How well does intuition serve us when managing data?

Andy Cairns
CGGVeritas (Data Services division)

FESAus Technical Meeting
June 2012
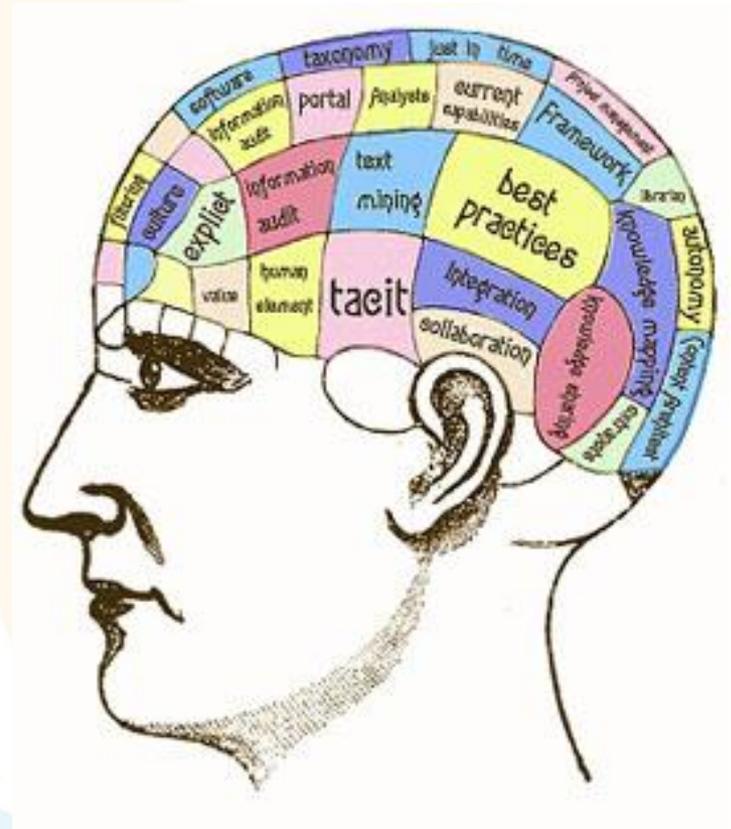
CGGVERITAS

Agenda

- Introduction

- What is Data Management?

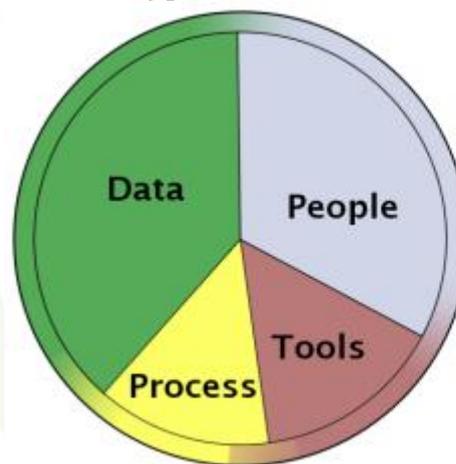- Legacy data Issues

- Some observations

- Some suggestions

CGGVERITAS

Introduction

Intuition:

Immediate insight or understanding without conscious reasoning.

CGGVERITAS

Introduction

- In 2010 a UK study (CDA) found that more than 70% of the value of E&P activities comes from understanding the subsurface.

- Of the 4 key factors, <u>data</u> was considered to be the most influential.

The elements that contribute towards understanding the sub-surface

- Ref: CDA/Schlumberger (2010): "The business value case for data management – a study" www.oilandgasuk.co.uk/datamanagementvaluestudy

Introduction

Some questions:

- How many people in the audience consider themselves to be data managers?

- Who feels that their work is regularly impacted by poorly managed data?

- Who feels that they have good work habits in creating deliverable products that are destined to be managed?

CGGVERITAS

Agenda

- Introduction

- **What is Data Management?**

- Legacy data Issues

- Some observations

- Some suggestions

CGGVERITAS

What is Data Management?

- DAMA - ten Data Management functions.
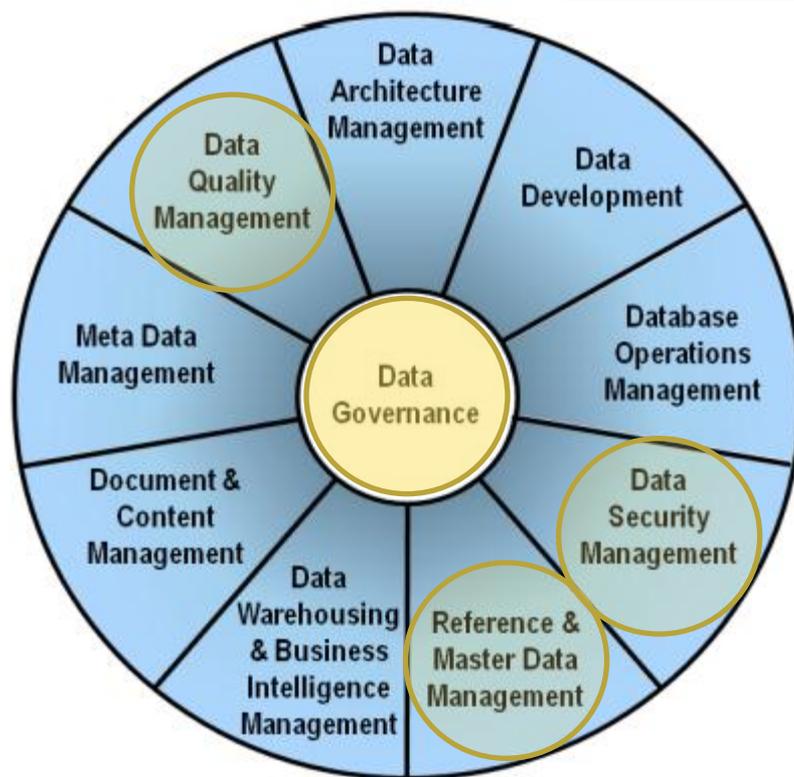
CDA report focused on four functions:

- Data <u>Governance</u>

- Data <u>Security</u> Management

- Reference & <u>Master Data</u> Management

- Data <u>Quality</u> Management

- Ref: DAMA – Data Management International www.dama.org

CGGVERITAS

What is Data Management?

- **<u>Data Governance</u>** is about having processes in place to ensure that data assets are managed formally throughout an organization.

- It addresses such issues as:
  - Architecture
  - Ownership
  - Strategies

CGGVERITAS
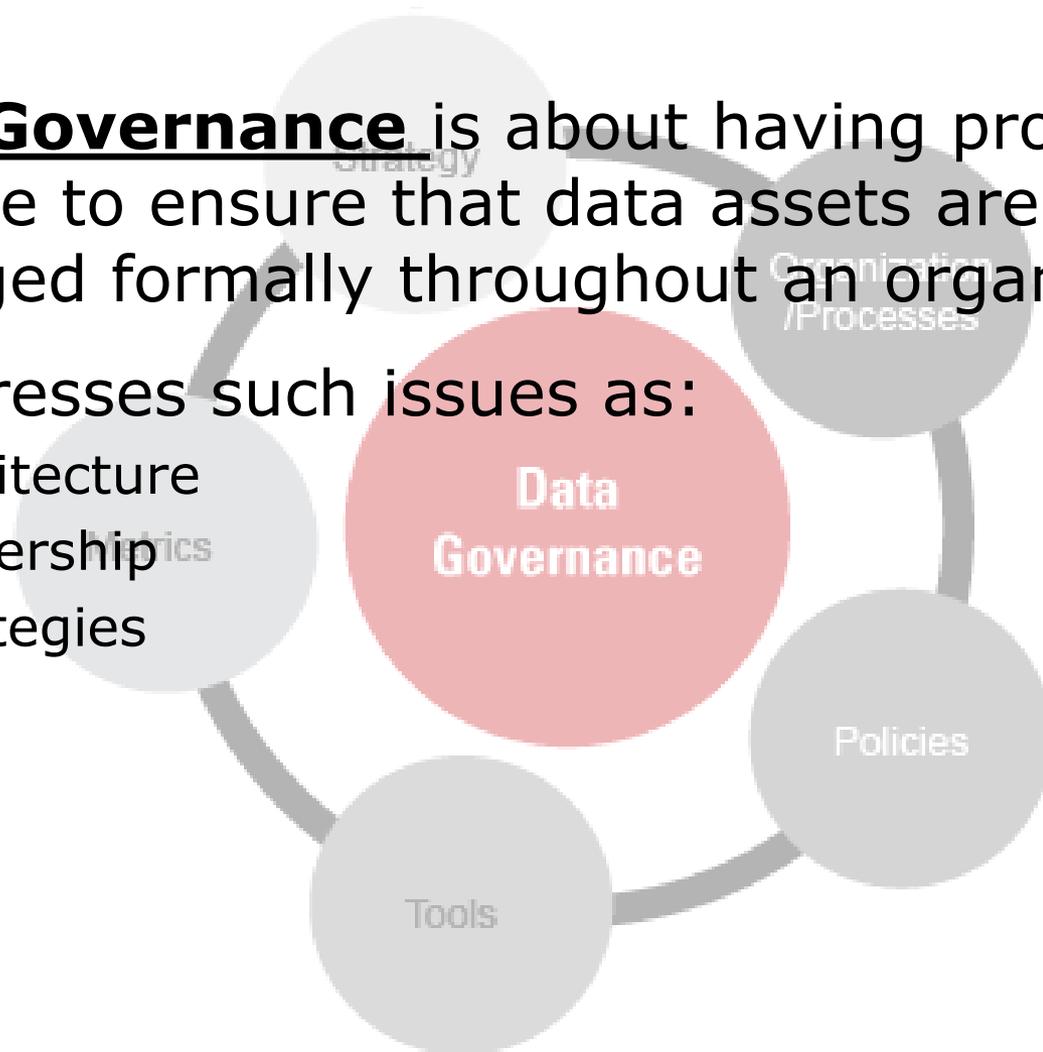
What is Data Management?

- **Data Security Management** is about having the data available to your organization as required but also secure from competitors and accidental loss.

- It addresses risk:
  - Risk of losing data
  - Risk of competitors having access to confidential data

CGGVERITAS

What is Data Management?

- **<u>Reference & Master Data Management</u>** is about exerting control over duplication and inconsistent versions of data, and working towards a single "approved" version.

- It deals with issues like:
  - Multiple repositories/databases
  - Multiple users
  - Version control

CGGVERITAS

What is Data Management?

- **Data Quality Management** relates to having processes in place to ensure that data meets certain quality standards and can be trusted.

- CDA report: *"The best judges of current data quality are the users and not the data managers"*

CGGVERITAS

What is Data Management?

- ***Clearly these four areas need to be addressed but in practice:***

- Different organizations approach data management in different ways – there is no simple off-the-shelf model.

- In general, smaller organizations have less formal processes in place and data management may be on an ad hoc basis.

- As organizations grow, there is typically a need for greater formality and control.

CGGVERITAS

What is Data Management?

- **_Is E&P data different?_**

- Unlike most industries, we have historically used <u>tape-based formats</u>.

- The need to <u>exchange</u> data between companies is commonplace.

- There are legal and financial reasons for us to preserve E&P data <u>indefinitely</u>.

- Datasets can be <u>heterogeneous</u>.

CGGVERITAS

What is Data Management?

- ***Regarding Quality of Technical Data:***

- If your data management team provides you with "bad" data, then presumably either:

- They were given the bad data in the first place, or

- They did something (deliberately or otherwise) to degrade the data

CGGVERITAS

Agenda

- Introduction

- What is Data Management?

- **Legacy data Issues**

- Some observations

- Some suggestions

CGGVERITAS

Legacy Data Issues

- ***What can we learn from management issues with legacy data?***

- Hopefully, quite a bit…
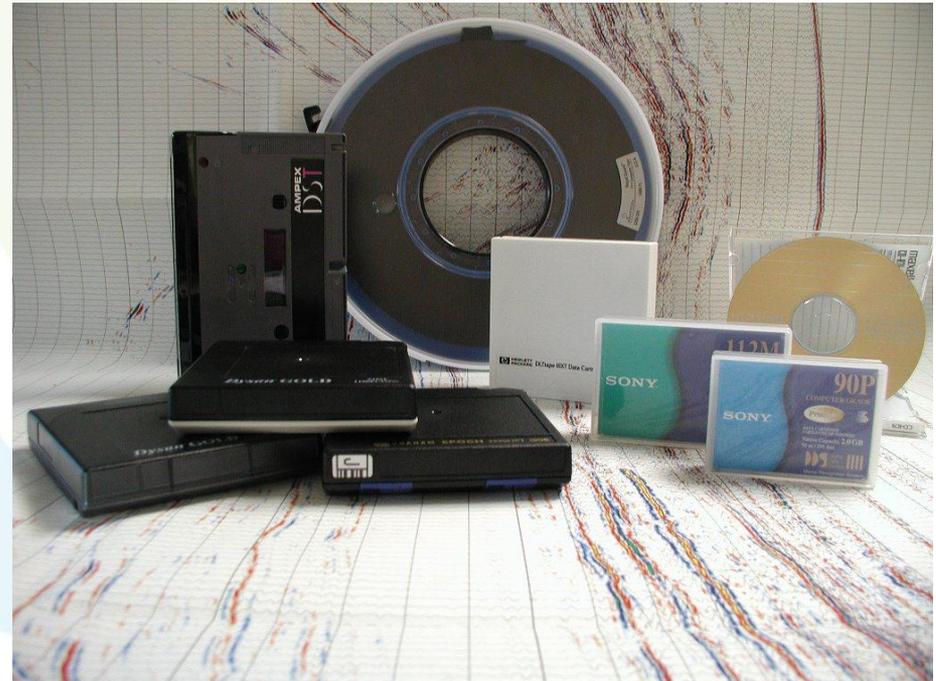
Legacy Data Issues

**"*Legacy" Data typically comes in one of four categories:*

1) <u>Tape-based</u> data (e.g. seismic, well logs).

2) <u>Hard-copy</u> documents (reports, maps, prints).
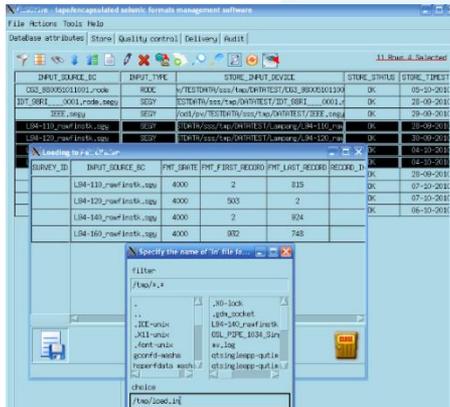
CGGVERITAS

Legacy Data Issues

"Legacy" Data typically comes in one of four categories:

3) Standalone
<u>digital files </u> on
disks, diskettes, etc.

4) Data controlled and
accessed via a <u>database</u>
or <u>application</u>.

CGGVERITAS

Legacy Data Issues

Managing this data means preserving its integrity and accessibility, and should be as simple as:

- <u>Knowing</u> what you've got and where it is (i.e. good cataloguing).

- <u>Copying</u> tape-based data to more modern tape media, or converting it to disk files if appropriate.

- <u>Scanning</u> hardcopy documents, possibly digitizing.

- <u>Migrating</u> disk-based and application data to new solutions as required.

CGG**VERITAS**

Legacy Data Issues

Certain tasks within this process tend to be undertaken as large or ad hoc projects that are often <u>outsourced</u>, for example:

- <u>Remastering</u> of older legacy tape data (format and media conversion).

- <u>Concatenation</u> of more recent tape data (media conversion).

- <u>Scanning</u> of hardcopy documents, possibly <u>digitizing.</u>

This sounds reasonable...

Question: _What problems might arise with this approach?_

CGGVERITAS

Legacy Data Issues

What can go wrong with reformatting data?

- **<u>Loss of Data integrity</u>** – either compromised or rendered useless. For example:

  - <u>Incorrect numerical or channel conversions</u> (e.g. when converting old seismic to SEGY)

  - <u>Incorrect handling of additional attribute or meta-data information</u> during conversion

  - <u>Incorrect sampling</u> (e.g. LIS to LAS)

CGGVERITAS

Legacy Data Issues

Tape to Disk issues?

For tape-based formats, "copying" to disk is not a simple copy process.

- "Encapsulation" is an option that permits the original tape content to be preserved

- File/Directory names need to be assigned

- The format may need to be checked/corrected

- Conversions may need to be made (e.g. EBCDIC to ASCII, blocked to unblocked, LIS to DLIS/LAS)

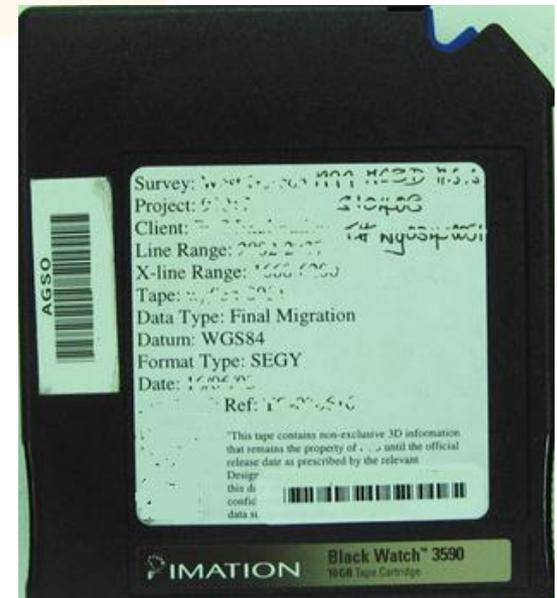**CGGVERITAS**

Legacy Data Issues

What else can we lose?

Loss of meta-data

When tapes are copied, decisions are made about how to handle input and output tape label info.

Meta data was often a combination of both file headers and tape label content.

Creators presumably assumed that the two would stay together.

CGGVERITAS

Legacy Data Issues

What about legacy well data?

Conversion from tape to disk files

There are different tools and approaches for converting well data from 9track or cartridge to files on disk – e.g. converting a LIS file to TAP or DLIS or LAS, etc.

Some conversions may well involve data loss (e.g. LIS to LAS) and the results are subject to the software in use and the operator-selected options.

CGGVERITAS

Catalogue Information?

Naming Conventions

Variation in naming is common and makes searching for data and sorting difficult.

SURVEY	LINE	TYPE
XYZ	ABC1	FLD
XYZ-81	ABC-01	PROC
XYZ	ABC01	Processed
XYZ-1981		Navigation

Legacy Data Issues

Other Issues?

Version Control

In remastering jobs involving processed data, it is not uncommon to find multiple versions of the same data – e.g. SEGY files with same EBCDIC headers, but trace data different and nothing to identify which is the correct version.

Legacy Data Issues

- The service providers undertaking reformatting and remastering have expertise in data recovery and legacy media/formats, but…

- Don't assume data _remastering_ involves data _cleanup_.

- Don't expect data cleanup for peanuts…

CGGVERITAS

Agenda

- Introduction

- What is Data Management?

- Legacy data Issues

- **Some observations**

- Some suggestions

CGGVERITAS

Some Observations

Data remastering will continue to be an issue:

The problems associated with updating media, moving data from tape to disk and occasionally reformatting will continue.

These efforts will continue to be undertaken by specialists in media and formats, but not seismic or well experts.

This function will not improve quality and will likely degrade it if not controlled and checked.

CGGVERITAS

Some Observations

Some worrying trends?

Creating and checking deliverables is a skill that is diminishing.

Time pressure often means that it is tempting to create and deliver products quickly.

Many customers do not have clear content standards.

CGGVERITAS

Agenda

- Introduction

- What is Data Management?

- Legacy data Issues

- Some observations

- Some suggestions

CGGVERITAS

Some Suggestions

Data creators play a role in minimizing quality problems:

Content standards should be sensible, understood and followed. In their absence, common sense should prevail.

All pertinent meta-data should be in the headers – not on tape labels or in email text, etc..

Products should be checked before delivery; versions should be controlled

CGGVERITAS

Some Suggestions

In the case of well data:

Remember – Petrophysicists are better judges of well data quality than data managers or transcription companies…

So be involved in decisions regarding the management of well data.

Make sure that deliverables that you create are complete and correct for ease of management.

Keep original data and audit trails.

CGGVERITAS

Some Suggestions

Intuition:

Immediate insight or understanding without conscious reasoning.

Does it serve us well?

CGGVERITAS

How well does intuition serve us when managing data?

QUESTIONS?

Andy Cairns
CGGVeritas (Data Services division)

FESAus Technical Meeting
June 2012

CGGVERITAS