

Data Stewardship: Conscious Data Management

Cynthia Dingman
Montana Dept of Environmental Quality
cdingman@mt.gov

1

Data Stewardship

Air.
Necessary for survival of organisms.

The intake of air is mostly an unconscious function. The oldest, most primitive part of the brain—the brain stem—regulates breathing.

Conscious Data Management

Data.

Necessary for survival of organizations.

The intake of data mostly an unconscious function. Data collection is the most "primitive" of organizational functions as it occurred before IT system evolution.

© Photo Edinburgh University Data Library

Data Stewardship

Just as poor quality air creates problems for organisms...



4

Conscious Data Management

Poor quality data creates problems for the health of your organization.



5

Data Stewardship

Sources of poor quality data

Which of the following most often contribute to data quality problems in your organization?

Inconsistent definitions for common terms	75%
Data entry by employees	75%
Data migration or conversion projects	46%
Mixed expectations by users	40%
External data	38%
Data entry by customers	26%
System errors	25%
Changes to source systems	20%
Other	7 %

Based on 1522 responses from 399 respondents in 2005.

Philip Russom, *Liability and Leverage A Case for Data Quality*, DM Review Magazine, August 2006

6

Conscious Data Management

Data Pollutants – the 4 'I's'

- Inaccurate Data
- Incomplete Data
- Inconsistent Data
- Identical Data

7

Data Stewardship

Inaccurate Data

- The values contained in the database fields are wrong, obsolete or superseded
- Inaccurate data lessens user confidence in the data

8

Conscious Data Management

Incomplete Data

- Missing data, e.g., blank values, no metadata
- Incomplete data may make the data meaningless and slows down business processes

9

Data Stewardship

Inconsistent Data

- No standard definition/format agreements

"Inconsistent data definition is a leading origin of data quality problems. Too often, the data itself is not wrong; it is just used wrongly. For example, multiple systems may each have a unique way of representing a customer....When good data is referenced wrongly, it can mislead business processes and corrupt databases downstream."

Philip Russom, *Liability and Leverage A Case for Data Quality*, DM Review Magazine, August 2006

10

Conscious Data Management

Identical Data

- Duplicate data, e.g., multiple records of identical data.
- Creates confusion among users.
- Creates "orphan" records.

11

Data Stewardship

Poor quality data enablers

- Average time spent verifying quality & accuracy of data is 12.1 hours (median time 8 hours)
- During 40-hour work week,
8 hours = 20%

12

Conscious Data Management

Poor quality data enablers

- Business analyst making \$60K/yr
- 20% of 60K = \$12,000/yr
- Multiply \$12,000 by number of analysts in your organization

Frank Dravis, *New Survey Supports Old Assertion*, EIM Blog, July 10, 2006

13

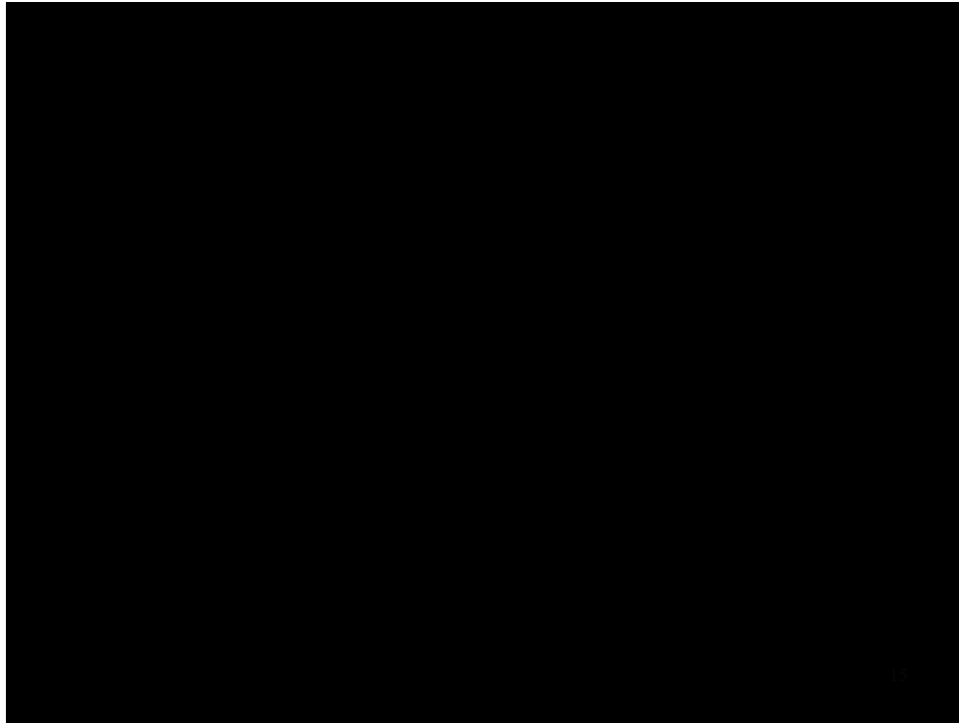
Data Stewardship

Poor-quality data enablers

"The problem, cost, and challenge of poor data quality is traditionally not visible to senior managers because it's most often cleaned up before they see it. They don't routinely see that the monthly sales-by-customer report takes three days because of the fiddling and rework that's required to get the data right."

Dan Pratte, *Poor-quality data can rob your company of information*, Tech Republic, May 30, 2001

14



Conscious Data Management

Data Quality Consciousness

- Recognizing consequences of poor quality data
 - Rework
 - Time loss
 - Increased liability
 - Faulty decision making
 - Substandard products
 - Lack of public confidence

16

Data Stewardship

Data Quality Consciousness

- Investigating root causes of other organizational problems may reveal poor data quality is to blame
 - Inadequate staff **or** too much time spent reconciling poor data?
 - Poor database design **or** poor data choking the database?

17

Conscious Data Management

Data Quality Consciousness

- Awareness of Data Management Costs
- Awareness of Poor Data Quality Costs
- Seeking data quality solutions through “Behavioral Science” not just Information Science

18

Data Stewardship

People Solutions

- Data Governance
- Must fit organizational needs
- Many good models
 - Google 500K results
 - Top-down, integrated, tiered

19

Conscious Data Management

Data Governance through Data Stewardship

Data Stewards work collectively to develop policies and procedures that promote

- Data accuracy
- Data consistency
- Reduce redundancy
- Encourage data sharing

20

Data Stewardship

Data Governance through Data Stewardship

Data Stewards work individually to

- Audit data quality
- Perform data cleanup
- Improve data management business processes
- Perform application testing
- Aid in application deployment

21

Conscious Data Management

Starting a Data Stewardship Initiative:
Is Management involvement required?

- Involvement as in "active participation"
 - NO
- Involvement as in "assign staff"
 - YES

22

Data Stewardship

Starting a Data Stewardship Initiative: 2 Keys to Success



Committed group leader

- Analyst type



Find/recruit those most affected by
poor data quality

- IT expertise not required
- Intimate with business process

23

Conscious Data Management

Starting a Data Stewardship Initiative: Step 1

- Identify data management challenges & needs
- Research Data Stewardship Models*
- Present to assembled team for discussion
 - To use as a basis—not a template
 - Tailor structure to meet your organization

24

Data Stewardship

Starting a Data Stewardship Initiative: Step 2

- Write a group charter
 - Define mission
 - Define structure (membership/representation)
 - Define roles/responsibilities
 - As committee AND as department representatives
 - Define group communication/meeting schedule

25

Conscious Data Management

Starting a Data Stewardship Initiative: Step 3

- Establish priorities
 - Initial priorities are foundation for committee work such as development of
 - Data Standards Policy
 - Process/procedures for data standards development
 - Database baseline
 - Data Standards
 - Data quality training/awareness
 - Data cleanup projects
 - Data audit training

26

Data Stewardship

Starting a Data Stewardship Initiative: Step 4

- Yearly review
- Annual Report to Management
 - Accomplishments, Lessons Learned
- Adjust priorities
 - Shewart Cycle – Plan, Do, Check, Act

27

Conscious Data Management

Nurture the Effort

Given time Data Stewardship values will infuse the culture of your organization with fresh "air".

28

Data Steward Blessing

May your data always be pure,



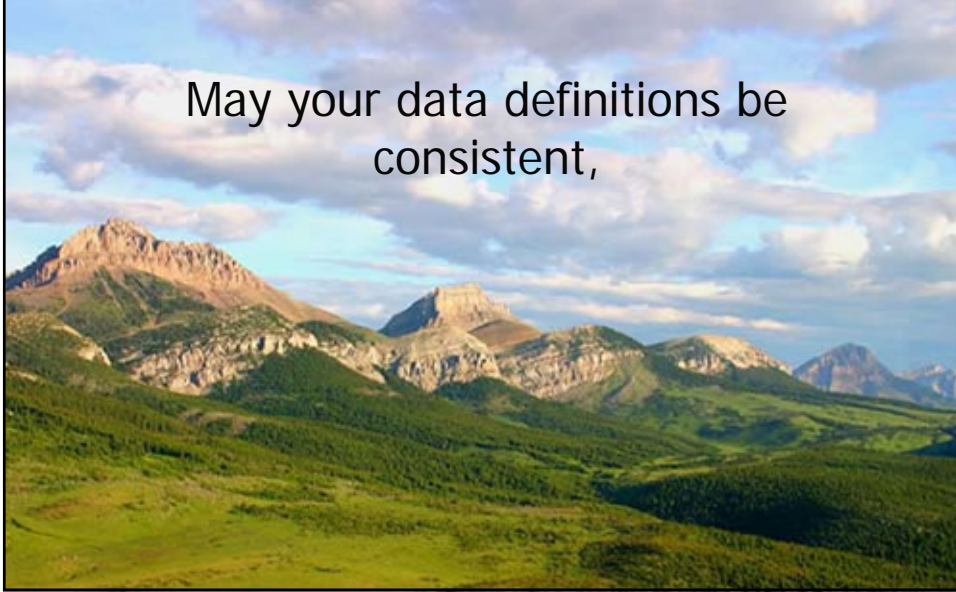
Data Steward Blessing

May your records always be complete,



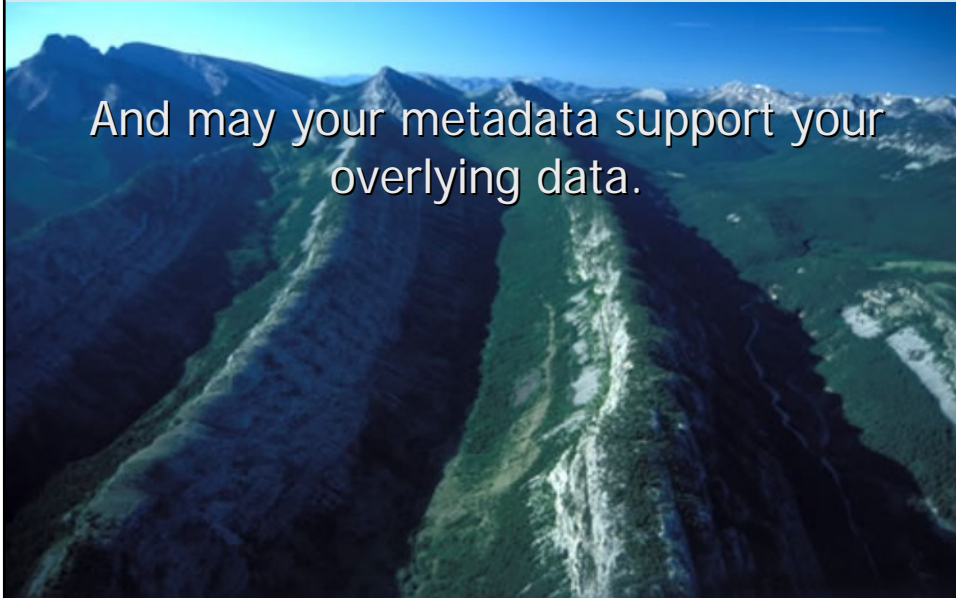
Data Steward Blessing

May your data definitions be
consistent,



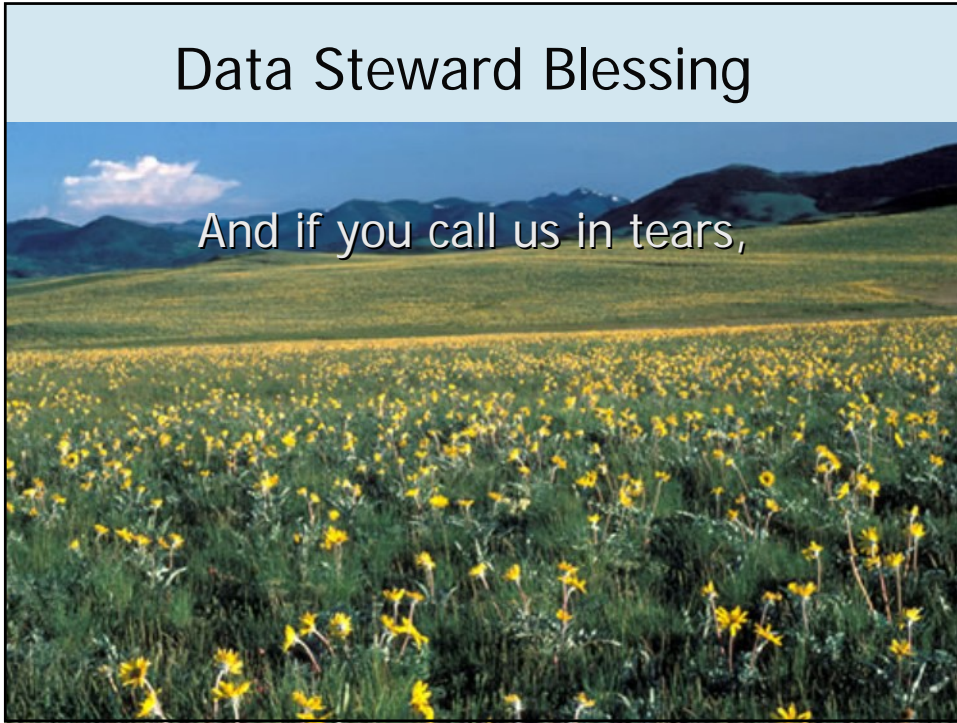
Data Steward Blessing

And may your metadata support your
overlying data.



Data Steward Blessing

And if you call us in tears,



Data Steward Blessing

May we always be ready to
lecture you about data quality.

