OrangeGen

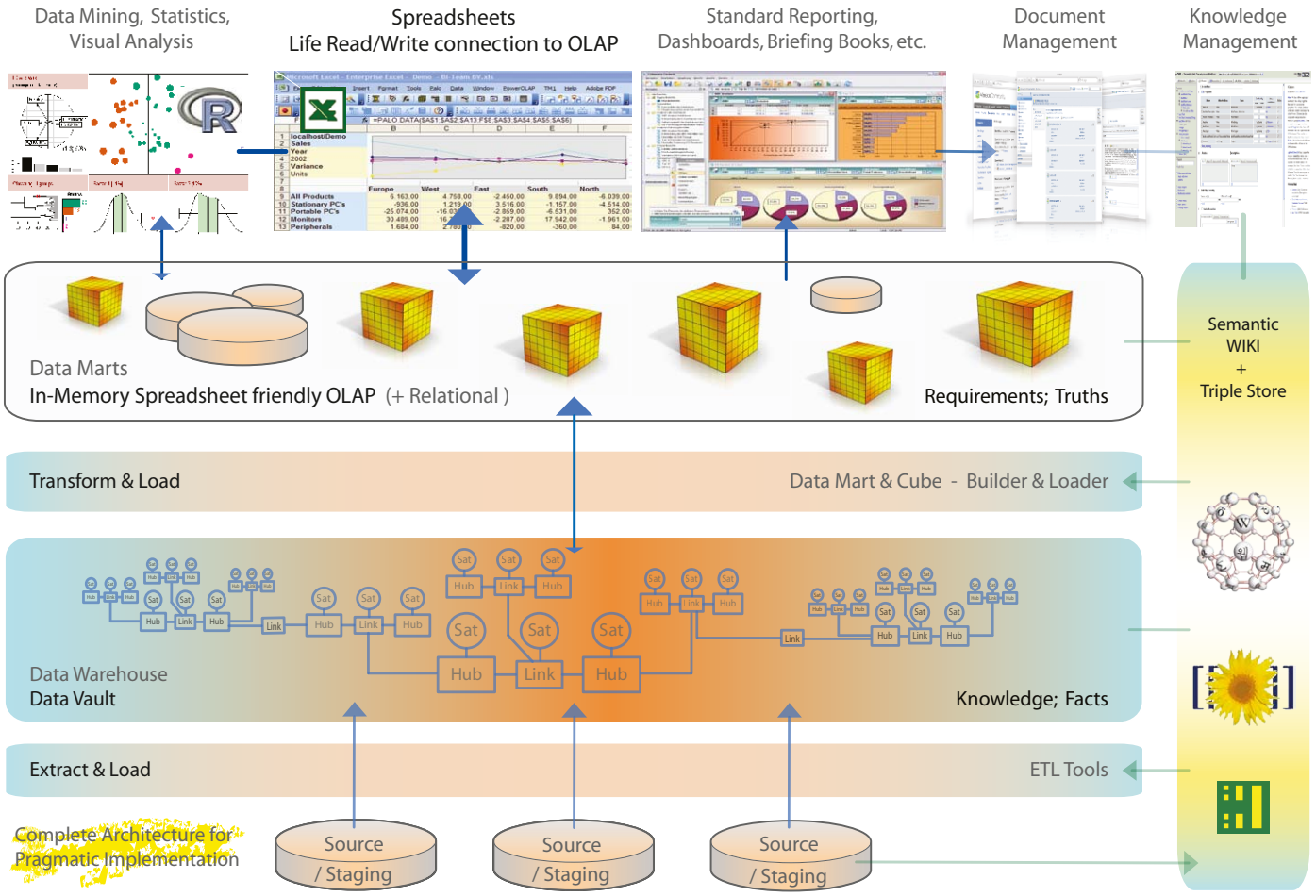
Meta- & Master Data Management

Semantic Wiki based Knowledge Management for Business Intelligence / Performance Management

Whitepaper

BI-Team
the Joy of Accomplishment

IBM Business Partner



The BI-Team Reference Architecture

This **universal and comprehensive architecture** provides for all conceivable Business Intelligence and Performance Management (BI/PM) functionalities. It offers optimal support for **evidence based management** activities, such as; Reporting, Dashboards, Planning, Forecasting, Quality Management, Capacity Based Costing, Scenario & Risk Analysis, Balanced Scorecard, etcetera.

This reference architecture consists of directives, the choice for specific techniques and best practices. The architecture can be realized with various products and is not dependent on suppliers. Specific advice for choices in your situation is available.

Distinguishing Choices

The well-known structure of a data warehouse with data marts on top, is realised with specific technologies:

1) Spreadsheet friendly OLAP for Data Marts
2) Data Vault as Data Warehouse concept
3) Semantic Wiki for Knowledge, Meta- & Master Data Management

The combination of a **Data Vault** data warehouse with **In-Memory Spreadsheet friendly OLAP** as prime technology for data marts, results in an unprecedented agile and structured environment. In addition the deployment of a **Semantic Wiki** caters for strong yet light-weight Knowledge, Meta- & Master Data Management.

Versatile & Widely Applicable

The integration of components for **Advanced Statistics** and **Data Mining** combined with **Document Management** and **Knowledge Management** makes it a complete environment.

Rapid Results

Number one critical success factor for BI/PM is the speed of implementation. This architecture is ideally suited for phased introduction and a surprisingly rapid implementation of projects. Directed towards a specific area of interest a Data Mart and (part of) a Data Vault are built. An application that immediately realizes value is implemented rapidly, while a strong foundation is being built. Return before investment is a real option!

Agile & Structured

This architecture offers a structured yet agile approach. Management by users is an important aspect. The delivery of quick answers on the always changing needs of the organisation, is the most important Critical Success Factor for the deployment of BI/PM. This design combines consistency and manageability with flexibility and adaptability.

Knowledge & Requirements

The distinction between *knowledge* and *requirements,* is an important aspect of this architecture. Knowledge and facts do not change, requirements and truths do.

In this vision ***a TRUTH is a statement about a FACT that is true in a specific CONTEXT***. For example, the turnover of customers are facts. The classification in customer categories based on turnover, creates a truth that may be useful for specific analysis. Facts are linked to knowledge, truths are linked to needs.

The **knowledge** in the organisation determines the **facts** that are stored in the Data Vault.

The **requirements** of the organisation determine which **truths** are available in the Data Marts.

The Foundations of this Architecture

This universal and comprehensive architecture offers all conceivable functionality to support effective Business Intelligence / Performance Management. It is vendor independent and not driven by new technologies, but by BI/PM experience and research into BI/PM reality. This architecture is build on the following corner stones:

1) Offer comprehensive functionality to support:
- working With the Data (Excel + OLAP)
- working Based on the Information (Web, pdf, etc.)

2) The Critical BI/PM Success Factors and Obstacles:
- Rapid Deployment, Rapid Results (< 6 months)
- Rapid Access and Data Delivery
- Requirements Change, are Unclear and yet Unknown
- The discovery and change of requirements must be supported by the approach and the software

3) BI/PM Reality:
- Almost always several BI/PM initiatives co-exist
- Incremental implementation is mandatory for BI/PM
- Active users must be able to do a lot themselves
- Everybody within the organisation must be able to use and contribute to the Meta- & Master Data

4) Non-Functional Requirements:
- *Auditability;* Transparency and traceability
- *Compliance*; Adherence to laws and regulations

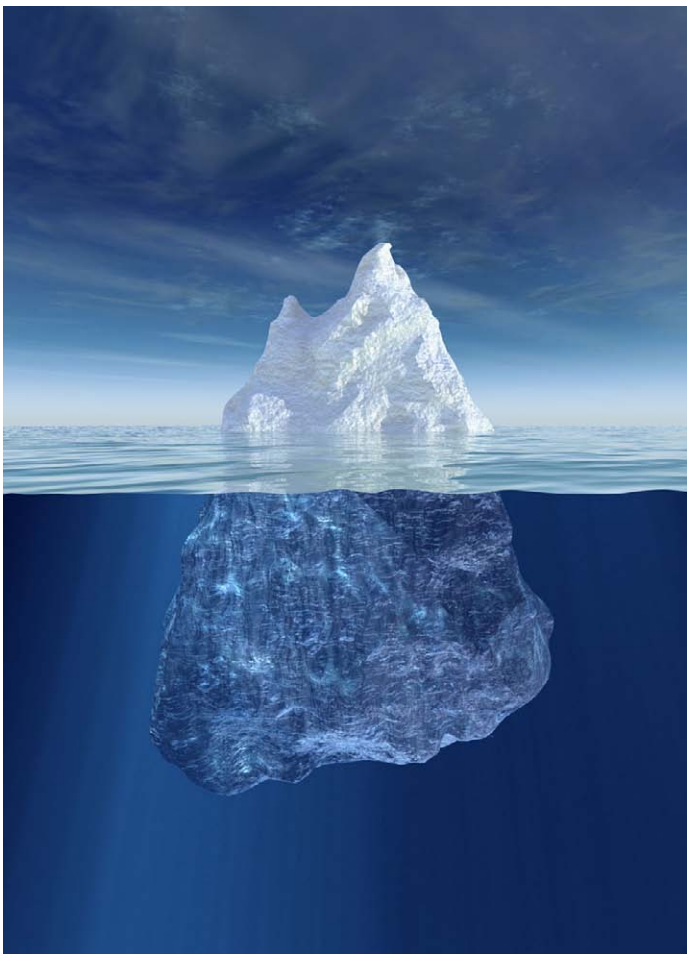
A Practical View

The stable, agile and versatile structure is the essence of this architecture, which is directly linked to specific technology choices. This architecture is not a nice dream based on new technologies. Spreadsheet friendly OLAP started in 1984 at Exxon and is recently fully adopted by IBM/Cognos. The general availability and proven stability of the recommended software make it a practical and reliable building plan.

The application layer supports all conceivable **Evidence Based Management** activities. This covers the **search for coherence** in the data, working **with the data** and working **on the basis of the information**. This is not the subject of this paper.

The choice of approach and technology regarding the Data Warehouse and Data Marts determine what is possible within a BI/PM environment, since these components form the foundation.

The layer for loading the data from the sources into the data warehouse and the layer for transferring the data from the data warehouse into the data marts are simple elements in the BI-Team Architecture diagrams. However, in the real world creating these layers is a complex and important issue. We will elaborate on that in this whitepaper.



The Tip of the Iceberg

In BI/PM the user applications, in particular reporting, constitute not more than the tip of the proverbial iceberg. Business Intelligence & Performance Management float on the invisible part under water: Data Access (ETL), Data Warehouse and Data Marts. Access to real-time, external and/or unstructured data creates ever greater challenges for IT to realise this invisible part.

The Whole Picture

If the requirements are taken care of in a sound manner, than the availability and quality of the needed data remain as the biggest project risk in BI/PM.

Building the data warehouse, data marts and thus data extract, transform and load (ETL) layers is the real challenge in BI/PM projects. If that is done right, than building reports and performing analyses is a relative simple task without real project risk.

Sales people may try to hide this part behind a cloud and try to make people believe that their boat is unsinkable. But ignoring this important part of BI/PM does not make it disappear. In the BI/PM world the *Titanic* is being rebuild on a regular basis, whereas a realistic and safe attitude and design do exist.

Meta- & Master Data Management is the Key to succesful Business Intelligence and Performance Management

The value of information stands or falls with knowledge of what the information means. That seems obvious, but it is often a stumbling block in the realisation of BI/PM applications and the use of BI/PM. Meta Data (knowledge about the information) and Master Data (reference information) are important in the whole information chain, from data entry to the interpretation of the information, for instance in reports.

Meta Data and Master Data

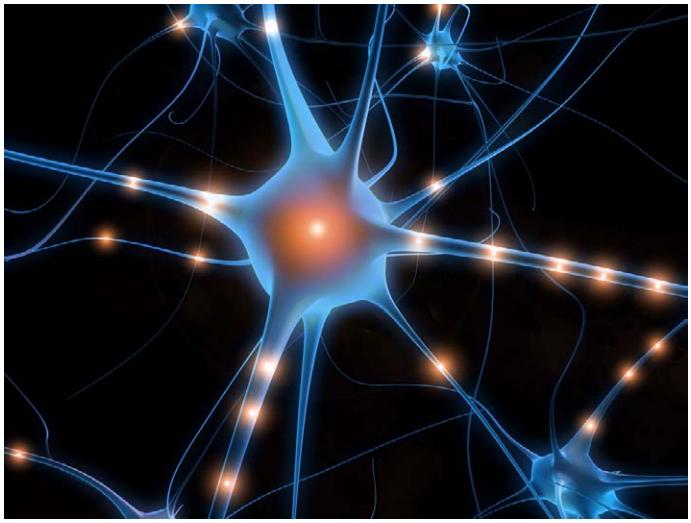
Just to be clear on this important subject, BI-Team has chosen to clearly name Meta Data and Master Data as separate, although intertwined, subjects.

These two aspects of knowledge management have their specific dynamics. Although it may look hard to distinguish in some cases, the difference is real, for instance in relation to consequences of changes.

Programmers would use the name *Objects* for Meta Data and *Instances* for Master Data, ontologists call this *Tbox* and *Abox, s*ales people rather do not talk about differences like this and choose one word.

- **Meta Data** is **data on data**, it describes data. Meta Data defines for instance which elements make up a complete and valid description of an object like a client, a product, a supplier, and what the data format of these elements in the database is.

- **Master Data** is **reference data**, it describes real *(Business)* ***Key Objects*** in an authorative manner. For instance the products, clients and suppliers. Also it contains reference tables of ID's for important objects like products, customers, suppliers. Most organisations work with more than one informationsystem. It is important to know what the '*gold standard*' of the information is, and where it lives.



At the Centre of the Nervous System

It is obvious that the quality of the information is a central issue in information delivery. It also is clear that a structured handling of the meaning of the data is the key to data quality. This is not only true for BI/PM or IT in general, but for the whole organisation and its stakeholders. Sharing a common body of knowledge and meaning is key to working together.

There are many aspects to the effort of gaining this common understanding and ground:

- **Technical**. Data is captured and kept somewhere with some technology that in the end defines what may be done with it.
- **Organisational** *(Business in USA terms).* The people that work with the data and on basis of the data, need to know exactly what they have to enter or what the data/information means.
- **Structural**. Data needs to be exchanged with stakeholders, government and regulatory bodies. An approach based on semantics constitutes a natural platform for cooperation. Also we want to be able to grow and change without having to start over again.

Web 2.0 and Pervasive Use

A great deal of the effort in the realization of Business Intelligence / Performance Management projects is spent on getting clear what the data mean, where the needed data can be found and on enhancing the quality of the data.

Most knowledge about the meaning of the data is in the heads of people with limited and unpredictable availability and a high price tag. They have to be able to add their knowledge in a very productive and flexible manner.

In the past ten years Internet based applications have revolutionized the way we work at a fast growing rate. Collaborative work has shown a lot of potential and MMDM is just such a case where this way of working is at its place. Experience shows that web 2.0 must replace something, it must not add work and it must create return immediate. Generating the Data Warehouse and Data Marts is such a case of return.

Furthermore for MMDM to be successful, involvement and commitment of the whole organisation is needed. Pervasive use must be possible without budget discussions and usage barriers.

The Encyclopedia of the Organisation

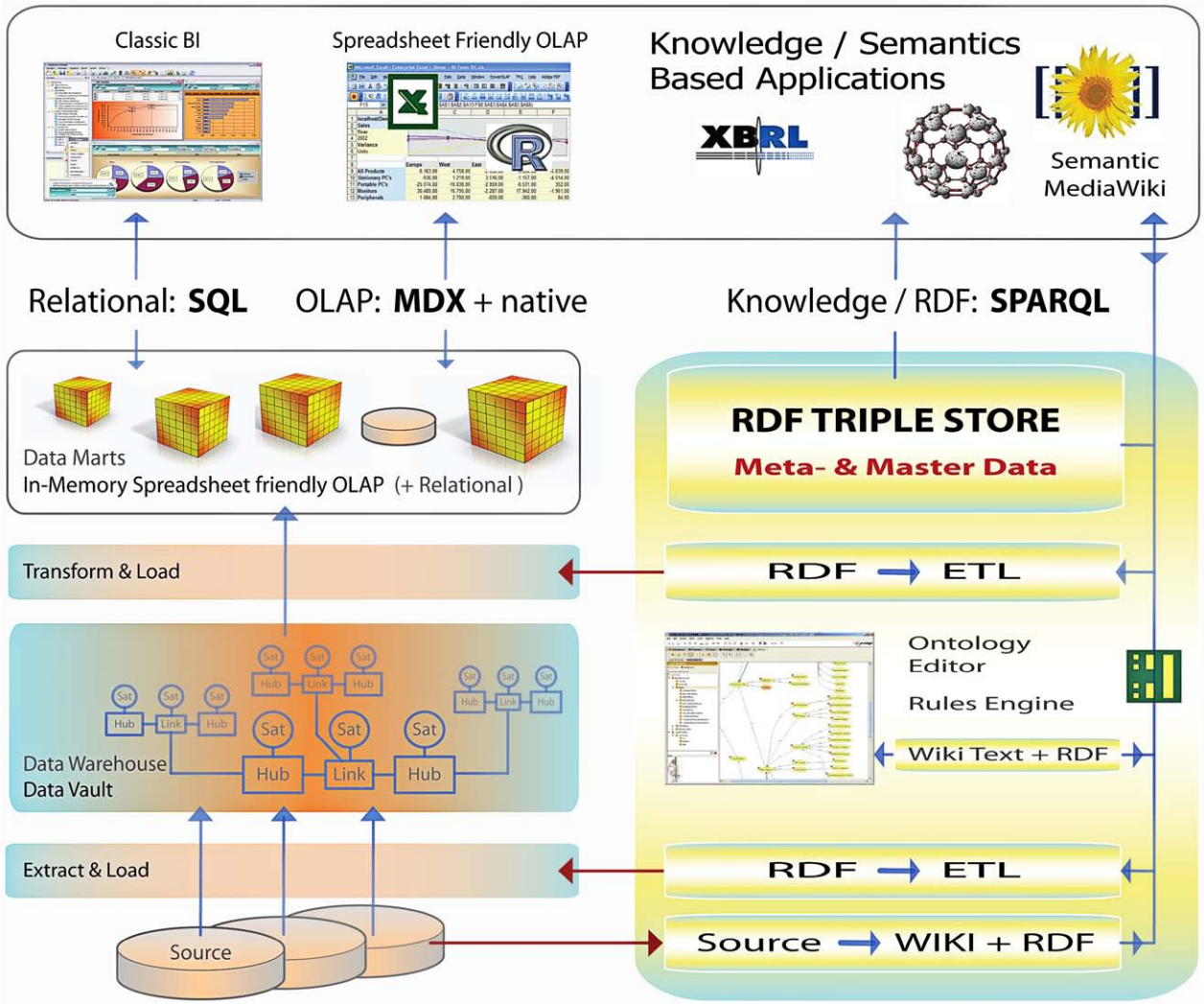
What we need is an Encyclopedia of the Organisation that can be build and expanded by all people in the organisation; we need something like Wikipedia.

Wikipedia is the well known public encyclopedia on the Internet. It started on January 15' 2001 and it has grown to more than 15 million entries in over 270 languages on January 15' 2010. It is recognized as being of comparable quality to well know professionally maintained encyclopedias. Wikipedia is used by a growing number of people as the reference of choice. This is a model for Wiki use in organisations.

Wikis offer a well know usage metaphor. Just a little training on how to use it within the organisation is sufficient to get the users going.

The proposed use of a Semantic Wiki for Meta- and Master Data Management (MMDM), is one of the most prominent aspects of the BI-Team Universal BI/PM Architecture.

Essential to a Wiki are the three A's;
Anybody can say Anything about Anything.
This has proven to be the basis for fast growth and high quality content. Fully open for additions and a guided way of changing, make a Wiki a very successful service. Discussion pages, an accessible history and version control features make it easy to let people freely add and change content, while still maintaining full control. Full Access Control is available in SMW+.



Wiki

Wiki Wiki, Hawaiian for speedy, quick, is exactly what is paramount in BI/PM. As can be seen from the diagrams; Knowledge, Master- & Meta Data Management constitutes an important part of this architecture, it is the integrating part. This is mainly embodied by a Semantic Wiki, a software service that is specially suited to be used for knowledge management in a professional environment.

A Wiki can be operated at low costs and is a low investment. Wiki pages can be created automatically. For example, for each product and customer a page with relevant information. Authorized persons can change and add to it. It is possible to include actual information and rich grapics in these Wiki pages. In this approach and architecture it is important that everyone in the Organisation may contribute to the Encyclopedia of the Organisation.

MediaWiki software

The Wikipedia foundation publishes MediaWiki, the software that Wikipedia runs on. It is licensed under the terms of the GNU General Public License (= Open Source). MediaWiki has proven to be robust, scalable, capable and is used day in day out by millions of users, most probably you are among them.

Extensibility is one of the great strengths of the MediaWiki software. The Semantic MediaWiki software is an example of an add-on to the MediaWiki software that greatly enhances its functionality. The MediaWiki software is a strong, stable, tried and tested foundation for our MMDM application.

Semantic Wiki

Semantics is the study of *meaning*. A semantic Wiki contains '*semantic*' tags in the text that make the *meaning* of the content accessible to a computer in a human readable manner. The *Semantically Tagged Information* can be queried with a *Semantic Query*.

Pages are at the heart of the structure of a Wiki, everything is a page with an URL as identification. Pages can contain links to other pages with the following tag: *[[link_to_page]]*. Pages that share some common aspect may be members of a category. This is identified by a link on the page: *[[Category:Name]]*.

Semantic MediaWiki adds one new tag: *[[semantic_statement::value]]* (*note the double colon*). Rather than creating a normal link from one article to another, the link is annotated with information what sort of relationship it represents. For example in a page with information on the city of Amsterdam these semantic tags could be present: a *Typed Link* represented by the tag: [[Capital of::the Netherlands]] and an example *Attribute*: [[population::750.000]]

The Benefits are:

- External reuse of the data in the Wiki
- Semantic templates for structured data entry
- Direct access to information instead of searching
- Generated Lists, no need for inflatory use of categories
- Inference, deduce implicit information
- Consistency and no redundancy
- A Semantic Wiki can act as a front-end to a RDF - Triple Store database management system, *see the next page*

Intellipedia

The very effective use of MediaWiki software by the American intelligence community (CIA, NSA, etc.) is an interesting and inspiring example of Wiki use.

In the aftermath of the 9/11 incident, the American intelligence community had to reconsider its way of working, the CIA asked their people to come up with ideas. The essay of *Calvin Andrus* with the title "*Towards a complex Adaptive Intelligence Community*" was a result and one of the inspiration sources for the Wiki initiative. Actually every manager should read it.

The intelligence Wikis went live in april 2006 and are called Intellipedia. In april 2010 its more than 250.000 users queried the Top Secret Wiki over 1 million times a week, the Secret Wiki over 1.3 million times a week.

One of the main differences with a public Wiki is that the users are known, so the reputation of authors can be established. Intellipedia is considered to be a great success, see http://en.wikipedia.org/wiki/Intellipedia

The Technology

For those interested how Meta- & Master Data Management with a Wiki works, we here describe the technology.
So called "*business users*" do not have to be aware of this, but after all it is tangible technology that we are implementing and not management theory.

RDF - Resource Description Framework

RDF is a world wide web consortium (W3C) standard. According to W3C: **"RDF is a framework for supporting resource description, or meta-data (data about data)".** According to *Shelly Powers* in the book *Practical RDF*: **"RDF is used to capture specific statements about a resource, statements that help form a more complete picture of the resource."**

In any language, simple facts can almost always be defined given three specific pieces of information: the subject of the fact, the property of the subject that is currently being defined, and its associated value. This correlates to what we understand to be a complete thought. Likewise a fact can be expressed as a RDF triple of the form **(Subject, Predicate, Object)**. It's like a short English sentence.

RDF may be seen as a triangulation of knowledge. Within the RDF specification, a RDF triple documents these three pieces of information in a consistent manner that ideally allows both human and machine consumption of the same data. The RDF triple is what allows human understanding and meaning to be interpreted consistently and mechanically.

These three facts describe RDF:

* A (RDF) triple is a 3-tuple, which is made up of a **subject, predicate** and **object.**
* Each RDF triple is a complete and unique fact.
* Each RDF triple can be joined with other RDF triples, but it still retains its own unique meaning, regardless of the complexity of the model in which it is included.

RDF statements are easy and natural to express in a Semantic Wiki, a Semantic Wiki can be regarded an embodiment of RDF. On a Semantic Wiki page that describes a specific Resource (the **Subject**), a tag is placed of the form: **[[Predicate::Object]]**.

These tags can be placed by hand in wikitext, with the aid of a semantic annotation editor or while creating or evaluating pages automatically.

RDF Triple Store

In this architecture, the Semantic Wiki operates as a Front-End to a RDF-Triple Store. A Triple Store is a Database Management System that is especially made to store RDF triples. Technically it may be an interface to a RDBMS, like for instance the Jena Triple Store is to MySQL or Oracle. Several companies supply a dedicated Triple Store, like the OntoBroker triple store from OntoPrise. SPARQL queries and a rules engine are the main functionality a RDF store adds. The BI-Team MMDM application communicates direct with a triple store and does not run without it.

SPARQL

SPARQL is a W3C standardised query language that can be used to access a triple store. It also defines the XML answer format and a Webservice. So every SPARQL provider (server) is a service oriented application. Just like SQL and MDX give access to relational and OLAP databases each providing their unique usability, SPARQL opens up a triple store with yet another set of unique strengths and new usage modes. SPARQL searches for and compares patterns in RDF graphs. It is designed to query non-uniform data and unlike SQL or MDX queries, always addresses the whole database at once. The user is not confronted with a particular database structure to handle.

Ontologies

Effective communication requires a common vocabulary. An ontology provides a description of the terminology, concepts and relationships for a particular area of interest. An ontology may be viewed as a declarative encoding of the meaning of the domain vocabulary terms, thus making it a key to enabling communication. For systems that are used by people whose understanding of a domain is not necessarily consistent, an explicit description of the important terms can be extremely useful.

Unlike a database, an ontology does not only contain facts, but also rules that are described in logic formula. This allows for the derivation of new facts from known facts and inference rules, it allows to deduce new implicit facts.

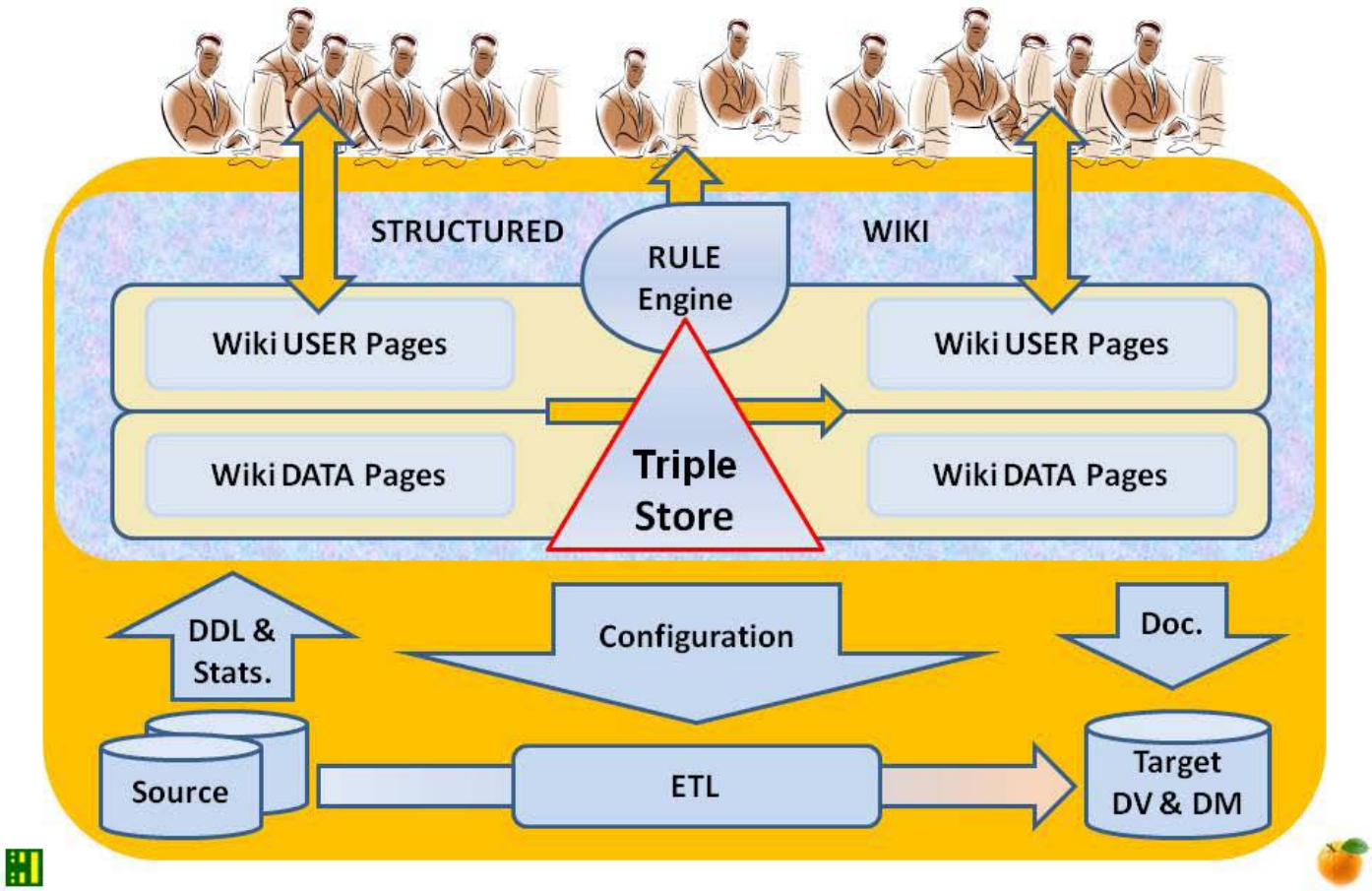
OWL

OWL stands for **Web Ontology Language**. OWL-Description Logic is a possible and likely candidate to express an ontology, but many more options exist. Whatever the language used, defining an ontology from scratch is not as easy as it may look. Luckily organisations are not an open world, thus a comprehensive ontology can be formulated in a finite amount of time.

Grow while you go

It is important that the Knowledge, Meta- & Master Data application is operational, while there is not yet a formal ontology present. Even better, an ontology that describes the organisation can be derived from the semantic Wiki. This allows starting at the bottom, users can immediately add information. However once an ontology is (partially) formulated, it is of great help to structure the information and create semantic forms that guarantee the right input of data into the Semantic Wiki.

www.OrangeGen.net



The BI-Team Software

The software is made up of 4 modules:

1) A module that creates semantically tagged Wiki pages from the data source's structure and content

2) A rules engine that evaluates the content of the Wiki and adds its conclusions to the Wiki

3) A module that reads the Wiki and creates a XML file that describes the target Data Vault or Data Marts

4) A module that reads the XML target-definition file and creates ETL code plus tagged Wiki pages that describe the target and its relation to the source

At any stage users can add their knowledge and influence results. Module two is used in an iterative loop. Each round interpreting the semantically tagged data and knowledge that users have entered. People that have knowledge of the data are asked for input until the right target definition can be constructed.

Generated Meta- & Master Data Content

The BI-Team MMDM application generates:
- Meta Data pages with the technical data on the database structure (schema / DDL) and a statistic fingerprint of the data in the current database (profile)
- Master Data pages per Business Key, for instance a page per customer, product, supplier, etcetera.

All Meta- & Master Data is now available on Wiki pages and can be found through Semantic Search and Lucene powered free search. The people involved can easy add their knowledge on the data and get information. Now it is also easy to see where data on reports is derived from.

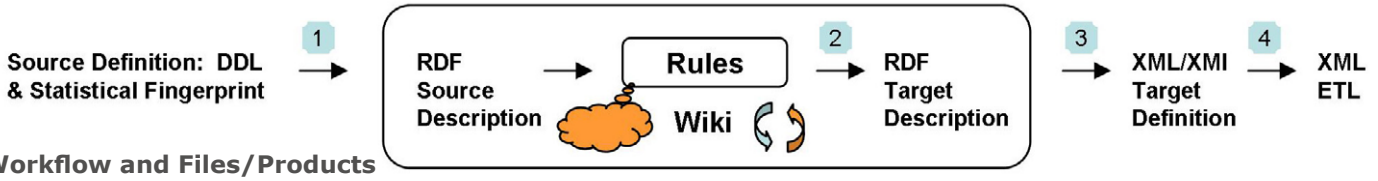
Standards based

The BI-Team software is written in Java/Groovy and will run on any platform. For deployment the Ontoprise SMW+ distribution of Semantic MediaWiki (SMW) is used. This is a professionally supported distribution by the people that as part of the HALO project create a lot of extra functionality for SMW, known as SMW+.

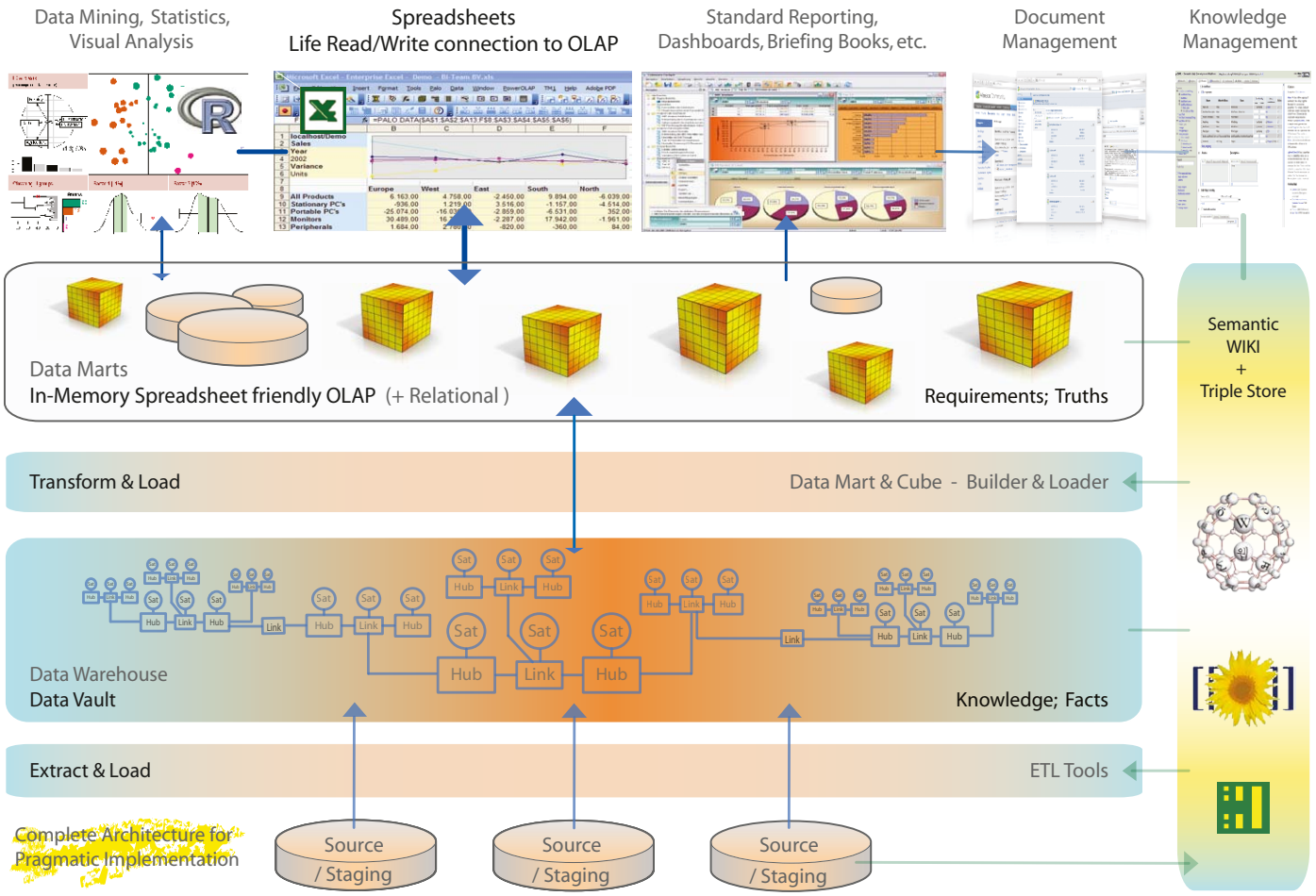
Knowledge Driven Configuration

An important aspect of this approach is the generation of the *"Part Under the Surface"*. The definitions of the Data Vault and Data Marts plus the loading code are generated from the knowledge that is encoded in RDF/OWL. **This is not Model Driven Software Design, we do not generate software. We use capable, tried & tested software that is configured by our application based on your knowledge.** That is why we call this ***Knowledge Driven Configuration***.

Generated Data Access makes BI/PM feasible for organisations and projects of any size.



Workflow and Files/Products



BI-Team Reference Architecture

This reference architecture describes a framework for the implementation of all aspects of Business Intelligence and Performance Management. This complete environment offers practical, clear and well-founded choices.

This architecture offers optimal support for **evidence based management** activities, such as; Reporting, Consolidations, Planning, Forecasting, Rolling Forecast, Quality Management, Capacity Based Costing, Scenario & Risk Analysis, Balanced Scorecard, Dashboards, etcetera.

Based on the needs and requirements of the organisation, those parts are implemented that will immediately yield a return on investment. Meanwhile a strong foundation is built and expanded based on the knowledge in the organisation.

Main benefits of this architecture are:

- Effortless scalable in scope, user count and data size.
- Structured & Agile, follows the organisation seamless to growth and divisions.
- Users are productive directly and without training.
- Spreadsheet friendly OLAP turns Excel into not only the most widely used, but also the most usable and useful BI/PM tool.
- Simple, practically functioning and affordable Knowledge, Meta- & Master Data Management is key to success in Business Intelligence and Performance Management.
- The Wiki approach provides a natural and structured way to build, manage and share knowledge. A well managed Semantic Wiki is of great benefit, it is:

The Encyclopedia of the Organisation.

The Foundations of this Architecture

This universal and comprehensive architecture offers all conceivable functionality to support effective Business Intelligence / Performance Management. The architecture is build on the following corner stones:

1) Offer comprehensive functionality to support:
- working With the Data (Excel + OLAP)
- working Based on the Information (Web, pdf, etc.)

2) The Critical BI/PM Success Factors and Obstacles:
- Rapid Deployment, Rapid Results (< 6 months)
- Rapid Access and Data Delivery
- Requirements Change, are Unclear and yet Unknown
- The discovery and change of requirements must be supported by the approach and the software

3) BI/PM Reality:
- Almost always several BI/PM initiatives co-exist
- Incremental implementation is mandatory for BI/PM
- Active users must be able to do a lot themselves
- Everybody within the organisation must be able to use and contribute to the Meta- & Master Data

4) Non-Functional Requirements:
- *Auditability;* Transparency and traceability
- *Compliance*; Adherence to laws and regulations

Only invest in what you need now...

Open to all future requirements...

BI-Team offers a versatile, complete and stable architecture, that can be implemented in parts, based on current requirements and a ***return before investment*** strategy.

Turn Excel from the most used, into the most usable and most useful BI/PM tool.

BI-Team
the Joy of Accomplishment

009h_BI-MDMWIKI
ISBN 978-94-90724-02-3

BI-Team BV
Minervalaan 23
1077 NK Amsterdam

tel. +31 (0)20 675 34 58
fax. +31 (0)20 675 72 35
www.bi-team.com