Data Point Modelling and XBRL

Part of the same supervisory reporting Information Supply Chain or competitive formats?

XBRL and DPM have coexisted in both bank and insurance supervisory reporting in Europe for over 10 years. Recent proposed changes are however set to generate conflict between DPM and XBRL. Under the name 'DPM 2.0', EBA and EIOPA are proposing to create a parallel definition language to the XBRL rules, plus for the first time, the DPM notation is being explicitly used in the XBRL taxonomies as part of an update to use the new XBRL-CSV format. The formalisation of the new DPM assertions creates a 'single source of the truth' issue. The use of the DPM-ID, an internal database identifier, with no grouping or semantic information, offers little value to NCAs or reporting entities in terms of optimising the generation or validation of XBRL reports. The result is expected to be increased costs and confusion within the NCAs and the thousands of banks and insurers that have heavily invested in XBRL.

DPM and XBRL: An Overview

XBRL (eXtensible Business Reporting Language) has become the standard for the exchange of complex business data, primarily for regulatory reporting, i.e., where a market regulator like the European Banking Authority (EBA) collects data for the European Capital Requirements Directive (CRD) from all banks in Europe. However, most people will be unfamiliar with Data Point Modelling (DPM). ISO describes the DPM model as '…a logical model oriented to data with business meaning and based on the Multidimensional Data Model (MDM) and in which the information requirements of European supervisors are reflected'.

For banks, CRD defines a set of predefined data reporting requirements commonly represented in tables. These are initially described in a spreadsheet so that business users can understand and review them. However, spreadsheets represent a very poor way to describe complex and large datasets and then resolve them into a consistent model, hence DPM notation was invented.

The DPM methodology is used to identify the individual data cells in spreadsheet templates, and then to deduplicate and model the dataset. The DPM output is then converted into a 'standards-based', XBRL Taxonomy for collecting data from numerous banks, each with its own different source systems and local storage mechanisms. The XBRL reports are collected and validated by the local country National Competent Authorities (NCAs) using certified XBRL validation engines and the data is loaded into analysis systems. The same reports are also shared with the EBA and the European Central Bank (ECB) to provide a common dataset for banking supervision and monitoring.

XBRL and DPM have been used for their specific roles in the EBA for over 10 years and delivered high-quality data for analysis. A similar system has been deployed by the European Insurance and Occupational Pensions Authority (EIOPA) for the supervision of the insurance market. However, recently proposed changes included under the name 'DPM Refit', now dubbed 'DPM 2.0', are the formalisation of a parallel set of definitions of the rules that are meant to check the quality of the data and for the first time, the EBA is proposing to

explicitly expose its DPM internal identifier in the new XBRL-CSV filing format. EIOPA has yet to decide which XBRL-CSV format they will use for the Solvency and Pension Fund collection system.

Why is the EBA mixing up the roles? What are the effects on the reporting framework? These questions are the focus of this article. Future articles will discuss how the XBRL specifications can be improved to fully meet the data modelling requirements of reporting systems like the EBA's and EIOPA's in XBRL.

A Brief History

The use of XBRL has become widespread among European financial market supervisors, starting with the EBA's common prudential reporting framework (Basel/CRD) and swiftly followed by EIOPA's Solvency II taxonomy. Other European financial supervisors, including the European Central Bank (ECB), and the Single Resolution Board (SRB) have both referenced the EBA CRD taxonomy in their own XBRL data collection frameworks, and local NCAs have successfully developed extensions to the common data dictionary, table definitions and quality rules to meet their own local regulatory reporting requirements. Similar approaches have been developed in other regions for financial market supervision.

The first data point models (DPM) began to be developed from the Matrix Schema in 2009, used by the Banca d'Italia and then further developed by several European supervisors, such as the Banco de España, to describe metadata without redundancy, in a coherent and unambiguous way. The primary objective was to develop a model from the Excel templates used by business analysts at the European Supervisory Authorities (ESAs) that could faithfully generate an XBRL taxonomy.

In August 2021, the DPM methodology was recognized as an ISO standard (ISO 5116-1:2021). One of the principles of the DPM methodology is that it assigns a column code and row code to each 'cell' in a Table, producing x and y coordinates. A Table may be made up of multiple sheets, effectively giving us x, y, and z coordinates for each cell. The Table is given a formal name, which gives us the combination that defines the specific cell or 'data item':

> Table name, sheet name, column id, row id.

These cell references can then be modelled to remove duplicates, determine members of a dimension (a breakdown list), etc. Once completed you have the definition of a set of unique 'data points'.

A specific DPM notation was created to enable business users to define the various accuracy and consistency rules that could be applied to the data. This follows the simple format of Excel functions and the DPM references – Table, sheet, row, and column codes.

> [F 02.00 (c010)] {r630} = +{r610} - {r620}

This formula basically states that in the sheet identified as F.02.00, the value in column 10, row 630 should be equal to the sum of the values in column 10, row 610 and row 620.

When the definitions are stored in a database, each data point is given a simple, unique id, known as the 'DPM-ID'. As far as we are aware, the DPM-ID unusually has no meaning or structure other than to act as an invariable identifier of a data point and as a unique key in the database, i.e., it is a total abstraction. It is without any semantic meaning or grouping, unlike XBRL concepts and contexts.

To enable the DPM model to be readily described in XBRL, XBRL International (XII) with the help of local NCA experts developed a new specification for describing table layouts, the XBRL Table Linkbase specification. It fully describes how a table of rows and columns should be rendered and split into sheets. From these specifications, software can automatically prepare templates mapped to the XBRL Taxonomy. Allowing users to load the data and then convert it to XBRL or allow a supervisor to render the data back into an Excel Template, potentially highlighting any validation errors for business users.

The DPM codes are currently not included in the XBRL data model but are referenced via labels in the Table Linkbase definitions (albeit, in a non-standard way). However, many vendors include the sheet names, row, and column codes in their templates to help users, who understand the DPM coordinates and can relate them to the original business templates and business concepts. This also enables software to transform data from DPM to XBRL, e.g., a user can supply a CSV that contains the Table id, (sheet name), column id, and row id, plus the value they were submitting, and the software will convert this to XBRL.

The DPM Refit and XBRL OIM

In 2019, the EBA started the Task Force for Evolving the Reporting Format (TFERF), in cooperation with EIOPA, the NCAs, and with the active collaboration of XBRL International. It wanted to address the problems it saw of high complexity and 'poor' performance of the XBRL-based reporting framework.

Three areas of concern identified by the TFERF were:
- The XBRL reporting format is unnecessarily verbose, generating large files.
- The XBRL validation rules are complicated, and existing validation engines cannot cope with large report files.
- The XBRL taxonomies are complex and difficult to maintain.

In 2020, the XBRL Standards Board (XSB) delivered a major update to XBRL that is directly aimed at reducing the size of XBRL files and at the 'opening up' of XBRL to submissions in other file formats.
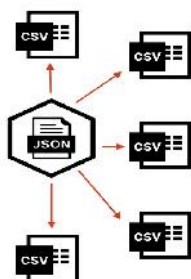
The Open Information Model (OIM) specifications included two new formats in which XBRL data could be provided, xBRL-CSV and xBRL-JSON. Tests showed that XBRL-CSV would greatly reduce the size of the CRD report files and would also enable the optimisation of data quality checks on record-based (multi-fact row) data.

In addition, the specification enables the submitter of the report to utilise the CSV file format of their choice, i.e., to suit the needs of their own source systems. They would be responsible for linking their format to the XBRL terms in the associated JSON file.

The issues identified by the EBA in XBRL Formulas are in part due to being defined in XML which is verbose and complex and, for many, difficult to develop and maintain. However, it also reflects the complexity of the domain and the model. This is further compounded by the low level of the DPM

business rules which are specified in terms of rows and columns at the table level, similar to the approach of spreadsheet formula. Consequently, the XBRL rules that are generated from the DPM definitions developed by business users are at an 'atomic level' and do not take advantage of any XBRL semantic relationships.

For example, in the XBRL model, a relationship between a total and its contributors is universal, and it applies across the whole taxonomy regardless of tables. Whereas a DPM rule will, almost always, specify the individual data points required per row or column within a given table. This is usually then replicated for every other row or column. There is, therefore, a trade-off between numerous, user-generated, simple table-specific rules versus fewer generic, more complex XBRL Formulas which use the underlying model to understand what needs to be calculated and therefore can be optimised by XBRL processors.

The concern about the processing time of large reports is mostly accredited to the 'Open' tables, in particular, those where there is an unlimited number of rows and the data is organised in a record format, i.e., multiple related facts per row. Performance on most tables, which contain relatively few, aggregate data points, has always been good for most XBRL processors. However, when you combine large datasets, expressed in XBRL-XML as a single fact per row, with a large number of low-level data quality checks then you do see processing times increase, however, analysis shows that they are being processed quite efficiently once the data is loaded into an optimised XBRL processor.

Find out more about OIM and XBRL Processing:

⟩ UBPartner XPE Certified for New XBRL OIM

⟩ XBRL Processing Engine

Unfortunately, the EBA's proposed approach for the collection of CRD data in xBRL-CSV does not help XBRL conversion or validation. Firstly, the EBA has decided against allowing the submitter to define the format of the CSV, but instead for the first time fully introduced DPM notation directly into the XBRL model by selecting the following fixed format:

DPM_ID, Value, Unit

According to the EBA, using the DPM-ID in the xBRL-CSV layout is a 'simple pivot' to the XBRL Table structures. However, if you are using XBRL as an open standards method to collect data, you would want to ensure that the significant changes that the NCAs and reporting entities are obliged to make to their systems to support xBRL-CSV brings tangible improvements or simplification to their reporting processes. The reality is that from a local NCAs and submitters' point of view, using a DPM-ID offers no advantages. Instead, it restricts validation performance and makes conversion between other formats more difficult.

For example, the EBA could have:

-) Selected a grid layout for the XBRL-CSV, as CRD is based upon tables, to help firms prepare their reports and relate their data directly to the spreadsheet structure that the NCAs and submitters understand. This format is ideally suited to 'aggregated' data.
-) If the tables that are 'record-based' were laid out according to their 'record-format', then the data could be read in as rows and some assertions processed inline, significantly improving the performance of XBRL formulas on large open tables, which per above has been analysed as causing the most issues.
-) Alternatively, if the EBA wants to keep a single fact per row (like xBRL-XML), then most users have become familiar with the table, sheet, column, and row ids. These references are much easier to locate and understand than the arbitrary DPM ID, so the XBRL-CSV could be based on these 4 reference ids. Most of the reporting entities' source systems are dimensional and can be readily mapped to such layouts.

It would be a relatively simple transformation for EBA's systems to process the above formats and help the NCAs and reporting entities. In part the model chosen, results from the fact that the EBA itself does not intend to use an XBRL processor to validate data but is building a proprietary tool to work directly on the CSV data loaded into its database. To support the DPM 2.0 model, the EBA is proposing an update to the existing expressions, originally termed 'DPM-ML', which could operate efficiently on the xBRL-CSV format. The EBA expects the DPM-ML definitions to be used to generate the XBRL Formula rules automatically.

The major concern here is that, even though the XBRL taxonomy is derived directly from the DPM model, there will inevitably be areas where they do not align, due to differences in what the models can express or simple taxonomy definition tooling errors. This creates the 'Single source of truth' issue.
-) XBRL is the source of truth for entities external to the EBA, i.e., the NCAs and the submitting entities (filers).
-) Whereas DPM is the source of truth for the EBA.

Today the position is simple, the XBRL Taxonomy is the single source of truth. The XBRL conformance suite supports this and is continually expanded when a lack of clarity or inconsistency is found. Given the complexity of the model and tooling, it is to be expected that there will be issues in the conversion between DPM-ML and XBRL formulas. The resolution of these, each time the models are updated, will cause costly time and effort to investigate and resolve by the EBA, NCAs, submitters or vendors.

The other issue, which is ignored in the decision to use the DPM-ID, is the human factor. Analysts and reporting experts at both NCAs and reporting entities have no idea what a DPM-ID is, per above it is just a technical code, e.g., "dp1234". They will need significant help in understanding what they need to report as well as in identifying and resolving issues. The costs of the reengineering and additional support are hard to estimate but are potentially large.

XBRL Taxonomy complexity versus DPM complexity is really in the eye of the beholder. However, there are probably fewer than 10 DPM experts in the world, whereas there is a whole XBRL community. UBPartner believes that a long-term look at improving the modelling capabilities of XBRL could reduce and simplify the development of XBRL taxonomies and the coding of quality business

rules. This would include items such as linking the independent specifications, such as the Table Linkbase and Formulas to provide table-based formula definitions; the versioning of elements to provide a lineage for taxonomy elements, enabling a full-blown master data management system for an XBRL taxonomy. These enhancements would enable vendors to provide tools to model complex frameworks like CRD4 and Solvency II in XBRL rather than in DPM and then converting. These ways of improving XBRL will be covered in another article, but for now, both EBA and EIOPA are facing a major turning point and set of decisions on which direction they head.

Moving Forward

Where does that leave the different players in this reporting supply chain:

- The EBA has used the DPM successfully in its internal systems for some time and will probably continue to do so. The decision to use a technical DPM identifier as the basis for the future XBRL-CSV collection system makes it simple to load into their DPM database, however, at what cost to the XBRL collection system ? The low-level DPM quality checks could be parallelised and provide an efficient internal EBA processing system, but again the XBRL Formula generated from them is inefficient. In earlier 'DPM Refit' presentations, there was also talk of providing the EBA DPM software, database, etc, as possibly 'open source' to NCAs. Such an initiative would create a single architecture across regulators, but the cost and maintenance would be significant o replace their existing XBRL systems. In addition, market supervisors have not shown the ability to fund or support the continuing maintenance of such initiatives in the past.
- Local country NCA's are in a difficult position, they can continue to use the XBRL software that they and their regulated entities have invested in or move to an EBA model of processing and push the XBRL validation requirement to the submitting entities. In addition, what does this mean for the NCAs that have built XBRL extensions on the CRD or S2 common dictionaries and models? Do they continue with these reporting frameworks in their current form or not? Either way, they are still at the centre of a major transformation between DPM and XBRL, which nobody would envy them for.
- Reporting entities remain at the 'mercy' of these decisions, being told to upgrade their systems to support the new XBRL-CSV format with absolutely no benefit to them. While their own source systems tend to be dimensional, and structured in terms that make business sense. Software developers will have to work extremely hard to hide the 'joins' between DPM and XBRL models where they do not fit very well.

The XBRL community has responded to the DPM 2.0 proposal, to help address some of the short-term issues:

- The XBRL Standards Board (XSB) has announced "Formula 2.0" which plans to complete the ongoing work on the OIM Rules supporting the adoption of xBRL-CSV (and xBRL-JSON), by removing the XPath requirements. It also intends to rapidly formalise the use of XF text-based Formula, which is a short-hand for XBRL Formula and can be freely converted between the XLink and XF syntaxes, into a specification. Both enhancements would significantly help frameworks like the EBAs and EIOPAs.
- XBRL Europe is proposing a 'bridge' between the DPM model and XF. A new notation for defining business rules, XF-DPM, could both help translate between DPM rules and XBRL Formulas but also improve the performance of the resulting XBRL Formulas. The proposal

suggests that XF is combined with the DPM notation. This would make it easier to automate, should make formula definition easier (less verbose), potentially make finding errors easier, and could be optimised for CSV validation performance. However, it is still a low-level way of defining formulas, so would still rely upon many of these, rather than using the semantics embedded in a dimensional XBRL model.

Conclusions

In summary, the concerns are that by adopting its proprietary DPM ID as the key identifier in the XBRL-CSV format and using the DPM-ML assertions to validate documents, the EBA is not providing any advantages to the NCAs or filing entities, while asking them to invest money and time in updating their existing reporting systems. EIOPA is studying the XBRL-CSV layout and will adopt it later than the EBA, however, it will use the jointly developed DPM Studio software and from that some small standardisation will be achieved.

Having consulted many other vendors and NCAs, most are pleased with the XBRL systems they have implemented since 2012. The systems work well for them and the EBA and EIOPA should be proud of this. However, the move to XBRL-CSV was supposed to make it even easier for NCAs and filers, but the EBA's selected XBRL-CSV format only increases the amount of work on the systems upgrade and introduces the 'single-source of truth' issue that may cause a serious headache in terms of resolving issues between DPM-ML and XBRL formula. The authors are also concerned that the selected layout may make it more difficult to leverage future updates proposed for the XBRL standard around OIM Rules 3.0, leaving them with a framework with EBA proprietary elements, which is difficult to extend and difficult to improve.

For NCAs and reporting entities, they need to quickly understand the impact of the format change and what the potential costs of the different options are. They will then be in a good position to provide feedback to the Supervisors if there is a consultation on the matter.

In the end, the real question is how committed to open standards are the European Financial Market Supervisors and whether they are willing to help the XBRL community to improve XBRL. At the same time, can the XBRL community respond to the requirements of such systems and deliver on many of the promises it has made over the last 10 years, but has still yet to deliver. The messages from the EBA around the 'DPM 2.0' does not bode well and will continue to worry many players in the sector. However, there are many discussions underway and there is always time to find a compromise.

The authors are David Bell, Kapil Verma and Martin DeVille of UBPartner. Please send comments, corrections, and any alternative ideas to info@ubpartner.com.