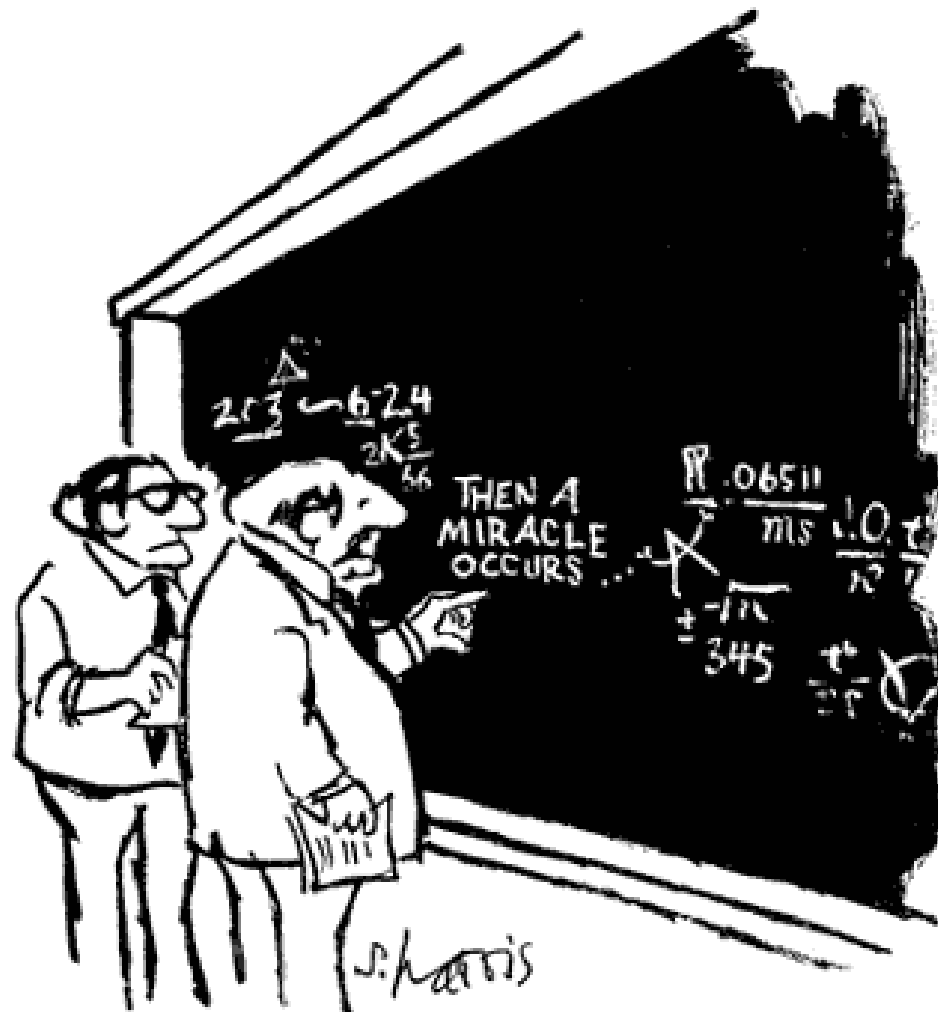


# A realistic look at open data

*Sharon Dawes*  
[www.ctg.albany.edu](http://www.ctg.albany.edu)

*Using Open Data Workshop*  
*Brussels, June 19-20, 2012*



"I think you should be more explicit here in step two."

Sidney Harris, 2012



# Sources of data problems

Conventional wisdom

Provenance

Practices

# Consequences of the problems

Underuse

Misuse

Non-use

Shifting costs and responsibilities



# Where do open data come from?

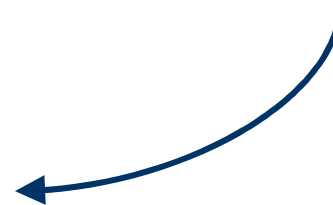
Administrative systems



Governed by specific policies and laws



By people with different kinds and levels of knowledge and expertise



# Case 1: Give me shelter

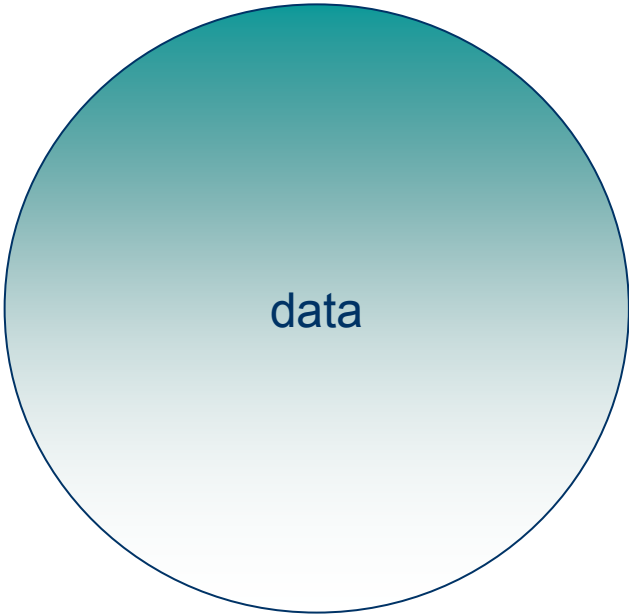


# Case 2: Cadastral records

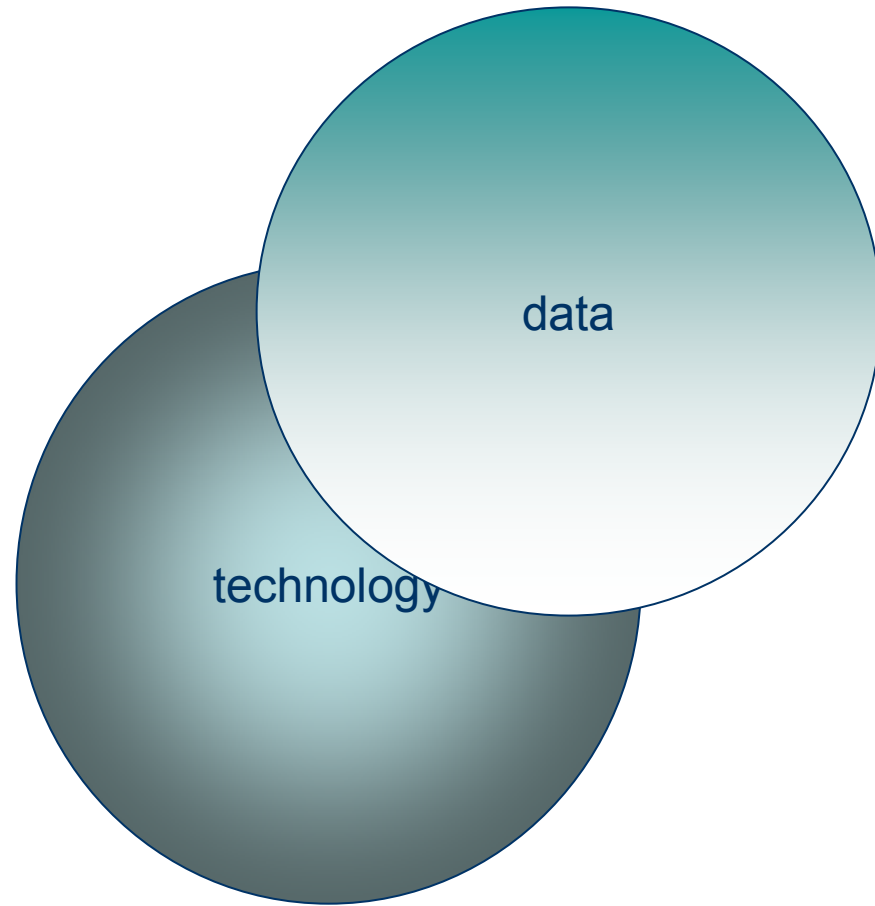


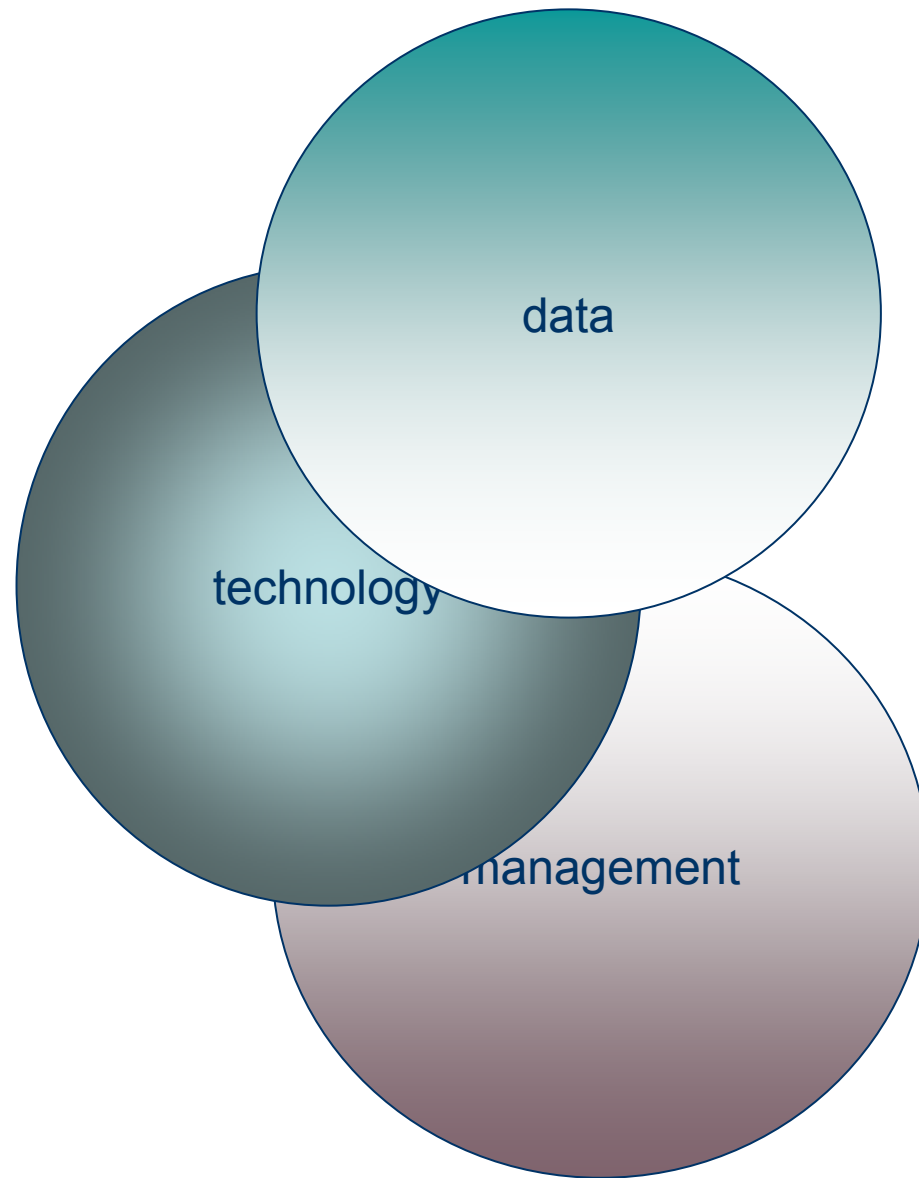
# Case 3: Where does the money go?

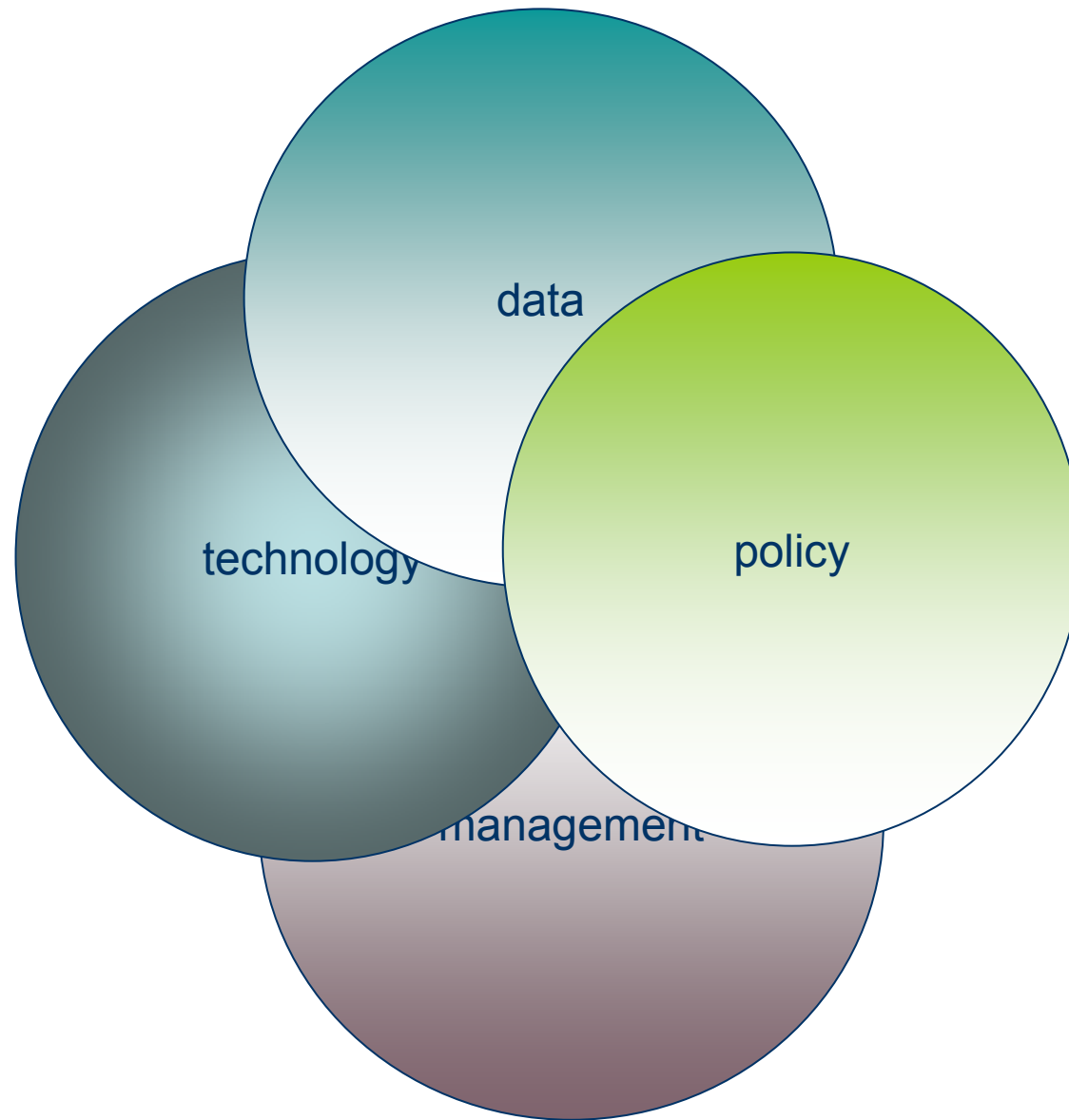


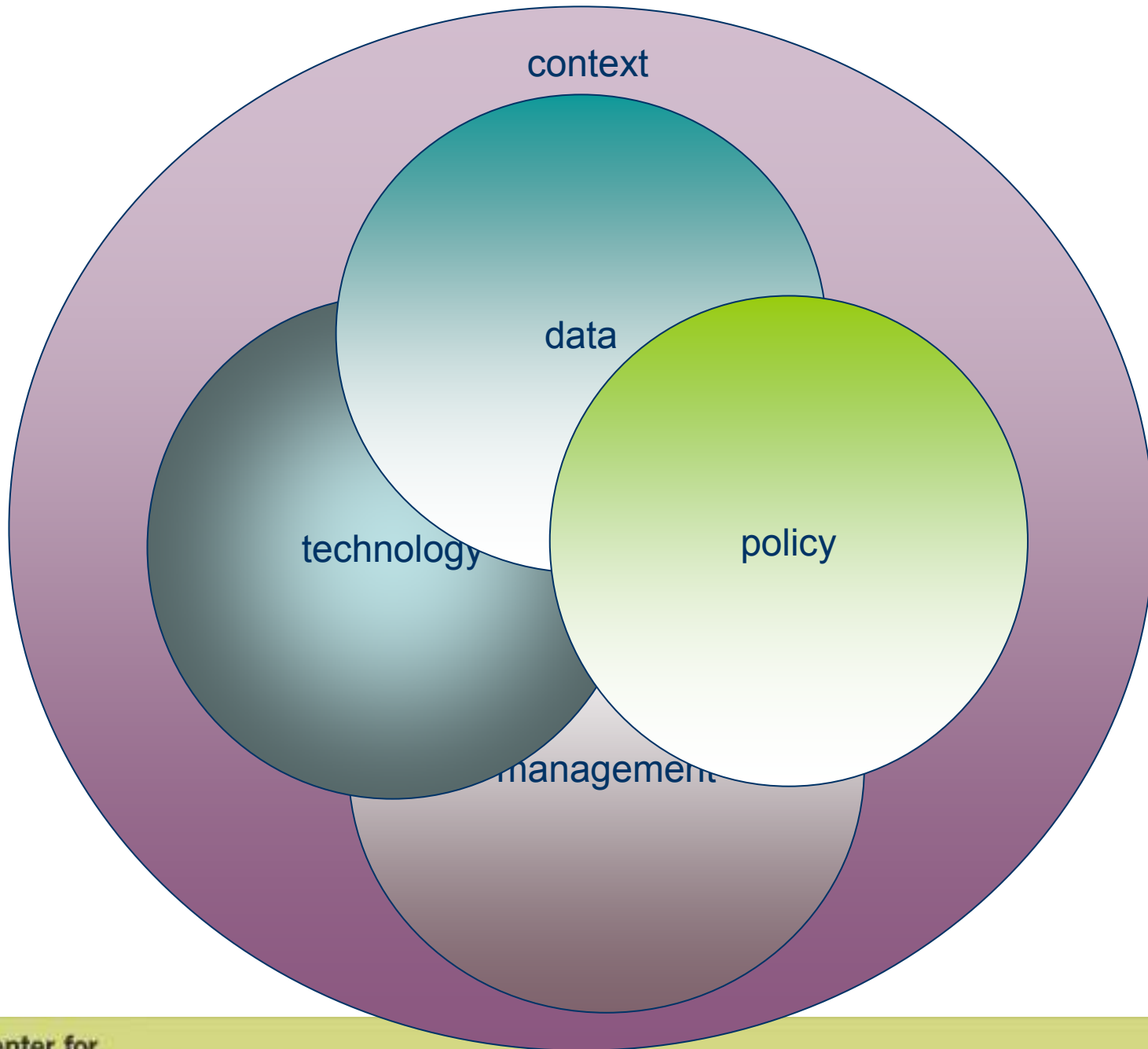


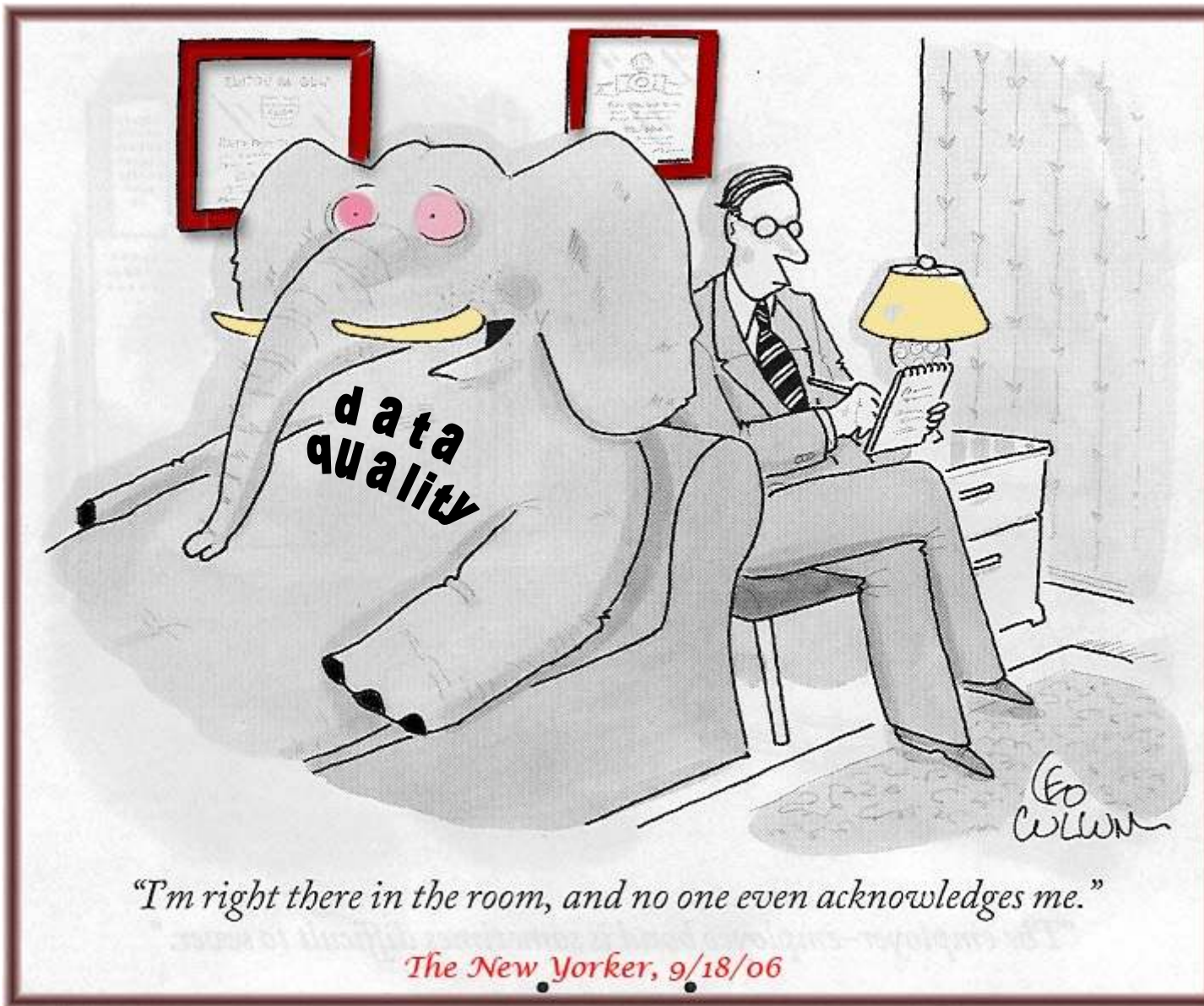












*"I'm right there in the room, and no one even acknowledges me."*

*The New Yorker, 9/18/06*



# Data quality = fitness for use

- Matters most from the *user's* point of view
- Depends on the user's purpose
- Four types of quality:
  - Intrinsic
  - Contextual
  - Representational
  - Access-related
- Usually involves trade offs
  - e.g., timeliness vs. completeness

(Wang & Strong, 1996, Ballou & Pazer, 1995)

# Dimensions of data quality

Accessibility	Extent to which data is available, or easily and quickly retrievable
Appropriate Amount of Data	Extent to which the volume of data is appropriate for the task at hand
Believability	Extent to which data is regarded as true and credible
Completeness	Extent to which data is not missing and is of sufficient breadth and depth for the task at hand
Concise Representation	Extent to which data is compactly represented
Consistent Representation	Extent to which data is presented in the same format
Ease of Manipulation	Extent to which data is easy to manipulate and apply to different tasks
Free-of-Error	Extent to which data is correct and reliable
Interpretability	Extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	Extent to which data is unbiased, unprejudiced, and impartial
Relevancy	Extent to which data is applicable and helpful for the task at hand
Reputation	Extent to which data is highly regarded in terms of its source or content
Security	Extent to which access to data is restricted appropriately to maintain its security
Timeliness	Extent to which the data is sufficiently up-to-date for the task at hand
Understandability	Extent to which data is easily comprehended
Value-Added	Extent to which data is beneficial and provides advantages from its use

Pipino, et al, 2002

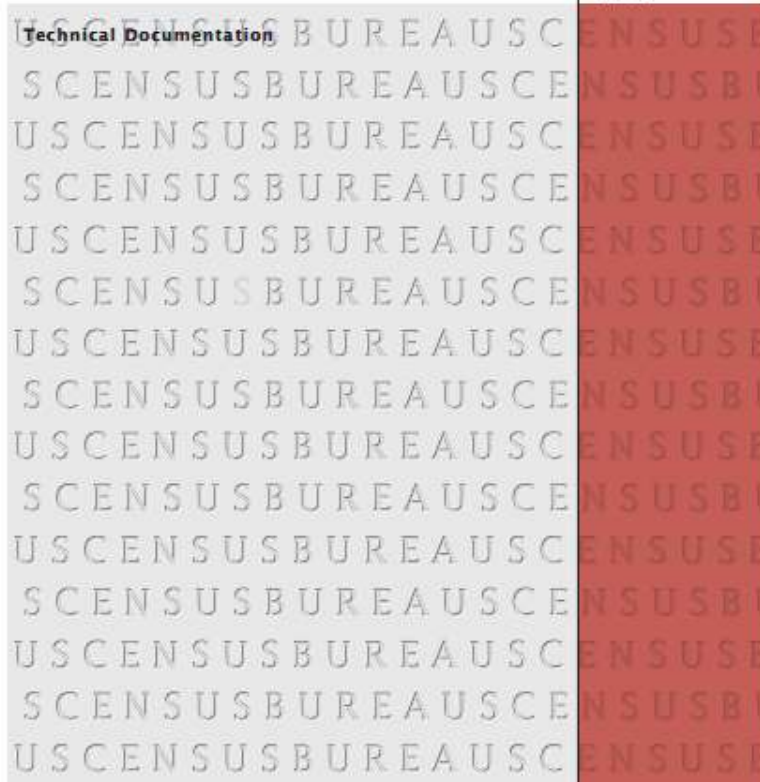
# Metadata

## Summary File 1

2000  
2000 Census of Population and Housing

2000  
Released July 2007

### Technical Documentation

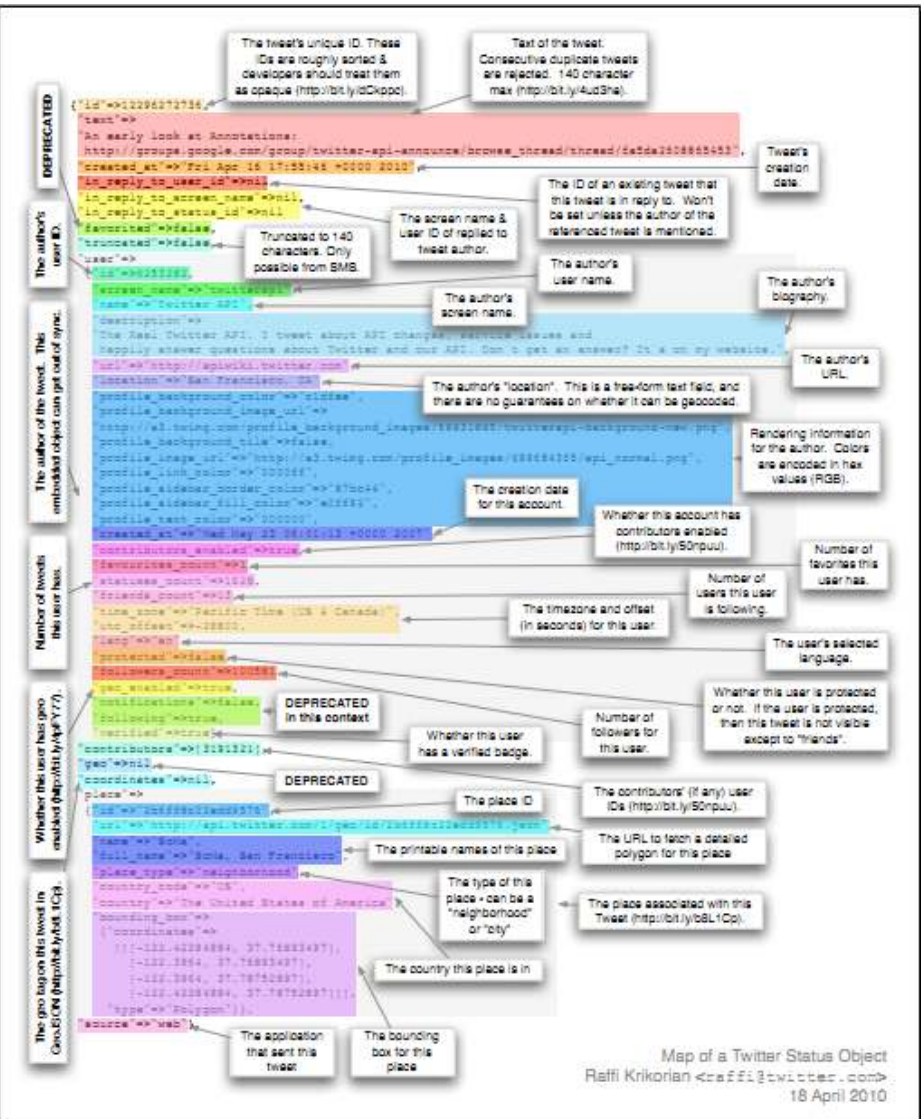


USCENSUSBUREAU  
Helping You Make Informed Decisions

U.S. Department of Commerce  
Economics and Statistics Administration

United States  
Census  
2000

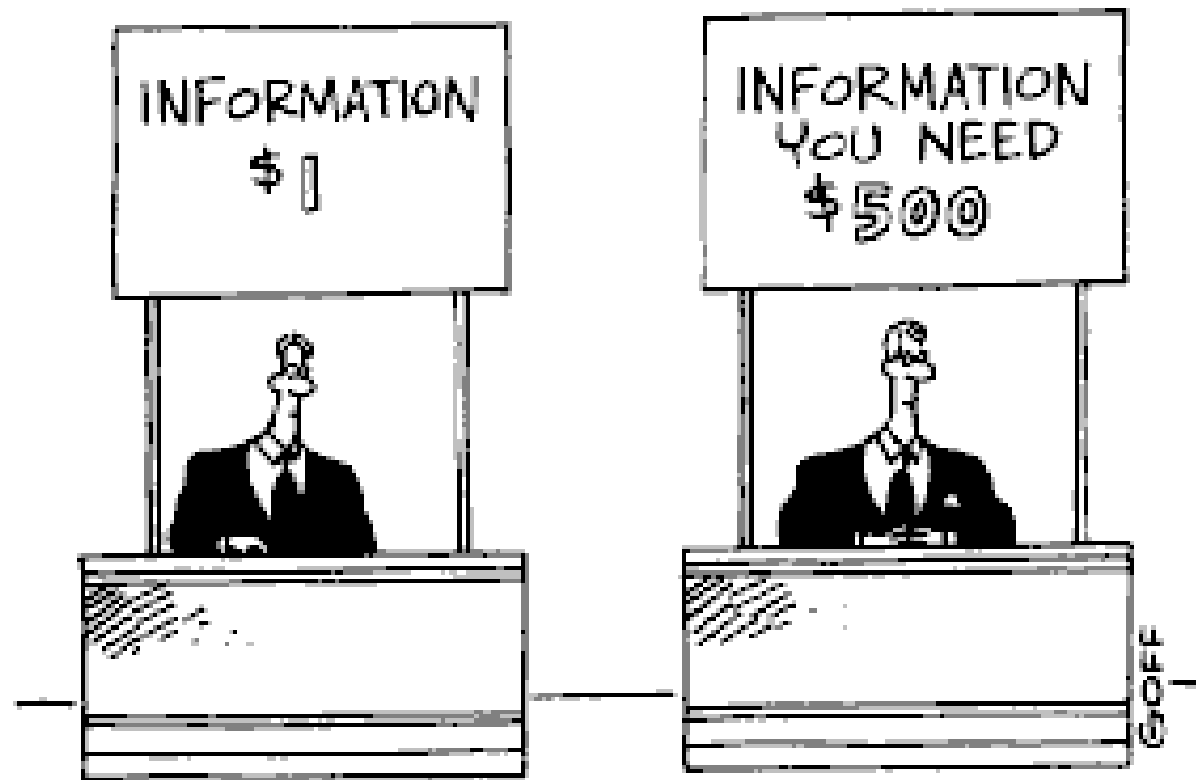
637 pages





# Data quality “tools”

- For providers
  - Appreciate data as an asset, a source of value
  - Adopt information policies to preserve and enhance value
  - Create and maintain metadata to support *unknown* users
  - Adopt stewardship practices
- For users
  - Be skeptical, ask questions
  - Understand the nature and context of the data
  - Use data sets with caution
  - Combine data sets with great caution



© 1996 Ted Goff

