Karen López, ISP

- Karen has 20+ years of data and information architecture experience on large, multi-project programs.
- She is a frequent speaker on data modeling, data-driven methodologies and pattern data models.
- She wants you to love your data.
About this Presentation

We will be using an interactive format - you will be participating in informal polls about data modeling issues and best practices.

This is not an introductory presentation - a good knowledge of data modeling issues will be assumed.

Why Contention?

...because people are contentious!
InfoAdvisors’ Discussion Groups

- E-mail, web, RSS and newsgroup based discussion group
- Data Modeling, Frameworks, Tools Groups
- Over 3000 subscribers
- Moderated
- No Charge
- [www.infoadvisors.com](http://www.infoadvisors.com)

Agenda

- Contentious Issues
  - Background
  - Poll
  - Results & Analysis
- Resources
Contentious Issues

Near-religious discussions and debates

- People rarely change their minds based on this discussion
- The most successful discussions are ones where both sides learn something new about the other viewpoint.

Logical and Physical

- **Logical Data Model**: A representation of requirements that are DBMS and Application independent
- **Physical Data Model**: A representation of requirements that are DBMS and usually application specific.
- **DDL**: Script that may be generated from a model or reverse engineered from DB
Our Process...

...vote early and vote often!

Vote!

I post a question

I give the answers a range.

You put the sticky in the area for your vote

We debate the answer and the results
Keys - Natural or Surrogate

• Just what you call them may show your true colors....
  – Surrogate, Non-intelligent, Unnatural, Dataless or Meaningless
  – Natural, Intelligent, Normal, Cluttered
Keys – Natural or Surrogate

Order Number
Order Date
Sales Person ID

Order Number
Order Line Number
Product Number
Quantity

ORDER ID
Order Number
Order Date
Sales Person ID

Order Line ID
ORDER ID
Order Line Number
Product Number
Quantity

Surrogate <> Identity Property

• Identity may or may not be used for surrogate key
• Surrogate key may be assigned some other way
Keys - Natural or Surrogate - Vote

What best describes your approach in an LDM / PDM?

1. Surrogate Keys? We don’t need surrogate keys - we have natural identifiers
2. 
3. 
4. 
5. Every entity deserves its own surrogate key.

Surrogate Keys

TIP!

Every decision is ultimately made based on cost, benefit, and risk. Be prepared to analyze all three.
Do we need Classwords?

- A traditional naming convention
- Usually means there’s a standard classword list
- Some tools can check for standards compliance
- Examples: Date, Amount, Count, Quantity

Attribute/Column Names

- Customer First Name / Customer First Name
- Order Effective Date
- Order Effective
- E-mail Address
- Email
- Item / Item ID
...not the same as

- Domains
- Data Types
- UDTs

Not Class Words

- Char
- Varchar
- Numeric (but Number might be)
- Text (but Description might be)
- Data (it’s all data)
- Info
Do we need Class Words- Vote

In a Logical and Physical data model, should we require class words?

1. Yes, they clarify what the data is
2. 
3. 
4. 
5. No, we don’t need them

When To Use Varchar?

- Variable in length
- Extra storage
- Some trick for managing
- Varchar(1)?
- Varchar(2)?
- Varchar(8)?
**Varschar - Vote**

What makes sense?

1. Whatever anyone wants (1, 2, 3 is fine)
2.
3.
4.
5. There is a minimum number where varchar should not be used

**Identity Property as a Surrogate Key**

- Often used as a surrogate key
- Limitations
- 1,2,3,4,5,6,7,8,9, etc.
Identity Property - Vote

What makes sense?

1. **Always** – Double Rainbows and Unicorns
2. 
3. 
4. 
5. **Never**, it is Evil Spawned over and over

How Long Should Object Names Be?

- Tables, Columns, etc.
- **18**?
- **3**?
- **128**?
Name Lengths - Vote

What’s the limit?

1. As short as physically possible
2.
3.
4.
5. Use meaningful, long names, even up to 128

NoSQL

• Not really about SQL
• Trend in non-relational databases
• Semantic Web/Technologies related
Non-relational Databases

- New approaches, methods, techniques
- Often large investments
- Often large returns
- Often large risks
- May be planned with or without data architectures… and architects

Vote: Non-relational DBs

How will these new approaches affect your DM goals?
1. More Work / Budget / Focus
2.
3.
4.
5. Less Work / Budget / Focus
Who Gets to Update the Model?

• Many marketing pitches portray a team of developers, modelers, DBAs, etc., happily working on the model
• Is it “More hands make for less work” or “Too many cooks spoil the soup”?

Who Gets to Update the Model

1. Only seasoned Data Modeling Professionals
2. 
3. 
4. 
5. Everybody – we’re all professionals here
Responsibility

• Current Project
• Current Boss
• Enterprise
• Shareholders
• Family
• ???

If Cost, Benefit, and Risk is our mantra…
• ....whose Costs, Benefits, and Risks do we use?
Vote: To Whom Do We Answer?

1. My Current Boss, on my Current Project, on this day…
2. 
3. 
4. 
5. Everybody (The Public Good)

What Are Your Issues?
Is that with one “L” or Two?

How is it spelled?
1. Modeling/Modeler
2. Modelling/Modeller
3. It depends on my mood…

What did we learn?
Contentious Issues & Collaboration

Near-religious discussions and debates
– People rarely change their minds based on this discussion
– The most successful discussions are ones where both sides learn something new about the other viewpoint.

…and that we have a long way to go….
Contentious Issues & Collaboration

If we can’t collaborate well with other data professionals, we don’t stand a chance of collaborating well with Project Managers, Developers, Testers – or users.

Karen López, ISP

- Karen has 20+ years of data and information architecture experience on large, multi-project programs.
- She is a frequent speaker on data modeling, data-driven methodologies and pattern data models.
- She wants you to love your data.
Thank you!

I’d appreciate feedback of any type about this presentation.

www.speakerrate.com
/karenlopez

karenlopez@infoadvisors.com