Using Business Architecture and Semantics to Drive Data Quality Improvement in Banking

Elisa Kendall, Thematix Partners LLC
By their own admission, about half of the Global-Systematically Important Banks (G-SIBs) are going to fail to comply with Basel III Risk Data Aggregation and Reporting Requirements (RDARR)*

* Basel Committee on Banking Supervision: Progress in adopting the principles for effective risk data aggregation and risk reporting, January 2015
Trouble spots

- The January report emphasized four principles with which G-SIBs will not comply

<table>
<thead>
<tr>
<th>Principles</th>
<th>No. of G-SIBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Governance</td>
<td>10</td>
</tr>
<tr>
<td>2. Data architecture &amp; IT infrastructure</td>
<td>9</td>
</tr>
<tr>
<td>3. Accuracy and integrity</td>
<td>9</td>
</tr>
<tr>
<td>4. Completeness</td>
<td>9</td>
</tr>
<tr>
<td>5. Timeliness</td>
<td>11</td>
</tr>
<tr>
<td>6. Adaptability</td>
<td>11</td>
</tr>
<tr>
<td>7. Accuracy</td>
<td></td>
</tr>
<tr>
<td>8. Comprehensiveness</td>
<td></td>
</tr>
<tr>
<td>9. Clarity and usefulness</td>
<td></td>
</tr>
<tr>
<td>10. Frequency</td>
<td></td>
</tr>
<tr>
<td>11. Distribution</td>
<td></td>
</tr>
</tbody>
</table>

* Basel Committee on Banking Supervision: Progress in adopting the principles for effective risk data aggregation and risk reporting, January 2015
The weakest link - data

- The *lowest* average compliance rating is for **Data Architecture & IT Infrastructure**
  - Runner-Up: “Only two G-SIBs reported compliance with the Governance Principle, and no G-SIBs fully comply with the Data Architecture and IT Infrastructure Principle.”

- **Action steps** from the Basel III Commission:
  - “identifying and defining data owners.”
  - “consistent and integrated vocabularies at the group and organizational level”

* Basel Committee on Banking Supervision: Progress in adopting the principles for effective risk data aggregation and risk reporting, January 2015
Nordea is the largest financial services group in the Nordic and Baltic sea region

Nordea = Nordic ideas

11 million customers
- 8 home markets
- Approx. 10 million retail customers
- 600,000 corporate customers, incl. Nordic Top 500

Distribution power
- Approx. 650 branch office locations
- Approx. 7 million Netbank customers

Financial strength
- EUR 2.645bn operating income (2015)
- EUR 646.9bn total assets (Q4 2015)
- EUR 288.2bn assets under mgmt (Q4 2015)
- AA credit rating
- Common Equity Tier 1 capital ratio of 16.5% (Q4 2015)

EUR ~41.3bn in market cap
- One of the largest Nordic corporations
- A top-10 European retail bank
# The value chain in capital markets

<table>
<thead>
<tr>
<th>Pre-deal process</th>
<th>Deal Capture process</th>
<th>Financial &amp; Risk control process</th>
<th>Settlement &amp; Confirmation process</th>
<th>Reporting process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer process</td>
<td>Deal capture</td>
<td>Credit risk</td>
<td>Collateral Management</td>
<td>Customer reporting</td>
</tr>
<tr>
<td>Issuer Services</td>
<td>Execution</td>
<td>Finance &amp; accounting</td>
<td>Confirmations</td>
<td>Financial &amp; legal reporting</td>
</tr>
<tr>
<td>People &amp; infrastructure</td>
<td>Position management</td>
<td>Liquidity Risk</td>
<td>Lifecycle management</td>
<td>Management reporting</td>
</tr>
<tr>
<td>Product process</td>
<td></td>
<td>Market risk</td>
<td>Liquidity Management</td>
<td>Regulatory reporting</td>
</tr>
<tr>
<td>Risk limits</td>
<td></td>
<td>Operational &amp; compliance risk</td>
<td>Settlement</td>
<td></td>
</tr>
<tr>
<td>Static data</td>
<td></td>
<td>P/L</td>
<td>Transaction validation</td>
<td></td>
</tr>
</tbody>
</table>

- **Customer process** includes Deal capture, Credit risk, Collateral Management, and Customer reporting.
- **Issuer Services** includes Execution, Finance & accounting, Confirmations, and Financial & legal reporting.
- **People & infrastructure** includes Position management, Liquidity Risk, Lifecycle management, and Management reporting.
- **Product process** includes Market risk, Liquidity Management, Regulatory reporting.
- **Risk limits** includes Operational & compliance risk, Settlement.
- **Static data** includes P/L, Transaction validation, Valuation.
Business facing aspects of data management

Data Management Strategy
- Data Management Strategy
- Communications
- Data Management Function
- Business Case
- Program Function

Data Governance
- Governance Management
- Business Glossary
- Metadata Management

Data Quality
- Data Quality Strategy
- Data Profiling
- Data Quality Assessment
- Data Cleansing

Data Operations
- Data Requirements Definition
- Data Lifecycle Management
- Provider Management

Source: CMMI Institute Data Management Maturity (DMM)SM Model, August 2014
Data governance objectives

- Establish effective principles and procedures to support business data owners in their data governance duties

- Create and support maintenance and management of a knowledge base (metadata) that facilitates data governance activities
  - Ensure that the metadata complies with the principles set out in BCBS 239 – Principles for effective risk data aggregation and risk reporting.

- Collaborate with IT to insure that the metadata is complete and accurate with respect to existing and planned data schemas and datasets.
Start from the business architecture

Terminology

- Policy
- Process
- Capabilities
- Strategy
- Information
- Value Maps
- Initiatives
- Organization
- Business Model
- Stakeholder Roles

The Business Architecture defines the business capabilities, information entities and value streams.
Governance metadata strategy

- Reuse of standard terminology where feasible
- Structure and content derived from the business architecture
- Policies and principles derived from best practices

Terminology (ISO 1087), Financial Instruments Business Ontology (FIBO), Financial Instrument Global Identifiers (FIGI), Legal Entity Identifiers (GLEIS)

Governance Metadata

- Business Architecture
- Application and data architecture
- CMMI DMM® Model

© Thematix Partners 2016
<table>
<thead>
<tr>
<th></th>
<th><strong>Data Asset Owner High-Level Activities</strong></th>
<th><strong>Metadata</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provide stakeholders analysis</td>
<td>Value streams and stakeholders (Biz Arch, DMM)</td>
</tr>
<tr>
<td>2</td>
<td>Describe business capabilities and business processes (upstream/downstream data flows)</td>
<td>Capability-Information map Value stream – Capability map (Biz Arch)</td>
</tr>
<tr>
<td>3</td>
<td>Create/update glossary for data elements identified via business architecture, maps, and flows (vocabulary + metadata)</td>
<td>ISO 704/1087 conformant glossary; terms linked to capability, information (FIBO, FIGI, GLEIS...)</td>
</tr>
<tr>
<td>4</td>
<td>Review relationships between terms</td>
<td>Information entity details and entity-entity relationships (FIBO, Biz Arch, and existing schema)</td>
</tr>
<tr>
<td>5</td>
<td>Documentation of business rules/business requirements (to ensure Data Quality)</td>
<td>BPMN/DMN, RuleML or related semantics standards as appropriate</td>
</tr>
<tr>
<td>6</td>
<td>Contribute to Business Logical Data Model</td>
<td>Links to logical schema, applications</td>
</tr>
<tr>
<td>7</td>
<td>Implement Data Quality standards (Controls, SLAs, Data security, etc...)</td>
<td>Prioritize by value stream impact of quality changes (Biz Arch, DMM)</td>
</tr>
</tbody>
</table>
How terminology fits in

- **BCBS 239 requires** “integrated data taxonomies and architecture across the banking group” – i.e., a **“ground truth”** for risk data

- **A vocabulary provides the basis for ground truth.** It is comprised of metadata and a unified, logical, consistent approach to concept identification, naming, representation

- An ISO 704 based approach provides the rigor required to develop the terminology required as the starting point for a compliant business vocabulary
  - As a part of preliminary work – collecting relevant documents and other references from stakeholders when starting requirements gathering
  - A term excerption phase – to extract keywords and phrases from documents to seed a term list (may be semi-automated) – including preliminary definitions and identification of the source for every term
  - A term list curation and reconciliation phase – identifying the highest priority terms from the extracted list for expansion
  - An augmentation phase – to identify the subset of the terms required to support the use cases (derived from the business architecture, including competency questions to ensure coverage)
  - A glossary development phase, whereby every term has exactly one definition, annotated with source and context details, explanatory and other notes, where used references, etc.
  - The resulting glossary becomes the starting point for business vocabulary / ontology development (the output of step 3 on slide 16)

- **Analysis of relationships and other business rules, mapping to FIBO as appropriate, and identification of gaps where further ontology development is needed is the result of step 4**
Initial work was to define concept of a trade – across stakeholder organizations (front office, product control, financial control and risk, as well as back office), applications and data sources

- Primarily manual with some interviews, SME review
- Identified ~200 related concepts, with focus on a limited number of instruments (basic equities, bonds)
- FIBO provided a number of the basics, but there were a number of gaps related to financial institutions, regulators, registration of identifiers, contract details, settlement concepts
- Extensions were contributed back to FIBO – in Foundations, Financial Business and Commerce, Indicators and Indices, Securities … with more to follow)
Subsequent pilots

- Second pilot initiated around market valuation – the process of verifying / revising the preliminary valuation of an instrument from market data and/or standard calculations
  - Reused business architecture given that the work was product control driven
  - Used automation to extract vocabulary from policy documents in addition to preliminary business architecture
  - Identified ~500 concepts with numerous formulae for calculating valuation under various conditions, times of day, settlement adjustments
  - Created preliminary vocabulary, with review process, integration into FIBO extensions and finalization planned
  - Significant gaps identified in FIBO content; some related to market data may be addressed in the next revision to FIBO IND, with increasing coverage in downstream FIBO specifications

- Current pilot has two goals: trade data high-level hierarchy and collateral optimization – optimization of delivery / recall / substitution processes under increasing eligibility requirements
  - Business architecture for collateral optimization is less mature, but some aspects initiated as part of a study for replacement of an existing system
  - Project is in preliminary phase after re-prioritization
  - Using automation to extract vocabulary from policy documents in addition to preliminary business architecture
  - Little to no legacy content in FIBO to leverage in this area, although securities and debt basics will be useful
Tooling challenges

- Few tools follow ISO 704 for terminology development
  - SMEs were comfortable with Excel (many in middle office are controllers), which was too rigid to create the connections we needed
  - Terminology extraction tools used either constructed their own taxonomy, without FIBO / relevant context, or created lists of terms with relevance factors but no hierarchy
  - Resulted in construction of a prototype Access-based tool for vocabulary management and review

- Few ontology tools / triple stores provide decent support for integrated, collaborative definition, annotation, provenance editing and review process for SMEs
  - Resulted in extensions to Stardog to generate dynamic, linked vocabulary pages that SMEs can search, review, and annotate in a controlled way in the corporate intranet, without requiring any desktop tooling

- Additional tooling for logical data schema mapping and validation using ontologies is also needed
  - Prototyped process for schema validation via a series of SPARQL queries against Stardog
  - The reverse – from ontology to schema, proved impractical
Terminology tool prototype

Term Entry Form

ID 178

Term Agreement sub accounts

Definition If collateral positions are grouped into Margin accounts top down rather than bottom up, it will be easier to understand the different margin balances than (as current) aggregating individual positions. The current data has grown too complex for the single column in algo positions that it inhabits.

Harmonized 360

Agreement sub accounts

Status

Create Harmonized Search Harmonized

Extended Issues BusArch Facts Notes

Document Source

09 PL Risk and MV05062012
NASDAQQOMX definitions.pdf
https://en.wikipedia.org/wiki/Liquidity_risk
http://www.boerse-frankfurt.de/en/glossary/b/bobl+future+584
https://en.wikipedia.org/wiki/IMM_dates
http://www.investopedia.com/terms/p/position.asp
Daily Settlement.pdf
Swedish Bankers Association
Finance Norway
Oxford Dictionary of Finance and Banking
http://www.investment-and-finance.net/derivatives/f/futures-style-
Harmonized terminology entry for trade

<table>
<thead>
<tr>
<th>ID</th>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>401</td>
<td>Trade</td>
<td>a basic agreement between parties participating in a voluntary negotiation and the exchange of goods and/or services for some consideration (e.g., desired goods and services that someone else possesses); a commercial transaction involving the sale and purchase of a good, service, or information</td>
</tr>
</tbody>
</table>
Definition of market valuation calculation as an extension of FIBO
Nordea Business Vocabulary

Concept Identifier: *thm-mv-mv:ListedEquityMarketValueCalculation*

Label: *Listed Equity Market Value Calculation Formula Expression*

Annotations

Definition

A mathematical expression that is used to calculate the controlled market value of a listed equity. The expression is quantity * quote, where quantity is the quantity of a single named listed equity in a position and quote is a published market price for that named listed equity at a specified date and time.

Formula

quantity * quote

Parent Hierarchy

- *thm-mv-mv:ListedEquityMarketValueCalculation* - has as a parent
  - *fibo-fnd-uti-alx Expression* - has as a parent
    - *fibo-fnd-rel-rel:Reference*
    - *fibo-fnd-rel-rel:Reference*

Relationships

Concept: *thm-mv-mv:ListedEquityMarketValueCalculation*  
has property *fibo-fnd-uti-alx:hasArgument* with values exactly 1 from concept *thm-mv-mv:p1quote*  
has property *fibo-fnd-uti-alx:hasArgument* with values exactly 1 from concept *thm-mv-mv:p1quantity*

Concept: *fibo-fnd-uti-alx Expression*  
has property *fibo-fnd-uti-alx:hasArgument* with values min 0 from concept *fibo-fnd-uti-alx:Variable*  
has property *fibo-fnd-uti-alx:hasArgument* with values min 0 from concept *fibo-fnd-uti-alx:Constant*
Next steps

- Continue refinement / build out of the vocabulary publisher with Complexible with preliminary deployment in Nordea Markets Intranet
  - Content focus on high-level trade-oriented concepts and collateral optimization
  - Design focus on SME research and review processes
  - Integration via Stardog’s extraction and indexing capability to link vocabulary to relevant policies, procedures, product catalogs, etc.

- Prototype schema validation for the trade data warehouse project via SPARQL queries against the Stardog-based vocabulary repository

- Investigate use of Stardog to support related Business Architecture and terminology work on the front end of the development pipeline, again leveraging built-in extraction and indexing capabilities
Lessons learned

- The sales process within the organization takes a tremendous amount of time and energy, but is key to success.

- Early, small deliveries that demonstrate value help get more and more acceptance → enthusiastic support → true engagement and active participation.

- Centralization of the management and governance of the vocabulary is critical to ensure harmonization, coordinated development, quality, continuous improvement.

- Roll out needs to start with one domain area / deliverable, then two in parallel, then scale up from there as policies, methodology, metrics and related processes are established and mature.

- Select deliveries based on high value use cases and regulatory importance, determined via business analysis.