Metadatamodels like DPM
Why it might be a building block regarding a potential AI strategy
#BelieveInYourself

Christian Ronacher
Christian.ronacher@erstegroup.com
We sit in here

← And not here

← Or here

← We sit in here
What we do with DPM

1. DPM is a metadata framework.
2. We extended FinRep Taxonomy to generate our own templates.
3. We use EBAs dictionary and add our dictionary to EBAs.
4. We connect Metadata to Data based on the regulator’s framework.
5. This enables querying data by natural language conversation.

What we do with DPM
Our goal is to provide self-explainable and curated data, which can be queried by conversation.
To achieve this, data must meet the following criteria.

Consistent  Transparent  Efficient  Validated  Simple accessible
We think metadatamodels are a cornerstone to provide meaning to reports and make them queryable **by conversation** (in an auditable way)
### Sidestep – What do we mean with querying by conversation

<table>
<thead>
<tr>
<th>GPT 4.0</th>
<th>The problem</th>
</tr>
</thead>
</table>
| what is the carrying amount?  
The carrying amount is not explicitly provided in the text you shared. However, the carrying amount for these derivatives can be calculated by taking the difference between the positive fair value and the negative fair value.  
For interest rate risk derivatives held for trading in the trading book, the carrying amount as of December 2022 would be:  
Carrying amount = Positive fair value - Negative fair value  
Carrying amount = 5,788 million EUR - 5,508 million EUR  
Carrying amount = 280 million EUR  
Please note that this is a simplified calculation, and there may be other factors to consider in a more comprehensive approach. | This needs to be trackable |
„reports follow the question“ vs. „I have know which report answers a certain question“

What's the population of France?

Used Wolfram

As of the 2020 estimate, the population of France is approximately 65.3 million people, making it the 22nd most populous country in the world. Here is a chart showing the recent population history of France:

![Population History Chart](chart.png)

The population density of France is estimated to be 309 people per square mile. The population growth rate is approximately 0.14% per year, and the life expectancy is 82.8 years. The median age of the population is 41.6 years.

Source: https://youtu.be/EOQV9VakBgE
**Where do we come from?**

### Template

**F 01.01**  
Balance Sheet Statement [Statement of Financial Position]: Assets

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Carrying amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>010</td>
<td>Cash, cash balances at central banks and other demand deposits</td>
<td>10898</td>
</tr>
<tr>
<td>020</td>
<td>Cash on hand</td>
<td>11713</td>
</tr>
<tr>
<td>030</td>
<td>Cash balances at central banks</td>
<td>15111</td>
</tr>
<tr>
<td>040</td>
<td>Other demand deposit</td>
<td>15111</td>
</tr>
<tr>
<td>050</td>
<td>Financial assets held for trading</td>
<td>11110</td>
</tr>
<tr>
<td>060</td>
<td>Derivatives</td>
<td>11112</td>
</tr>
<tr>
<td>070</td>
<td>Equity instruments</td>
<td>11100</td>
</tr>
<tr>
<td>080</td>
<td>Debt securities</td>
<td>11119</td>
</tr>
<tr>
<td>090</td>
<td>Loans and advances</td>
<td>11114</td>
</tr>
<tr>
<td>100</td>
<td>Trading financial assets</td>
<td>49321</td>
</tr>
<tr>
<td>101</td>
<td>Derivative liabilities</td>
<td>64320</td>
</tr>
</tbody>
</table>

### Map each cell

**Documentation:**  
In this cell the carrying amount for cash and cash equivalent positions

### Write Code for the cell

**Cell 10898:**

```
SELECT *  
FROM xyz Table  
Inner Join ....
```

**Documentation:**  
This select statement retrieves BET_100 (carrying amount) for WER_ART_103 („Cash“ or „Cash equivalent“ positions … )
We use the metadata information provided by EBA

**Template**

<table>
<thead>
<tr>
<th>Code</th>
<th>Dictionary Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>APL2576</td>
<td>Accounting Portfolio „Cash“</td>
</tr>
<tr>
<td>ATY1177</td>
<td>Metric „Carrying Amount“</td>
</tr>
<tr>
<td>BAS1506</td>
<td>Base „Assets“</td>
</tr>
<tr>
<td>MCY1881</td>
<td>Category „Cash on Hand“</td>
</tr>
</tbody>
</table>

**Use its metadata**

Show me the Metric „Carrying Amount“ (ATY1177) of all Accounting Portfolio „Cash on short notice“ (APL 2576) of Basetype „Assets“ (BAS1506) which are in the Main Category „Cash“ (MCY1881)
We use the metadata information provided by EBA

**Template**

<table>
<thead>
<tr>
<th>Code</th>
<th>Dictionary Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>F 01.01</td>
<td>Balance Sheet Statement (Statement of Financial Position) - Assets</td>
</tr>
</tbody>
</table>

| Cells in FinRep~10.000 |

| Approx 300 members |

**Use its metadata**

**e.g. for a the cell**

Show me the Metric „Carrying Amount“ (ATY1177) of all Accounting Portfolio „Cash on short notice“ (APL 2576) of Basetype „Assets“ (BAS1506) which are in the Main Category „Cash“ (MCY1506)
# Mapping the Members

<table>
<thead>
<tr>
<th>Cell-Information</th>
<th>Map each member</th>
<th>And get the keys</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Cell Image" /></td>
<td>We: How can I find the Assets (Bas1506) in System A?</td>
<td>System A: Here are they keys</td>
</tr>
<tr>
<td></td>
<td>System: Please provide a Rule for finding all „Assets“</td>
<td>AccountNumbers:</td>
</tr>
<tr>
<td></td>
<td>We: Here ist the rule! Assets = „XYZ“ or „something else“. Please give us the keys which follow this rule</td>
<td>{12345, 34567}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>System B: Here are they keys</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TradeIDs:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{A1, A2, A3, A4}</td>
</tr>
</tbody>
</table>

Show me the Carrying Amount (ATY1177) of all Assets (BAS1506) which are of Type Cash (MCY1506) in the Accounting Portfolio „Cash on short notice“ (APL 2576)
Tagging the Data

Create a Big Table

<table>
<thead>
<tr>
<th>Posting-ID</th>
<th>Hundred Other Dimensions</th>
<th>Carrying Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12345</td>
<td>A1</td>
</tr>
<tr>
<td>2</td>
<td>23456</td>
<td>A2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1.394.023</td>
<td>34567</td>
<td>A3</td>
</tr>
</tbody>
</table>

Tagging according to the Rules

<table>
<thead>
<tr>
<th>BAS</th>
<th>MCY</th>
<th>APL</th>
<th>GroupString</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets</td>
<td>Cash</td>
<td>Cash on hand...</td>
<td>APL2576BAS1506MCY1881</td>
</tr>
<tr>
<td>Liabilities</td>
<td>something else</td>
<td>APL2576BAS1506MCY1881</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Assets</td>
<td>Loans</td>
<td>Amortized Cost</td>
<td>OTH9999KEY8868</td>
</tr>
</tbody>
</table>

The cell as a search criteria

Show me all Assets (BAS1506)
Producing the Numbers

The final templates

Searchcriteria

SearchkeyA

Show me all „Derivatives“ of Risk-Type „Interest Rate Risk“ for „Financial Assets“ in the Accounting Portfolio „Held for Trading“.

SearchkeyB = A+ another Key

Show me all „Derivatives“ of Risk-Type „Interest Rate Risk“ for „Financial Assets“ in the Accounting Portfolio „Held for Trading“.

but only Instrument Type „OTC Options“
Sidestep – some Experiments with GPT 4.0

GPT 4.0

Now show instead of the categorisation keys the explanation according to the dictionary. Always start with the metric first.

Here's the table with explanations according to the dictionary. I've used Rust format.

```
<table>
<thead>
<tr>
<th>Metric</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrying amount for Financial assets</td>
<td>'Carrying amount for Financial assets' (ATY1177)</td>
</tr>
<tr>
<td>Held for trading</td>
<td>'Held for trading' (APL2566ATY1177BAS1506)</td>
</tr>
<tr>
<td>Interest rate</td>
<td>In this cell there is the 'carrying amount' (ATY1177) for all 'Assets' (BAS1506) with the 'Accounting Portfolio' of type 'Financial assets' but only namely for 'Derivatives' (MCY1994) which are of 'Type of Risk' 'interest rate risk'</td>
</tr>
</tbody>
</table>
```

Show me the Metric „Carrying Amount“ (ATY1177) of all Accounting Portfolio „financial assets held for trading“ (APL2566) of Basetype „Assets“ (BAS1506) which are in the Main Category „Derivative“ (MCY1994) of Risktype „Interest Rate“ (TRI2701)

e.g. for a the cell

<table>
<thead>
<tr>
<th>Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivatives: Trading</td>
</tr>
<tr>
<td>010</td>
</tr>
<tr>
<td>APL2566</td>
</tr>
<tr>
<td>ATY1177</td>
</tr>
<tr>
<td>BAS1506</td>
</tr>
<tr>
<td>MCY1994</td>
</tr>
<tr>
<td>TRI2701</td>
</tr>
</tbody>
</table>

F 10.00

Show me the Metric „Carrying Amount“ (ATY1177) of all Accounting Portfolio „financial assets held for trading“ (APL2566) of Basetype „Assets“ (BAS1506) which are in the Main Category „Derivative“ (MCY1994) of Risktype „Interest Rate“ (TRI2701)
Yes, it does help. Based on the Categorisation-Key APL2566ATY1177BAS1506MCY1994TRI2701, the reported information for F10.00, r010, c010 is as follows:

- APL2566: Accounting Portfolio of type "financial assets held for trading"
- ATY1177: Carrying amount
- BAS1506: Assets
- MCY1994: Derivatives
- TRI2701: Type of Risk "interest rate risk"

So, for F10.00, r010, c010, the reported information is the carrying amount of financial assets held for trading, specifically for derivatives with interest rate risk.
Sidestep – some Experiments with GPT 4.0

GPT 4.0

Comparing Cells

The difference between F10.00, r010, c010 and F10.00, r010, c020 lies in the type of financial instruments they represent:

F10.00, r010, c010:
- APL2566: Accounting Portfolio of type “financial assets held for trading”
- BAS1506: Assets

F10.00, r010, c020:
- APL2566: Accounting Portfolio of type “financial liabilities held for trading”
- BAS1513: Liabilities

Both cells report the carrying amount (ATY1177) of derivatives (MCY1994) with interest rate risk (TRI2701), but one represents assets held for trading (c010) while the other represents liabilities held for trading (c020).
Guidelines

Guideline 1

Before you think about correctness think about built in consistency

Use the build in datastructures within the taxonomies

DPM has built in Consistency Checks

- Show me all „Derivatives“
- … for assets
- … but only Instrument Type „OTC Options“

Consequence

- 8
- 0
- 4
Hierarchies, CatKeys, Dimensions and its members, appearances within the templates, cell definition (as intersection of a row and column).

relationship between members across dimensions, this „arghh-Dimension“ „MCY“ validations work only for numbers not for datasets and not just numbers
## Validation for dataset

<table>
<thead>
<tr>
<th>Guideline 1</th>
<th>DPM has built in Consistency Checks</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why validation for numbers Aren’ t enough?</td>
<td>… Show me the assets</td>
<td>This is a problem though numbers are the same and validation rules do still apply</td>
</tr>
<tr>
<td>For all „Derivatives“</td>
<td>… but only Instrument Type „OTC Options“</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>is a 4</td>
<td></td>
</tr>
</tbody>
</table>
Validation for dataset

**Guideline 1**

Why validating numbers is not enough?

**DPM has built in Consistency Checks**

- … Show me the assets
- For all „Derivatives“
- … but only Instrument Type „OTC Options“

**Consequence**

It should look like this because OTC Options is a subset of „derivatives“
Guidelines

Guideline 2

Add words to the dictionary if the meaning is not precise enough

Example

Hierarchy 33

1638 Counterparties other than financial corporations
  1631 Central banks
  1649 General governments
  1650 Households
    1657 Non-financial corporations
  1644 Financial corporations
    1649 Credit institutions
      - 611 Credit institutions - Own shares Erste Group Bank AG
      - 612 Credit institutions other than Own shares Erste Group Bank AG
    3063 Financial corporations other than credit institutions
## Guidelines

### Guideline 3

Go for the meaning and find every *meaningful* error where consistency is guaranteed:

| Meaningful means errors above 2€ |

### Example

Definition: $C = A + B$

- **Option 1)**
  
  $C = A \text{ union } B$

- **Option 2)**
  
  A is mapped to data
  B is mapped to data
  C is mapped to data

$\Rightarrow$ Weird technical problem

$\Rightarrow$ Additional analyses
**Advantage 1: Metadata gives Data a meaning**

<table>
<thead>
<tr>
<th>DPM as a Metadatamodel</th>
<th>Other Metadatamodels</th>
</tr>
</thead>
<tbody>
<tr>
<td>... DPM Metadata provides meaning to data e.g. we know exactly what the number is composed of.</td>
<td>DPM provides this, without tagging or doing anything AFTER the data is prepared. It could be done before!</td>
</tr>
</tbody>
</table>
**Sidestep – some Experiments with GPT 4.0**

<table>
<thead>
<tr>
<th>GPT 4.0</th>
<th>Wrong answer</th>
</tr>
</thead>
</table>
| can you guess what the connection between MCY46166 and MCY46167, MCY46168 and MCY 46169 is | * Cash and cash equivalents  
  - Cash on hand  
    - MCY46167  
  - Cash balances at central banks  
    - MCY46168  
  - Other demand deposits  
    - MCY46169  
  - Cash and cash balances at central banks and other demand deposits  
    - MCY46166  
    - APL2876 -> 2876 |
| Based on the member labels, it seems that MCY46166 represents the total amount of cash and cash balances at central banks and other demand deposits, while MCY46167, MCY46168, and MCY46169 represent the individual components that make up this total amount. MCY46167 represents the amount of cash on hand, MCY46168 represents the amount of cash balances at central banks, and MCY46169 represents the amount of other demand deposits. |
How we want to use LLMs and other tech

„Conversational Intention Recognition Machine“ like LLM

„Computational Knowledge Engine“ Like Wolfram
LLMs eg. ChatGPT

Prompt Configuration

Semantics (Metadata)

Concept Store(s) like DPM

Mappings

Supervising Conversation

Documents

Querying and Extraction Tool

Data

ERSTE
Reversing the flow – instead of querying data we can also gather data.
Disclosures in accordance with EU Taxonomy: The key performance indicators in this section were determined based on Commission Delegated Regulation (EU) 2021/1778 in conjunction with the International Financial Reporting Standards applicable for the Consolidated Financial Statements. Thereby, revenue, capital expenditures and operating expenses were predominantly directly allocated to an economic activity listed in Delegated Regulation (EU) 2020/852; in determining capital expenditures and operating expenses, allocations were also made based on the revenue of the Taxonomy-eligible activities. To avoid double counting, the allocation was always made to one economic activity only. Taxonomy-eligible revenue accounted for 20% of revenue according to the Consolidated Statement of Income in the reporting year. In the reporting year, Taxonomy-eligible capital expenditures accounted for 40% of additions (including additions from business combinations) to other intangible assets and property, plant and equipment in accordance with Note 13 to the Consolidated Financial Statements. Taxonomy-eligible operating expenses accounted for 14% of the corresponding expenses recognized in the Consolidated Financial Statements in the reporting year. The remaining portions of the key performance indicators are not Taxonomy-eligible. Our main Taxonomy-eligible economic activities are derived from the manufacture of low-carbon transport and energy-efficient building technologies (mainly from Mobility and Smart Infrastructure operations), transport infrastructure (from Mobility operations) and the service of energy-efficient building technologies (from Smart Infrastructure operations), as well as the Group’s own real estate portfolio. The majority of Taxonomy-eligible capital expenditures result from the latter economic activity. The above-mentioned economic activities refer to chapters 5, 6 and 7 of Annex I of Delegated Regulation (EU) 2020/852.

Some reports text only

Many create their own layout

Real world examples

And other issues (language, cut offs,...)
We want to share this data
Explored Benefits

Errors are easier to find

Changes are much easier to introduce

Easy Code maintenance (metamodel contains most of the information)

Validation follows meaning

Separation of Concerns (Semantics, Rules, Data)

Analytics included

OOB maintenance by the regulator

DPM is agnostic to models and formats

Insane increase in Dataquality (by following the guidelines)
Yet Unexplored Potentials

universal „conversational data retrieval interface“ (versus linear Chatbots)

Don‘t Search – Ask! which creates improvable knowledge

Creating a marketplace for data

Connecting different models via metadata models
Sidestep – some Experiments with GPT 4.0 regarding connecting different data models

FUNKTION XEN_Sektor_Finanzen(A_Mandant_MA, A_Einheitsnummer_ID emzl, A_Stichtag_Datum repGtm)
aktion_sektor = SELECTIERE XEM:MA49_Sektor_fuer_Malzedeczky_CodA VON EM_Einheit_MS
internet_org = SELECTIERE EO:EO40_Internationale_Organisation_OS_Code VON EO_Einheit_OS VON EM_Einheit_MS
k_gem_finan = SELECTIERE EM:EM441_KreditinstitutGem_Finan, Kennzeichen VON EM_Einheit_MS
WENN(aktion_sektor ISTGLEICH "Zentralbank (1210)" UND NICHT(internet_org ENTHÄLT ("Bank für Internationalen Zahlungsausgleich (BIB)", "Internationaler Währungsfonds (IWF)"))) Dann
\(\text{Value} = \text{"Central banks (1)"}\)
SONST WENN(aktion_sektor ENTHÄLT ("Zentralbank (1311)", "Länder (inkl. Landeshöchstzins, Landesfonds) (1312)", "Gemeinden (inkl. Gemeindefonds und -verbände) (1313)", "Sozialversicherung (1314)", "Staat nicht zuordenbar (133002)") OR internet_org ENTHÄLT ("Sonstige internationale Org (UI)")) Dann
\(\text{Value} = \text{"General governments (2)"}\)
SONST WENN(k_gem_finan ISTGLEICH Währ) Dann
\(\text{Value} = \text{"Credit institutions (3)"}\)
SONST WENN(aktion_sektor ENTHÄLT ("sonstige finanzunternehmen, GP")) UND internet_org ISTGLEICH NULL Dann
\(\text{Value} = \text{"Financial corporations other than credit institutions (4)"}\)
SONST WENN(aktion_sektor ISTGLEICH "Nicht-finanzable Unternehmen (1700)") Dann
\(\text{Value} = \text{"Non-financial corporations (5)"}\)
SONST WENN(aktion_sektor ENTHÄLT ("Selbständige Haushalte (mit u. ohne Arbeitnehmer) (1400A)", "sonstige private Haushalte (1400B)", "private Haushalte nicht zuordenbar (14002)", "private Organisationen ohne Erwerbszweck (1550)")) Dann
\(\text{Value} = \text{"Households (6)"}\)
ENDE
RUECKGABE \(\text{Value}\)
Sidestep – some Experiments with GPT 4.0 regarding connecting models

GPT 4.0 Conversation

DPM Description

F 10.00
Derivatives: Trading

Assets
Liabilities

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>010</td>
<td></td>
</tr>
<tr>
<td>020</td>
<td></td>
</tr>
</tbody>
</table>

- APL2566
- APL2566
- AY1117
- AY1117
- BA51506
- BA51513
- MCY1994
- MCY1994
- TRI2701
- TRI2701

Interest rate

Please note that this list is not exhaustive, and there may be other relevant dictionary entries in the context of financial reporting. However, these are the main ones we have discussed so far.
Yet Unexplored Potentials

1. Creating a universal "conversational data retrieval interface (versus Chatbots)"
2. Don’t Search – Ask! which creates improvable knowledge
3. Creating a marketplace for data
4. Connecting different models via metadatamodels
5. Centralizing meaning and using distributed data
6. Self Tutoring and Learning
7. Connecting data (metadata, data, taxonomies, reports)
### Potential 7: connect FinRep and IFRS Taxonomy and self tagging!

<table>
<thead>
<tr>
<th>Challenge</th>
<th>LLM and Metadatamodel</th>
<th>Technical Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports should be machine readable and comparable. While machine readability is not a problem, comparability is a hard task?</td>
<td>...comparability gets possible even if different words or tags are used? What if words and data is connected, not via html links but via an ontology.</td>
<td>IFRS Taxonomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESEF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FinRep Taxonomy</td>
</tr>
</tbody>
</table>
Knowledge and understanding and therefore fun increases

Explored but completely underestimated benefit
“... LET THE MACHINES GET ON WITH THE ADDING UP,"

“(No)...WE DEMAND RIGIDLY DEFINED AREAS OF DOUBT AND UNCERTAINTY!”

Discussion before Deep Thought started to calculate the Meaning of Life From Douglas Adams Book “The Hitchhiker’s Guide to the Galaxy”