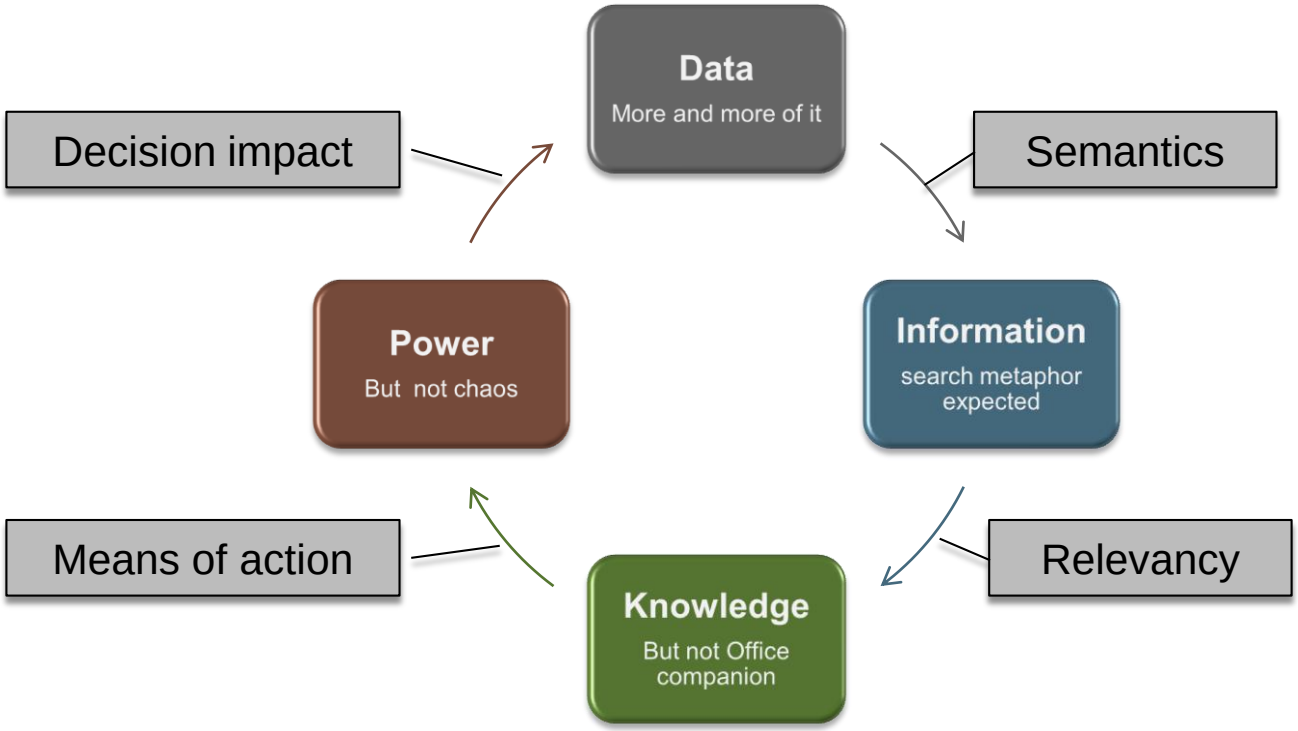
Turning raw data into actionable knowledge: Known challenges and new complexities

Yannick Cras, Chief Development Architect, Core BI Technologies
CSDM 2010

The opinions developed in this presentation are personal to the author and do not necessarily reflect the views of SAP.

Business Intelligence: a double-entendre

- Of course, as in « **Intelligent Business** »
  - Help businesses make intelligent decisions
- But also as in « **Intelligence Service** » and « **vivre en bonne intelligence** »
  - Help business users acquire intimate knowledge and understanding of their business, and the ability to act on it at their level.



SAP BusinessObjects

From Data to Knowledge: why this matters

Former Federal Reserve chairman Alan Greenspan testified before Congress that **insufficient data was one of the causes of the recent financial crisis**. Although Greenspan has previously praised computer technology as a tool that can be used to limit risks in financial markets, yesterday he said the **data submitted to the financial system was often a case of garbage in, garbage out**. Greenspan said that business decisions by financial services firms were supported by major advances in computer and communications technology. "The whole intellectual edifice, however, collapsed in the summer of last year because the data inputted into the risk management models generally covered only the past two decades--a period of euphoria," Greenspan said. **If the risk models were built to include historic periods of stress, capital requirements would have been higher and the financial world would be in better shape today**, he said.

Computerworld (10/23/08) Thibodeau, Patrick

SAP BusinessObjects

Some BI Challenges

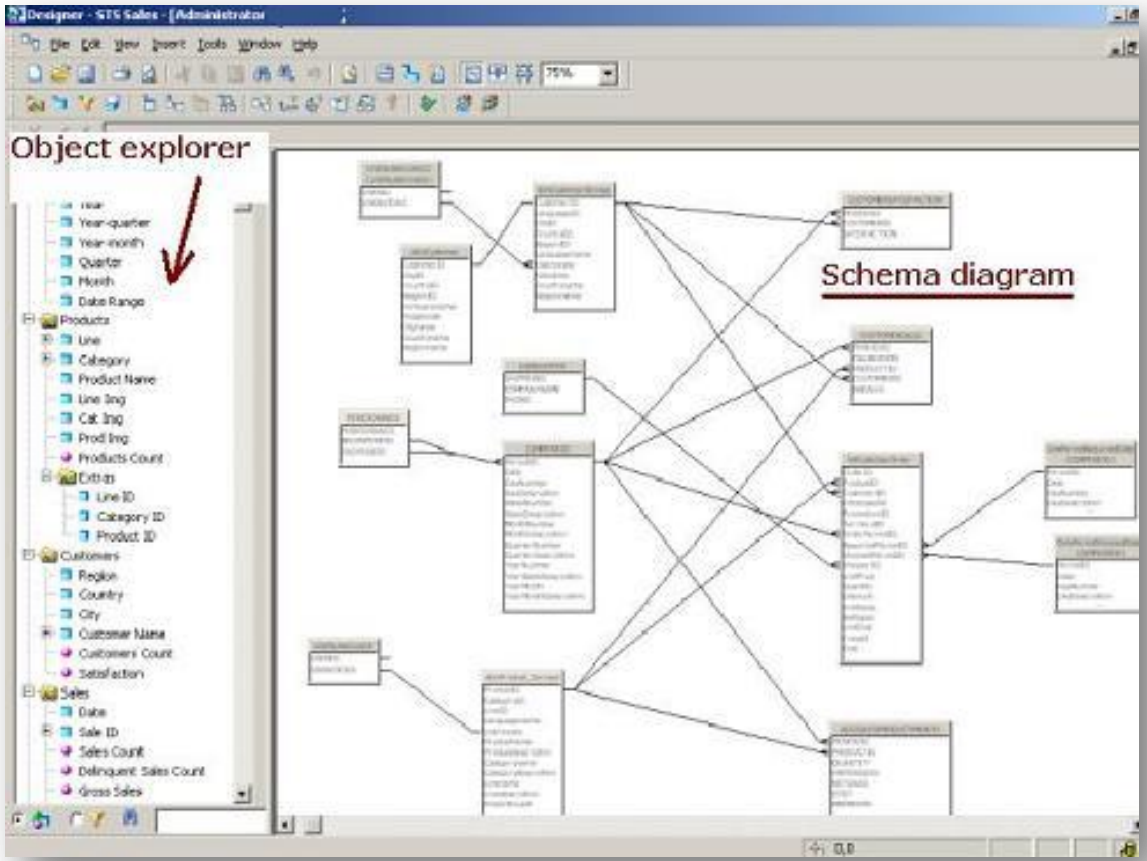
- Extracting and maintaining **data semantics**
  - **Reverse-engineering Business Semantics**
  - **Asking and answering the right questions**
  - What can we expect from unstructured data
  - **Fighting Semantic leaks**
- Retrieving **relevant information**
  - Personalization, modeling situations
  - What does colaborative ranking or social computing mean in the Enterprise
  - **Data Quality and the identity problem**
  - **Identifying what matters**
- Building **actionable knowledge**
  - Designing ad-hoc processes
  - Declarative representation of processes
  - **Towards User-designed Information Systems**

- Dealing with massive amounts of data with expected sub-second answers
  - « in-memory » cloud computing
  - From search to complex question answering

SAP BusinessObjects

Reverse-engineering Business Models

- The business intent of an information system is quite often lost after the design and implementation phases.
- We are forced to reverse-engineer the business semantics from the schema
- How to do it with minimal intervention of users?

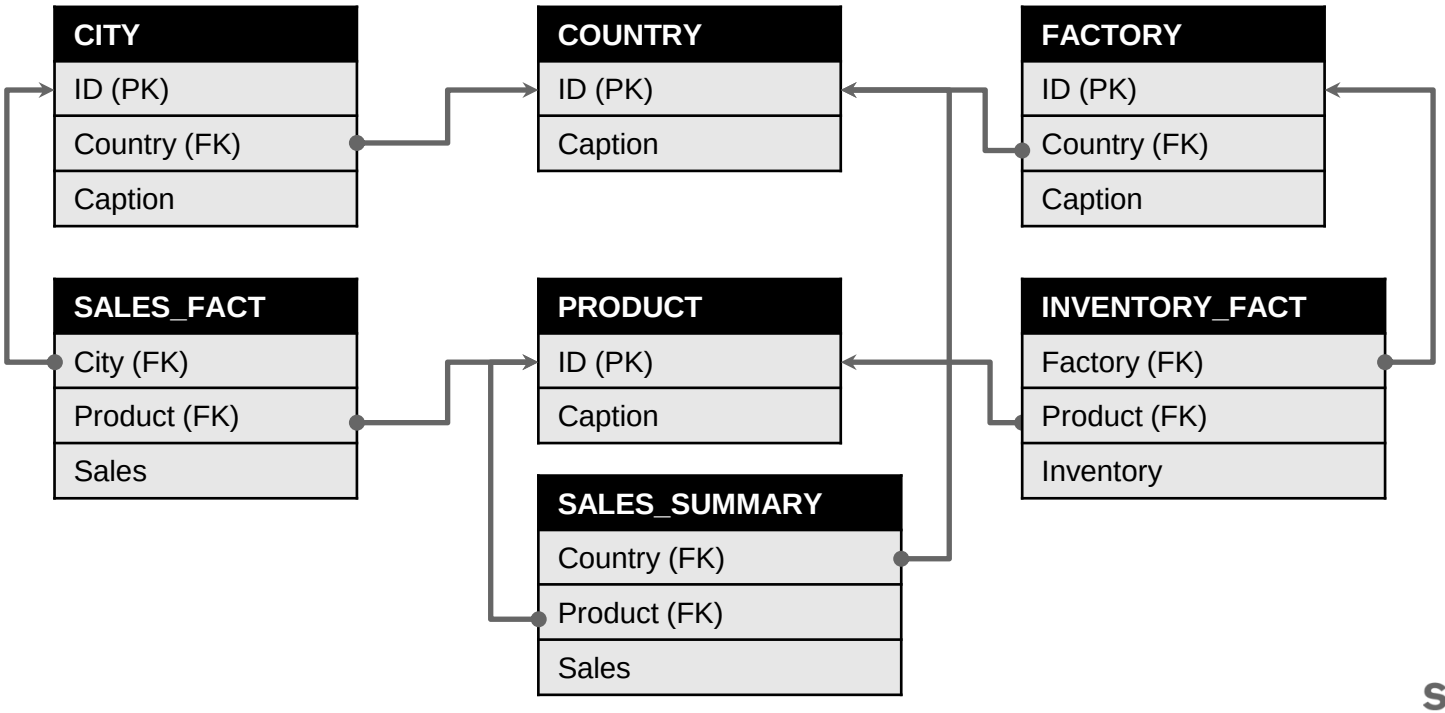


SAP BusinessObjects

Typical algorithms and their imits

- Identify primary and foreign keys and N-1 join paths
- Root tables as fact tables, leaves as dimensions
- Facts as measures or details
- Alias « fan » attributes, eliminate loops…
- Coping with de-normalized schemas, summary tables, many-to-many relationships etc.

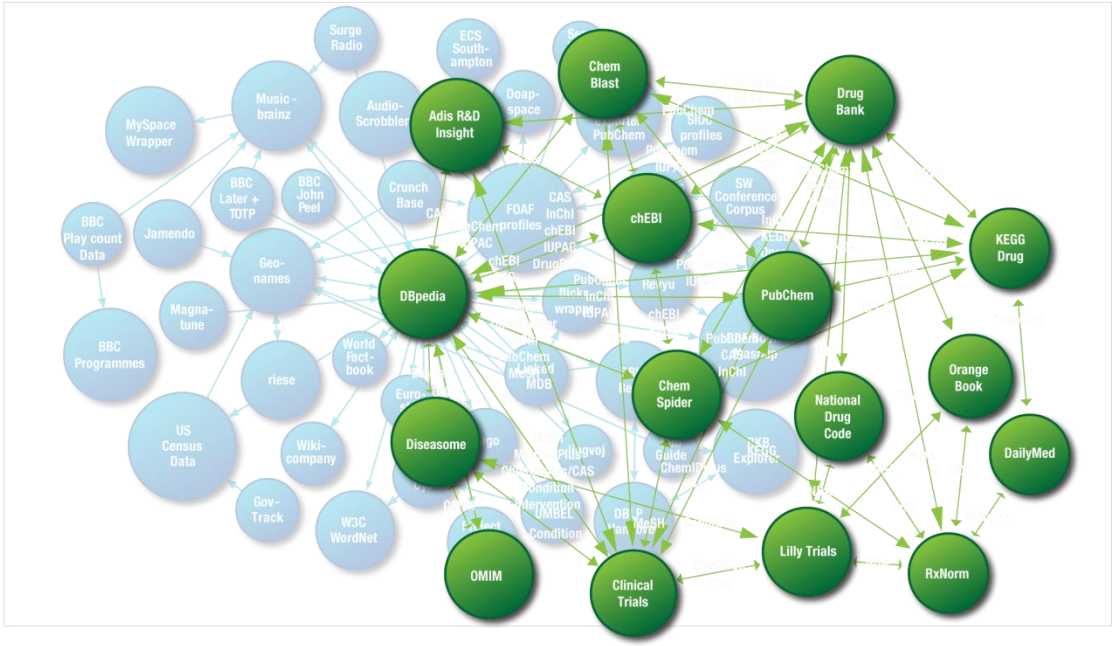
- Always a « Best we can do » approach. **Human validation is unavoidable**.



SAP BusinessObjects

Is « This » also « That »? The identity problem

- Reconciliating multiple views of the same entity
  - Global Unique ID/URI, surrogate key, stantard catalogs and taxonomies etc.
- A huge practical issue in the enterprise…
  - After M&As
  - Master Data Management underused
  - Half of datawarehousing projects fail in 1st year
- … and on the web
  - Merging ontologies
  - SameAs.Org



- **The complexity is economic and organizational**

SAP BusinessObjects

Giving answers is (relatively) easy…
… asking the right question is not

For each Product, average yearly sales and current inventory level per factory

- Average over all times, or since the product started selling?
- Only factories where this product is manufactured, or all?
- Can't be expressed naturally in SQL or MDX (client needs to compensate)

US Sales for 2010 and 2011

- 'and' is an inclusive 'or'
- Microsoft's English Query

For each customer, sales over the top-3 products sold in the country of this customer

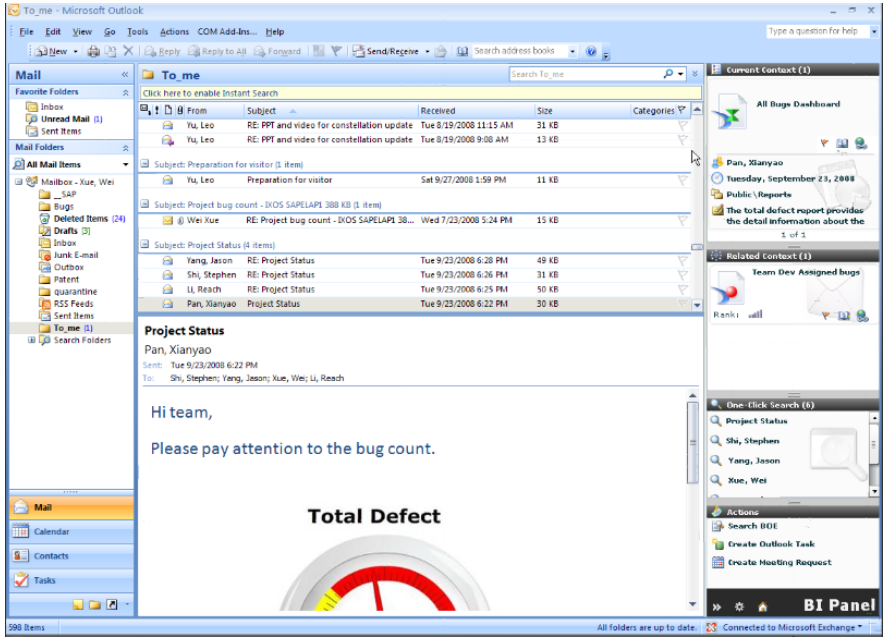
- Human beings just don't accomodate nested subqueries well

- **What must be the metaphor and semantics of user-friendly, intuitive yet expressive interfaces to ask business questions?**

SAP BusinessObjects

Identifying the relevant facts

- « wisdom of the crowd » does not work quite the same way in the enterprise
  - E.g. collaborative ranking, pagerank algorithm etc.
- More specialized data for less people
  - but more information about both
- How can we identify what is of interest to a business user?
  - Entropy footprint of data sets – outlier detection
  - Advanced visualization
  - Situational applications



- Complex interaction of multiple aspects (user, structured data, unstructured data)

SAP BusinessObjects

Fighting Semantic leaks

- Each time a piece of data goes into Excel, it stays there for good.
  - By far the most widely used BI front-end
- Excel accomodates data, not really information
  - 123 in B18.
  - How do I know that this is a headcount figure?
  - How do I know when and how it was computed?
  - How do I annotate the **fact**, not the cell, not the number?
  - How do I prove that I complied with Sarbanes-Oaxley?
- Protecting semantics
  - Give a « semantic URI » to any piece of knowledge (even transient)
  - Watermark it
  - Share the semantic definition, not only the value.
- Defining interoperable metamodels, shared by all applications

SAP BusinessObjects

User-designed information system: Business-Model Driven Architecture?

- There would be no need to deal with semantic leaks if the information systems were automatically derived and maintained from the business model in the first place.

- Our belief: **Business users should drive the IT, live, upon need – full stop.**
  - They own the business, they know the need, they generate the value.
- Our vision: **Give them the power to specify their needs.**
  - Business users generate the logic and models. IT governs, provisions and secures.

- Our belief: **IT should spend more time creating value than fighting entropy, not the opposite.**
  - Integration code, multiple sources of truth, poor MDM adoption, failing DWH projects…
  - Trying to make ice sculpture, but the fridge door is open.
- Our vision: **Create and preserve meaning, not just data**.
  - « $M42 » is data. « Net Revenue for 2007 in the US » is meaning.
  - Data can be duplicated, but its meaning is unique.
  - Move it. Share it. Combine it. Secure it. Watermark it. Compile it.
  - Why should I ever have to enter my home address more than once?
    - And why should it take more time to propagate than a domain name?

SAP BusinessObjects

Turning raw data into actionable knowledge: Known challenges and new complexities

Yannick Cras, Chief Development Architect, Core BI Technologies
CSDM 2010

The opinions developed in this presentation are personal to the author and do not necessarily reflect the views of SAP.