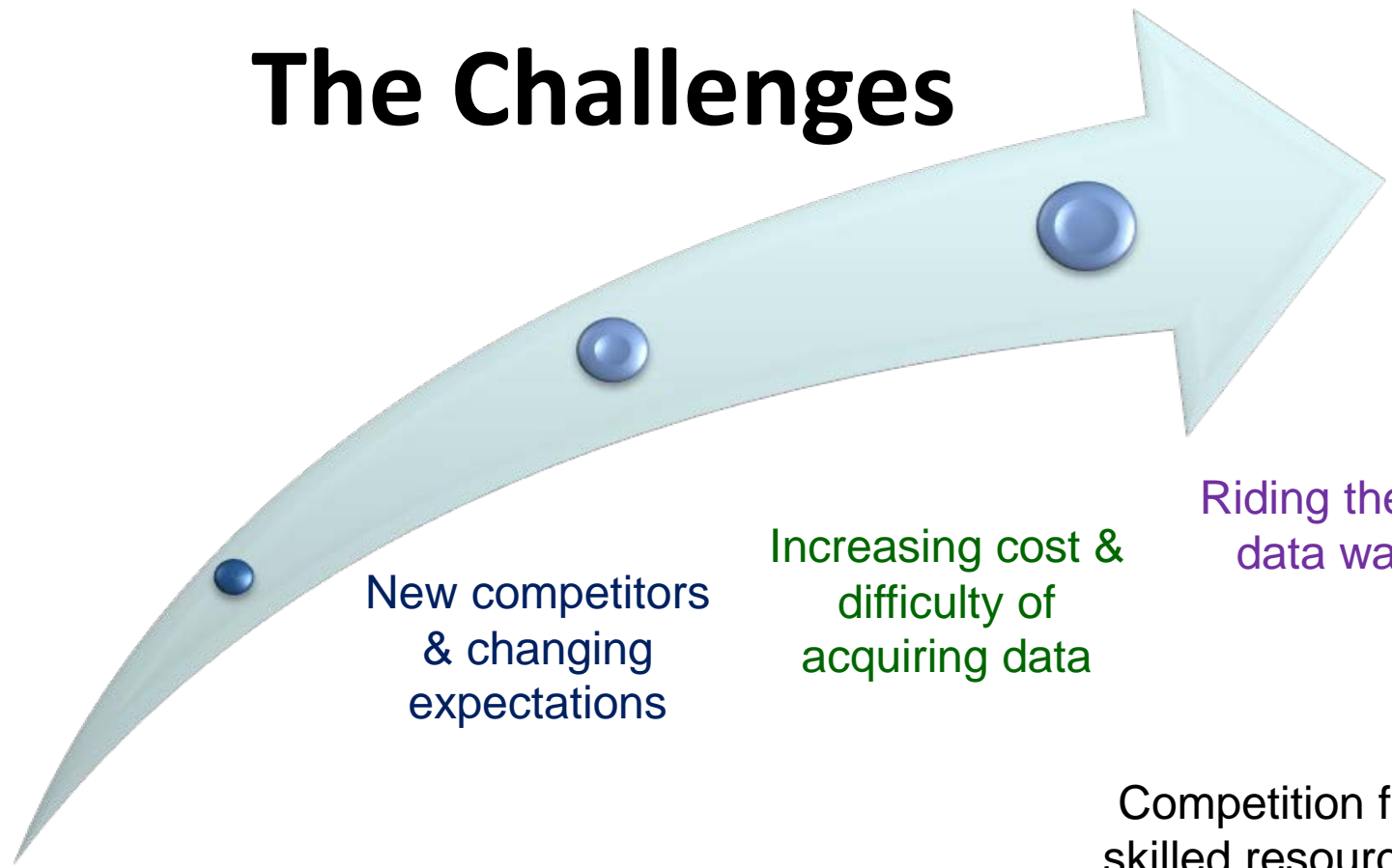
Generic Statistical Information Model (GSIM)

Thérèse Lalor and Steven Vale
United Nations Economic Commission for Europe
(UNECE)





The Challenges



Rapid changes in the environment

New competitors & changing expectations

Reducing budget

Increasing cost & difficulty of acquiring data

Competition for skilled resources

Riding the big data wave



GSIM

These challenges are too big for
statistical organisations to
tackle on their own.

We need to work together



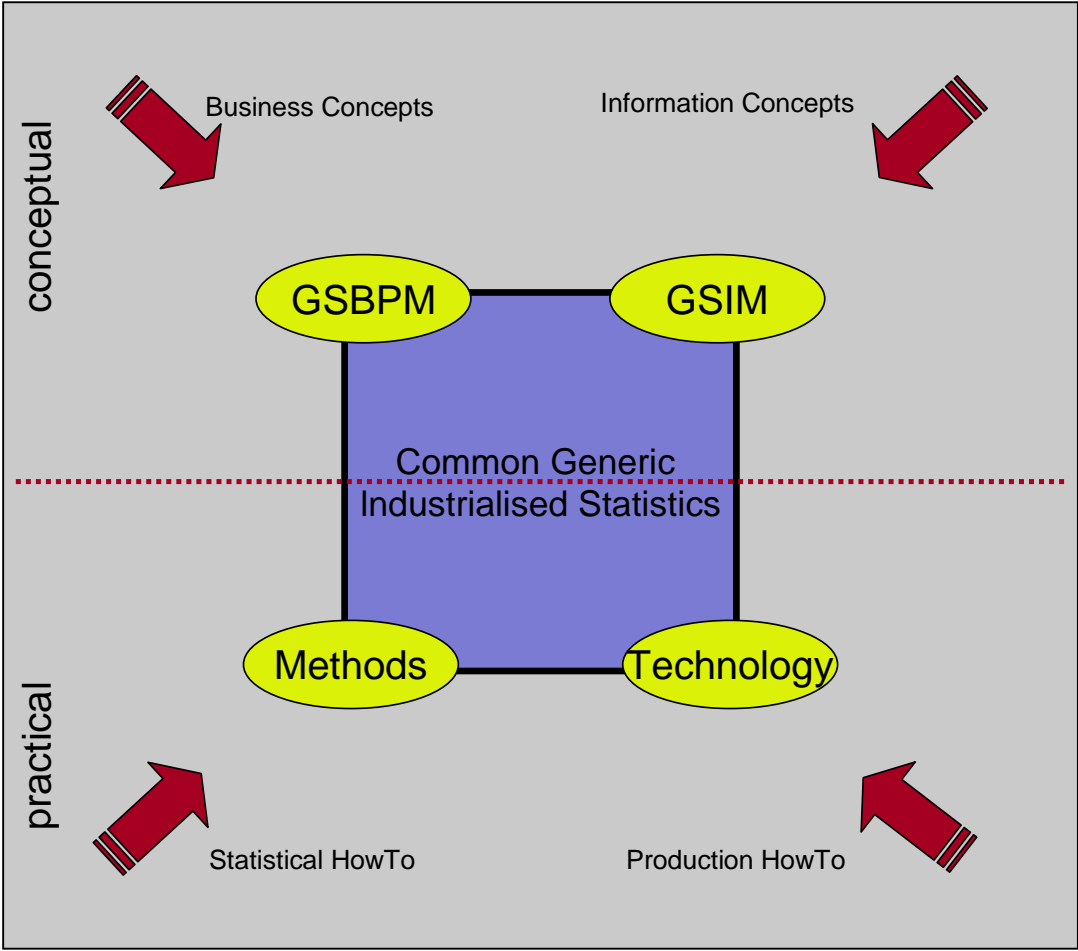
Response from Official Statistics

- A High Level Group consisting of 10 heads of national and international statistical organizations was created



**Using common standards, statistics can
be produced
in a more efficient way
No domain is special!**





GSIM

The GSBPM

Quality Management / Metadata Management

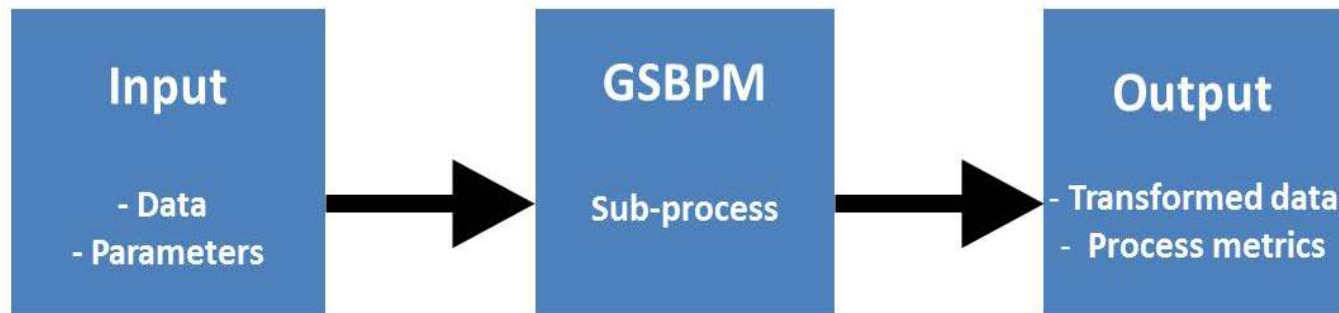
1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Archive	9 Evaluate
1.1 Determine needs for information	2.1 Design outputs	3.1 Build data collection instrument	4.1 Select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Define archive rules	9.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Manage archive repository	9.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design data collection methodology		4.3 Run collection	5.3 Review, Validate & edit				6.3 Scrutinize & explain
1.4 Identify concepts	2.4 Design frame & sample methodology	3.3 Configure workflows	4.4 Finalize collection	5.4 Impute	6.4 Apply disclosure control	7.3 Manage release of dissemination products	8.3 Preserve data and associated metadata	
1.5 Check data availability	2.5 Design statistical processing methodology	3.4 Test production system		5.5 Derive new variables & statistical units	6.5 Finalize outputs	7.4 Promote dissemination products	8.4 Dispose of data & associated metadata	
1.6 Prepare business case	2.6 Design production systems & workflow	3.5 Test statistical business process		5.6 Calculate weights		7.5 Manage user support		
		3.6 Finalize production system		5.7 Calculate aggregates				
				5.8 Finalize data files				

**The GSBPM is used by more than
50 statistical organizations
worldwide to manage and
document statistical production**



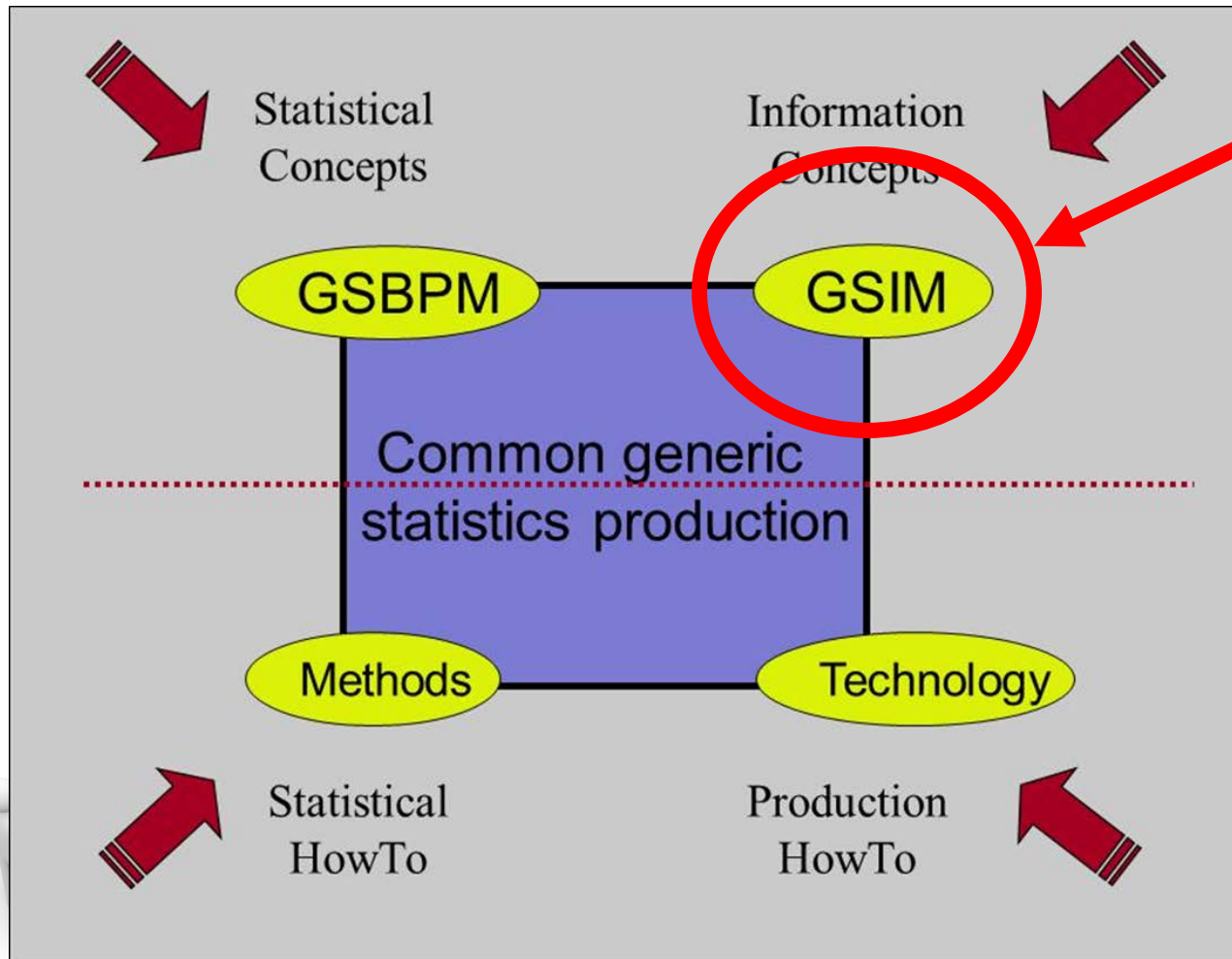
GSIM is complementary to GSBPM

Another model is needed to describe information objects and flows within the statistical business process



GSIM

Introducing the GSIM



You are here

What is GSIM?

- A reference framework of information objects
- It sets out definitions, attributes and relationships regarding information objects
- It aligns with relevant standards such as DDI and SDMX



Purposes of GSIM

- Improve communication
- Generate economies of scale
- Enable greater automation
- Provide a basis for flexibility and innovation
- Build staff capability by using GSIM as a teaching aid
- Validate existing information systems





GSIM is a conceptual model:
It is a new way of thinking for
statistical organizations





GSIM enables:

- Communication
- Coordination
- Cooperation
- Collaboration





HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.

YEAH!



SOON:

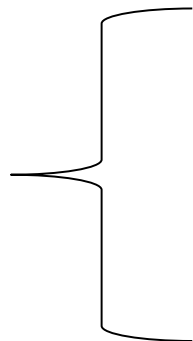
SITUATION:
THERE ARE
15 COMPETING
STANDARDS.



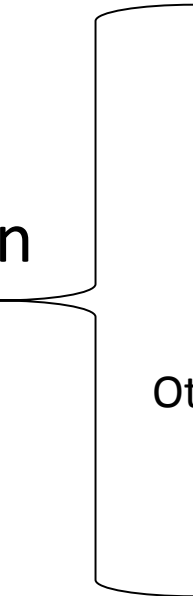
GSIM



Conceptual model



Implementation standards



DDI

SDMX

Other relevant standards

Geospatial standards

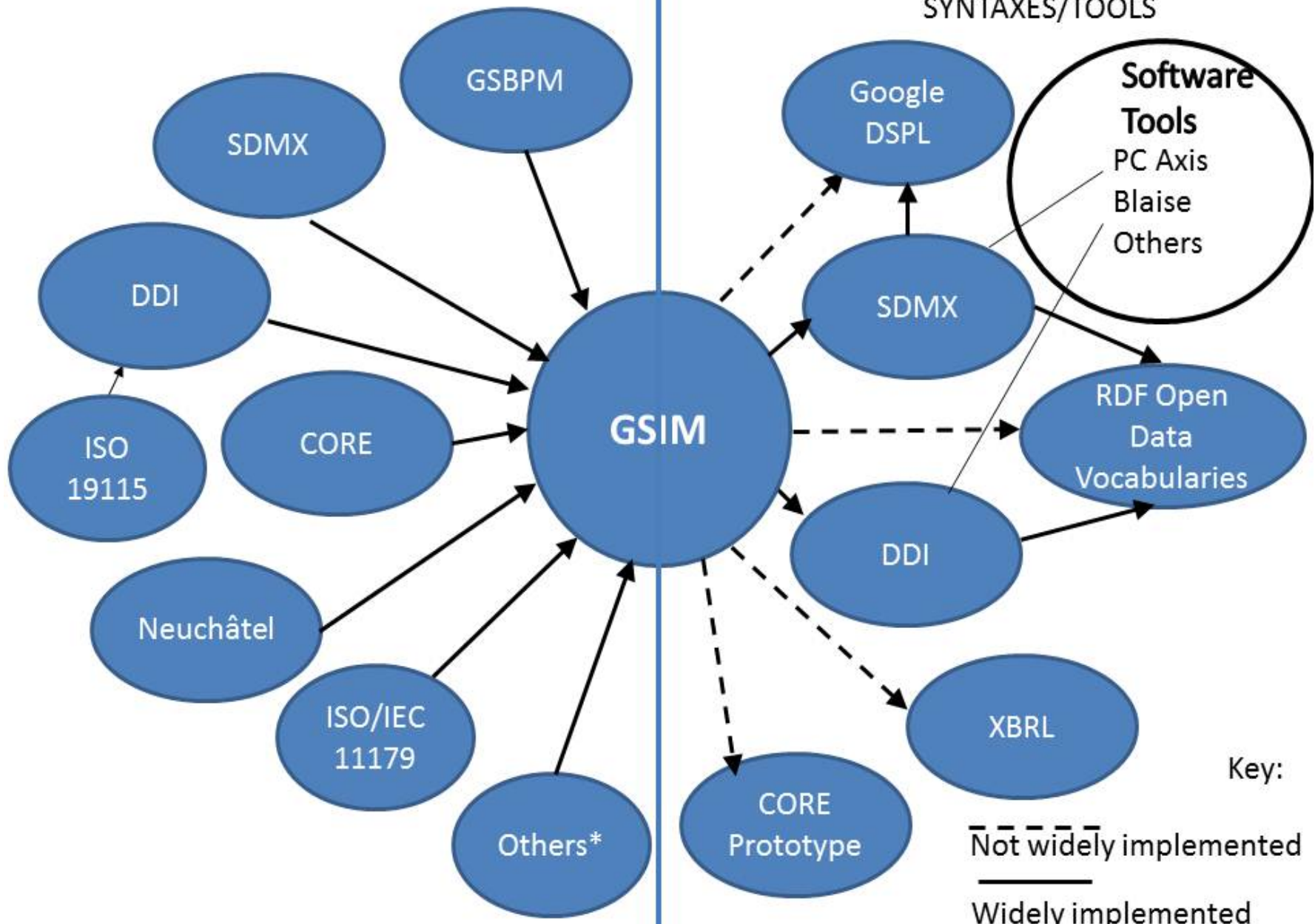


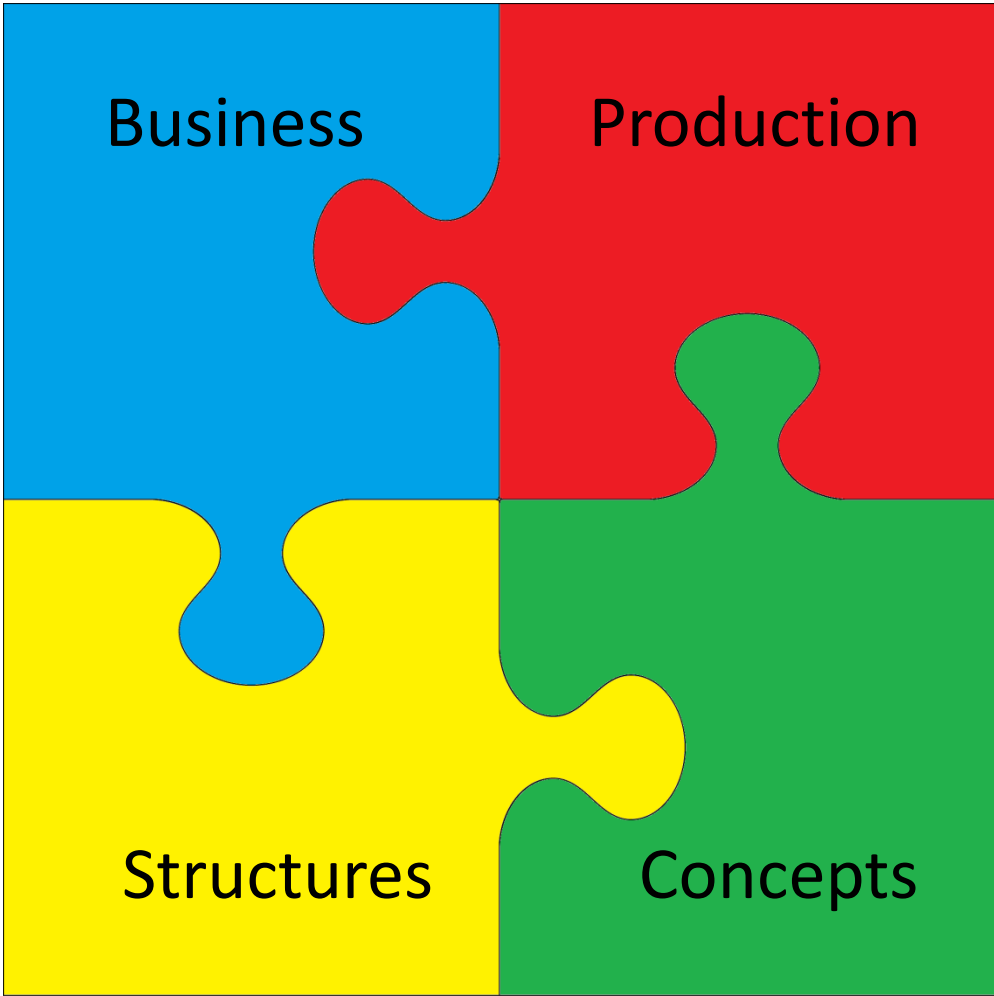
GSIM



REFERENCE STANDARDS/MODELS

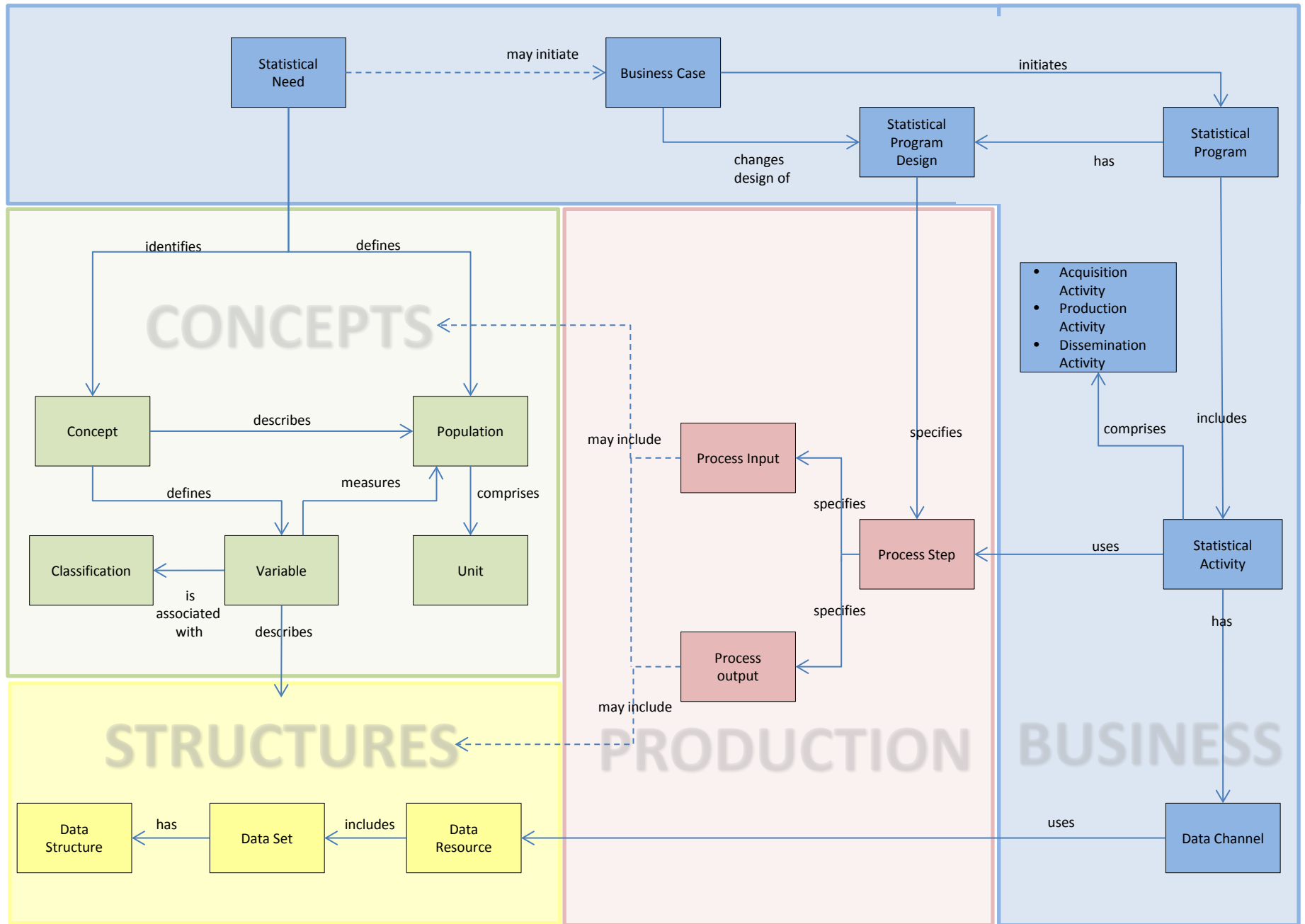
IMPLEMENTATION SYNTAXES/TOOLS





GSIM







GSIM: The “sprint” approach

- The HLG-BAS decided to accelerate the development of the GSIM
- “Sprints” – 2 week workshops for 10-12 experts (IT, methodology, statistics, ...)
- Sprint 1 – Slovenia, February 2012
- Sprint 2 – Republic of Korea, April 2012
- Integration Workshop, Netherlands, September 2012



Moving to GSIM in practice

GSIM could lead to:

- A foundation for standardized statistical metadata use throughout systems
- A standardized framework to aid in consistent and coherent design capture
- Increased sharing of system components





Moving to GSIM in practice

- Common terminology across and between statistical agencies.
- It allows NSIs and standards bodies, such as SDMX and DDI, to understand and map common statistical information and processes.



GSIM v1.0

- Released in December 2012
- We need people to use GSIM “in anger”. Then we will know how best to improve it.



More information

□ GSIM

□ [http://www1.unece.org/stat/platform/display/metis/Generi
c+Statistical+Information+Model+\(GSIM\)](http://www1.unece.org/stat/platform/display/metis/Generi
c+Statistical+Information+Model+(GSIM))



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

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, 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?

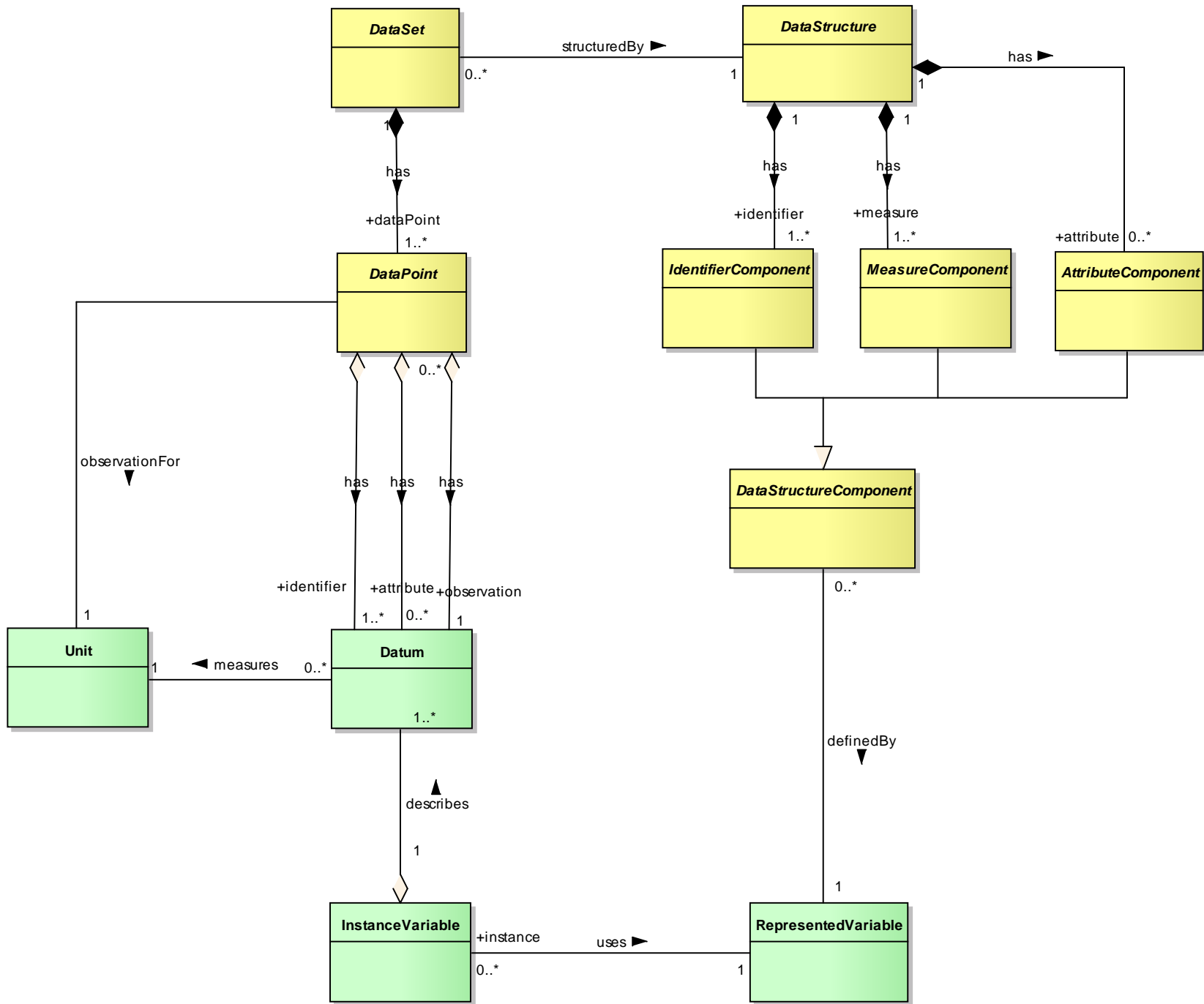
- To make applications work together on all levels, we will need to map existing application models to each other
 - On the basis of DDI
 - On the basis of GSIM
- From a technical perspective, this can be very difficult
 - Having an agreed base model at the conceptual and implementation level makes it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



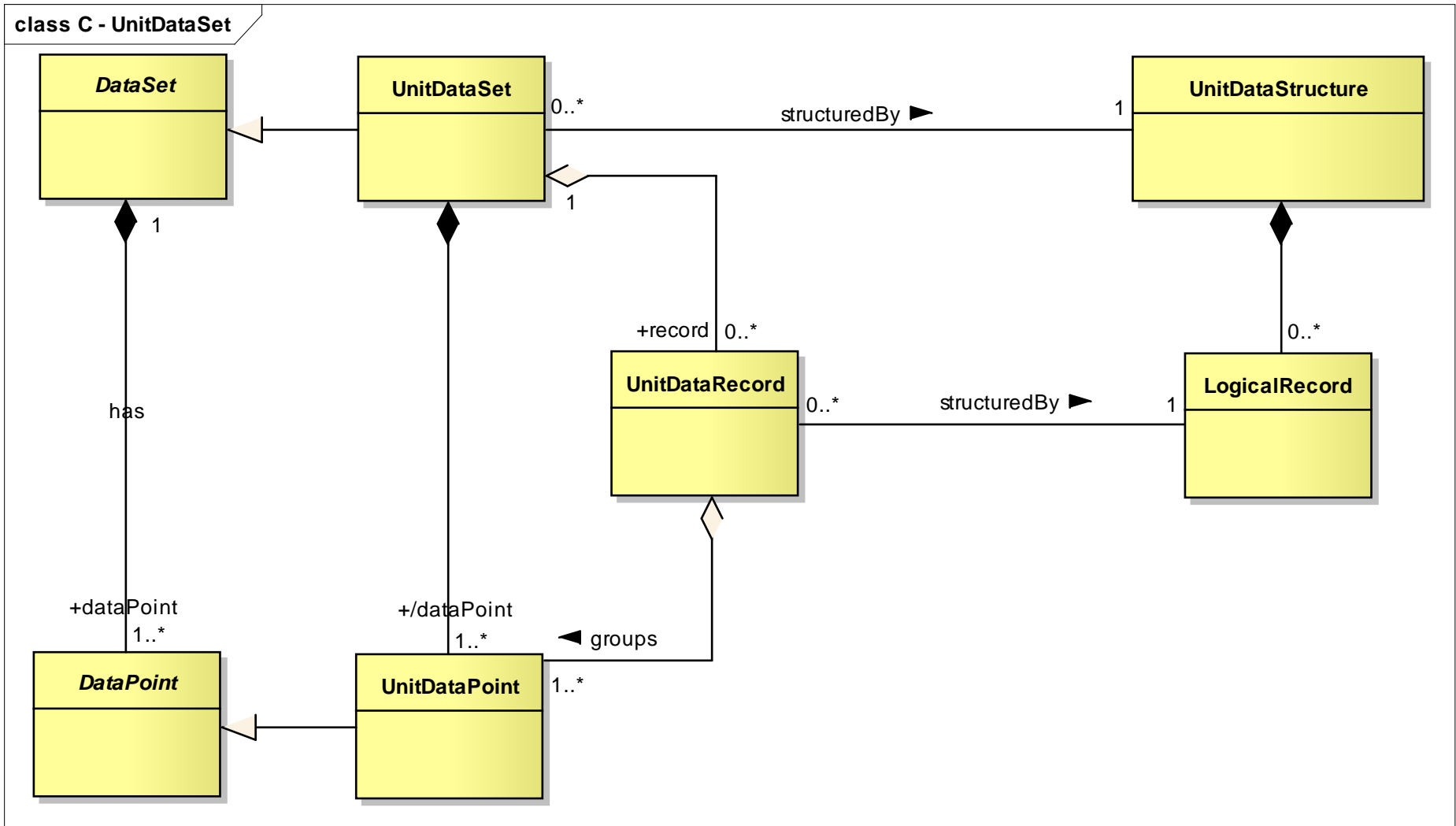


Other GSIM Constructs

- 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

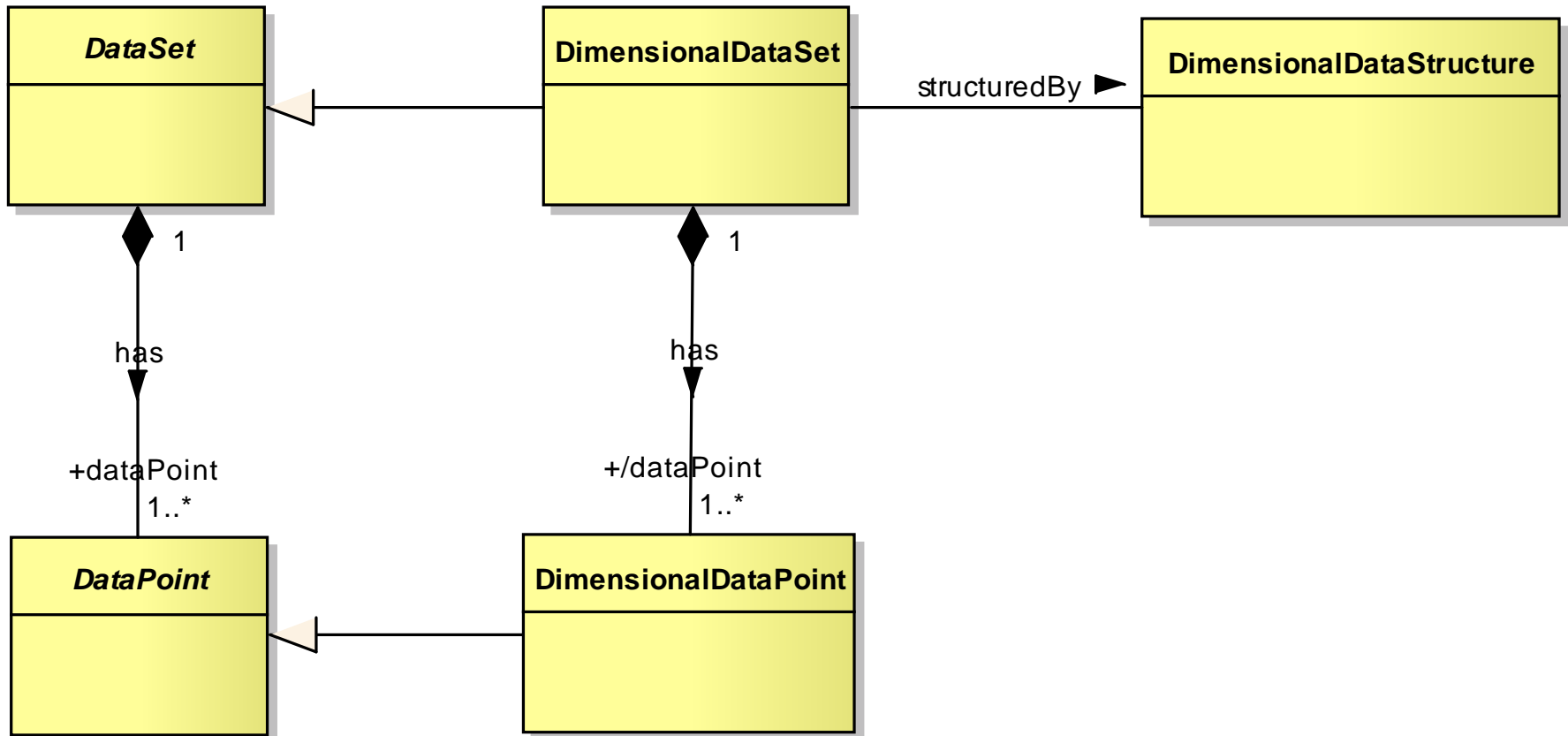


Unit Data Set



Dimensional Data Set

class E - DimensionalDataSet



Describing Questionnaires

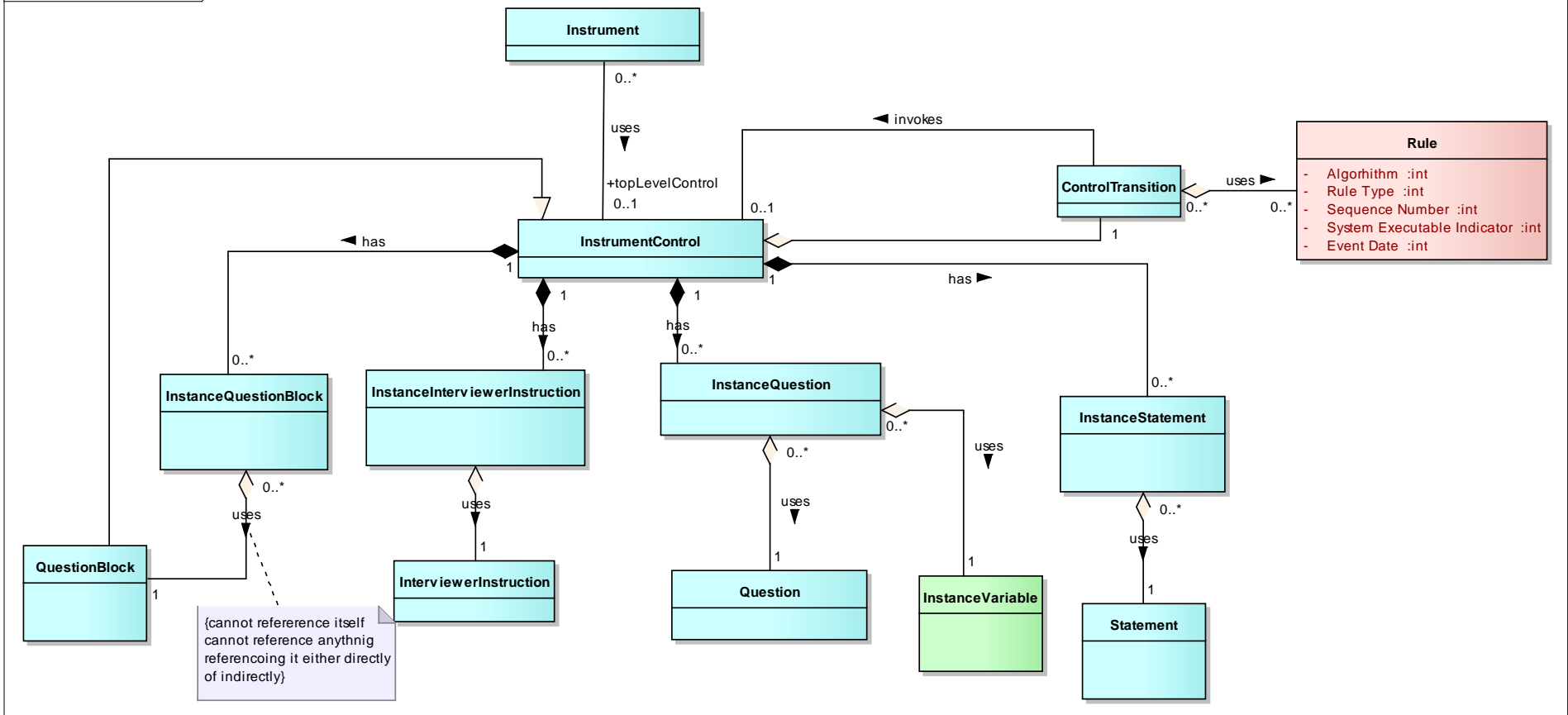
- DDI – Lifecycle has a very complete description of a questionnaire/instrument
 - Includes the mode and specifics of the instrument
 - Includes the questions, statements, and instructions used
 - Includes the flow logic of the questionnaire
 - Can have multiple-question “blocks”
- GSIM does the same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



GSIM

GSIM Survey Instrument

class D - Instrument Control

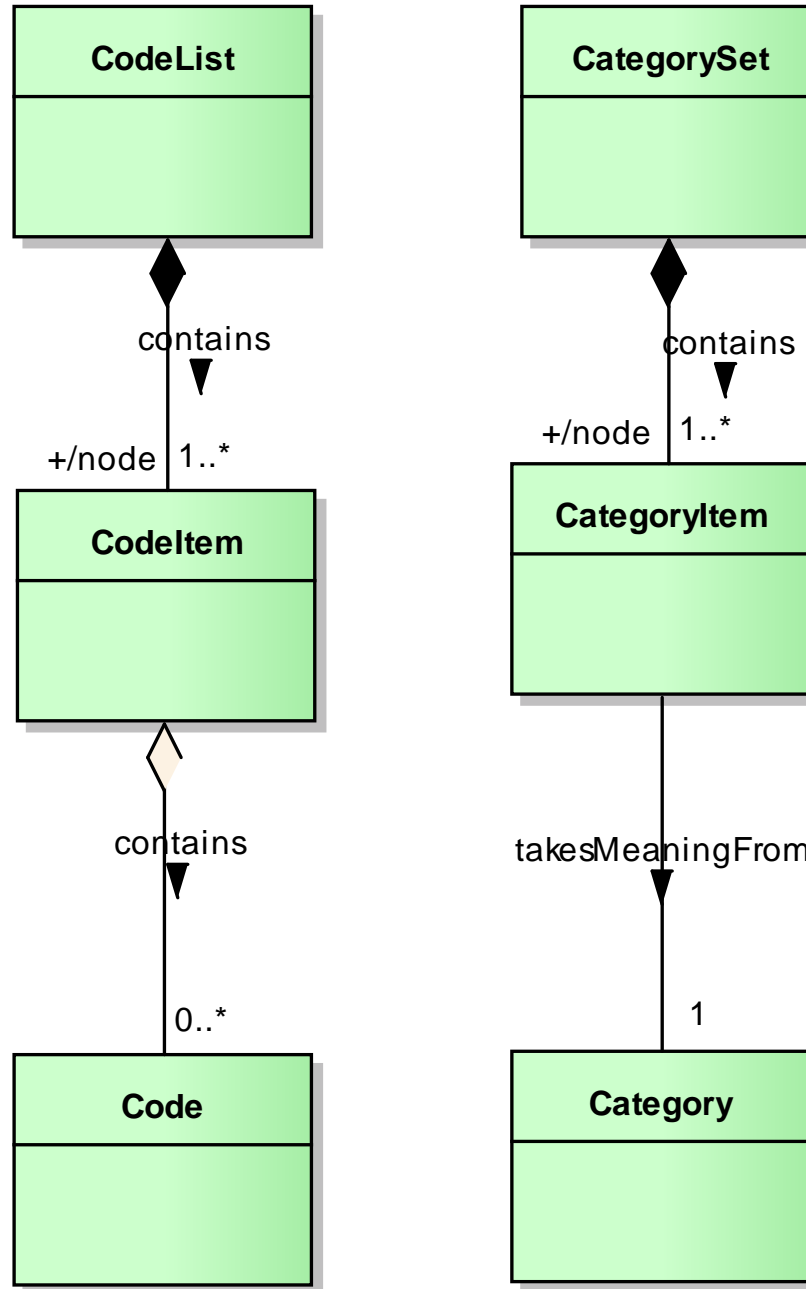


Classifications, Codelists, Categories and Concepts

- 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)

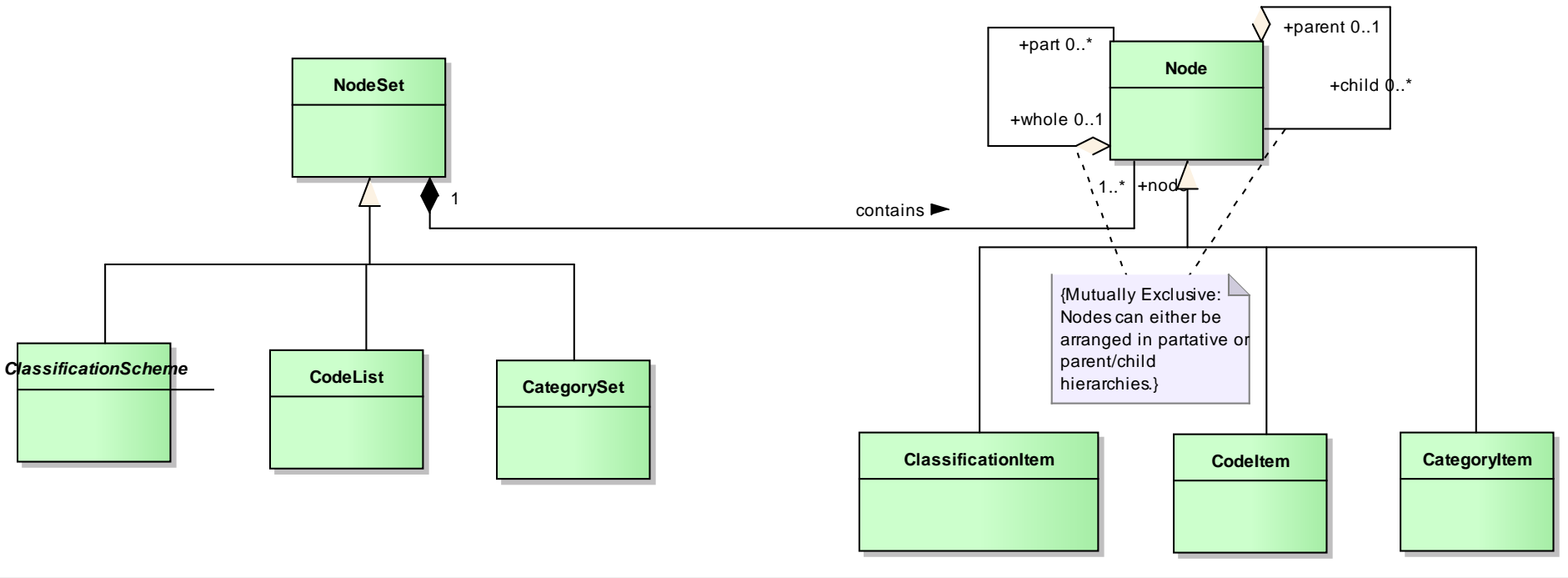


class C - Category-Code



Nodes and Node-Sets

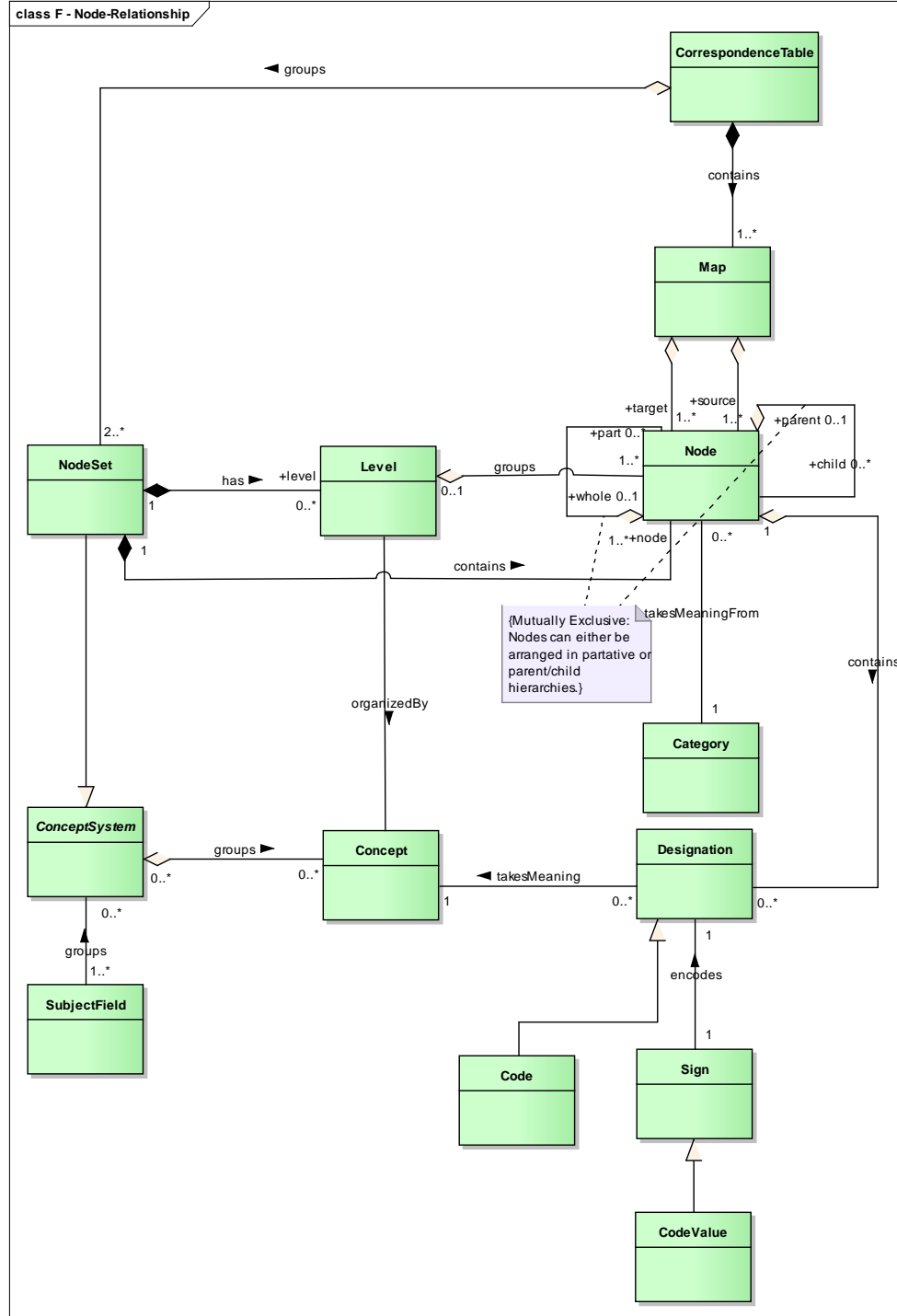
class E - Node-Inheritance



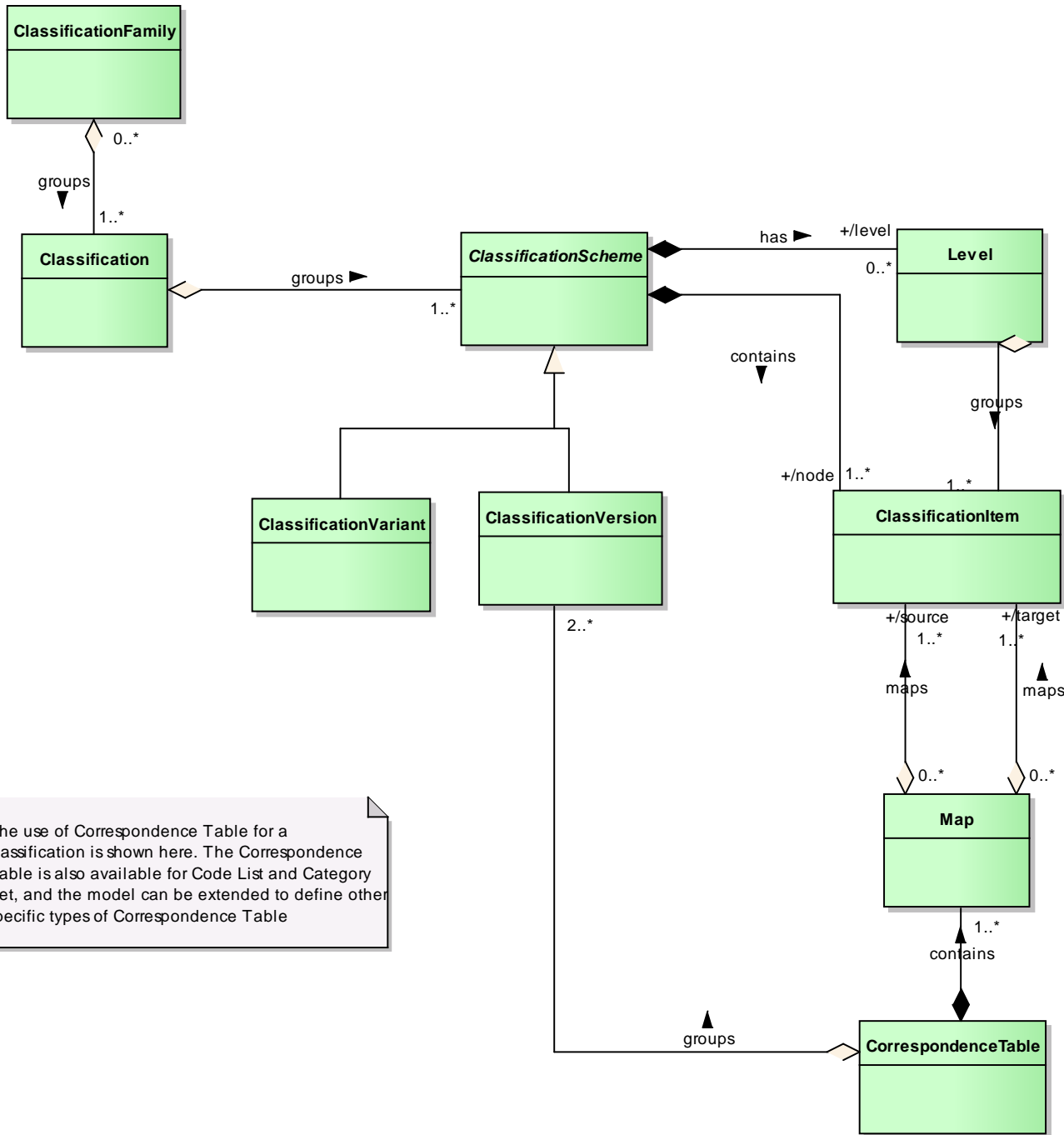
GSIM



GSIM



class B - 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



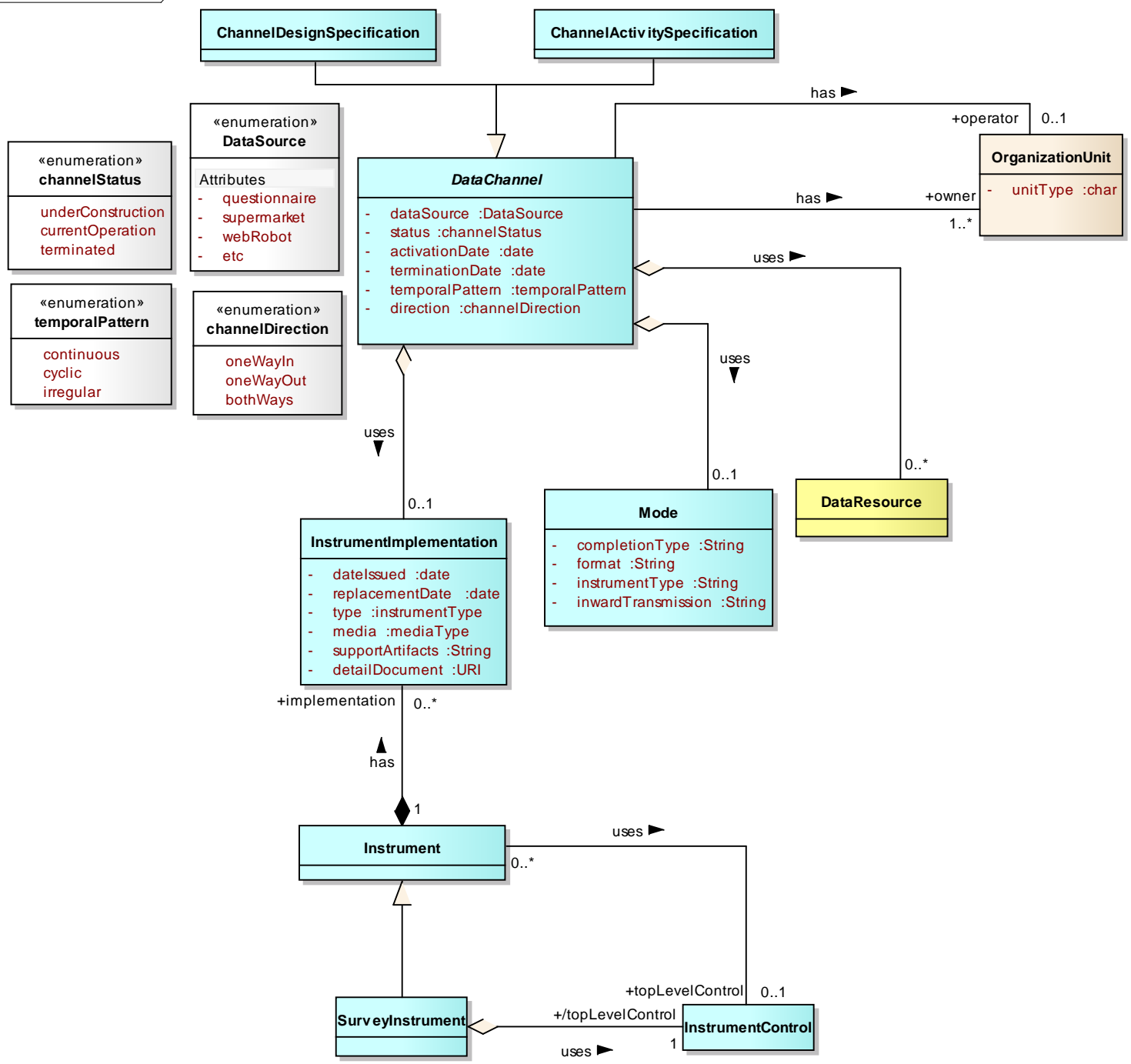
Events and Processing

- 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



GSIM

class C - Data-Channel



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?

