



# First, Firster, Firstest

## Three lessons from history on information overload and technology

Strata Conference  
September, 2011

Mark R. Madsen  
<http://ThirdNature.net>

# Sivowitch's Law of Firsts

*“Whenever you prove who was first, the harder you look you will find someone else who was more first. And if you persist in your efforts you find that the person whom you thought was first was third.”*

*- Eliot Sivowitch*



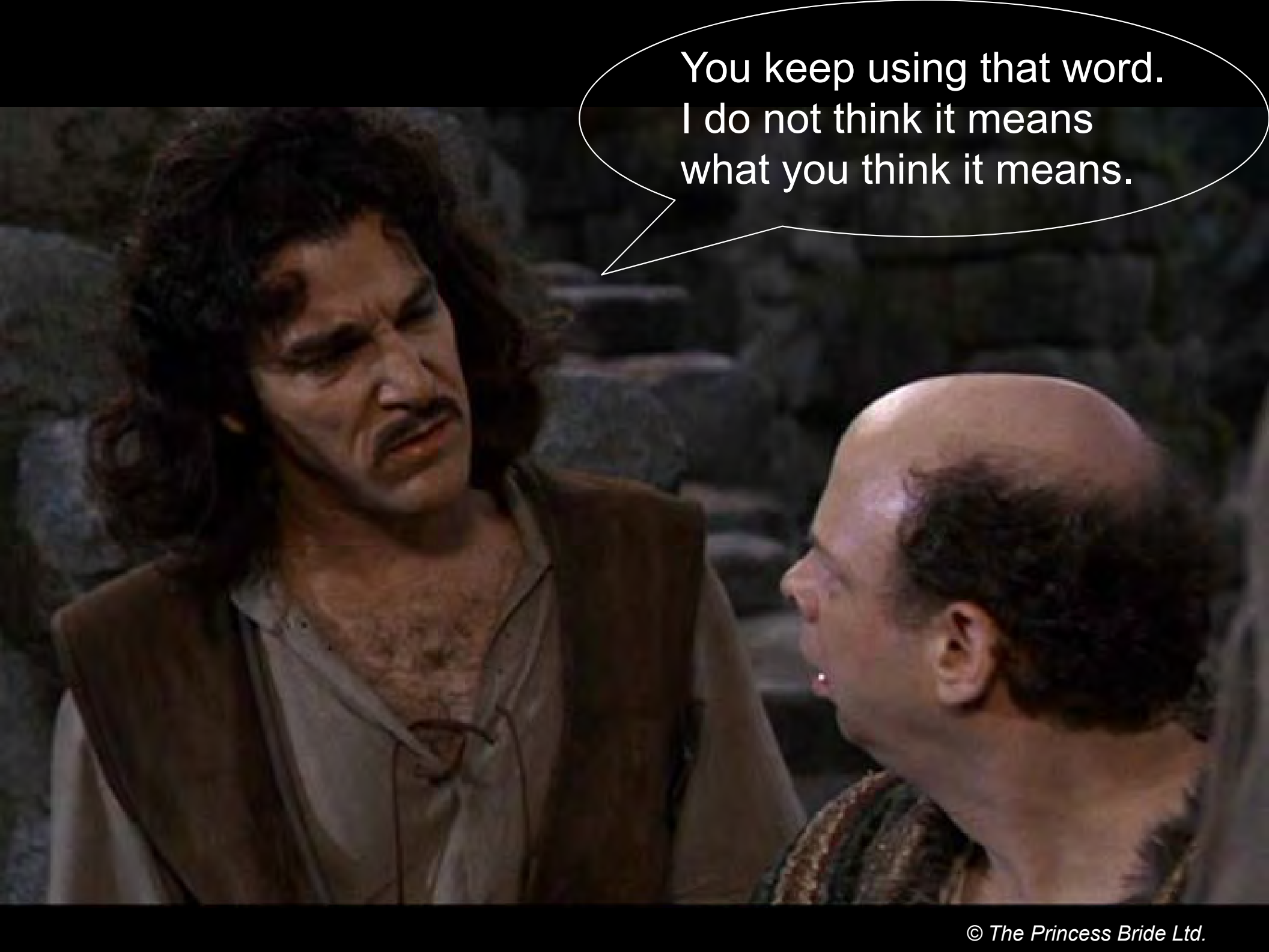
"Those who cannot remember the past are condemned to repeat it."

*George Santayana*

If there's one lesson we can take from history, It's that nobody learns any lessons from history.

**The future of data is the relational database**





You keep using that word.  
I do not think it means  
what you think it means.

# Good conceptual model, bad implementation



The relational database is the franchise technology for storing and retrieving data, but...

1. Single, static schema model
2. No rich typing system
3. Limited API in atomic SQL statement syntax

# Big Data: The SQL vs noSQL argument



There's a difference  
between having no past  
and actively rejecting it.

“There is nothing new under the sun  
but there are lots of old things we  
don't know.”

Ambrose Bierce





# The Elizabethan Era

Automated printing.

Information explosion:

- 8M books in 1500
- 200M by 1600
- Commoditization

Data management tech:

- Perfect copies
- Indices
- Topical catalogs
- First real encyclopedia
- Font standardization

# The Elizabethan Era: Storage and Retrieval



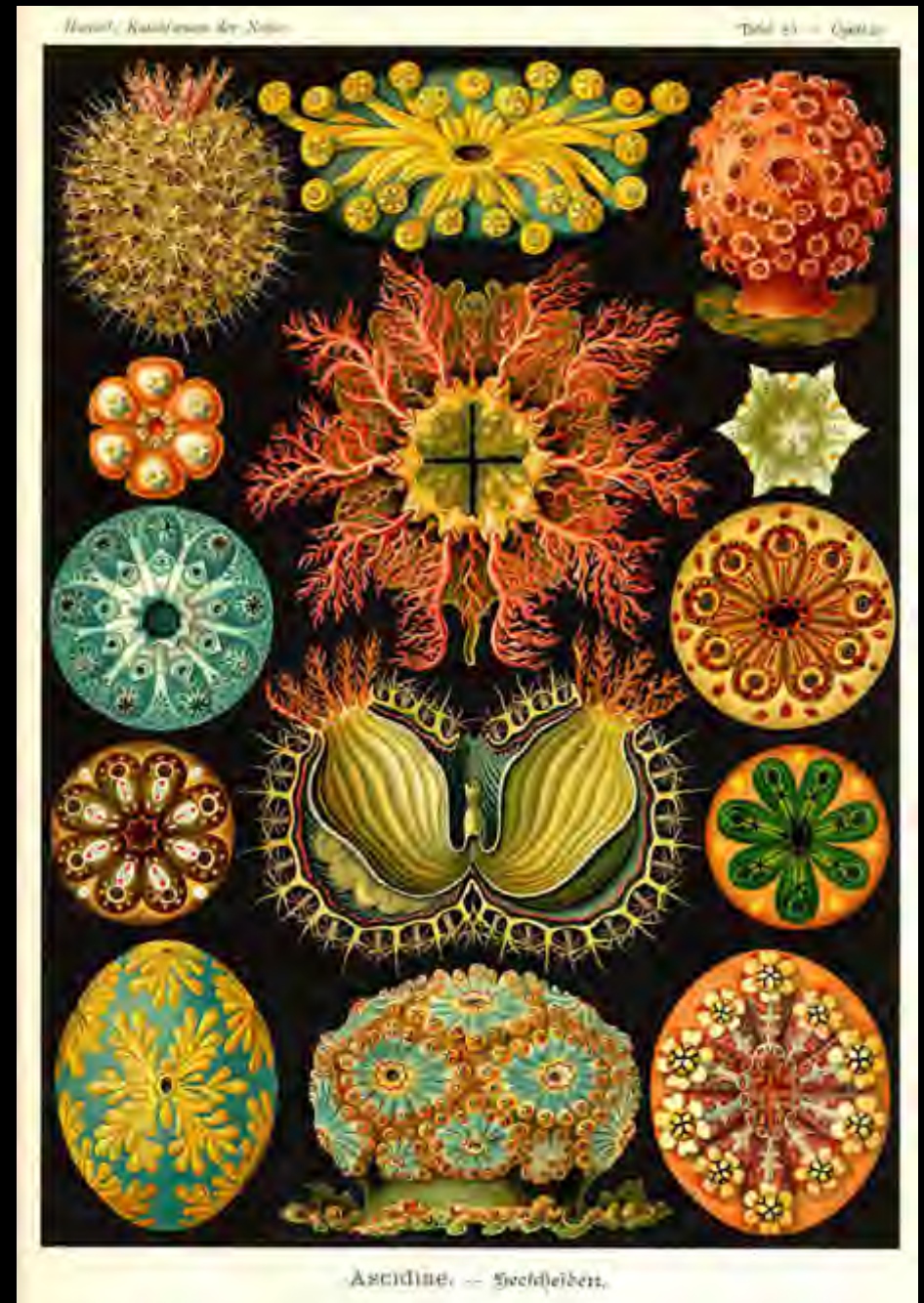
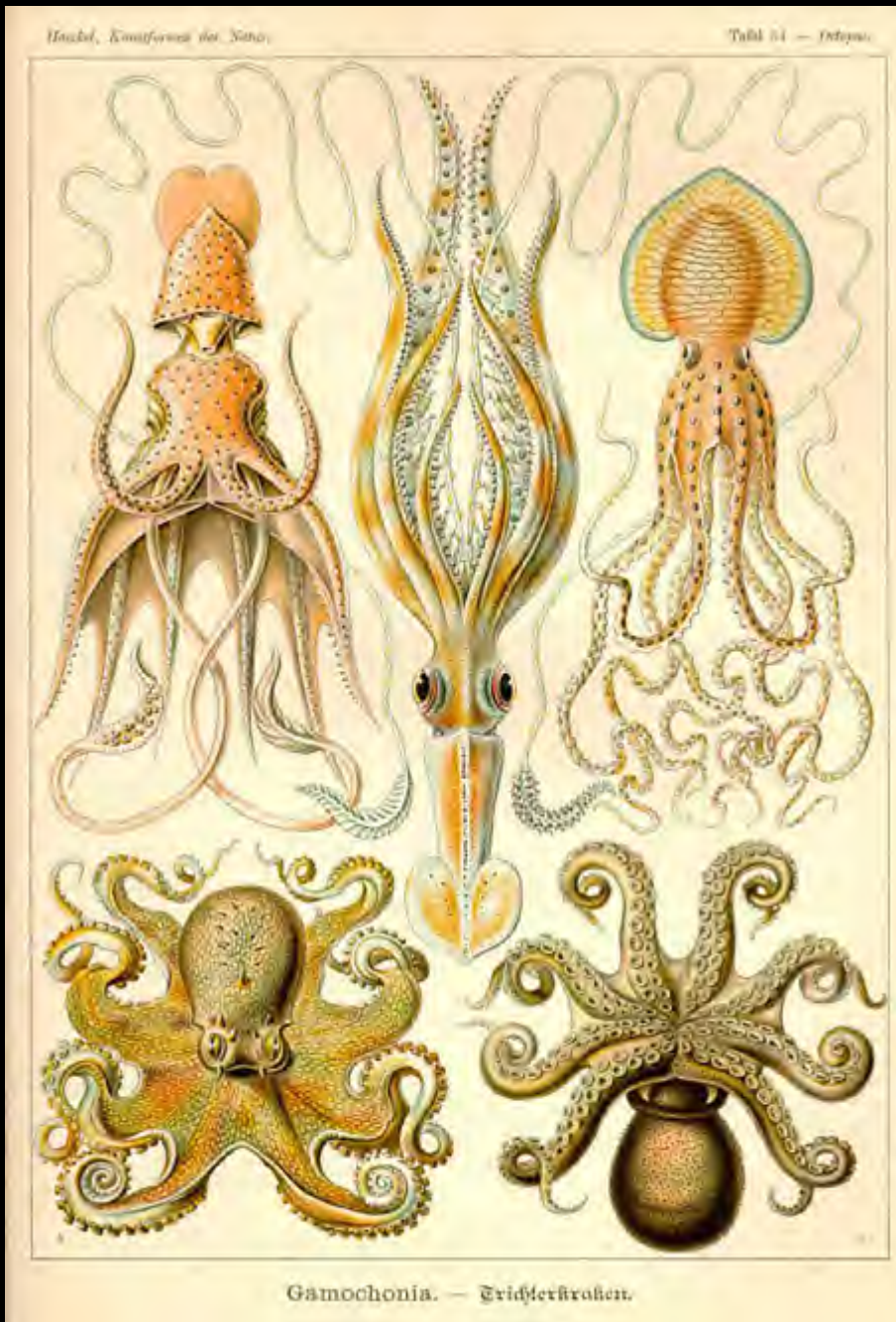
# The Elizabethan Era: Storage and Retrieval



# The Elizabethan Era: Storage and Retrieval



# The Georgian Era: The Explosion of Natural Philosophy



# Buffon



Bottom up orientation  
Flexible structure  
Explanatory, descriptive

*Faceted classification*

# Linnaeus



Top down orientation

Static structure

Descriptive rather than  
explanatory

*Taxonomic classification*

# The Theory of American Degeneracy



VS



# The Theory of American Degeneracy

# The Theory of American Degeneracy



*Linne'*



VS



TKO!

# The Victorian Era



# Charles Ammi Cutter



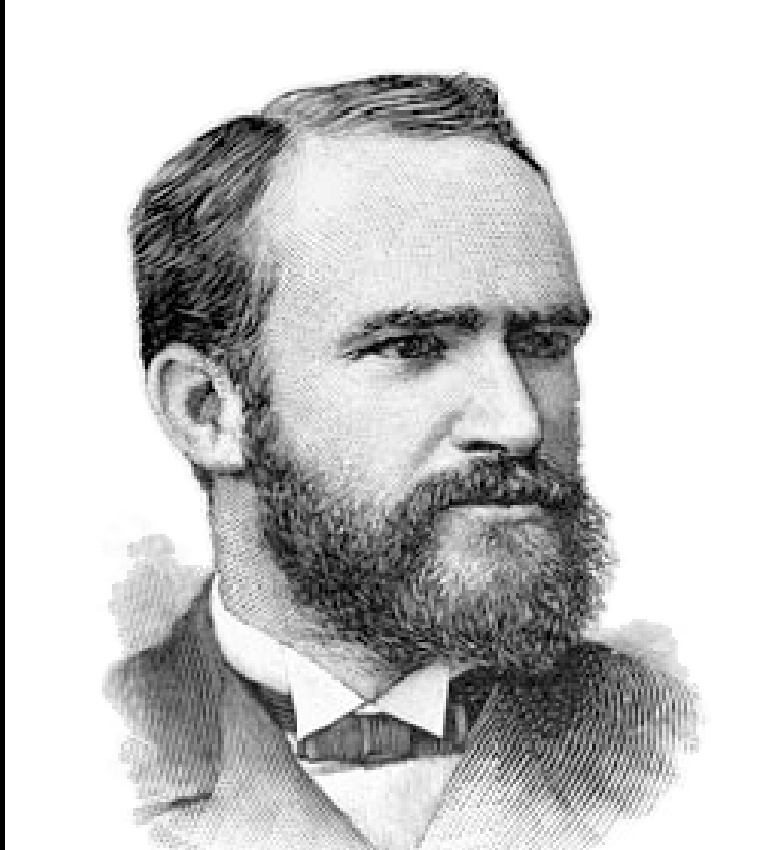
Cutter Expansive  
Classification System  
(~1882)

Bottom up orientation

More flexible structure

Explanatory, descriptive

# Melvil Dewey

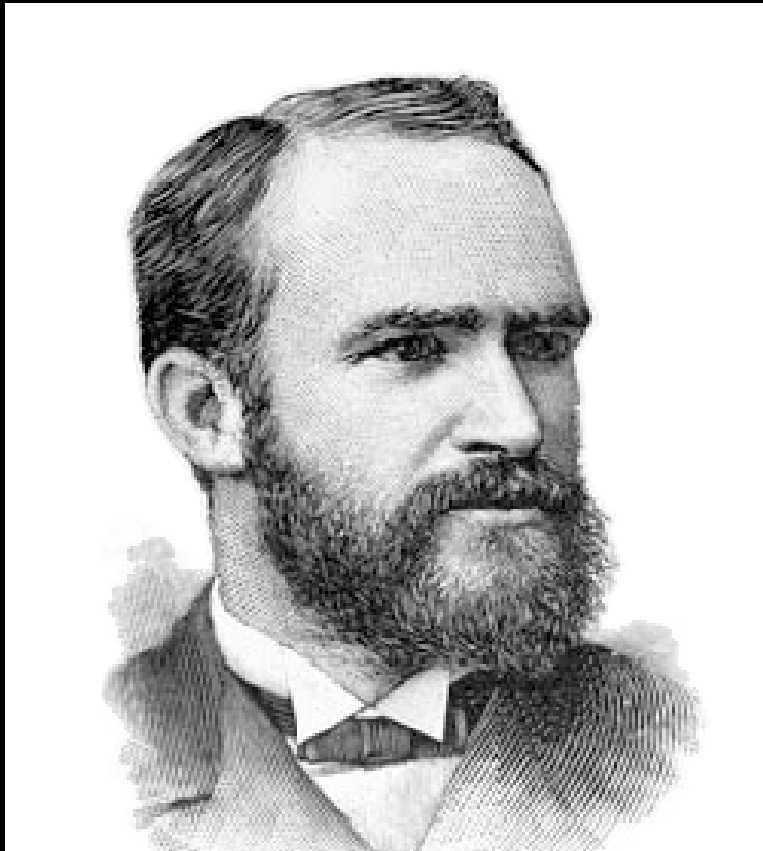


Dewey Decimal System

Top down orientation

Static structure

Descriptive rather than  
explanatory



VS



# Every technology is a tradeoff between something

History is always the same:

- Top down vs. bottom up
- Authority vs. anarchy
- Bureaucracy vs. autonomy
- Control vs. creativity
- Hierarchy vs. network
- Power vs. ease
- Dynamic vs. static



*In every choice, something is lost when something is gained.*

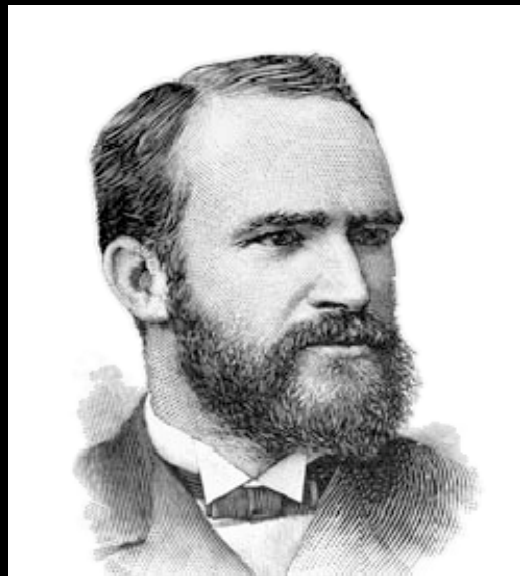
# So why did Linnaeus and Dewey win?



**Good enough  
wins the day**



**KO!**

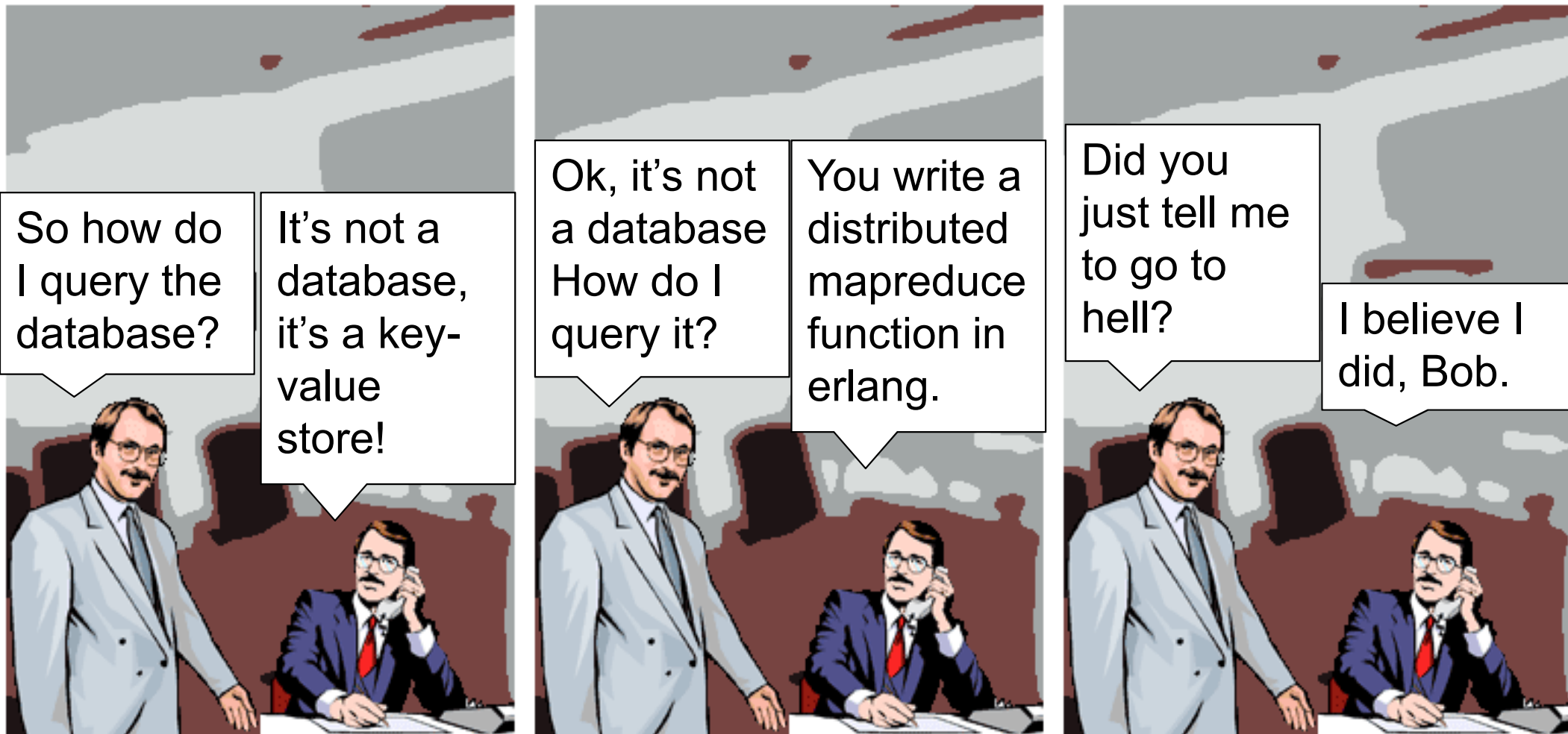


**It wasn't solving  
the problem you  
thought it was.**



# What lesson might we apply from this?

## Fault-tolerance



*Perhaps you should think about pragmatism a little bit.*

# Dealing with data in the industrial era



*Paul Otlet at his desk*

# 19<sup>th</sup> Century Data Loading



# Writing to the Database, Note Multi-processing



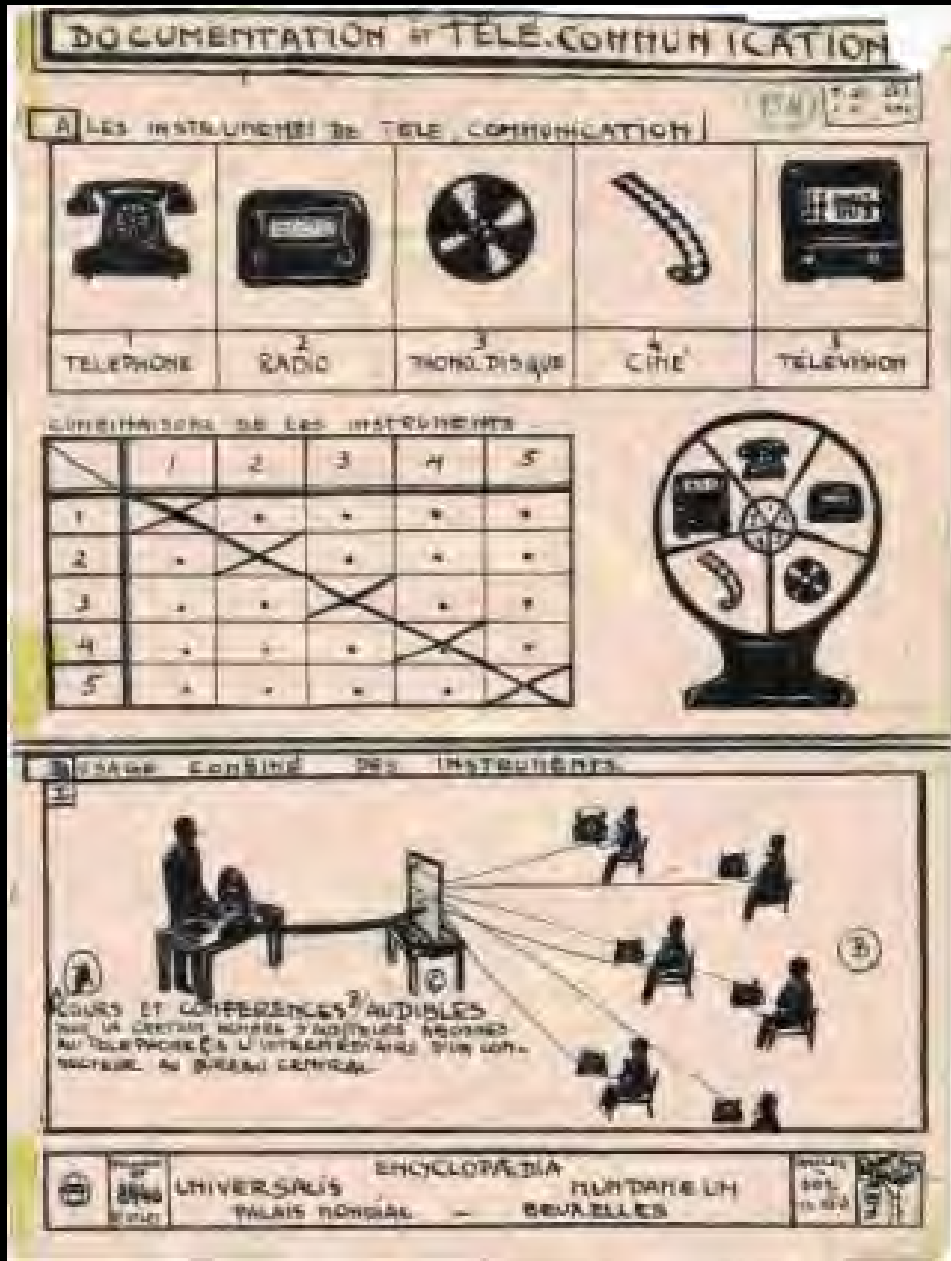
# Large Scale Information Storage



# Information Retrieval



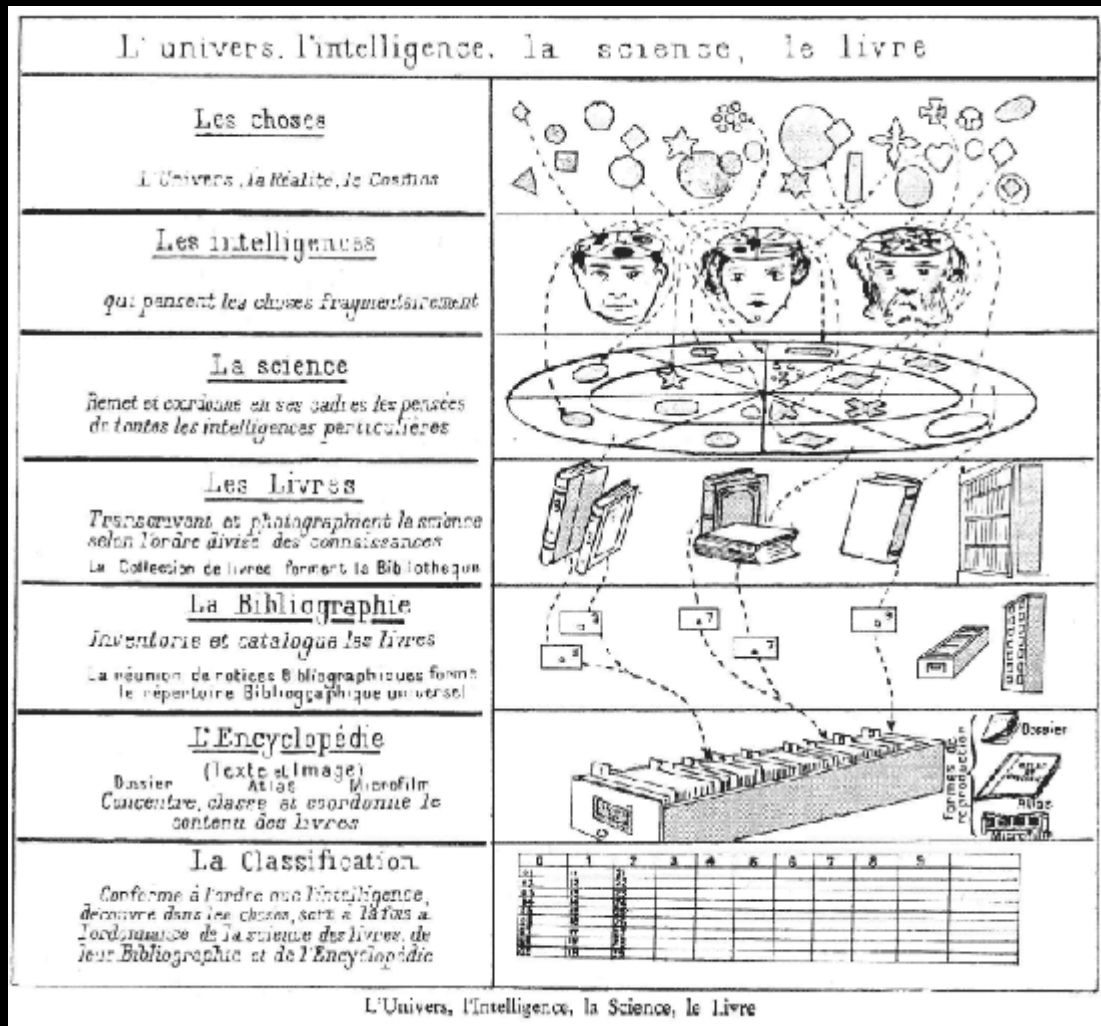
# The Computer & Internet Were Invented in 1934



Otlet's future vision:

- Technological developments will improve the ability to manage information
- Current technologies can be integrated to provide individual discovery, access and collaboration

# The Mundaneum Worked, For a While



Two primary flaws of the Mundaneum:

- Static, top-down classification system
- Loading could not keep up with data production rates

*Sounds familiar*

# Information Management Through Human History

**New technology development**

creates

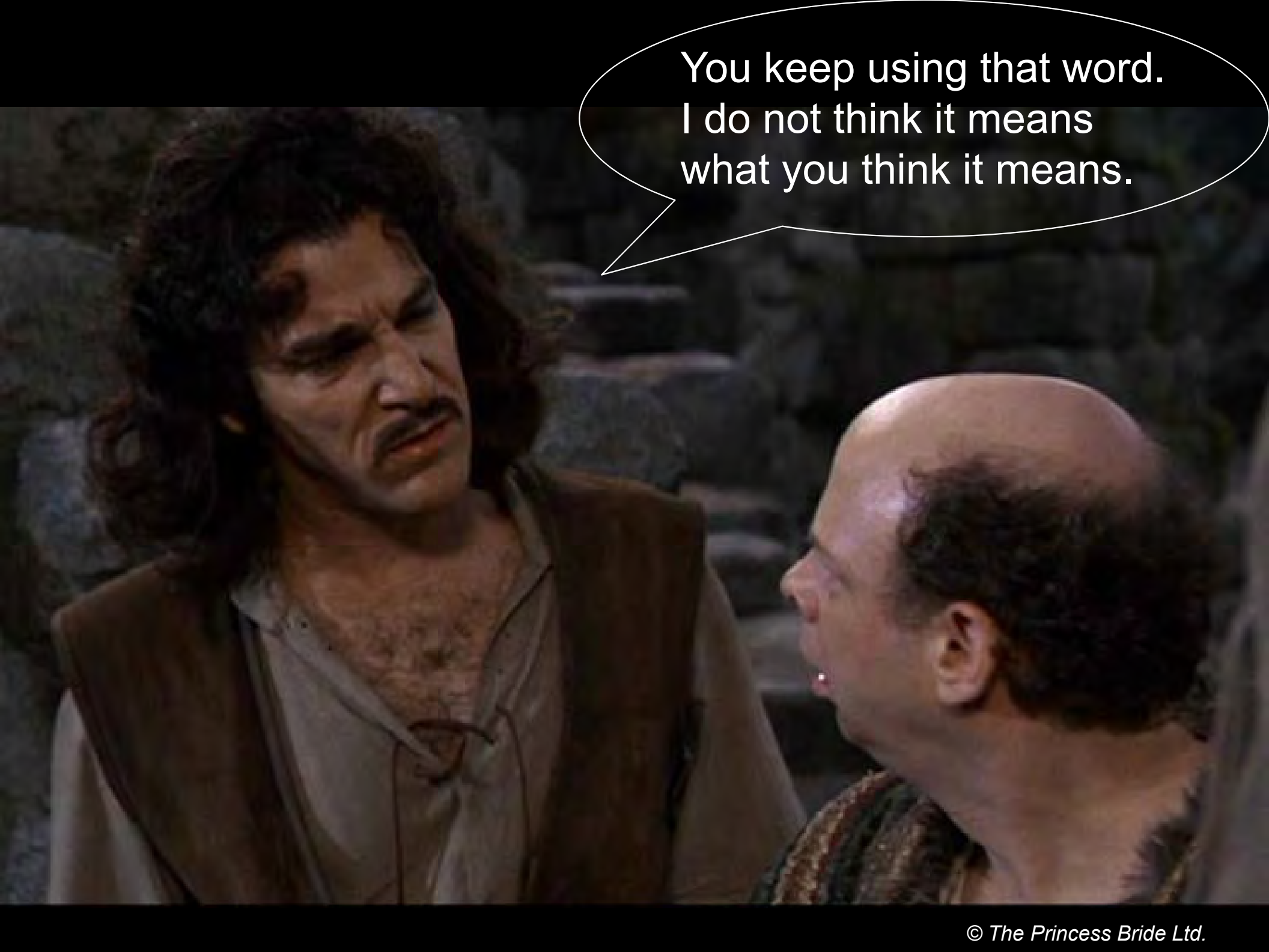
**New methods to cope**

creates

**New information scale and availability**

*creates...*

# Big Data



You keep using that word.  
I do not think it means  
what you think it means.

# Big data?

Unstructured data isn't really unstructured.

The problem is that this data is **unmodeled**.

# The future of data is the relational database



**SQL**



**noSQL**

# The future of data is the relational database



**SQL**



**noSQL**

**The false dichotomy can be removed by technology**



***Code defines what's possible now - maybe it's time to recode***

# Conclusion

IF YOU  
PROCRASTINATE  
LONG ENOUGH  
MOST PROBLEMS  
SOLVE THEMSELVES



# CC Image Attributions

Thanks to the people who supplied the creative commons licensed images used in this presentation:

manuscript\_page.jpg - <http://www.flickr.com/photos/calliope/306564541/>

manuscript\_illum.jpg - [http://www.flickr.com/photos/diorama\\_sky/2975796332](http://www.flickr.com/photos/diorama_sky/2975796332)

bookshelf by spectrum.jpg - <http://flickr.com/photos/santos/1704875109/>

moose.jpg - <http://www.flickr.com/photos/stephenandjes/4286949510/>

Vatican library - <http://www.flickr.com/photos/paullew/1550844955>

Little girl and fire – Dave Roth

Procrastinate – <http://www.cracked.com>

Fault tolerance - <http://browsertoolkit.com/fault-tolerance.png>

# About the Presenter



Mark Madsen is president of Third Nature, a technology research and consulting firm focused on business intelligence, analytics and information management. Mark is an award-winning author, architect and former CTO whose work has been featured in numerous industry publications. During his career Mark received awards from the American Productivity & Quality Center, TDWI, Computerworld and the Smithsonian Institute. He is an international speaker, contributing editor at Intelligent Enterprise, and manages the open source channel at the Business Intelligence Network. For more information or to contact Mark, visit <http://ThirdNature.net>.

# About Third Nature



Third Nature is a research and consulting firm focused on new and emerging technology and practices in business intelligence, data integration and information management. If your question is related to BI, open source, web 2.0 or data integration then you're at the right place.

Our goal is to help companies take advantage of information-driven management practices and applications. We offer education, consulting and research services to support business and IT organizations as well as technology vendors.

We fill the gap between what the industry analyst firms cover and what IT needs. We specialize in product and technology analysis, so we look at emerging technologies and markets, evaluating the products rather than vendor market positions.