

Data Modelling in the EBA context. Introductory explanations.

On 17 February 2011ⁱ, the European Central Bank (ECB) and the Committee of European Banking Supervisors (CEBS) published a new classification system aimed at reducing over time the reporting burden placed on entities that are required to deliver data to the Eurosystem as well as to supervisory authorities. It is a bridging manual linking the ECB's monetary and financial statistics requirements with the supervisory reporting templates (FINREP and COREP) developed by CEBS, and a database that will help reporting agents and other users identify the similarities and differences between data for ECB statistical purposes and data for financial supervisors in the context of the CEBS framework.

At a much more modest level, but with exactly the same aim, the regulators and IT experts in the EBA are working on Data Models. Reconciliation at the accounting level is always complex, as the scope of FINREP (financial), COREP (risks) and ECB (statistics) are different. As an example, FINREP is to be reported according to IFRS scope and/or Capital Requirements Directive (CRD).

I. Common Breakdowns

However, an IT-like harmonisation approach looks more feasible. The idea is to review the data models of COREP, FINREP and ECB Statistics frameworks to discover common elements. Some breakdowns are evidently common, as currencies or countries. Others are to be investigated in full. At the end of the exercise, a common set of breakdowns is shown to be applicable to all the frameworks, but others breakdowns are specific. This is the current draft:

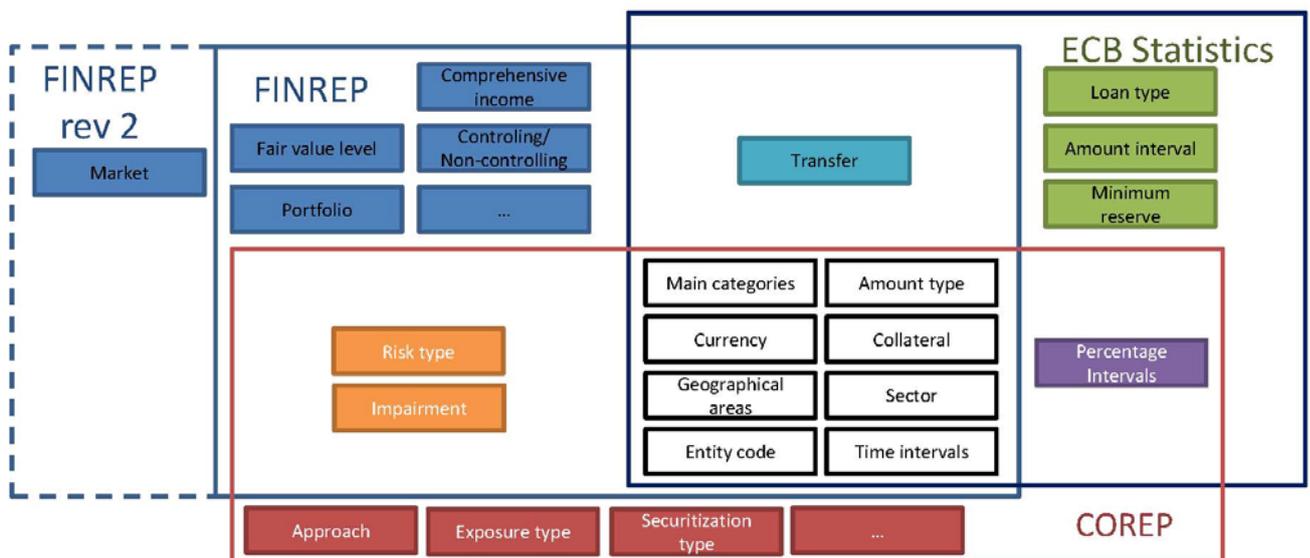


Figure 1: COREP, FINREP and ECB statistics share 8 breakdowns (Main categories, Amount type, Currency, Collateral, Geography, Sector, Entity code, Time intervals). COREP/FINREP also share Risk type and Impairment. COREP/ECB also share Percentage Intervals. FINREP/ECB also share Transfer. Solvency II has not been studied yet.

This data model reconciliation would be very useful in designing databases, both in reporting institutions as well as NSAs. Even when the elemental items have different scopes, the design of the data warehouses would be similar, and easier to link with existing Information Systems. There are economies of scale, more accuracy in the reported information, and reduction in the regulatory burden.

The problem in defining the formal data model is that the IT experts are basically unable to fully understand the regulatory frameworks, while the regulators are usually not trained in formal data modelling. Most of the effort in the EBA context has been invested in getting to work together regulation and IT experts on data modelling.

II. Reporting the Full Context of Breakdowns:

When the regulators design forms, they use bi-dimensional Excel spreadsheets, in which a large number of breakdowns are collapsed, and therefore hidden. What actually is the full set of breakdowns applicable to each cell is, for sure, in the brain of the regulation experts, but not always clear on the paper for everybody.

A strong preference has been voiced by reporting entities in recommending that the same cell of the same template should have identical interpretation, definition and use, and therefore the same value for same circumstances, irrespective of the NSA. Clarifying the exact content of each cell with a full description of the context of breakdowns and scope is therefore of the utmost importance.

However, the bi-dimensional-only Excel cell representation cannot capture easily the full set of breakdowns. See an example obtained with an Excel Add-in tool:

The screenshot shows an Excel spreadsheet with a table in the bottom left corner. The table has the following content:

Debt securities	1
Central banks	2
General governments	3
Credit institutions	4
Other financial corporations	5
Corporates	6
Loans	7
Central banks	8

The 'Coordinate Values' box on the right contains the following information:

- ZAxis coordinate values:**
 - Portfolio : Measured at amortised cost
 - Reporting scope : CRD consolidation
- Row coordinate values:**
 - Category of assets : Debt securities held
 - Counterparty sector : General governments
- Column coordinate values:**
 - Abstract : Assets
 - Amount type : Unimpaired
 - Delinquency : > 90 days ≤ 180days
- Cell code: 6

Figure 2: Reading the reporting template only, the full context applicable to the single cell is not evident at all. However, in the rectangle, the coordinate values box shows exactly the full context of coordinates or breakdowns applicable to the cell.

This Excel Add-in is proof of concept Open Source tool, freely availableⁱⁱ for supervisors and for the public.

The Excel Add-in tool was originally developed as a pure auxiliary help, but proved to be very fruitful when creating a data model for ECB statisticsⁱⁱⁱ. The development of the tool is aligned with the latest proposals on ESCB shared software development.

The Excel spreadsheet containing the data model is basically a data dictionary for (in this case) ECB BSI-MIR statistical reporting, presenting a full description of the breakdowns and basic scope applicable to each cell. This Excel representation is purely conventional as the data model is basically an Excel spreadsheet with some kind of data dictionary. The rationale of this data model is to represent the exact set of disaggregations applicable to each cell in the regulatory framework.

	A	B	C	D	E	AR	CZ	EM	FI	GI	GQ	GV	HD	HG	HO	HT	IA
1						D	D	D	D	D	D	D	D	D	D	D	D
2	#	Table	Cell code	Base item label	Base item code	Counterparty	Geographical area	Currency	Time interval	Amount type	Amount interval	Related to minimum reserve	Entity code	Loan type by purpose	Transfer	Collateral Guarantee	Percentage interval
2474	2470	EST-2	0118	Business volume	md34	x	x	x	x	x	x	-	-	-	-	-	-
2475	2471	EST-2	0119	Business volume	md34	x	x	x	x	x	x	-	-	-	-	-	-
2476	2472	EST-2	0120	Business volume	md34	x	x	x	x	x	x	-	-	-	-	-	-
2477	2473	EST-2	0163	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	x	-	-	-
2478	2474	EST-2	0168	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	x	-	-	-
2479	2475	EST-2	0162	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	x	-	-	-
2480	2476	EST-2	0172	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	-	-	-	-
2481	2477	EST-2	0180	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	-	-	-	-
2482	2478	EST-2	0188	Annual percentage rate of charge	pd33	x	x	x	x	x	-	-	-	-	-	-	-
2483	2479	EST-2	0229	Narrowly defined effective rate of	pd35	x	x	x	x	x	-	-	-	x	-	x	x
2484	2480	EST-2	0230	Narrowly defined effective rate of	pd35	x	x	x	x	x	-	-	-	x	-	x	x
2485	2481	EST-2	0231	Narrowly defined effective rate of	pd35	x	x	x	x	x	-	-	-	x	-	x	x

Figure 3: Data Model for ECB BSI MRI statistic. Each entry (Column A) links a cell (Columns B and C) with its characterization of "Base Item Code" (Column D) and up to 12 applicable breakdowns (Columns AR to IA).

Another more basic exercise, using a basic Excel artefact, with no need for any specific tool, has also been made public on COREP 2010-11-10 as an intermediate release.^{iv} In this exercise, the original templates, showing supervisory information (left side: "credit counterparty" or "total exposures") are also prepared to show IT information (right side: "SCC" or "254" linking with the data dictionary), while maintaining exactly the original format.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
1	CR SA Total	CREDIT AND COUNTERPARTY CREDIT AND DILUTION RISK											CR SA Total	CREDIT AND COUNTERPARTY CREDIT AND DILUTION RISK						
2		Original exposure pre conversion																		
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11	Total exposures																			
12																				
13	On-balance sheet items																			
14																				
15	Off-balance sheet items																			
16																				

Figure 4: The original content of the cells is replaced manually by links to the Data Model, including a button selecting "supervisory information" or "IT information".

III. Summary

As it has been explained, to create a data model is mainly an issue for regulation experts. Of course, IT experts may help by providing tools and expertise. But the semantic definition of regulatory frameworks is definitely a task for the regulators.

Once the data model is ready, the work of IT experts starts on a sound basis. Note that the acronym XBRL has not yet been used, as the the Data Model is IT generic.

All these development^v goes back to February 2005, The kick-off meeting in February 2005, under the guidance of Pierre-Yves Thoraval as COREP Chair, Adrian Abbot of the UK-Financial Services Authority and Frédéric Marié of the Banque de France, along with international XBRL experts, such as Charlie Hoffman and Walter Hamscher. It agreed on the use of XBRL and discussed how to apply XBRL to support COREP. Missing required functionalities were analysed, especially, more dimensional features. A solution path was found, CEBS XBRL Network started operations and the standard was accordingly extended by the XBRL International Consortium. Academic research^{vi} supported the way forward.

During the initial COREP and FINREP development of XBRL taxonomies, it was never evident if breakdowns in different tables were equivalent or not. A first step was the concept of Data Matrix, showing the breakdowns applicable to the different tables. An excellent refinement came when the Bank of Italy contributed the much more robust Matrix Schemas (Excel files with breakdowns in a structured form) that have since been routinely published with each XBRL taxonomy release, thus providing an important step for quality control. Thanks to this contribution to the process, the Data Model much facilitates the creation of the corresponding XBRL taxonomy.

The available experience on XBRL formulas applied to XBRL taxonomies designed according to the Data Model is that they are more regular, efficient, and easier to review.

Linking bi-dimensional *human-oriented* representation with formal *computer-oriented* data models has been an open issue since the 70's, when the relational *computer-oriented* Data Base model was introduced, competing with the initial, more intuitive and *human-oriented* hierarchical model^{vii}. But that is another story.

ⁱ See the press release about the results of the Joint Expert Group on Reconciliation at Joint Expert Group on Reconciliation of credit institutions' statistical and supervisory reporting requirements (JEGR), established jointly by the European System of Central Banks' (ESCB) Statistics Committee and Banking Supervision Committee together with CEBS, at <http://www.eba.europa.eu/News--Communications/Year/2010/New-classification-system-between-the-reporting-fr.aspx>

ⁱⁱ See the Open Source tool at http://www.openfiling.info/?page_id=67 OpenFiling is a parallel initiative to EuroFiling, focused in Open Source for filing.

ⁱⁱⁱ See the ECB BSI MIR statistics taxonomy at <http://www.eurofiling.info/bsi-mirTaxonomies/taxonomy.html>

^{iv} See the complete Excel and instructions at <http://www.eurofiling.info/corepTaxonomy/taxonomy2012.html>

^v Boixo, I. y Schmehl, K. 2010. "Collaborative Development of IT Supervisory Frameworks", XBRL Global, 1, 2, 37-44 <http://www.eurofiling.info/documents/XBRLglobalVol2SB.pdf>

^{vi} Boixo, I. y Flores, F. 2005. "New Technical and Normative Challenges for XBRL: Multidimensionality in the COREP Taxonomy", The International Journal of Digital Accounting Research, 5, 2005, 79-104.

^{vii} See an explanation about hierarchical and relational at http://en.wikipedia.org/wiki/Database_model