One-To-Many ...or Much Too Much?

ABSTRACT

A tutorial about efficient, effective requirements analysis that promotes information quality by recycling some best practices, reconditioning some others, and debunking and discarding the rest.

BIOGRAPHY

Joe Maguire
Principal Analyst and Consultant
O'Kelly Associates

A 28-year veteran of the computer industry, Joe Maguire is an analyst and consultant specializing in data management and requirements analysis. His hard-won perspective is informed by broad experience including twelve years in product development for software vendors (Digital, Lotus, Microsoft, Bachman Information Systems); thirteen years as an independent consulting data modeler and requirements analyst for clients (ranging from small startups to Fortune-10 behemoths); and three years as an industry analyst for Burton Group and Gartner specializing in best practices in data management. He is a much-published author whose books have been praised by a wide range of media outlets including The Mathematica Journal, The Data Access Newsletter, The Boston Sunday Globe, and National Public Radio. A frequent public speaker, Mr. Maguire returns to MIT IQIS for the third consecutive year.

One-To-Many
… or…
Much Too Much?

MIT IQIS 2011
Joe Maguire
joe.maguire@okellyassociates.com
josephmaguire.blogspot.com

Joe Maguire

- Analyst/Consultant:
  – Data modeling; requirements analysis; nexus of data and content; collaboration
- Previous Work
  – Decade + in product development for SW vendors
  – Decade + consulting (data + process modeling)
  – Industry analyst (Burton, Gartner)
- Publications
  – *Mastering Data Modeling* (Carlis and Maguire)
  – Dozens of papers: industry analysis, best practices
  – *http://josephmaguire.blogspot.com*

One-To-Many … or … Much Too Much?

- A tutorial about efficient, effective requirements analysis that promotes information quality by recycling some best practices, reconditioning some others, and debunking and discarding the rest.
- As we move forward you—the attendees of this tutorial—will choose which "best practices" we consider.

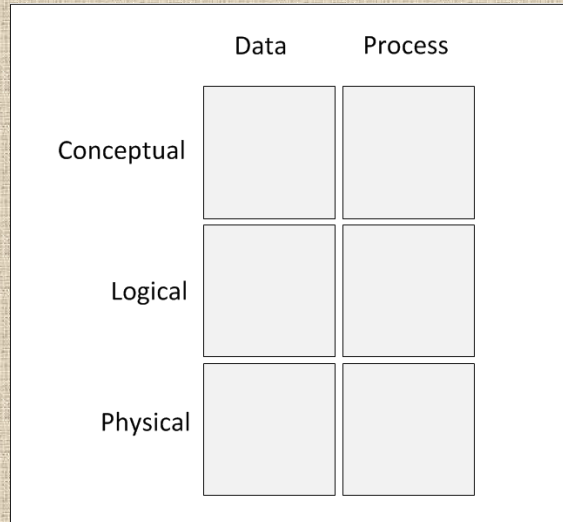13 July 2011   © 2011 O'Kelly Associates   3

Agenda

- Quick Refresher on Modeling
- Checklist: How to Assess a "Best Practice"
- Assessing Some Typical Best Practices
- Best Practices For Collecting Data Requirements

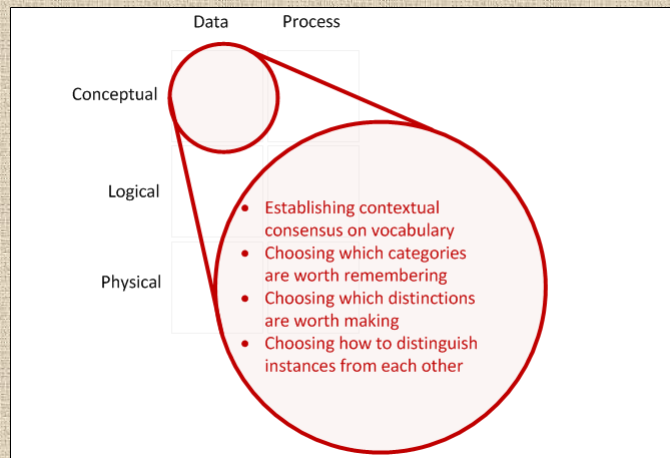13 July 2011   © 2011 O'Kelly Associates   4

Quick Refresher on Modeling

	Data	Process
Conceptual		
Logical		
Physical		

13 July 2011   © 2011 O'Kelly Associates   5

Quick Refresher on Modeling

Conceptual
Logical
Physical

Data   Process

- Establishing contextual consensus on vocabulary
- Choosing which categories are worth remembering
- Choosing which distinctions are worth making
- Choosing how to distinguish instances from each other

13 July 2011   © 2011 O'Kelly Associates   6

Quick Refresher on Modeling
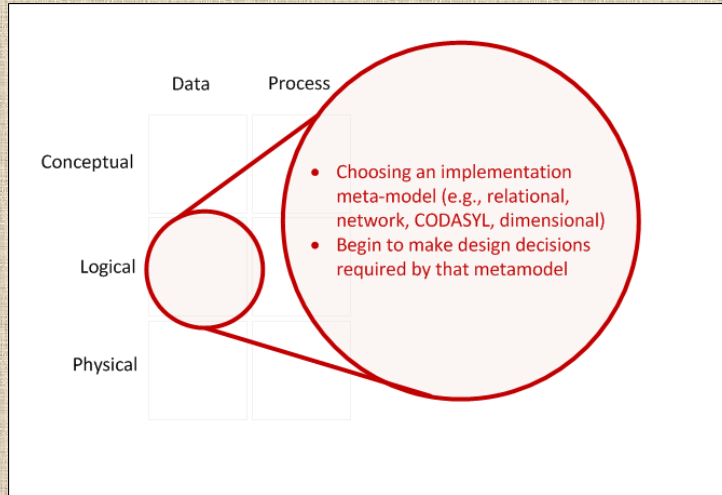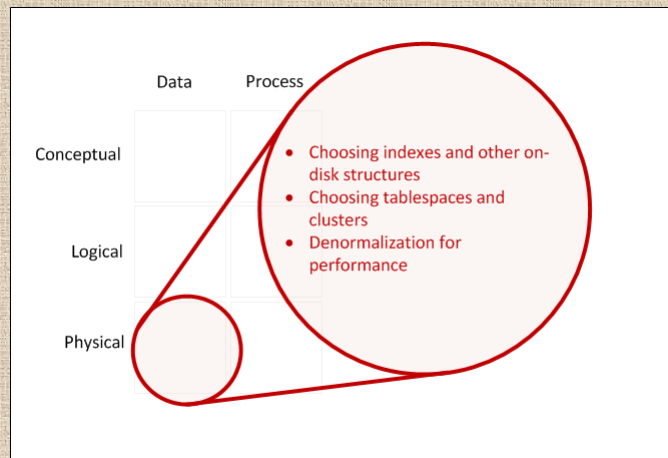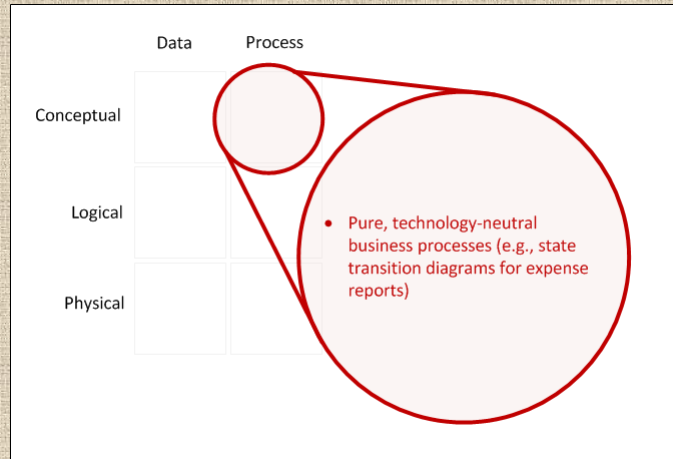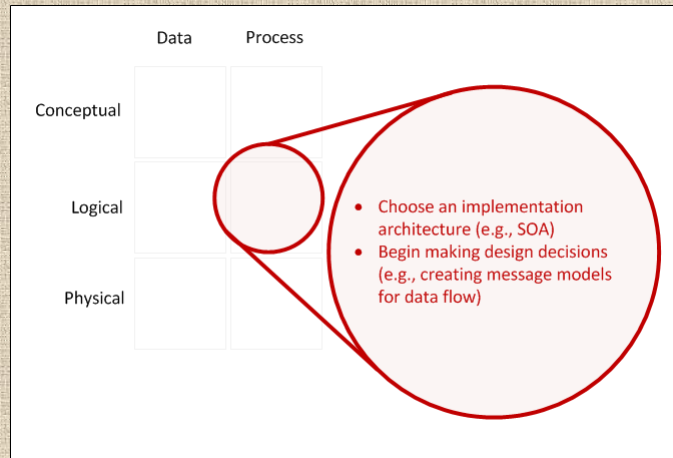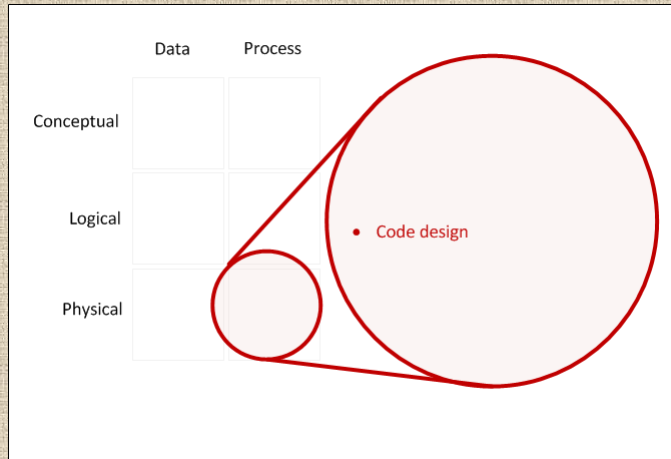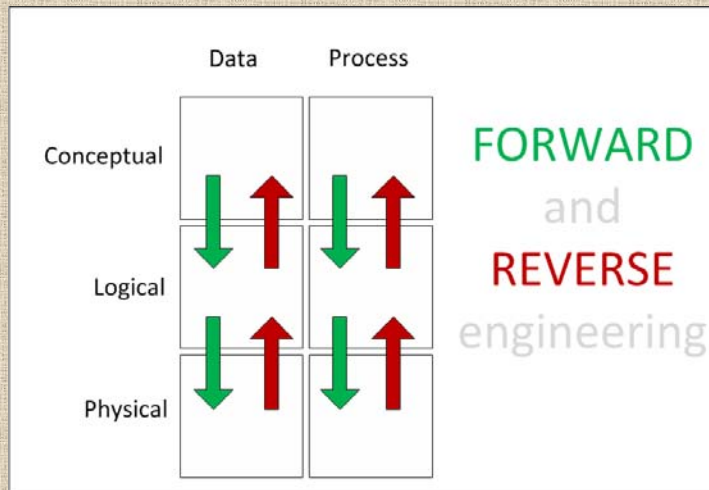
Quick Refresher on Modeling

Quick Refresher on Modeling

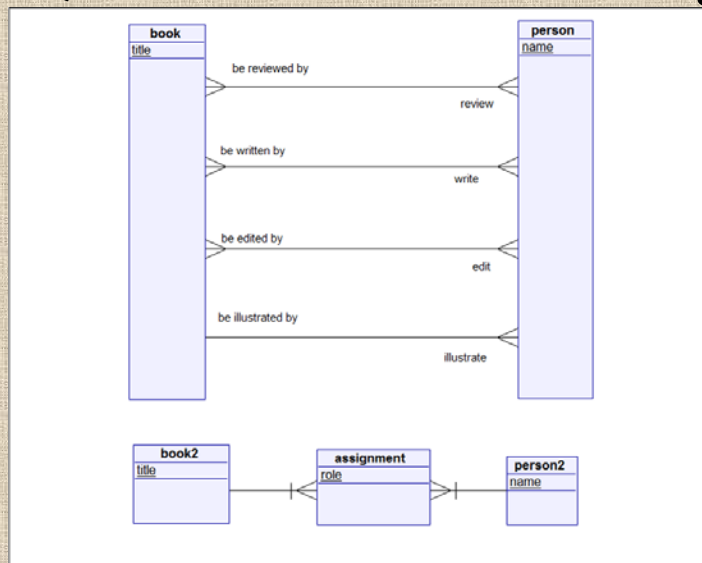- Three Axes
  - Data vs. Process
  - Conceptual – Logical – Physical
  - Greenfield Development vs. Maintenance
- Best practices will vary depending on where you are in this 3-D space.
  - Beware of proffered "best practices in data modeling" that do not acknowledge these three axes.

Quick Refresher on Modeling

Agenda

- Quick Refresher on Modeling
- *Checklist: How to Assess a "Best Practice"*
- Assessing Some Typical Best Practices
- Best Practices For Collecting Data Requirements

How To Assess a "Best Practice"

- Does it honor the distinction between conceptual, logical, and physical?
- Does it honor the distinction between data and process?
- Does it honor the distinction between greenfield development and maintenance?
- Does it support the process of creating data models—that is, does it apply to in-progress models as well as to completed models?

Distinguish Conceptual, Logical, and Physical

- Alleged, much-cited "Best practice"
  - Generalize as much as possible. This minimizes tables and creates a model that is responsive to change.
- Reality
  - Do NOT generalize when performing conceptual modeling, because this conceals the important vocabulary from the users.

13 July 2011 © 2011 O'Kelly Associates 17

Distinguish Data and Process

- Alleged, much-cited "Best practice"
  - Seek a rich notation, one that can express lots of constraints.
- Reality
  - Most constraints are motivated by process or by policy, not by the fundamental structure of the data. If you aspire to produce a data model that honors app/data independence, postpone the contemplation of constraints until later.

13 July 2011 © 2011 O'Kelly Associates 18

Distinguish Greenfield Development and Maintenance

- Alleged, much-cited "Best practice"
  - Use normalization to improve your models.
- Reality
  - Normalization is a way to repair broken designs. For greenfield development, it is preferable to find a best practice that produces high-fidelity models in the first place.

Acknowledge the *Process* of Modeling

- Alleged, much-cited "Best practice"
  - Model entities and relationships before attributes.
- Reality
  - Users do not prioritize based on syntactic distinctions. For conceptual modeling, better to prioritize based on significance as perceived by the users. For example:
    - Model things with mass before events and theories.
    - Model things and places before things-in-places.

Agenda

- Quick Refresher on Modeling
- Checklist: How to Assess a "Best Practice"
- ***Assessing Some Typical Best Practices***
- Best Practices For Collecting Data Requirements

Assessing Some Best Practices

- You get to choose which best practices we assess.
- As you choose, keep in mind these typical categories of best practice:
  - Model Expressiveness
  - Notation, Layout, and Drawing Conventions
  - The Process of Modeling
  - Words, Language, and Naming Conventions
  - Data-model documentation and sample instances

Assessing Some Best Practices

- Possible topics:
  - Agile data modeling
  - Exclude attributes from conceptual models?
  - Include as many constraints as possible?
  - UML class diagrams for expressing data requirements?
  - Limit diagrams to 7 ± 2 entities?
  - Use box-in-box notation for subtypes/supertypes?
  - Generalized vs. detailed models?

13 July 2011   © 2011 O'Kelly Associates   23

Assessing Some Best Practices

- Possible topics:
  - Refactoring databases
  - Rules for naming entities
  - Rules for naming attributes
  - Rules for labeling relationships
  - Should all IDs be arbitrary? Are IDs necessary?
  - Expressing data types on models
  - How to best use data-model shapes?
  - What is the *process* of collecting data requirements from users?

13 July 2011   © 2011 O'Kelly Associates   24

Assessing Some Best Practices

- Possible topics:
  - Use of color on model diagrams
  - Should coarser entities always be near the top of top of the diagram? (The "no dead crows" rule)
  - Should different line styles be used for different relationship types?
  - Should different box styles be used for different entity types?
  - How should verbalization of in-progress models be performed?

13 July 2011   © 2011 O'Kelly Associates   25

Agenda

- Quick Refresher on Modeling
- Checklist: How to Assess a "Best Practice"
- Assessing Some Typical Best Practices
- ***Best Practices For Collecting Data Requirements***

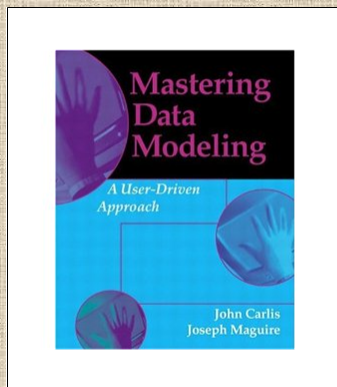13 July 2011   © 2011 O'Kelly Associates   26

Best Practices for Collecting Data Requirements

- Content for this section will be built up during the seminar.

Best Practices for Collecting Data Requirements

- *Mastering Data Modeling: A User-Driven Approach*, by John Carlis and Joseph Maguire, Addison-Wesley, 2000.

joe.maguire@okellyassociates.com