



Business Semantics Management at the Flemish Public Administration

*Information Maturity in Large Organisations through Business Semantics:
a Business Case at the Flemish Public Administration*

By

Pieter De Leenheer, Ph.D (pieter@collibra.com)

Aldo de Moor, Ph.D (aldo@collibra.com)

Stijn Christiaens (stijn@collibra.com)

March 15, 2010

White paper

Collibra nv/sa

<http://www.collibra.com>

sales@collibra.com

(T) +32 2 263 21 18 (F) +32 2 263 21 59

Adequate information management requires more than persistently storing data, a function databases already provide. Despite the rigorous structure that may have been imposed on data, if it cannot be disclosed to third parties, their value is practically zero. The ICT-outsourcing partnership between the Flemish Ministry of Education and Training (FMET) and EDS-Telindus, now HP, recognizes the need for **data governance**. This is shown by initiatives such as the development of a Data Warehouse and an Information Governance Organization.

In this article, we describe the support of the Flemish Ministry of Education and Training (FMET) in their next big leap towards information maturity. Using a Metadata Roadmap we combine metadata technology, methodology, culture, and organization that will allow FMET to semantically unlock their earned information and so revalue it.

Closed World Syndrome

The ability to unlock data is related to the ability to understand the hidden information in it. We aim to give a meaningful answer to questions about the following mission critical data:

1. **semantics:** what is the meaning of my data?
2. **utilisation:** How is my data used?
3. **provenance:** where does my data come from?
4. **governance:** who is responsible for what data?
5. **quality:** what is the quality of my data?

Many information systems suffer from a *closed-world syndrome*. They were designed from a naive assumption they have already stored all possible facts about the world, hence there never will be a need for large-scale data exchange with other systems. Moreover, a database is usually designed from a strong IT/IS perspective. Consequently, only the designer is familiar with its internal structure and rules. The technological nature of the syndrome lies in vendor lock-in of data caused by the fact that they are usually stored in proprietary (read: *closed*) formats. Obviously this does not make any sense in today's networked value constellations (like the World Wide Web) where online information exchange across business processes becomes central.

Just-in-Time Information

To answer the above questions, the ICT-outsourcing partnership aim at **just-in-time information (JITI)**, we define as *the ability to interpret the latent information in exchanged data in the right context and in a timely manner, without the help of the original designer*. The need for JIT information follows from the fact that information forms arguments during strategic decision making. Consider, for example, the minister who asks a report on the possible influence of the mother's educational profile on her children's school performance. People are in constant need for relevant JITI in order to analyse this correlation and finally to allow the minister to make well-founded political decisions. This calls for a *pragmatic* approach.

Metadata Management

For JITI to be effected, the ICT-outsourcing partnership is convinced that they, as earlier for data itself, must pay attention to the structured recording and publishing of data about data, also called metadata. The Flemish regulation describes metadata as:

“Documentation that describes the content and frequency of updating an authentic data source, and technical manner in which that resource can be unlocked.”

This definition remains vague when it comes to defining metadata, but clearly hints that metadata should have a dual utility (Akkermans and Gordijn). On one hand, it should refer to a formal specification for computer-based applications. On the other hand, it should refer to real-world objects in the business domain.

Enterprise applications, such as master data management and business intelligence claim to provide a total solution for metadata management. In practice, however, they produce redundant or anomalous metadata because they do not take into account the interest of other applications (again the closed world syndrome). Furthermore, the metadata is of a mere technical nature, which at best provides only a partial answer to those questions mentioned earlier. Indeed, as we will see, there are inherent limits on just how much technology can really help in this regard.

In current metadata management practices, the underlying methodological principles are usually ignored (Cardoso, 2007). Moreover, as proven by the narrow focus on technical metadata, they systematically ignore the subtle gap that looms between

(i) **knowledge sharing** between people on the business/social level (based on business metadata), and

(ii) **data exchange** between computerized information systems at the operational/technical level (based on technical metadata).

According to De Leenheer et al. (2010), the basic principle of metadata management lies in capturing the co-evolution of (a) social knowledge sharing and information needs that emerge from it, (b) the supporting computerized information systems, and (c) the metadata that enables data exchange between these systems. Doing so, it imposes a conceptual bridge between business (metadata) and technology (metadata).

Business Drivers

Based on our interviews, within the ICT-outsourcing partnership metadata management is fuelled from seven business drivers.

Documentation

“Sometimes people do not know what data is out there. Knowing the whereabouts of data starts with many calls within and between business units over and over again.”

(Anton Derks)

Business (intelligence) metadata is key to the provision of relevant documentation. Employees must document their data in a systematic manner in order to minimize the loss of know-how in case they would leave the organization. This is especially important in the context of the ICT-outsourcing partnership in which a regime of high turn-over dominates. Business Analysts must define relationships between documents that document the reasoning of their strategic advices. The collective practice of documentation will progressively decrease the overhead brought by repetitive calls for often the same latent documentation.

⁴ For a full list of semantic conflicts, see the book “Adaptive Information” by Jeffrey T. Pollock and Ralph Hodgson.

Communication

“No communication without metadata!”

(Jan Dejonghe)

Metadata facilitates communication both internally as externally. E.g., vendors of administrative school software should be able to understand the semantics the software should obey when checking in enrolment data with FMET. This can be formulated in the form of a technical data specification (e.g., in UML or XSD) that was generated from business metadata that is on its turn compliant to certain administrative regulations. Doing so, metadata provides a “language” that can cope more effectively with communication problems between business and the in-and outflow of external ICT consultants.

Reuse

Metadata accelerates the retrieval of assets and promotes their reuse. Currently assets, including reports, queries, data, architecture, technology, and licenses are defined ad hoc. In the planned service-oriented architecture, metadata will facilitate the retrieval and reuse of software services.

Codification strategy where data is codified in a common format that makes it easy to exchange and reuse is not the silver bullet. Personalisation complements this via right tooling and culture, which allows next to codification, the emergence of personal networks, engendering reputation among its peers. Therefore, it is equally important to address the provenance of data. If one knows the owner of a data asset, one can use the personal network (with Web 2.0 tooling) to gain additional know-how through socialisation.

Impact Analysis

“You hear now and again, “they have changed something to the data table, and now it does not work anymore.”

(Frans Decuyper)

Metadata is crucial to capture complex dependencies between different systems, people and applications, and to calculate the impact in the event of a change transparently. A precise impact analysis allows a more precise cost-benefit analysis. Moreover, it reduces the likelihood that they are subsequently surprised by unexpected side effects. The ICT-outsourcing partnership as-is often can't see the forest for the trees.

Disambiguation

“There are 180 000 teachers, more than one million students and thousands of educational institutions. FMET forms a large part of our society, hence it is certainly important that labels are attributed the same meaning.”

(Martin Maesen)

Metadata helps to get rid of inconsistencies or ambiguities. It is very valuable to know that a term has an unambiguous meaning. For example, the term “family” is fairly easy to understand for most people, but in the software application it has a strict sense, inferred from the legislation. Another problem is caused by (naturally occurring) homonyms. E.g., if one would check all decrees ranging from primary to higher education, the word “study area” is often attributed contradictory meanings, while one would expect this in FMET frequently used term should be intuitively obvious. There are also concepts that are poorly defined, and require additional interpretation. An example is a term “part exemption” that is applied in regulation but apart from that has few leads to its definition. This brings about the question whether we should take the definition into account are not. These are issues that continually recur in discussions with institutions about changing data models (“yes, but... what about the part exemptions”). Several discussions are repeated over and over again, and would be unnecessary if the relevant terms would be disambiguated properly.

Uniformity

“In my early years I have worked on a glossary: the most difficult part was to obtain consensus.”

(Marleen Deputter)

Unambiguous metadata is not sufficient. It is crucial that metadata uniformly applies to the entire organization and its stakeholders. It is very valuable for everyone to be sure that what is said in one place, is also valid elsewhere. A major source for the lack of uniformity is the wave pattern that FMET follows when it evolves: first, at the business level, there is a new decree (or set of related decrees). Next, this decree is “implemented” in a computer application. Inherent to its nature, legislation may change, giving rise to an organic growth in the implemented applications. An example is their Salary System and more specifically the lack of meaningful codes that are contained therein.

Compliance

Metadata plays an important role in regulatory compliance, a field that is not much explored so far (Ryan et al., 2004):

- Authentic sources: where data are spread over different systems, it is difficult to determine which of the systems is the original data source.
- Privacy: some data are covered by privacy legislation. When it is impossible to learn whether a piece of data is covered by privacy rules, it is difficult for an organization to comply with such regulation.
- Security: FMET has currently a Security Officer to verify regulatory compliance. Using metadata provenance of data can be logged.

Metadata Landscape SWOT Analysis

As with information management (Maes, 1999), is the strategy, structure, and operationalisation of metadata management not trivial because it must be able to align the complex and rapidly evolving information needs to the dito business needs

Technological support for metadata management in terms of software is necessary, but certainly not enough. Therefore we distinguish, alongside technology, three other dimensions in the metadata landscape: methodology, organization, and culture. Crucial is the development of a teachable and repeatable methodology consisting of a number of coordinated methods and techniques that allows the organization to perform different metadata management activities effectively and efficiently. This methodology should be anchored in an organizational arrangement of social roles and responsibilities, appropriate to the establishment of the Information Governance (see Decuyper, 2009). Finally, in order to apply the methodology properly, it is of great importance that there is a right culture of joint understanding, managing and using metadata in the cross-process information chain. In the context of knowledge management, Kankanhalli indicated that social incentives brought about by cultivation are essential to increase the usage of tools and methods.

Designing a solution requires a metadata landscape analysis that deepens each of the four dimensions, and consequently focuses on the internal weaknesses and external threats. To this end, we have interviewed more than twenty people on both sides of the partnership. Based on our findings, we performed a strength-weakness (SWOT) analysis (De Leenheer et al., 2009).

The resulting SWOT diagram is shown in Figure 1. An important strength is that metadata culture is already fairly mature. This is shown by the intention of its management, the curiosity of its employees (taking grass root initiatives providing success stories), and the targeted conceptual training. Another important strength is the outsourcing partnership that already successfully exists for many years.

On the contrary, the weakness lies in the fact that the current metadata technology, methodology, and organization are not adequate. Other threats are the constantly evolving policy changes that impede a smooth long-term rollout of information management. This is aggravated by the high turnover, we already discussed. Moreover, if FMET does not act fast, there is a threat that other public administrations will enforce their own metadata standards, making FMET a mere dependent entity.

The opportunities include the growing availability now of commercial semantic technology that can be deployed to gain information maturity; and the demand for exchange of experiences, especially strong in the context e-government. Moreover, the ICT-outsourcing partner is present throughout all ministries, which provides a wide platform for exchange of best practices.

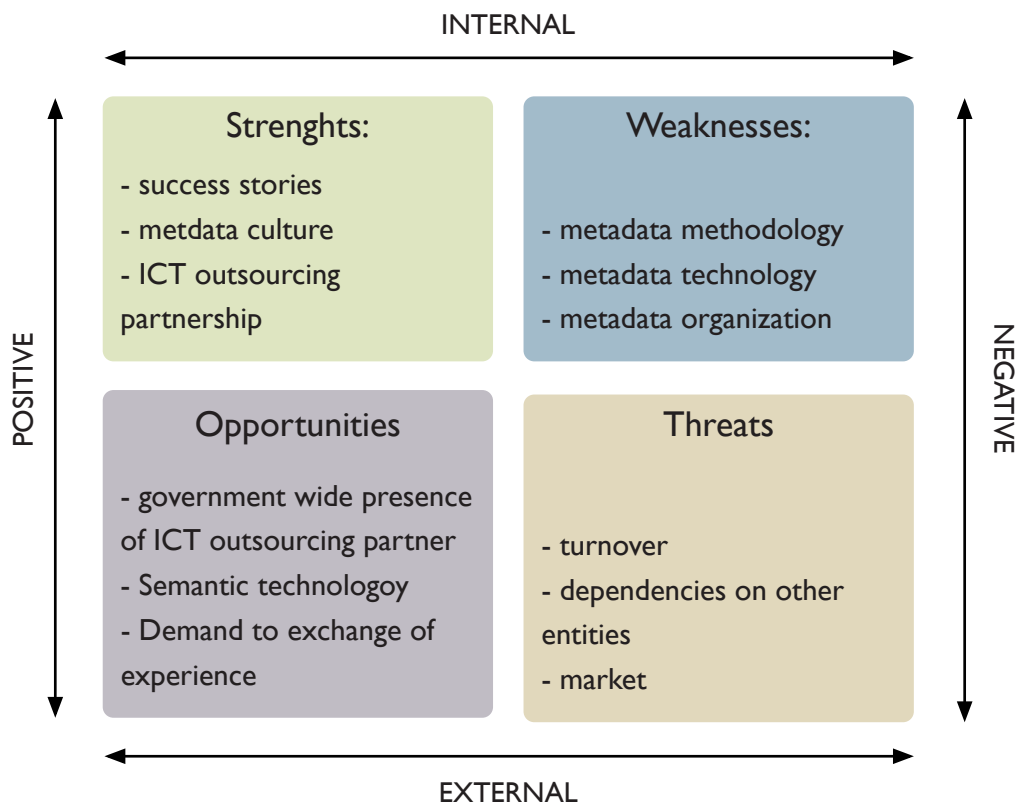


Figure 1 SWOT analysis of the metadata landscape (De Leenheer et al. 2009)

Business Semantics Management

Business Semantics Management (BSM) (De Leenheer et al, 2010) provides methodology, technology, culture and organization that enable parties to (i) obtain consensus on (the semantics of) key business concepts, and (ii) apply this consensus uniformly throughout the organization. Respectively, BSM consists of two complementary cycles: semantic reconciliation and semantic application (see Figure 1) that each group a number of activities.

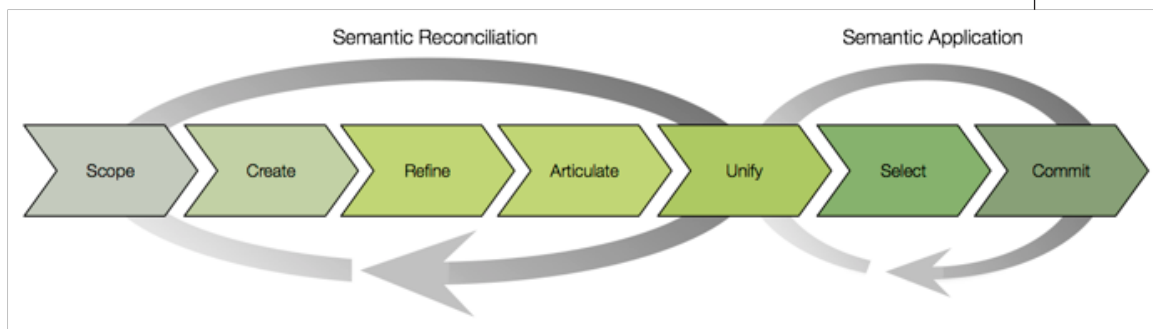


Figure 2 Business Semantics Management consists of two complementary cycles: semantic reconciliation and semantic application. Both cycles communicate via the unify activity

Contrary to what some middleware vendors claim, it is practically impossible to implement and maintain a central “metadata repository “. There are several reasons for this:

- the historically grown inconsistent and difficult to unlock collection of metadata sources;
- the structural independence of the business units (“entities”) within FMET;
- the intended independence of FMET towards its ICT-outsourcing partner HP;
- the general economic trend towards dynamic value networks;
- the increasing presence of “Web 3.0” where data and services are decentralized and accessible to each other via URIs. A vision shared with Semantic Web and Web Science communities (cf. IEEE CS Jan. 2010, etc.).

However, note that a fully decentralized approach is not feasible within FMET where business semantics are determined by regulations. Alternatively, BSM stands and falls with two initiatives: a business semantics glossary (BSG), and an enterprise information model (EIM). BSG provides a real-world reference and its derivate EIM provides a formal specification to build computational implementations.

Business Semantics Glossary

BSG provides a single point of reference for FMET’s business vocabulary and rules. To this end, we adopt OMG’s Semantics for Business Vocabulary and Business Rules (SBVR) standard.

The BSG supports the semantic reconciliation process.

1. Scope sets out the core terms that are actually needed to improve the information chain. Specific business drivers that want to resolve a weakness or threat in a certain application context fuel this activity. E.g., a communication breakdown in an IT/IS context may be caused by incomplete transformation of incoming personnel data from the more than 1,500 educational institutions to the data specs of the central salary system. The breakdown here is caused by a lack of specification of terms as “personnel” and “salary”. The derived need for manual translation (e.g., using XSLT) introduces a weakness, as defining the translation requires know-how about the resp. formats. Second in a business context, the lack of a uniform and unambiguous meaning of the term “study area” following externally imposed rules may form a legal threat. This observation initiates a reconciliation cycle where metadata related to “study area” are being reconciled. Note that we have oversimplified the scoping process here..

2. Create Every core term is syntactically described. E.g., Figure 3 illustrates the concept definition for term “Street Address” (specialisation of the type “Address”) in the BSG. Apart from its gloss, it is described in terms of facts (eg “business address located in street”) and rules (eg business address has exactly one zip code). For each term (within its context) there are certain roles to play as a “steward concept and a number of relevant stakeholders. Each change is carefully logged in order to be able go back any time. The activity is fed by implicit know-how of domain experts, by to automatic extraction of facts from existing metadata (see De Leenheer (2009) for a review of extraction techniques). According to our metadata landscape analysis, many FMET entities have isolated “grassroots” metadata initiatives. They manifest themselves in various forms such as taxonomies, keyword systems, glossaries, information about database fields, metadata in Web pages and content management systems (CMS). These use many proprietary formats as well as open formats such as XSD and UML used. Thanks to SBVR (OMG, 2007), the business concepts can be defined in a natural way, while at anytime a formal enterprise information model can be automatically generated in any format.

3. Refine Fact types created in the previous activity are refined so they are understandable to both business and technology. E.g., The somewhat technical term “Empl” becomes “Employee” or “EmplAddr” is revealed as the fact type “Employee residing in Residential address.

4. Articulate Since multiple users parallely render their perspective on a term, it may be that after the refine activity some facts and rules impose contradicting statements. During this activity, conflicts and inconsistencies are removed. Specifically designed algorithms may help here. Eg. in the Netherlands an address is uniquely identified by a combination of postcode and house number, while in Belgium a combination of postcode, street name and street number is required.

5. During unification, a new version of the business model is generated.

The screenshot shows the 'Home Address' term page in the Business Semantics Glossary. The page is structured as follows:

- Navigation:** Home, Vocabularies, Speech Communities, Semantic Communities, Statistics, Support Center.
- Search:** search...
- Breadcrumbs:** Flemish Public Administration > School Indications > Location (concept)
- Term:** Home Address (concept)
- Actions:** Add a Description, Add a Definition, Add an Example, Add a Note, Add a Synonym, Set Type.
- Vocabulary Entry Information:**
 - Vocabulary: Location
 - Steward: Felix Van de Maele (Change)
 - Articulation Completeness: 62.5%
 - Concept type: Object Type
- Stakeholders:** There are no stakeholders. Click here to add one.
- Taxonomy:**
 - Address (Parent)
 - Home Address (Specialized Term)
- Descriptions:** An Address where a company or Legal Entity is located. Added By Felix Van De Maele On 2010-03-15 17:56:27.0.
- Fact Types:**
 - Home Address has Postal Code
 - Home Address has Street Number
 - Home Address is located in Street
- Examples:** The Legal Entity "Flemish Public Administration of Education" is located at the home Address "Koning Albert II-iaan" with Street number "15" and Postal Code "1210". Added By Felix Van De Maele On 2010-03-15 17:57:41.0.
- Notes:** A valid Address is actually only an indication to stipulate that place. It is not a unique identification of a (cadastral) parcel, Legal Entity, or map position. It's purpose is to ease finding that place, and more importantly, to Address the location in order to increase the change of successfully reaching it. Added By On SNote.Date.
- Synonyms:** Locatie.Vestigingsadres (Delete). Add a Synonym.

Figure 3 Screenshot of the term “Home Address” in the Business Semantics Glossary. Even though the concept definitions look like natural language, one can automatically generate an enterprise information model that provides a formal specification.

Enterprise Information Model

An enterprise information model (EIM) is a “flattened” version (e.g., in UML or XSD) of the BSG that is generated in a timely manner. This turning point in the BSM methodology from reconciliation into application of the business semantics happens during the unification activity. The EIM serves as a uniform technical specification to implement semantic applications. We distinguish two activities.

1. **Select** relevant concepts are selected from the EIM for a particular application. It may be required to add additional application-specific constraints.
2. **Commit** Information systems in the scoped application are improved using the selected concepts. E.g., improve information systems’ interoperability through semantically describing and mapping the underlying data structures.

Once semantic applications are deployed, there is a possibility to feed back unexpected side effects or failures into the EBK during a new iteration of BSM. The cycle is repeated until an acceptable balance of differences and agreements is reached between the stakeholders that meets the requirements of the business. Gradually, closed divergent metadata sources are replaced with metadata sources that follow an open standard, and are kept coherent via BSG. The overall picture of this full-cycle BSM is illustrated in Figure 4.

Metadata Architecture and Governance

Of particular importance is that BSM becomes structurally and gradually embedded in FMET’s Enterprise Architecture by means of a Metadata Architecture. The basis for this is the EIM. Metadata Governance is concerned with establishment, modification, and implementation and monitoring of the Metadata Architecture by using BSM so that the ensuing business information systems will optimally contribute to the desired business results. While fueled by business drivers, the implementation of BSM is determined by a metadata charter, principles and policies.

- A **metadata charter** is a Memorandum of Understanding that provides motivation, goals, and key stakeholders. It provides a framework of roles and responsibilities and it identifies certain authorities.
- **Metadata-principles** establish start points for metadata management that must be respected. E.g., metadata must always be made publicly available and in an open standard format.
- **Metadata policies** contain clear guidelines for relevant actors within the organization to implement BSM in all its facets with sufficient quality and according to the principles.

Subject of these principles and policies are evolving concepts, metadata applications, methodologies and culture, but also the relationship with the ICT outsourcing partner. E.g., a policy that implements the above principles is a clear choice for the RDF or XSD format for publishing metadata.

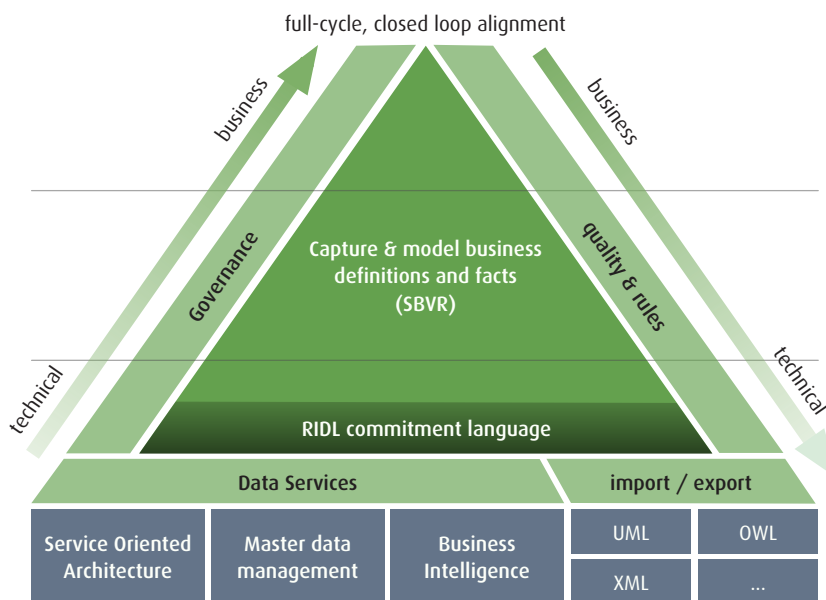


Figure 4 Full cycle business semantics management

Conclusion

The embedding of BSM in FMET requires a planning to implement a coherent set of projects in line with the ICT Strategic Plan 2010-2014. Eventually this should bring the partnerships information management in 2014 to an acceptable level. BSM constitutes a powerful catalyst to align and fuel just-in-time information management processes from the business and supporting technical data management processes. Doing so, ICT can be used effectively and efficiently.

The yardstick that we use to measure information maturity is the Information Maturity Model (IMM)³ (Figure 5). From implemented proof-of-concept we have reviewed some aspects BSM. If we project our findings from these PoCs on the IMM, we conclude that FMET was at IMM level 2 at the beginning of our analysis, and that the organization is not far off from achieving level 3. The five years plan aims Level 4.

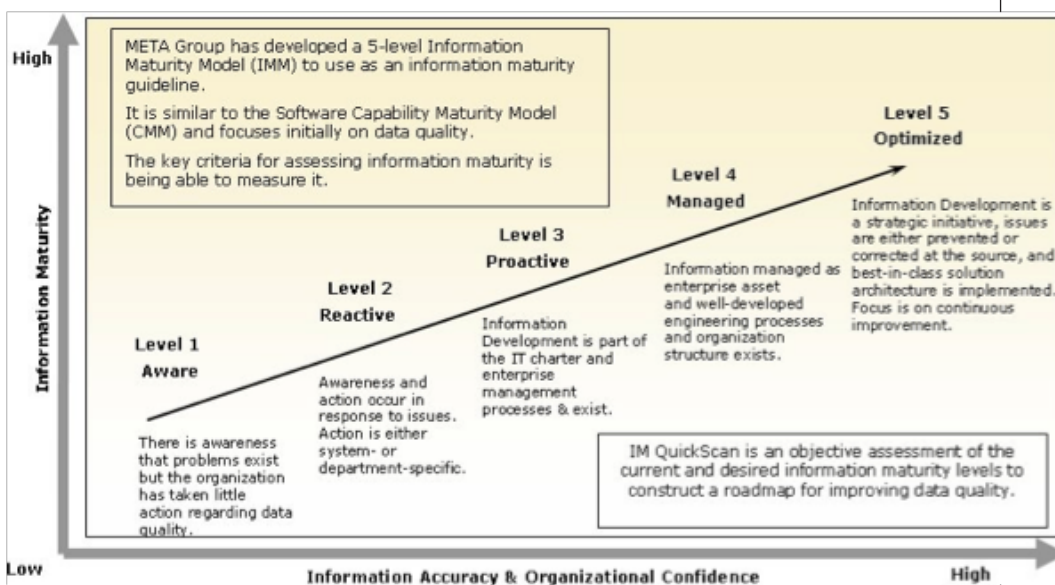


Figure 5 Information Maturity Model: a practical yardstick to qualify information maturity (by courtesy of Sean McClowry)

³ by Meta Group (Gartner) - http://mike2.openmethodology.org/wiki/Information_Maturity_Model.

Achieving Level 4 IMM will provide a platform with many new capabilities such as the development of Semantic (Business) Intelligence and Semantics-driven SOA. Moreover, the outreach of best-practice applications and associated metadata deliver a unique reputation to FMET. E.g., the European ISA program tries via its Semic.eu platform to promote sharing and standardization of metadata for public administrations.

References

- F. Decuyper. Information governance. Technical report, O&V, September 2009
- P. De Leenheer, S. Christiaens, and F. Van de Maele. Studierapport Metadatalandschap bij O&V, 2009
- P. De Leenheer. Ontology Elicitation. In Encyclopedia of Database Systems, eds. Liu, L. and Ôzsu, T., Springer, 2009
- P. De Leenheer, S. Christiaens, and R. Meersman. Business semantics management: a case study for competency-centric HRM. Journal of Computers For Industry, forthcoming, 2010
- R. Maes. Reconsidering information management through a generic framework. PrimaVera Working Paper 99-15, 1999
- OMG. Semantics of Business Vocabulary and Business Rules (SBVR), Version 1.0 (formal), 2008