

Metadata as a byproduct of Object Oriented Design

Professor John A. Lehman
University of Alaska Fairbanks

What is metadata?

- Literally “beyond data”
- In practice, data about data
- ALA: “structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities”

Sources of alternative views

- Different disciplines develop independent models of the same phenomena
- Tools from one discipline can help in others

Alternative views of Metadata

- Library view
- Computer Science
 - Database management view
 - Object view
- First order vs. second order

Library view

- Source: cataloging
- Influence of batch technology (MARC)
- Standard category set

Previous computer science view

- Influence of batch technology
- Standard category set
- Examples
 - Early “interactive” systems (e.g. SABRE)
 - EDI

Current Computer Science view

- Source: entity modeling, object modeling
- Main new application: data warehousing
- Assumptions
 - Form follows function
 - Data structure derived from problem
 - Importance of use cases

“English” definitions

- Entity: what do we need to know about (conceptually)?
- Attribute: what do we need to know about it?
- Instance: a real life example of an entity
- Value: the data (attributes) describing an instance

Sources of conceptual models

- Linguistic convention
- Intellectual (disciplinary) consensus (e.g. “standard” metadata sets)
- Data analysis
 - Factor analysis
 - TEI
- Practical utility

Alternative views of underlying “reality”

- Different conceptual models (entities)
- Different sets of metadata (attributes)

Sources of metadata attributes

- Design documents
 - Entity models
 - Object models
 - Other system documentation
- Algorithmic derivation

Representation of metadata

- Database schemas
- Markup languages (e.g. SGML/XML)
- Use of DTD to describe entities and their attributes

EDI as an example of paradigm shift

- Batch EDI problems
 - Need for universal (or industry-wide) common set of metadata
 - Inflexible custom software
- Current approach: XML + DTD
 - Still require common data architecture
 - Use of XML parser as middleware