



EUROPEAN CENTRAL BANK

EUROSYSTEM

Value of a Logical Data Model

Why we model (and
how)

Presentation for Eurofiling 2023

June 2023

Arjan Bos
Co-ordinator BIRD Work Stream on Data Modelling



Outline of the presentation

- 1 What is a logical data model
- 2 Logical data models are everywhere
- 3 Intro into Banks' Integrated Reporting Dictionary (BIRD)
- 4 Full-scale data approach
- 5 Some design principles
- 6 Developing BIRD is a community process
- 7 Questions?

What is a logical data model

First: What is a Model

- Abstract representation of reality
- Only tackle that what is in scope
- Leave everything else out

Logical Data Model

- Abstract representation of data
- No content, only structure
- Technology independent
- Deals with soundness, consistency, completeness, derivability of data
- Mathematical correct representation of the requirements
- Based on Set theory and Predicate logic.

Logical data models are everywhere

Store data? We need structure

- Then there is a data model
- Every app on your phone stores data and has a data model
- Even every excel table has a data model

Send data? We need structure

- Reported data has a data model
- Each data interface has a data model
- Every API has a data model

Analyse data? We need structure

- Each type of analysis and tool has its own data model
- Analysis has a different data model than interfaces

Multiple reporting frameworks in one logical data model

Multiple reporting frameworks

- Different slices of the same data
- Same concepts repeated in multiple reports
- BIRD puts them into one non-regulatory model

For Banks – By Banks

- Banks need to report
- BIRD helps banks to harmonise their reporting into one model
- ESCB, Consultancy companies and software vendors also participate

Full Banks' Integrated Reporting Dictionary (BIRD) Logical Data Model (LDM) is extensive

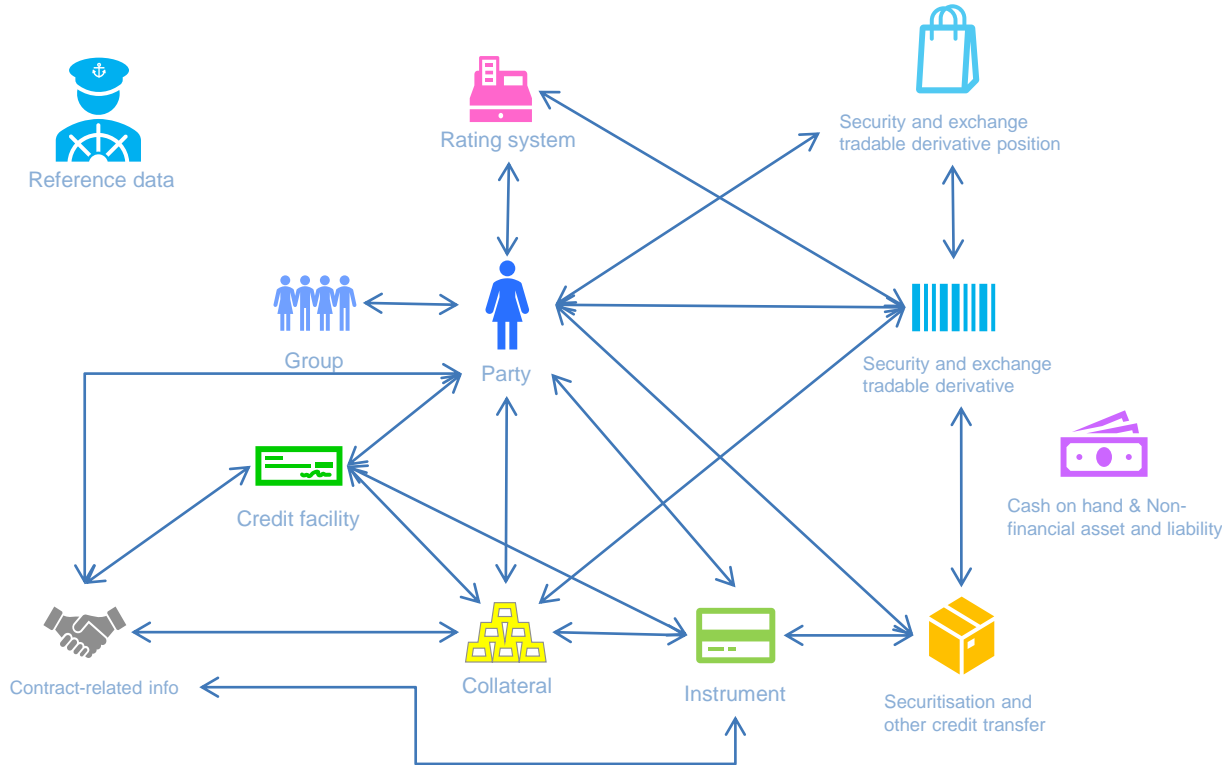


Instead of thousands of pages and 150+ templates, we have:

- 500+ entity types
 - Of which 360 subtypes
- 1327 attributes
 - Of which 861 foreign keys
 - 127 primary keys
 - 608 “observations”

The foreign keys and primary keys are data quality checks, or validations, that are “baked into” the structure of the model.

The concepts of the BIRD LDM are straightforward



Modelled frameworks

- AnaCredit (100%)
- FINREP (85%)*
- Asset Encumbrance (90%)*
- Security Holdings Statistics (95%)

*FINREP and Asset Encumbrance are subject to verification from supervisors

Unmodelled frameworks are also partially in

- Monetary Financial Institutions Interest Rate (60%)
- Resolution planning (40%)
- Balance Sheet Items (35%)

BaFin Feasibility Study – granular to aggregates

Example FinRep F5 Template from granular data based on the BIRD LDM

**This shows capabilities of BIRD. Supervisors still need to verify results for fit for purpose*

3 attributes on granular Instrument and 2 relationship types to granular Collateral with 1 more attribute

1 statutory metric (gross carrying amount) on granular Instrument to be aggregated using the common rule set

			Carrying amount						
			Central banks	General governments	Credit institutions	Other financial corporations	Non-financial corporations	Households	
			0005	0010	0020	0030	0040	0050	0060
0009 By product	On demand [call] and short notice [current account]	0010							
	Credit card debt	0020							
	Trade receivables	0030							
	Finance leases	0040							
	Reverse repurchase loans	0050							
	Other term loans	0060							
	Advances that are not loans	0070							
	LOANS AND ADVANCES	0080	42						
0089 By collateral	of which: loans collateralized by immovable property	0090							
	of which: other collateralized loans	0100							
0109 By purpose	of which: credit for consumption	0110							
	of which: lending for house purchase	0120							
0129 By subordination	of which: project finance loans	0130							
			Sum						

89 data points can be derived by pivoting granular data

Potential **Anchor Value** to validate plausibility of all other sub-aggregates and ensure responsibility

1 relationship type from party to instrument

Granular data – Giant game of telephone

- Within an information chain

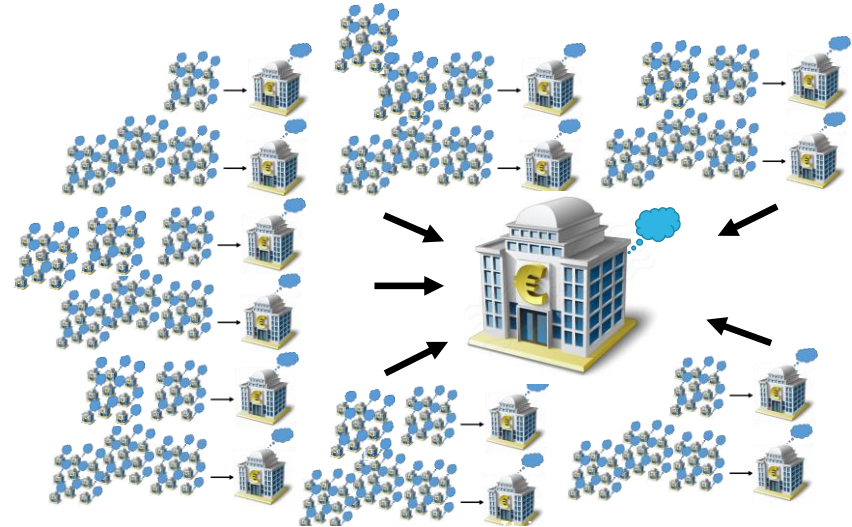
 - From front-office input to ECB report is a long way with many hand-offs.
 - Each hand-off requires interpretation

- Across information chains

 - Can we compare instruments from multiple reporting?

- Harmonizing is a must

 - Definitions must be clear
 - Automation removes interpretation where possible



Full Scale Data Approach

Source: Ronald Damhof – Full Scale Data Architects Meetup 09-2019

1. Start from Law / Regulation / Policy (FINREP/AnaCredit/...)

2. Create reference with terms, definitions and term-relations (Ontology)

3. Create logical data model

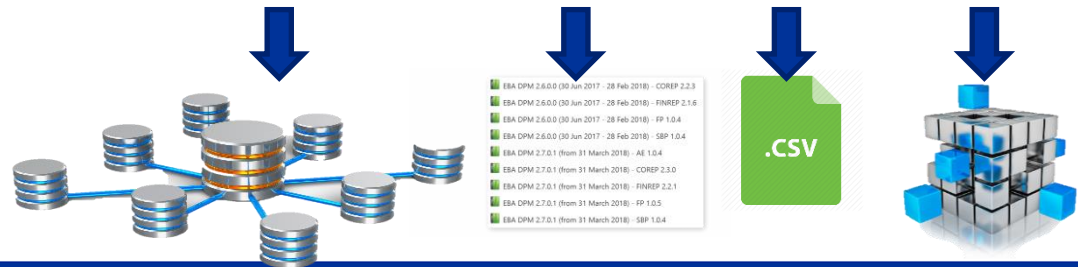
4. Create implementations (interfaces, data bases at banks)



definition ▶ noun 1 a statement of the exact meaning of a word, especially in a dictionary. *a dictionary definition of the verb.* ■ an exact statement or description of the nature, scope, or meaning of something; our

definition ▶ noun 1 a statement of the exact meaning of a word, especially in a dictionary. *a dictionary definition of the verb.* ■ an exact statement or description of the nature, scope, or meaning of something; our

definition ▶ noun 1 a statement of the exact meaning of a word, especially in a dictionary. *a dictionary definition of the verb.* ■ an exact statement or description of the nature, scope, or meaning of something; our



Define once



Report once



Separate concerns



Model only required data



Least granular option that covers all requirements



- 44 members help data modelling
- 100+ members help the process
- Members come from Commercial banks, Central banks, Consultancy companies and Software vendors
- See <https://bird.ecb.europa.eu> for more information and *all* the content of the BIRD model

Data models reflect the real world



Yulan Lin
@y3l2n

 Follow

"If you want your data model to be simple, go out and make the world simple, and then come back to me." @datachick
[#JSCDataScienceDay](#)

RETWEETS 10 LIKES 17



10:20 PM - 26 Apr 2017

Questions?



Multiple Frameworks to LDM – Separation of Concerns

Regulation / Framework

- Regulation describes in words what is to be regulated and reported
- Regulatory framework is a container that holds the rules and reporting requirements described in the regulation

Taxonomy

- Each Framework is analysed and decomposed into a Taxonomy
- A taxonomy is a systematic arrangement of terms from a framework showing the relations between the terms.

Ontology

- Multiple taxonomies create a common classification. The taxonomies are semantically integrated into one ontology.
- An ontology is a systematic arrangement of the important categories of terms from taxonomies, showing the relations between those terms.

Logical Data Model

- Relations in the Ontology structure the data in an LDM where words change into model
- Terms that are used in other terms become entity types; Terms that use other terms become attributes
- The Ontology is made mathematically correct into an LDM by applying normalization
- Collection of data does need more stringent rules than compilation of statistics, thus they require a stricter LDM
- An LDM is a detailed representation of data requirements and is still independent of any implementation constraints.

Technical implementation

- The LDM is transformed into suitable technical representation (Input Layer; Data Base; SDMX; Datamart; Cubes; Templates; CSV;...) – dependent on the technical considerations

Please note that each step analyses and solves additional concerns. This separation focusses the analysis. The solutions of the earlier steps remain valid!