

Modeling Unstructured Data

Be prepared to analyze and design complex data requirements

About this seminar

As data requirements mature, as megabytes become cheaper, as CPU speed becomes faster, we as analysts and modelers will be faced with more complex requirements. Many of these requirements will depend on the efficient storage and retrieval of unstructured data. Merrill Lynch recently estimated that over 85% of all business information exists as unstructured data. The sheer quantity and complexity of unstructured data opens up many new opportunities for the analyst and modeler. Imagine requirements such as:

- Show me consumer feedback on my product from all *website discussion groups* for the last six months.
- Show me all photographs taken of the fountains in Rome from the summers of 2002 through 2007.
- Show me all contracts which contain a particular liability clause.

As with many areas of information technology however, there is ambiguity and confusion on the definitions and therefore the business and information technology implications of unstructured data.

This seminar will clearly explain unstructured data as well as related concepts such as ontologies and taxonomies in a context that will help you identify new business intelligence opportunities and design efficiencies for your organizations. Through a combination of lecture and exercises, we will reinforce unstructured data concepts and challenges.

You will learn to

- Distinguish between structured, semi-structured, and unstructured data
- Separate *Content* from *Requirements*
- Build relational and dimensional models which leverage text tagging, ontologies, and taxonomies

Outline

1. The necessity of 'Wayfinding' in today's rapidly changing business intelligence world
2. The need to expand our understanding of traditional modeling concepts such as meta data, domains, and class words
3. The distinguishing factor between structured, semi-structured and unstructured data
4. The recurring dream every CEO and CIO share
5. The necessity to separate content from requirements
 - Data Content Cube
 - Data Requirements Cube
 - Mapping Content to Requirements
6. The challenges that unstructured data will surface
7. The expanded set of activities the analyst and modeler will perform in the next 2-5 years
8. Let's practice! We will reinforce our understanding of structured, semi-structured, and unstructured data by analyzing and modeling:
 - A contact database
 - A photograph organization application
 - A course evaluation system
9. The power of text tagging
10. The difference between a taxonomy and an ontology

Who should attend

- Analysts
- Architects
- Data modelers
- Database administrators
- Project managers

Speaker bio

Steve Hoberman is a trainer, consultant, and writer in the field of data modeling. He balances the formality and precision of data modeling with the realities of building software systems with severe time, budget, and people constraints. Steve focuses on templates, tools, and guidelines to reap the benefits of data modeling with minimal investment.

After graduating Phi Beta Kappa from Queens College and completing a Master of Science in Information Networking at Carnegie Mellon University, Steve joined Bell Communications Research (Bellcore) in 1990 and started his data modeling career on an information engineering project. He was an analyst and data modeler on a team building an enterprise data model for the entire telephone industry. This project offered broad exposure to many cutting-edge disciplines of the 90's – object-oriented modeling techniques, Computer Aided Software Engineering (CASE) tools, and first-generation metadata repositories.

Building on the Bellcore foundation, Steve developed an interest in the human side of data modeling and the “next generation” techniques. In 1994 he went to work on Wall Street, performing data modeling work for many financial applications. Many of his “do it fast but right” modeling techniques came from the high pressure projects of Wall Street. In 1997 Steve joined Mars, Inc. as their Data Warehouse Architect. During his nine year tenure at Mars, Steve filled a variety of roles including Lead Data Modeler, Developer Team Lead, and SAP Functional Analyst. Each of these roles provided opportunities to grow and evolve his unique, experience-based approach to data modeling.

Steve taught his first data modeling class in 1992 and has taught over 10,000 people data modeling and business intelligence techniques since then. He has presented at over 50 international conferences in every format from short presentations to full-day classes, and has been selected to deliver keynote addresses at major industry conferences in North America and in Europe.

Steve is a columnist and frequent contributor to industry publications. He is the author of several data modeling books including *Data Modeling Made Simple*, *Data Modeler's Workbench*, and *Data Modeling for the Business*. With interest in building a data modeler's community, he founded the Design Challenges group, which today boasts more than 3,000 data management practitioners who tackle monthly data modeling puzzles. (Add your email address to join this group at www.stevhoberman.com.) Steve is an innovator in data modeling and the inventor of the Data Model Scorecard®, which has quickly become the standard for data model quality.