

**XBRL EUROPE**

**25<sup>th</sup> XBRL EUROPE  
DIGITAL WEEK**

**Online conference**

**15-18th June 2020**



# TAXONOMY DESIGN - Thinking validations & analysis during the modeling phase

---

- Julien REBER
- Invoke
- Member of XBRL France

# Agenda

- Data Point Modeling on dynamic tables
- Usage of data points vs Dimensions

# Data Point Modeling on dynamic tables

- The DPM is a **structured representation of the data**, identifying all the business concepts, relations and validation rules.

# Data Point Modeling on dynamic tables

- For a given reporting model, the **chosen modeling approach** has impacts on:
  - The **size** of the instance documents.
  - The **data capture** from the instance documents.
  - The **time required to filter the data** for a validation or an analysis.

# Data Point Modeling on dynamic tables

- To illustrate our point, let's take the example of **Table T 03.01 - Intragroup liabilities**, LDR DPM 4.0.3, Single Resolution Board

A		B	C	G	H	I	J	K	L	M	N	O	P	Q	R	S
T 03.01 - Intragroup liabilities																
Columns																
Intragroup liabilities																
		Insolvency ranking	Contract identifier	Entity name of lending entity	Identifier of Lending Entity (preferably LEI)	Type of identifier	Relationship of lending entity with issuing entity	Is lending entity included in the resolution group of reporting entity?	Governing law	If third country, contractual recognition	Outstanding principal amount	Accrued interest	Currency	Issuance date	Earliest date	
		0030	0040	0045	0050	0053	0055	0057	0060	0070	0080	0090	0100	0110	0120	
Rows	Row ID (Intragroup liability)	999	418077 [Ranking in insolvency]	418084 text	423757 text	410400 text	428661 [Type of identifier]	428670 [Relationship of lending entity with issuing entity]	429083 TRUE/FALSE	418090 [Country the law of which governs the contract / master agreement]	418097 [Contractual recognition of bail-in powers]	418101 IC\$	418105 IC\$	418108 [Currency in which the transaction / contract is denominated]	418112 yyyy-mm-dd	
	Metric	(ei598) Ranking in insolvency [ei:ZZ:ZZ28]	(si599) Contract identifier [si]	(si168) Name of entity [si]	(si289) Entity code [si]	(ei706) Type of identifier [ei:AT:AT93]	(ei707) Relationship of lending entity with issuing entity [ei:RP:RP3]	(bi764) Entity is included in the resolution group of reporting entity [bi]	(ei600) Country the law of which governs the contract / master agreement [ei:GA:GA4]	(ei601) Contractual recognition of bail-in powers [ei:ZZ:ZZ37]	(mi602) Outstanding amount. Principal [mi]	(mi603) Outstanding amount. Accrued interest [mi]	(ei604) Currency in which the transaction / contract is denominated [ei:CU:CU3_31]	(di605) Date of issuance [di]	(di606) Earliest redemption date [di]	
	(BAS:BA) Base	(BA:x8) Liabilities and Equity (MC:x29) All equity, All liabilities	(BA:x8) Liabilities and Equity (MC:x29) All equity, All liabilities													
	(MCY:MC) Main category															
	(DFS:OF) Own funds															
	(MCJ:MC) Main category of collateralised item															
	(MCK:MC) Main category of item that is the object of the transaction / (IDL:ID) Row ID (Intragroup liabilities)			(MC:x29) All equity, All liabilities	(MC:x29) All equity, All liabilities	(MC:x29) All equity, All liabilities	(MC:x29) All equity, All liabilities	(MC:x29) All equity, All liabilities								
	(RPB:RP) Related parties/Relationships (RUL:RP) Role in the transaction (RPE:BT) Eligibility for resolution	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group	(RP:x34) Intra-group
				(RP:x43) Investor, creditor	(RP:x43) Investor, creditor	(RP:x43) Investor, creditor		(RP:x43) Investor, creditor								

# Data Point Modeling on dynamic tables

- In this table, data points are defined using **metrics and functional dimensions at column level**, plus a **technical dimension** “eba\_dim:IDL – Row ID” placed at the **row level** to dissociate the data as breakdown.
- The same table could have been designed using a different modeling approach by using a **one (different) metric per column** (without functional dimensions) and the **same technical dimension** “eba\_dim:IDL – Row ID” at the **row level** as breakdown.

# Data Point Modeling on dynamic tables

- What are the **consequences** of the two modeling approaches in term of instance **document size** ?

# Impact on the size of the instance documents

- Here are the differences between the two approaches regarding the **XBRL (XML)** instance document composition and size:

	Current DPM approach	The least dimensional possible approach
Number of context for 1 row	8	1
Number of context for 1 000 rows	8 000	1 000
Number of context for 80 000 rows	640 000	80 000
Instance size for 1 row	8 KB	4 KB
Instance size for 1 000 rows	5 578 KB	1 903 KB
Instance size for 80 000 rows	452 453 KB	156 188 KB



# Impact on the size of the instance documents

- With its “concept” of aliases for the contexts, the **OIM-CSV** drastically reduces the size of the instance documents.

## Aliases in Json

```
"documentInfo": {
  "documentType": "http://xbrl.org/CR/2020-05-06/xbrl-csv",
  "namespaces": {
    "eba": "http://www.eba.europa.eu",
    "eba_dim": "http://www.eba.europa.eu/xbrl/crr/dict/dim",
    "eba_met": "http://www.eba.europa.eu/xbrl/crr/dict/met",
    "eba_typ": "http://www.eba.europa.eu/xbrl/crr/dict/typ",
    "eba_BA": "http://www.eba.europa.eu/xbrl/crr/dict/dom/BA",
    "eba_BT": "http://www.eba.europa.eu/xbrl/crr/dict/dom/BI",
    "eba_MC": "http://www.eba.europa.eu/xbrl/crr/dict/dom/MC",
    "eba_OF": "http://www.eba.europa.eu/xbrl/crr/dict/dom/OF",
    "eba_RP": "http://www.eba.europa.eu/xbrl/crr/dict/dom/RP"
  }
},
"tableTemplates": {
  "T_03-01": {
    "dimensions": {
      "eba_dim:IDL": "",
      "eba_typ:ID": "#rowNumber"
    },
    "columns": {
      "rowNumber": {
      },
      "dp418071": {
        "dimensions": {
          "concept": "eba_met:ei597",
          "eba_dim:MCK": "eba_MC:x29"
        },
        "eba:documentation": {
          "logicalDatapointId": "418071",
          "templateId": "T_03.01",
          "SheetId": "n.a.",
          "rowId": "r999",
          "columnId": "c0020"
        }
      },
      "dp418077": {
        "dimensions": {
          "concept": "eba_met:ei598",
          "eba_dim:BAS": "eba_BA:x8",
          "eba_dim:MCY": "eba_MC:x29",
          "eba_dim:RPR": "eba_RP:x34"
        },
        "eba:documentation": {
          "logicalDatapointId": "418077",

```

## Data in CSV

```
T_03-01.csv
1  rowNumber, dp418071, dp418165, dp418077, dp418084, dp423757, dp410400, dp
2  111, eba_ZZ:x238, eba_ZZ:x291, eba_ZZ:x262, "Row1", "ACME ltd", "LegalIde
3  222, eba_ZZ:x239, eba_ZZ:x292, eba_ZZ:x263, "Row2", "Demo Inc.", "Anothe
```

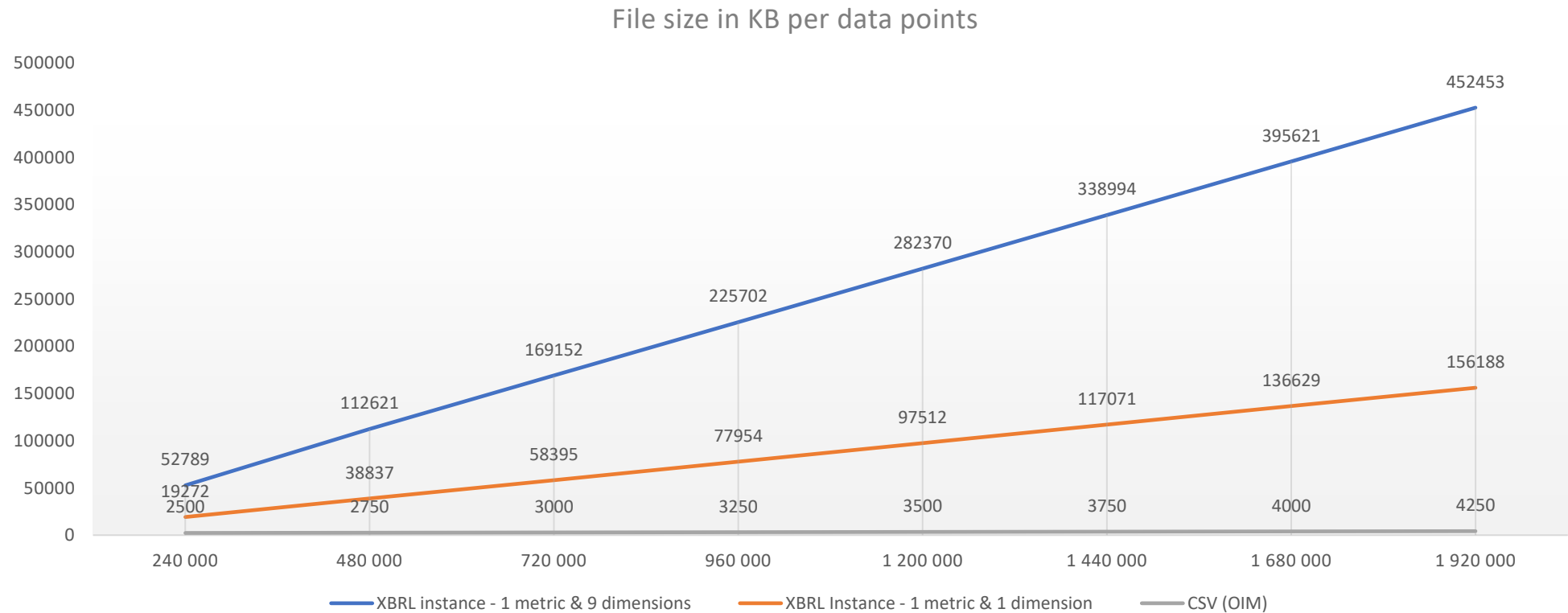
# Impact on the size of the instance documents

- The benefit in term of size is the following:

	Current DPM approach	The least dimensional possible approach	OIM-CSV, csv data file
Instance size for 1 row	8 KB	4KB	<b>&lt;1 KB</b>
Instance size for 1 000 rows	5 578 KB	1 903 KB	<b>250 KB</b>
Instance size for 80 000 rows	452 453 KB	156 188 KB	<b>4 250 KB</b>

# Impact on the size of the instance documents

- Shown into a graphic view:



15-18th June, 2020

25th XBRL Europe Day Digital Week

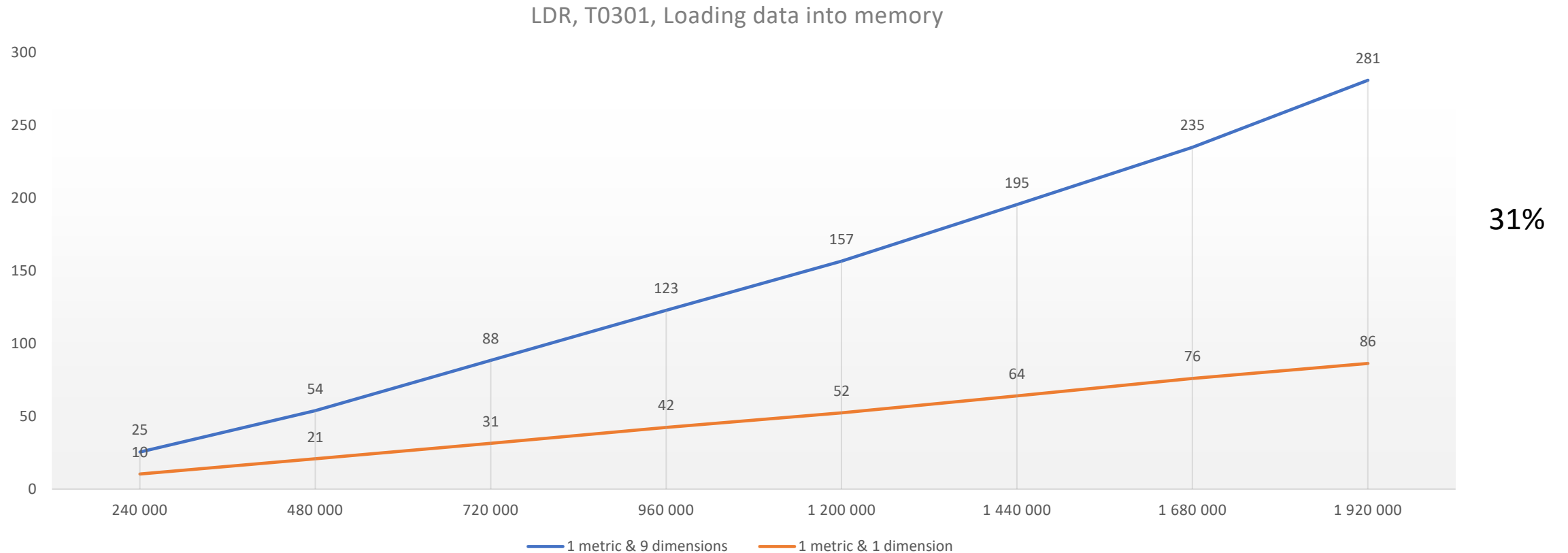
11

# Data Point Modeling on dynamic tables

- What are the **consequences** of the two modeling approaches in terms of the **data capture** ?

# Impact on the data capture

- The benefit in terms of time regarding **data capture** is the following:



15-18th June, 2020

25th XBRL Europe Day Digital Week

13

# Data Point Modeling on dynamic tables

- What are the consequences of the two modeling approaches in term of **data filtering** for **validation or analysis** ?

# Impact on the filtering of the data

- To illustrate our point, let's take the example of the XBRL assertion **srb\_v6511\_m**, LDR DPM 4.0.3, Single Resolution Board :
  - Test: "iaf:numeric-greater-equal-than(iaf:sum((\$a, \$b)), \$c)"

A	B	C	P	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
T 03.01 – Intragroup liabilities			\$a + \$b					>=						\$c	
			Columns												
			Outstanding principal amount	Accrued interest	Currency	Issuance date	Earliest redemption date	Legal maturity	Secured/unsecured	Amount of pledge, lien or collateral	Guarantor if applicable	Structured or other Non-Standard Terms	Amount meeting the conditions for Internal MREL eligibility	Qualifying as own funds	Amount Included in own funds, taking into account phase-out as applicable
			0080	0090	0100	0110	0120	0130	0140	0150	0160	0170	0175	0180	0190
Rows	Row ID (Intragroup liability)	999	418101 I€\$	418105 I€\$	418108 [Currency in which the transaction / contract is denominated]	418112 yyyy-mm-dd	418116 yyyy-mm-dd	410216 yyyy-mm-dd	418120 [Collateralisation status]	410142 I€\$	410401 text	418124 [Type of contractual terms]	428279 I€\$	418127 [Type of regulatory capital]	410199 I€\$
			(mi602) Outstanding amount, Principal [mi]	(mi603) Outstanding amount, Accrued interest [mi]	(ci604) Currency in which the transaction / contract is denominated [ci:CU:CU3_3]	(di605) Date of issuance [di]	(di606) Earliest date of redemption [di]	(di157) Legal final maturity date [di]	(ci607) Collateralization status [ci:Z2:Z229]	(mi53) Carrying amount [mi]	(zi283) Entity code [zi]	(ci608) Type of contractual terms [ci:Z2:Z230]	(mi308) Outstanding amount [mi]	(ci609) Type of regulatory capital [ci:DF:DF8]	(mi81) Amount including transitional provisions [mi]
			(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(MC:669) Collateral posted	(MC:29) All equity, All liabilities	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category	(BAS:BA) Base category
			(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities	(MC:29) All equity, All liabilities
			(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:44) Guarantor	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(RP:34) Intra-group	(BT:12) Eligible

# Impact on the filtering of the data

- On the **current modeling** approach, the assertion `srb_v6511_m` **filters the data** as follow:

Common filter

	Global filters on assertion	Filters on variable \$a	Filters on variable \$b	Filters on variable \$c
<code>eba_dim:IDL</code>	distinct values			
<code>eba_dim:ROL</code>	<code>eba_RP:x0</code>			
<code>eba_dim:MCJ</code>	<code>eba_MC:x0</code>			
<code>eba_dim:MRE</code>	<code>eba_BT:x0</code>			
concept		<code>eba_met:mi602</code>	<code>eba_met:mi603</code>	<code>eba_met:mi81</code>
<code>eba_dim:BAS</code>		<code>eba_BA:x8</code>	<code>eba_BA:x8</code>	<code>eba_BA:x11</code>
<code>eba_dim:MCY</code>		<code>eba_MC:x29</code>	<code>eba_MC:x29</code>	<code>eba_MC:x275</code>
<code>eba_dim:RPR</code>		<code>eba_RP:x34</code>	<code>eba_RP:x34</code>	<code>eba_RP:x0</code>
<code>eba_dim:MCK</code>		<code>eba_MC:x0</code>	<code>eba_MC:x0</code>	<code>eba_MC:x29</code>
<code>eba_dim:OFS</code>		<code>eba_OF:x0</code>	<code>eba_OF:x0</code>	<code>eba_OF:x10</code>

Variable specific filters

- The **contexts** are different per row



# Impact on the filtering of the data

- On the **least dimensional approach**, the assertion `srb_v6511_m` **filters the data** as follow:

Common filter {

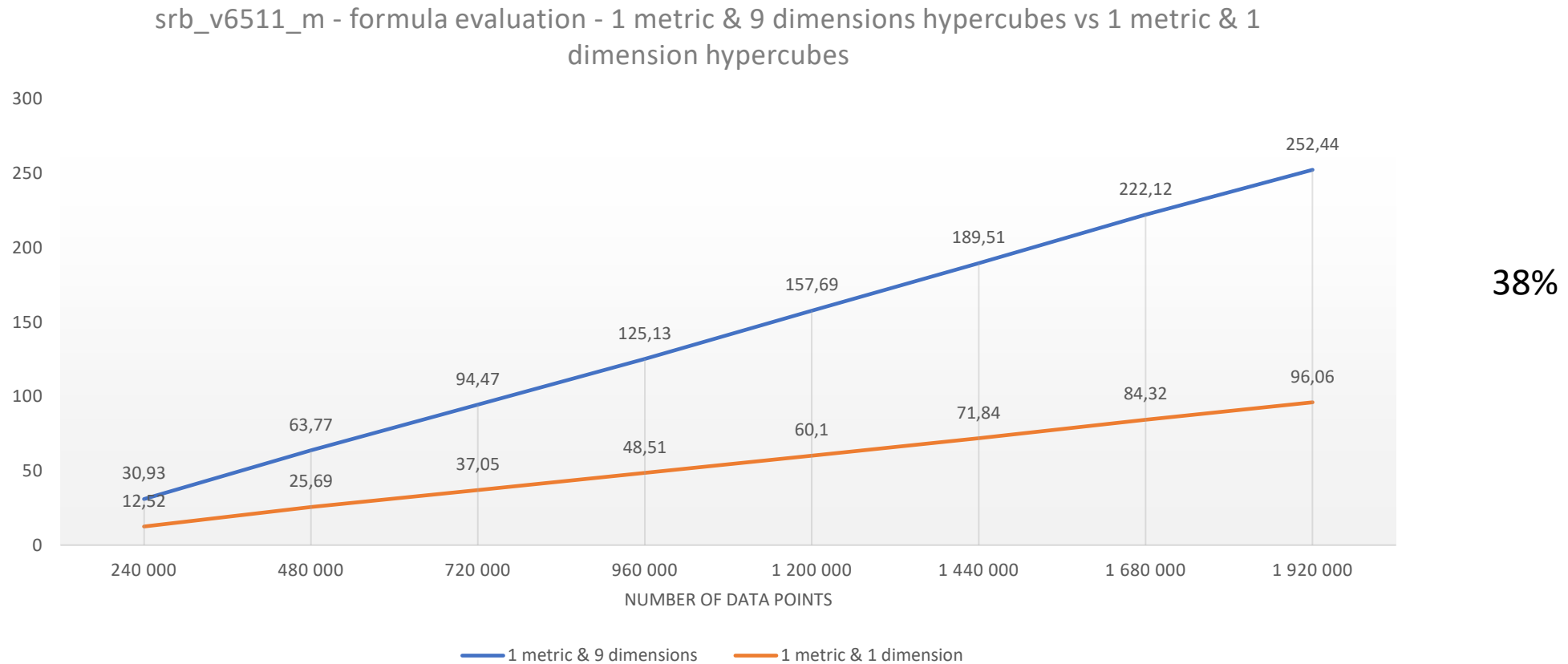
	Global filters on assertion	Filters on variable \$a	Filters on variable \$b	Filters on variable \$c
<code>eba_dim:IDL</code>	distinct values			
<code>concept</code>		<code>eba_met:mi602</code>	<code>eba_met:mi603</code>	<code>eba_met:mi81</code>

Variable specific filters

- The **contexts** are the same per row

# Impact on the filtering of the data

- The **difference between those 2 approaches** on the `srb_v6511_m` assertion for the “**same**” data points controlled is :



# Impact on the filtering of the data

- The **OIM-CSV**, with its principle of **creating aliases** dissociating the data from its “*context*” information, has the same problem for **data filtering** when using the current modeling approach.

```
"dp418084": {  
  "dimensions": {  
    "concept": "eba met:si599",  
    "eba_dim:BAS": "eba_BA:x8",  
    "eba_dim:MCY": "eba_MC:x29",  
    "eba_dim:RPR": "eba_RP:x34"  
  },  
  "eba:documentation": {  
    "logicalDatapointId": "418084",  
    "templateId": "T 03.01",  
    "SheetId": "n.a.",  
    "rowId": "r999",  
    "columnId": "c0040"  
  }  
},
```

```
"dp410400": {  
  "dimensions": {  
    "concept": "eba met:si289",  
    "eba_dim:MCK": "eba_MC:x29",  
    "eba_dim:ROL": "eba_RP:x43",  
    "eba_dim:RPR": "eba_RP:x34"  
  },  
  "eba:documentation": {  
    "logicalDatapointId": "410400",  
    "templateId": "T 03.01",  
    "SheetId": "n.a.",  
    "rowId": "r999",  
    "columnId": "c0050"  
  }  
},
```

# Agenda

- Data Point Modeling on dynamic tables
- Usage of data points vs Dimensions

# Usage of data points vs Dimensions

- At the beginning of **Solvency II DPM**, **EIOPA** prepared 2 modeling approaches :
  - **Highly Dimensional DPM-based XBRL Taxonomies** : high quality of the model, explicit dependencies between concepts, ...
  - **Moderately Dimensional DPM-based XBRL Taxonomies** : concatenation of HD Metrics and those dimensions that are not necessary from table rendering perspective
- EIOPA asked the software vendors **which approach should be promoted**:
  - The **Moderately Dimensional approach has been chosen** by the great majority of the market, thinking it would be easier to manage fewer dimensions.

# Usage of data points vs Dimensions

- The report « **S.06.02.01 - List of assets** », DPM 2.4.0 EIOPA, is defined with **one metric at the column level** and **technical dimensions at the row level** as breakdowns

Line identification	Asset ID Code and Type of code	Fund number	Matching portfolio number	Portfolio	Asset held in unit linked and index linked contracts	Asset pledged as collateral	Country of custody	Custodian	Quantity	Par amount	Valuation method	Acquisition value
CO001	CO040	CO070	CO080	CO090	CO090	CO100	CO110	CO120	CO130	CO140	CO150	CO160
*artificial key*   *mandatory*	*foreign key to S.06.02.01.02*   *mandatory*	*optional*	*optional*	Metric: Portfolio (investment, securities lending and repo)[210]	Metric: Held in unit linked and index linked funds	Metric: Asset pledged as collateral	Metric: Country of custody (including not applicable)	Metric: String	Metric: Decimal	Metric: Monetary	Metric: Valuation method (Full scope)	Metric: Monetary
XA: S.06.02.01.01 line identification	UI: URI	NF: Number of fund	MP: Matching portfolio number					TS/Custodian	DC/Quantity	TA/Notional amount BC/Assets VG/Solvency II		TA/Acquisition value BC/Assets VG/Solvency II

Asset ID Code and Type of code	Item Title	Issuer Name	Issuer Code and Type of code	Issuer Sector	Issuer Group	Issuer Group Code and Type of code	Issuer Country	Currency	CIC	SCR calculation approach for CIU	Infrastructure investment	Holdings in related undertakings, including participations
CO040	CO190	CO200	CO210	CO230	CO240	CO250	CO270	CO280	CO290	CO292	CO300	CO310
*natural key*   *mandatory*	Metric: String	Metric: String	Metric: String	Metric: Issuer sector - NACE	Metric: String	Metric: String	Metric: Issuer Country (including not applicable)	Metric: Original currency of exposure/transaction/instrument	Metric: String	Metric: SCR calculation approach for CIU	Metric: Infrastructure investment [210]	Metric: Participation (Full scope Individual) [240]
		TS/Name of										

- From a **business validation and analysis** point of view this report contains **problems on filtering**

# Usage of data points vs Dimensions

- Let's look at the assertion **s2md\_BV784-5**:

- Test: "if (not(empty(\$a))) then iaf:numeric-equal(\$b, iaf:sum((iaf:numeric-multiply(\$a, \$d), \$e))) else (true())"
- Label: BV784: The "Total Solvency II amount" in S.06.02 - List of assets should be equal to the product of "Par amount" and "Unit percentage of par amount Solvency II price" plus "Accrued interest". →table 1: S.06.02; Filter: not{{c0290} like '##71' or {c0290} like '##9#' or {c0290} like '##09'}; Expression: if {c0140} <> empty then {c0170} = {c0140} \* {c0380} + {c0180}

	\$a			\$b	\$e
Quantity	Par amount	Valuation method	Acquisition value	Total Solvency II amount	Accrued interest
C0130	C0140	C0150	C0160	C0170	C0180
Metric: Decimal	Metric: Monetary	Metric: Valuation method (Full scope)	Metric: Monetary	Metric: Monetary	Metric: Monetary
DC/Quantity	TA/Notional amount BC/Assets VG/Solvency II		TA/Acquisition value BC/Assets VG/Solvency II	VG/Solvency II BC/Assets	VG/Accrued interests BC/Assets

if c0290 <> '##71' or '##9#' or '##09' Then  
 $c0170 = c0140 * c0380 + c0180$

	\$f									\$d
CIC	SCR calculation approach for CIU	Infrastructure investment	Holdings in related undertakings, including participations	External rating	Nominated ECAI	Credit quality step	Internal rating	Duration	Unit Solvency II price	Unit percentage of par amount Solvency II price
C0290	C0292	C0300	C0310	C0320	C0330	C0340	C0350	C0360	C0370	C0380
Metric: String	Metric: SCR calculation approach for CIU	Metric: Infrastructure investment [210]	Metric: Participation (Full scope individual) [240]	Metric: String	Metric: Nominated ECAI [SCR simplification included] [240]	Metric: Credit quality step [240] (Full scope)	Metric: String	Metric: Decimal	Metric: Monetary	Metric: Pure
TS/CIC code				TS/External rating			TS/Internal rating	DC/Residual modified duration	TA/Unit price VG/Solvency II without accrued interests BC/Assets	VG/Solvency II without accrued interests PP/Percentage of par value

# Usage of data points vs Dimensions

- The **complexity** of this assertion is linked to the **indirect filtering** of the variable **\$f** (i.e. CIC codes must be different than '##71' or '##9#' or '##09') on the other variables of the assertion :

**Select** each data point C0290 different than ##71 or ##9# or ##09

**Retrieve** the **list** of values for the technical dimension **s2c\_dim:UI** in their context

**For each value** in the list of dimension **s2c\_dim:UI**

**Select** each other data point : C0140, C0170, C0180 and C0380;

**Perform** : if {c0140} <> empty then {c0170} = {c0140} \* {c0380} + {c0180}.



# Usage of data points vs Dimensions

- Thanks to a **modeling** using HD approach, i.e. considering the column **C0290 – CIC** as a **dimension** of the **other data points**, the **filtering will be direct**:

**Select** the list of context with with dimension CIC different than ##71 or ##9# or ##09.

**For each** context in the list

**Select** each data point : C0140, C0170, C0180 and C0380;

**Perform** : if {c0140} <> empty then {c0170} = {c0140} \* {c0380} + {c0180};

- In DPM 2.4.0, on **63 assertions** linked to the **report S.06.02**, **53 assertions** are using the **CIC** data points **as a filter** of the variables.

# Usage of data points vs Dimensions

- In general we find the **same problem** on every modeling **using the MD approach**, for which **functional and/or technical dimensions** are used instead of using “**business**” dimensions to describe data points.

Examples :

- Solvency 2 – S0602: on **31 data points** at least **15** could have been business dimensions
- LDR – T0301: on **24 data points** at least **10** could have been business dimensions

# Usage of data points vs Dimensions

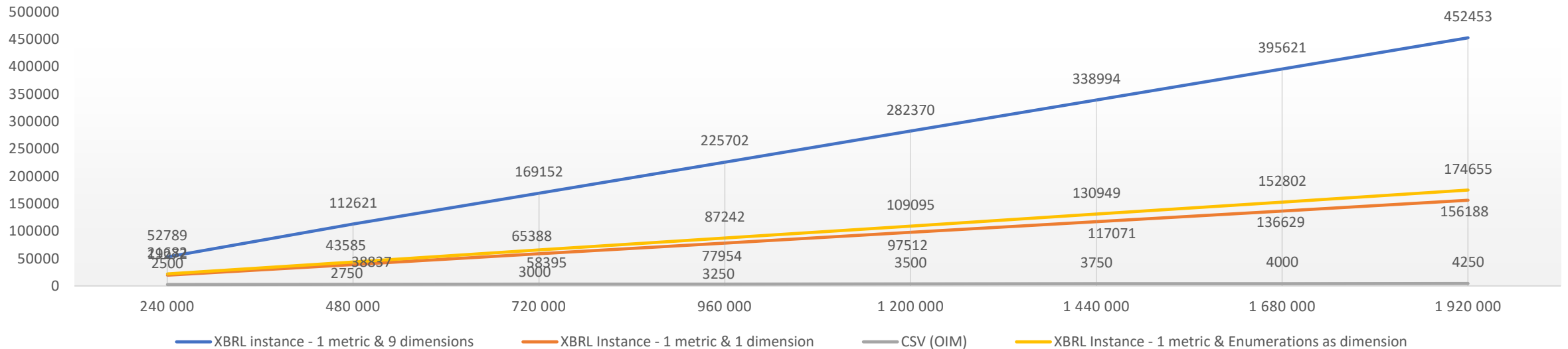
- How to **identify** the current data points which should have been **dimensions** ?
  - Basically each data point defined as “**enumerationItemType**” is a dimension;
  - Some data points like “**stringItemType**” (like CIC code) or “**dateItemType**” may be considered as dimensions.

# Impact on the size of the instance documents

- There is a **minimal impact** in term of file size :

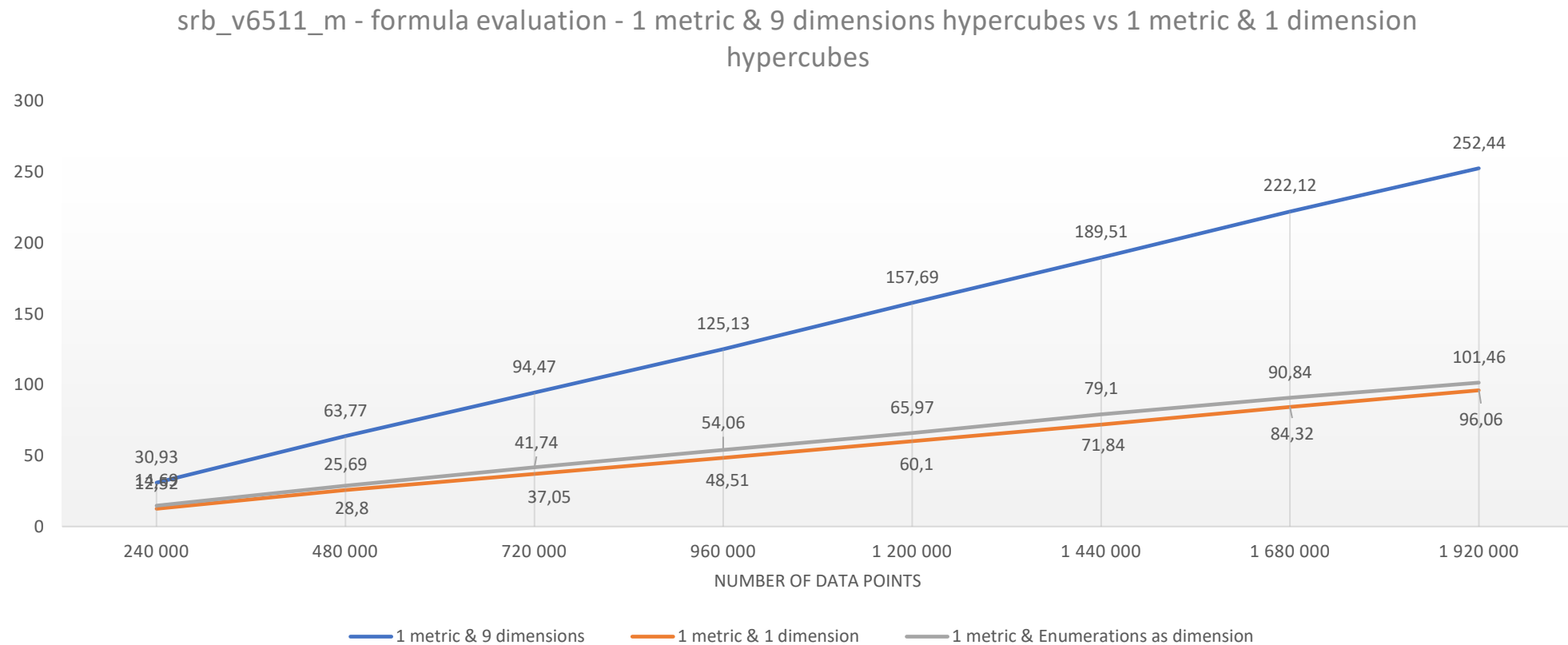
	Current DPM approach	The least dimensional possible approach	With Enumerations as dimensions	OIM-CSV, csv data file	OIM-CSV, csv with enumeration as dimensions
Instance size for 1 row	8 KB	4KB	<b>4 KB</b>	<1 KB	<b>&lt;1 KB</b>
Instance size for 1 000 rows	5 578 KB	1 903 KB	<b>2 204 KB</b>	250 KB	<b>250 KB</b>
Instance size for 80 000 rows	452 453 KB	156 188 KB	<b>174 655 KB</b>	4 250 KB	<b>4 250 KB</b>

File size in KB per data points



# Impact on the filtering of the data

- There is a **minimal impact on the filtering** of the srb\_v6511\_m assertion because **all the data points share the same context for a row** :



# Usage of data points vs Dimensions

- Thanks to the **OIM-CSV** we have a good opportunity to **reduce** the processing times regarding **data capture**.
- Even on XBRL (XML) instance, the **modeling approach** can already **enhance the performances and the usage of the resources** needed to perform the whole process.
- For the **validation and data analysis** we need to define “**business**” **dimensions**, share the **same context** for one row, in order to have a **direct data filtering by using dimensions** instead of data points.

**Thank you for your attention !**