Enterprise Data Governance
Enterprise Data Governance

Reference & Master Data Management, Semantic Modeling

Pierre Bonnet
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testimonials from the MDM Alliance Group</td>
<td>xiii</td>
</tr>
<tr>
<td>Foreword</td>
<td>xxv</td>
</tr>
<tr>
<td>Preface</td>
<td>xxix</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xxxix</td>
</tr>
<tr>
<td>Introduction to MDM</td>
<td>xli</td>
</tr>
<tr>
<td><strong>PART ONE: THE MDM APPROACH</strong></td>
<td>1</td>
</tr>
<tr>
<td>Chapter 1. A Company and its Data</td>
<td>3</td>
</tr>
<tr>
<td>1.1. The importance of data and rules repositories</td>
<td>3</td>
</tr>
<tr>
<td>1.2. Back to basics</td>
<td>6</td>
</tr>
<tr>
<td>1.2.1. Past differences</td>
<td>9</td>
</tr>
<tr>
<td>1.2.2. The rich data model</td>
<td>11</td>
</tr>
<tr>
<td>1.3. Reference/Master data definition</td>
<td>12</td>
</tr>
<tr>
<td>1.3.1. Data initialized before use by transactional systems</td>
<td>14</td>
</tr>
<tr>
<td>1.3.2. Duplicate data</td>
<td>14</td>
</tr>
<tr>
<td>1.3.3. Data exchanged with third parties</td>
<td>19</td>
</tr>
<tr>
<td>1.4. Searching for data quality</td>
<td>19</td>
</tr>
<tr>
<td>1.4.1. Data quality</td>
<td>20</td>
</tr>
<tr>
<td>1.4.2. The quality of data models</td>
<td>23</td>
</tr>
<tr>
<td>1.4.3. The level of maturity of data quality</td>
<td>25</td>
</tr>
<tr>
<td>1.5. Different types of data repositories</td>
<td>27</td>
</tr>
</tbody>
</table>
4.4.3. The MDM as a springboard for transformation of IS ............................... 81
4.5. Summary of the return on investment of MDM ............................... 83

PART TWO: MDM FROM A BUSINESS PERSPECTIVE ............................... 87

Chapter 5. MDM Maturity Levels and Model-driven MDM ......................... 89
  5.1. Virtual MDM ........................................ 89
  5.2. Static MDM ........................................ 92
  5.3. Semantic MDM ..................................... 95
      5.3.1. Improved administration by business users .................. 97
      5.3.2. A greater reliability in the data repository .................. 97
      5.3.3. Preparation for MDM integration with the rest of a system .......... 98
  5.4. The MDM maturity model .................................. 100
  5.5. A Model-driven MDM system .................................. 103
      5.5.1. Variants ....................................... 103
      5.5.2. Hiding join tables .................................. 104

Chapter 6. Data Governance Functions ............................................ 109
  6.1. Brief overview ........................................ 109
  6.2. Ergonomics ......................................... 111
  6.3. Version management ..................................... 112
  6.4. The initialization and update of data by use context .................... 114
      6.4.1. The affiliation of contexts ............................ 115
      6.4.2. The automatic detection of shared data .................... 117
  6.5. Time management ....................................... 118
      6.5.1. Data history tracking ............................... 119
      6.5.2. Business transaction .................................. 120
      6.5.3. Validity period ..................................... 121
  6.6. Data validation rules ..................................... 122
      6.6.1. Facets .......................................... 123
      6.6.2. Integrity constraints .................................. 124
      6.6.3. Business rules ..................................... 126
  6.7. The data approval process .................................... 128
  6.8. Access rights management .................................... 129
  6.9. Data hierarchy management .................................... 130
  6.10. Conclusion ........................................... 131
### Chapter 7. Organizational Aspects

7.1. Organization for semantic modeling

7.1.1. The foundations of the organization

7.1.2. Data owners

7.1.3. The Enterprise Data Office

7.1.4. Does this organization involve risks?

7.2. The definition of roles

7.2.1. Data owner

7.2.2. Data analyst

7.2.3. Data architect

7.2.4. Data cost accountant

7.2.5. Data steward

7.3. Synthesis of the organization required to support the MDM

### Chapter 8. The Semantic Modeling Framework

8.1. Establishing the framework of the method

8.1.1. The objectives of semantic modeling

8.1.2. The lifecycle of semantic modeling

8.2. Choosing the method

8.2.1. The Praxeme method

8.2.2. Choosing another method

8.3. The components of Enterprise Data Architecture

8.3.1. The business object

8.3.2. The data category

8.3.3. Business object domains

8.3.4. Data repository architecture

8.4. The drawbacks of semantic modeling

8.4.1. The lack of return on investment

8.4.2. Lack of competency

8.4.3. The blank page effect

8.5. Ready-to-use semantic models

8.5.1. Software packages

8.5.2. Industry specific models

8.5.3. Generic data models
Chapter 9. Semantic Modeling Procedures ............ 187

9.1. A practical case of semantic modeling:
   the address ............................................. 187
   9.1.1. Non-compliant version of semantic modeling .... 188
   9.1.2. First draft of semantic modeling .............. 191
   9.1.3. Modeling of the lifecycle of the address ....... 192
   9.1.4. Complete semantic modeling of the address .... 197

9.2. Example of Enterprise Data Architecture .......... 199

9.3. Semantic modeling procedures ..................... 202
   9.3.1. Extended business operation ................. 202
   9.3.2. Elementary business operation ............. 207
   9.3.3. Single-occurrence and multi-occurrence
   business operations .................................. 208
   9.3.4. Fostering the upgradeability of data models .. 209
   9.3.5. Other principles ............................. 213

Chapter 10. Logical Data Modeling ............... 215

10.1. The objectives of logical modeling ............... 215
10.2. The components of logical data modeling ........ 216
10.3. The principle of loose-coupling data .............. 217
10.4. The data architecture within categories .......... 221
10.5. Derivation procedures ............................ 221
   10.5.1. Derivation of the semantic classes .......... 221
   10.5.2. Examples of derivation of semantic classes .. 224
   10.5.3. Derivation of elementary business operations 227
   10.5.4. Derivation of extended business operations .. 228
   10.5.5. Derivation of inheritance .................. 228
   10.5.6. Identifier management ....................... 229
   10.5.7. Calculated information ...................... 229
10.6. Other logical modeling procedures ............... 229
   10.6.1. Enumeration data type ....................... 229
   10.6.2. User message ................................ 230
   10.6.3. User interface components .................... 230
   10.6.4. Data documentation .......................... 231
   10.6.5. Naming rules .............................. 231

Chapter 11. Organization Modeling ............... 233

11.1. The components of pragmatic modeling .......... 234
11.2. Data approval processes ........................ 235
   11.2.1. Process example ............................ 235
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.4. Performance</td>
<td>282</td>
</tr>
<tr>
<td>A.5. Lifecycle of the Address business object</td>
<td>282</td>
</tr>
<tr>
<td>A.6. Insight into the XML schema</td>
<td>283</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>285</td>
</tr>
<tr>
<td><strong>Index</strong></td>
<td>287</td>
</tr>
</tbody>
</table>
Testimonials from the MDM Alliance Group

“Master Data Management and Information Management are key disciplines in any Company Architecture and Service Oriented Architecture initiatives. The MDM Alliance Group is delivering some solid added value in these areas by releasing these procedures in the public domain. Excellent work.”

Didier Boulet
Director SOA - THALES Corporate - KTD/Software

“Pierre Bonnet’s book on Master Data Management and Semantic modeling is a timely and comprehensive guide to creating solid foundations for Master Data Management within the company. The semantic approach to modeling master data represents an important step toward building industry specific standards and therefore massively reducing the risk and cost of Master Data Management projects. To have Pierre’s extensive architectural knowledge and vision within this area in print is a must for anyone embarking on a Master Data Management initiative.”

Owen Lewis
Director Product Development - Agile Solutions Ltd
“The data explosion in terms of volume and transactions has crystallized new paradigms such as Event Driven Architecture (EDA), Cloud Computing and other models such as SaaS. In such a massively distributed environment, completely virtualized and ubiquitous, the quality of data, their localization and transactional integrity will be dramatically critical. Another dimension of business that is magnifying the effect of the vectors described above is the breathtaking acceleration in the swiftness of business change that is forcing companies to rethink the manner in which they manage their critical data and information.”

“Rethink it means making a conceptual leap into this new economic model which means that companies will need to be guided in order to successfully achieve their transformation, thus increasing their competitive advantage.

The MDM Alliance Group is an initiative that will guide any company of any size to go through the transition phase in order to reach their goal. The modeling coupled with the semantic approach is an extremely powerful tool to increase the transparency of critical and valuable data while at the same time reducing the complexity of the architecture and multiplicity of interactions. The MDM alliance also provides tools and business templates for accelerating the learning and the operational effectiveness of companies. Bear in mind that data, since the beginning of mankind, has made and broken up empires and, from that viewpoint, nothing has changed except that today data is more important than ever due to the speed of business change, which means that any company must safeguard its highest competitive asset: data.”

Didier Mamma
Director EMEA Strategy & Development - Progress Software S.A.S.
“The majority of companies have taken steps towards a better management of their data capital or reference data. They are all looking for reassurance and want to benefit from the thoughts of others.

The MDM Alliance Group therefore brings an essential networking aspect, which is at the base of all good business practice for companies wishing to start a reference based management project. Atos Origin firmly supports the MDM Alliance Group.”

Laurent Schapira
BI, CRM & MDM Solutions Manager - Partner Atos Consulting

“The MDM Alliance Group is a debating forum, a think tank that prevents us from re-inventing the wheel. The modeling procedures and the ready to use data model help progress from state of the art to the reality of company projects.

The MDM Alliance Group is an accelerator helping to free the players involved be they users in industry or IT. Its contribution to the maturing process in the MDM market is well and truly established.”

Clément Roudeix
Director, BI & MDM, Financial Services – SOPRA group

“I have come from a traditional data modeling background where languages were not commonly used, so I have been trying to find some unbiased guidelines that would enable me to express my intentions in a more formal and universal way.

The explanation and examples given within the MDM Alliance Group documents have greatly helped me in this
goal. They have enabled me understand just how potentially wide ranging and valuable MDM is.”

Graham Chapman
Senior Enterprise Domain Architect: Information Enterprise Architecture at Inland Revenue, New Zealand

“Time (or money) to market a solution seems to shorten as we age. Should we waste in modeling? Think of the data silos “galaxy” accumulated in less than 20 or 30 years of data processing within your own organization, and ask why? It has been disorganized to such extent because nobody wanted to pay the bill for modeling upfront...so, everybody had to pay anyway, but afterwards (data transformation, data migration, data redesign etc.).

Data (and even more master data) is your asset. It is like your money in the bank. You may not have plenty initially but you most probably want it there for as long as possible. If you leave your safe on weak foundations, you will keep rebuilding and soon discover that everyone around preferred to use the bank as a service instead. Can you still afford to remain alone in your data mastering? Standards bodies like OASIS and others progressively deliver prebuilt vertical models like UBL, CIQ, ACORD, HL7 etc. However, those are intended for data flows and do not fit for data implementation. Packaged solutions (like SAP, Oracle Application, etc.) also provide out-of-the-box configurable data implementation schemas.

However those implementations tend to embed most of your master data (from the package prospective only), thus preventing us from keeping an homogeneous level on data quality, control and governance in the overall IT ecosystems. The same pitfalls apply if you try Corporate Services in SOA before data mastering. This book is a great opportunity for re-using from past experiences and
capitalizing on state-of-the-art pragmatic modeling
techniques. Beyond is the MAG initiative which is a way to
not only avoid the blank page but mostly to provide added
value from all of us, as long as we all play the fair game of an
open collaborative effort.”

_Xavier Fournier-Morel_
Co-author of the SOA Architecture Guidelines

“MAG is a community Pierre Bonnet founded to share
MDM Modeling procedures and pre-built data models. The
MDM Alliance Group publishes a set of pre-built data
models that include the usual concepts (Location, Asset,
Party, Party Relationship, Party Role, Event, Period [Date,
Time, Condition]) downloadable from the website. And some
more interesting models like Classification (Taxonomy) and
Thesaurus organized across three domains.

Although we may disagree about the semantics I do agree
with him that adopting this approach can help us avoid
setting up siloed master databases...unfortunately often
evident when using specific functional approaches such as
PIM (Product Information Management) and CDI (Customer
Data Integration) modeling.”

_James Parnitzke-James_
James is a hands-on technology executive, trusted partner,
advisor, software publisher and widely recognized database
management and company architecture thought leader

“The MDM apparatus (it’s a lot more than just
technologies), is a fundamental component to guarantee that
a Company Architecture is translated to an efficient IT
system. Moreover, without a correct MDM vision
implemented at the four levels (semantic, logical,
organizational and technological, see the Praxeme aspects) a
Company Architecture may be a complete failure and a total waste of money. The MAG work (method and pre-built models) is an invaluable step in our search for rationality for the IT system we are building for the near future.”

_Fabien Villard_
Secretary of the Praxeme Institute

“The Praxeme methodology, “Sustainable IT Architecture” and especially Master Data Management have helped me in consulting with various organizations in service-oriented architecture, as well as in implementing the process of IT system overhaul. I recognize the MDM Alliance Group as a precious source of information and exchange platform for methodology.”

_Jay Zawar_
Independent Consultant in SOA

“The correct management of a unique set of data repository is key to the company’s agility and financial performance. The MDM Alliance Group clearly demonstrates that this is a business opportunity.”

_Emmmanuel Laignelet_
Director of Evolan Solutions – SOPRA Group

“A reference book, that gives new found importance to companies’ IT heritage. In this book, Pierre Bonnet answers the concerns of IT architects giving a strong methodological framework for the management of data repository. He puts the data at the heart of the business in perspective and gives an efficient and pragmatic approach resting on a sure-footed base, the corner-stone of business data repositories.
In this book IT architects will find, the keys, a guide in the analysis and structuring in four layers (semantic, pragmatic, logic and software) and a collaborative approach (business/IT) which will help make their first MDM project a ‘success story’.

Olivier Sommerard
Technical Director – KHIPLUS

“Today a large amount of energy is expended to maintain, for better or worse, the quality of data. Whether this be in terms of data cleansing or data integration. But it is also necessary to take into account from the SOA point of view the wasted effort in the access management to the data. One can only fear that duplications of codes on the lower layers of SOA can only lead to the same in the upper layers.

From the point of view of reference data (data used or produced by several applications), from the moment a drive is made to increase quality and reduce costs, standardization is mandatory. Time and time again this principle has been shown to be true in many industries. Without doubt it is time that the software industry takes into account this standardization at the level of reference and master data management.

The logical consequence of this statement is that we can only wish for the normalization of reference and master data.

In this framework, the processes of the MDM Alliance Group, that is to say the semantic modeling of the MDM and the pre-built data models, will almost certainly become, progressively, a must. An ambitious goal, certainly difficult but promising, being that it offers off-the-shelf models.

Even supposing that these pre-built models cannot be used as they are with our respective IT systems, they do have the advantage of giving us the opportunity to
personalize our model without having to start from scratch, an approach which is often costly and reliant on a knowledge base for re-use. The approach certainly merits a trial.”

Jean-Pierre Latour  
Company Architect – SMALS

“There is an intense worry about the reliability of the KPIs and the reliability of the risk indicators. Because of this, it is not unusual that the same KPI on the one hand supplied from the production data base and on the other from BI do not have the same value. Which one is correct? The explanations by Pierre, on the quality of the data repository and by extension on the quality of operational data are totally convincing: It is possible to correct this state of affairs thanks to MDM. Starting an MDM project is attainable: Both methods and tools exist. Pierre explains them very clearly. It is clear that this work is a contribution that cannot be ignored and which is of great value, to the good management of an MDM project.”

Antoine Clave  
Information Systems Consultant – FIABILIS

“Master Data Management is a key technology to ensure the consistency and integrity of IT systems. Pierre Bonnet, who is a renowned expert in this field, provides a global view within the company and its IS. In his book “MDM and semantic modeling”, he introduces modeling techniques in order to define the best architecture and MDM usage for each IS, thus paving the way to IS maturity improvement.”

Philippe Desfray  
Director R&D – SOFTEAM
“I have met Pierre Bonnet “virtually” on an MDM-related Linkedin group. This allowed me to discover and join the MAG initiative wholeheartedly. Although a long time database professional (from DBA to Data-base Architect), my exposure to MDM started only a few years ago, with an IBM MDM “draft” solution, while I was involved in “architecting” an analytical DWH model, based on an operational ODS/ETL model, integrating a 3rd party CRM package, and complying with global, corporate wide reporting requirements. My first impression, outside the “silod” legacy (mainframe) world, was the semantic “chaos” introduced by the brave, new and open, distributed platforms, applications and database management systems. Beyond the versatile XML, the MDM (metadata management) was not standard, not accepted, but even had a lukewarm reception from the majors (IBM, MS & Oracle). It was clear that MDM was something else, a few abstraction layers higher, definitely aimed at the business alignment of the data and application semantics. SOA was hot and it promised relief to IT of all the legacy (mainframe, that is) pains. It is less hot now, but it is more mature and it has become absolutely clear there is no IT alternative to it (sic!).

While becoming an architect (an alliteration for a seasoned systems engineer, with the stress on both terms) entering the marvelous Company Architecture (Zakman’s) World I’ve realized that architects have missed one point: legacy (including new technologies, from MS, Oracle and IBM, among others) was not present explicitly, while (passive) data or (actionable) information were persisted tremendously, all over the place, like in a (flat) Babel Tower. The (world of) IT was (is) flat (courtesy of T. Friedman). The answer to that lack of “dimensions” was (already) there: Master Data Management - that business's own lingua franca, that IT should translate into local platform MDM “dialects”, ensuring the long-time aimed “integration” and
“interoperability” of heterogeneous databases and applications.

To my “beginners” experience, Pierre Bonnet has provided the SOA “basics” and he has promised the next and complementary book on MDM “all-you-can-eat”. The MDM Alliance Group (MAG) is the “place to be” and to discuss the future and mature MDM, tied to help business to seamlessly integrate and inter-operate data and information. Thank you Pierre!”

Nick Manu
Architect and DBA DB2 zOS & Linux, Crossroad Bank for Social Security, Belgium (eGovernment)

“Pierre is an expert in the MDM domain and understands well the intersection of SOA and MDM which is a rapidly emerging topic in Company Architecture. His work on “Sustainable IT Architecture” is an important contribution to the field. As more companies seek to extract the maximum business value of the existing and ongoing investments in IT, the sustainability model helps to coordinate stakeholders and to establish a higher level of functioning for today’s much maligned IT department. The integration of MDM into the SOA conversation reflects a mature understanding of the reality of Company complexity, but also provides a path forward for Architects and Practitioners alike.”

Miko Matsumura
Vice President and Deputy CTO at Software AG – author of SOA Adoption for Dummies

“Pierre has delivered, over a few short years, an impressive amount of guidance and best practices, be it with his colleagues at the Praxeme Institute, or by founding the Sustainable IT Architecture and MDM Alliance Group communities.
His vision tackles courageously the problems that IT has been facing for well over a decade now and he is offering an innovative, clear and proven path towards agility. He was one of the first ones to associate SOA and MDM, and more than that to provide a complete articulation between MDM, BPMS and BRMS as part of the Agility Chain Management System.”

Jean-Jacques DUBRAY
Co-author of OASIS ebBP, SCA,
Author of SDO Specifications and Composite Software Construction

“An essential step for those who wish to significantly improve (in other words, modernize) the agility of their IT system, the creation of a master data management foundation, an apparently simple issue, can come up against problems which could cause the failure of its implementation. Functional architecture, IS integration (exchanges) modeling; but also organization around data (governance) and potential added value exploration are some of the elements to be seriously considered once taken on. Through their recommendations, fully shared and added to by Micropole Univers, the MDM Alliance group delivers the Best Practice enabling the try conversion!”

Lilian Jouaud
Director, Company Information Management - Micropole-Univers
Foreword

If Master Data are the DNA of your business, MDM with Data Governance is its genetic engineering.

**Why governance of Reference and Master Data must be addressed as a pro-active business initiative and should not be considered as a curative technology project**

Reference and master data are the DNA of any organization. They define all the facets of your business and reflect the value and differentiators you provide within the market. Products, customers, channels, locations, geographies, accounts, organization, employees, suppliers, etc., are the critical assets at the heart of your business. The definition of DNA on Wikipedia can easily be applied to reference data: “(DNA) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints or a recipe, or a code, since it contains the instructions needed to construct other components of cells.”
As with DNA, reference and master data are the codification of your business, shared across all business lines, consumed by all IT systems.

So, if reference and master data are the DNA of your business, data governance is the genetic engineering. It means that the main purpose of a data governance initiative is to improve the quality, consistency and relevance of this data across the entire organization, not to fix issues after they have already occurred.

As in biology, improving the quality of your data cannot rely only on curative techniques. While data quality and data integration solutions are a key foundation for cleansing and connecting your data, you need to provide your business users with an active control on their shared data. Data governance is a pro-active business initiative that has a real benefit to enabling efficient and effective business initiatives or compliance requirements.

**Semantic data modeling and Model-driven MDM**

If your goal is to gain a real control over your data, you cannot avoid the data modeling exercise. Without a common and unified description of your data, how could business users share the same concepts?

In this book, Pierre Bonnet introduces the concept of a Model-driven MDM based on semantic data modeling. Far beyond traditional models, semantic models describe your data in meaningful terms for all stakeholders, including business users. This means you can design a rich description of your reference and master data and hide or bypass the usual constraints of IT relational oriented modeling such as join tables or frozen cardinality links. Then it becomes possible to define complex data objects, mix hierarchical,
relational and object-oriented concepts, configure business rules and validation controls, add documentation, etc.

Semantic data modeling associated with the Model-driven MDM allows business users to be involved from day one in your data governance program. With a model-driven solution they can easily collaborate on data modeling and quickly achieve a description of their data in their shared business language because “what you model is what you get”.

While data modeling requires effort, the realization of a mutual and shared understanding for the whole business will become of recurring importance for a pro-active data governance program. It is the first step to building the best version of the truth and establishing a unique reference and master data repository with active data governance capabilities.

Power to business users

Once data models have been designed, your data governance journey is not over. Building a unique description of your data is useless if business users cannot gain control of data itself. This means that to be active, an MDM/data governance solution must provide not only a central repository for storing the truth, but also a full set of data management features and a user experience for collaboration that maximizes adoption.

Pierre Bonnet proposes an exhaustive description of the core capabilities that business teams need to apply for pro-active governance of their data.

It starts with a rich user experience in order to provide data owners, stewards and managers with a collaborative environment for managing data and improving quality over time. It also addresses key issues such as version control,
security, business processes and rules and finally integration of master data across information systems.

Based on his extensive experiences at Orchestra Networks but also the MDM Alliance Group and Sustainable IT Architecture communities, Pierre proposes an unbiased perspective on MDM/data governance methodology, in order to help you build a truly pro-active data governance program.

Christophe Barriolade
CEO, Orchestra Networks