Pressure Point Ltd

Left Field Solutions
Lateral and elegant IM solutions

Collateral Damage Brewing inside e-Government?

Or

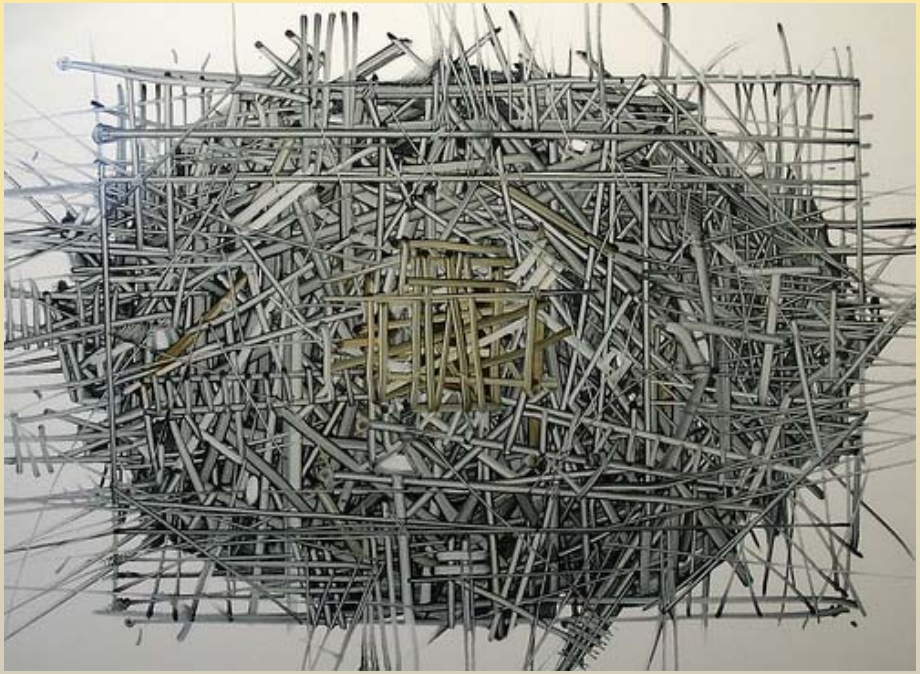
"*A Treatise on the Nature and Consequences of Ephemera*"

Three Contentious Statements

1. Data is inherently untrustworthy
2. Systems that *create* records are seldom suitable for *keeping* records
3. Destruction of data is badly managed

The outcome is a high risk of collateral damage

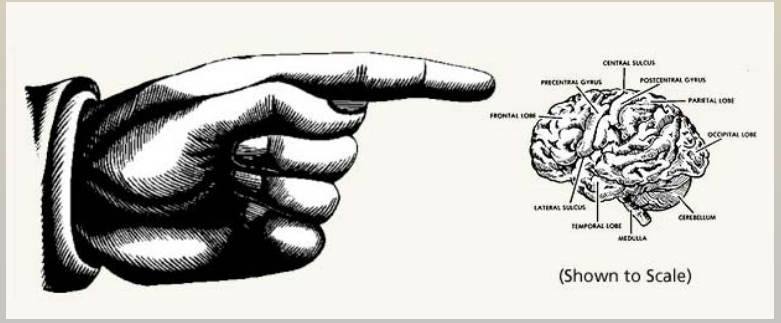
Data in government - Current state



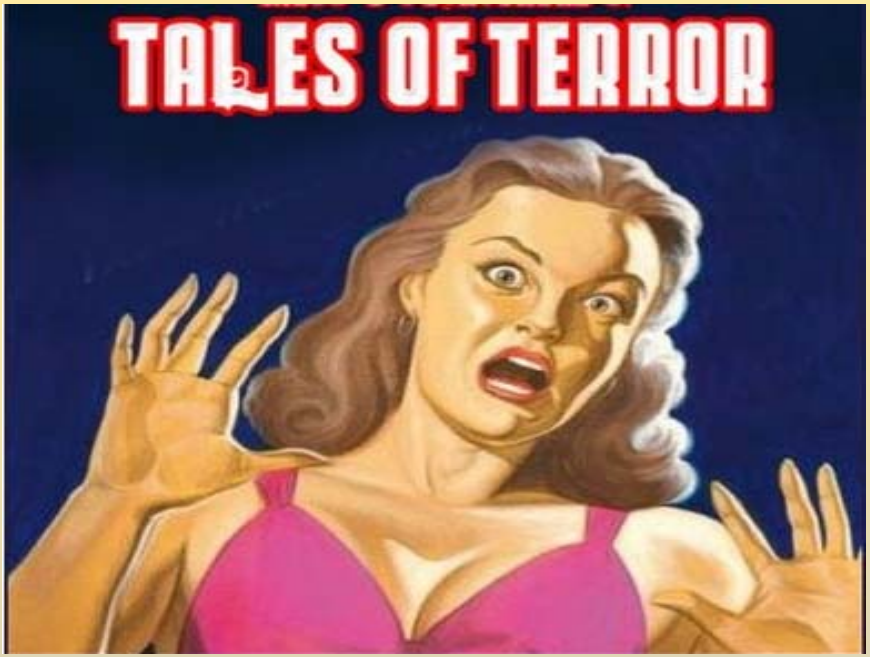
- Where is the data?
  - LOB databases
    - Oracle, SQL,
    - Mainframes
    - Other proprietary databases
  - Collections of databases
  - Datamarts
    Data warehouses
  - MS Access "databases"
- Plus
    - Spreadsheets, MS office documents, PDFs, videos,

Who’s managing the data?

- Who's responsible, and to what extent?
  - Business manager
  - IT (by default)
  - Records manager for data about records
  - No-one?



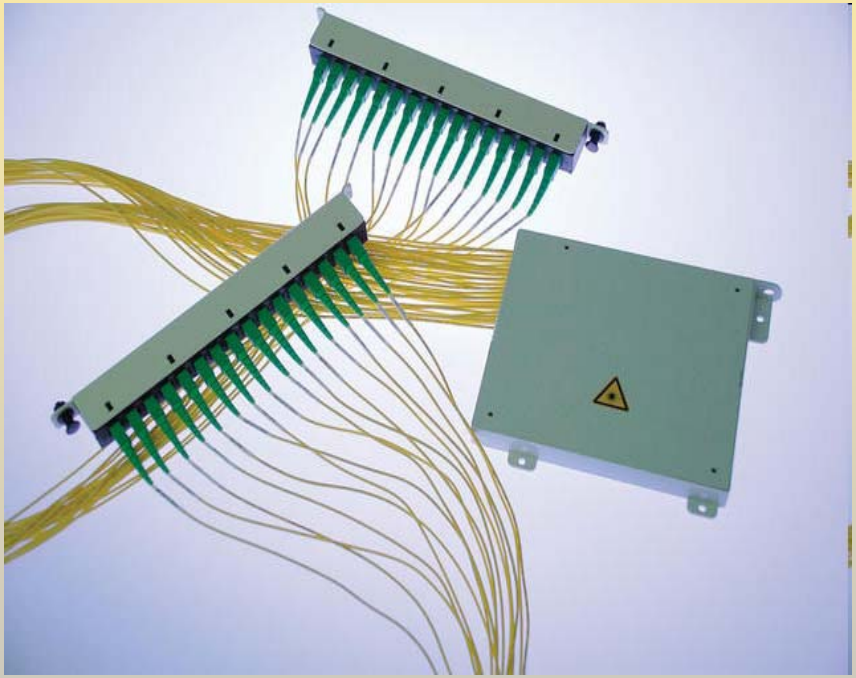
- Variable:
  - Ownership/stewardship
  - Timeliness of delivery
  - Processing Accuracy
  - Currency of data
  - Sources of data
  - Data exchange - inside and between organisations
  - Disposal of "redundant" data
  - Analysis
  - Reporting



or

“Who cares if there's more than one Ms Jones?”

Splitting and recombining data



- "*Here we go again*"

  Organisation split up into several, then brought back into one organisation.....
- One database split into many
  - Organic growth, no unique IDs
  - Then recombined
- *Where is the one source of truth?*

Old data problems persist when linked to new systems

- "*Still wrong?*"
  Increased visibility reveals the warts...
  - Wrong addresses
  - People linked to wrong contact details
  - Single-sided scanning for double-sided documents
  - Inappropriate comments handwritten on scanned documents

Poor processes allow poor data

- "*We regret…*"

  Ever written to a dead person?"

  Social security numbers identify individual but …
- "Unique" Identifier may be replicated across offices and agencies
- Document author linked to login, but who is that?

Poor systems design

- *"Tssk, tssk"*
  Date fields in different formats
- Accidentally swapping month and day on import
- Abbreviations & acronyms,
  - PDA, anyone?
- Misuse of a data field,
  - e.g. disposal rules in description field
- Heights of adults
  - 16cm tall?
- 32 ways to record gender

Poor usability testing

- *"Duh…"*

  Triggering incorrect answers
  - Inconsistent user interfaces
  - Double negatives
    - At Wellington airport "Yes" means I have not packed any dangerous goods,
    - At Auckland airport "Yes" means I have packed dangerous goods.
  - Right data in wrong field
    e.g. DOB vs Conviction #

Humans make mistakes

- "*Oops*"
- Accidental miss-spelling and accidental miss-entry
- Time recording codes
  - Why are the mail-handling staff recording time against Strategic Policy Development?
- Ever entered the current year in your D.O.B?
- Server clocks reset while applications running
- System settings changed by administrators in ignorance

Deliberate sabotage

- "*How could they!*"
  Deliberate miss-spelling and deliberate miss entry
- Users persistently pick the first item on a drop down list
- Old habits die hard
  - e.g. ignoring structured fields in favour of free text
- Contact details deliberately entered incorrectly so they can't be traced
- Multiple identities legitimately created in order to create confusion

Not following due process

- *"Sloppy…"*
  Property marked as sold, but
  - There was no property to sell
  - Title deeds not transferred

- Registering with a government department
  - By paper triggers a follow up system,
  - Register on line and no follow up system?

Bit Rot and Link Rot

- "*It's off, dear*"
  Bit Rot
  - Photo cards "wear out"
  - Disk storage becomes corrupted
    - This calls for Team Digital Preservation!
- Link Rot for databases
  - Databases link to data in other databases to run transactions
  - Data retention isn't always checked against linkages to other processes/systems

Three Contentious Statements

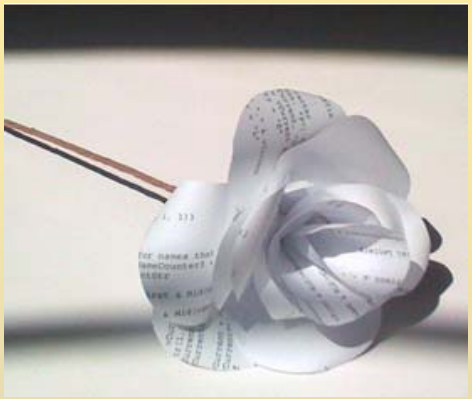
1. Data is inherently untrustworthy
2. Systems that *create* records are seldom suitable for *keeping* records
3. Destruction of data is badly managed

Recordkeeping for data

or

"Is there nowhere to hide from recordkeeping?"

Attributes of recordkeeping systems:

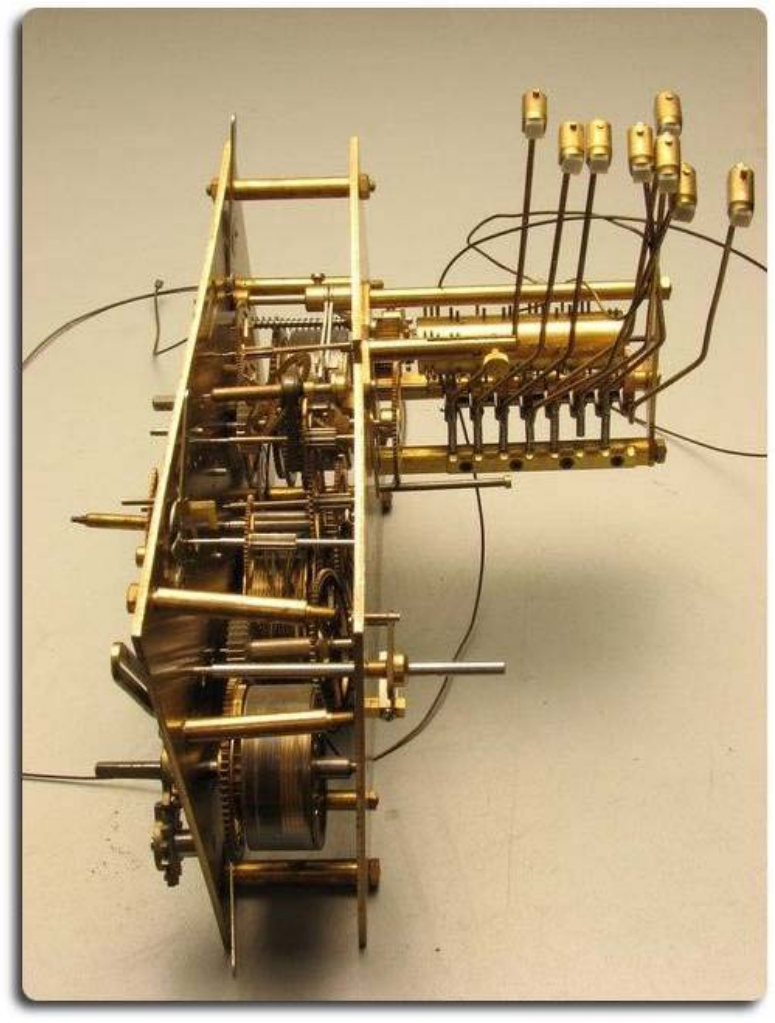


- What's a record?
  - Doesn't mean permanent retention
  - Does mean controlled management
  - Does mean controlled disposal
- See Archives NZ "Create and Maintain Standard"
  - Principle 1: Recordkeeping must be planned and implemented
  - Principle 2: Full and accurate records of business activity must be made
  - Principle 3: Records must provide authoritative and reliable evidence of business activity
  - Principle 4: Records must be managed systematically

How do records attributes apply to data?

- Can the transaction deliver a self-contained record?
- Can you demonstrate that records captured are
  - Comprehensive
  - Authentic
  - Reliable, have integrity and remain unaltered
  - Complete
  - Useable
  - Tamperproof

- Records (datasets!) must be useable, accessible and retrievable for the entire period of their retention.
  - Retention by agreement with Archives NZ
  - Maintenance processes ***(documented!)*** mitigate risk
  - Preserve evidential admissibility
  - Retention and disposal actions must be applied systematically.



Transactional databases

"No place for a complete record?"

Triumph of hope over experience

- Belief in the reliability of IT "archiving"
  - Is there a restore process?
  - Back up failure notification?
  - Where are the backup tapes or disks?
  - Restore process tested?
  - Time to access?
  - Granularity?
  - Recycle period?

- "We d*on't need to keep that"*
  - Horizon is often restricted
- Common Omissions
  - Disposal, whether deliberate or accidental
  - Application to data of the disposal authority that you signed with Archives NZ
  - Identification of active and static data
  - Long-term retrievability

Here today, gone tomorrow…

- Data can be ephemeral
  - Retrieval of data from disparate databases results in a unique record but is it retained anywhere?
  - Is the final complete output available over time?
- Is data valid three months later ?
  - Car ownership transferred but was the vehicle de-registered?
  - Customer address is a previous address

Are we too clever for our own good?

- We now separate format from content.

  Website forms:

  – Does the Agency store data and form?

  – Are they persistently linked?

  – What is the retention period for data? for form?

- De-commissioning databases.....

  – What happens to the records?

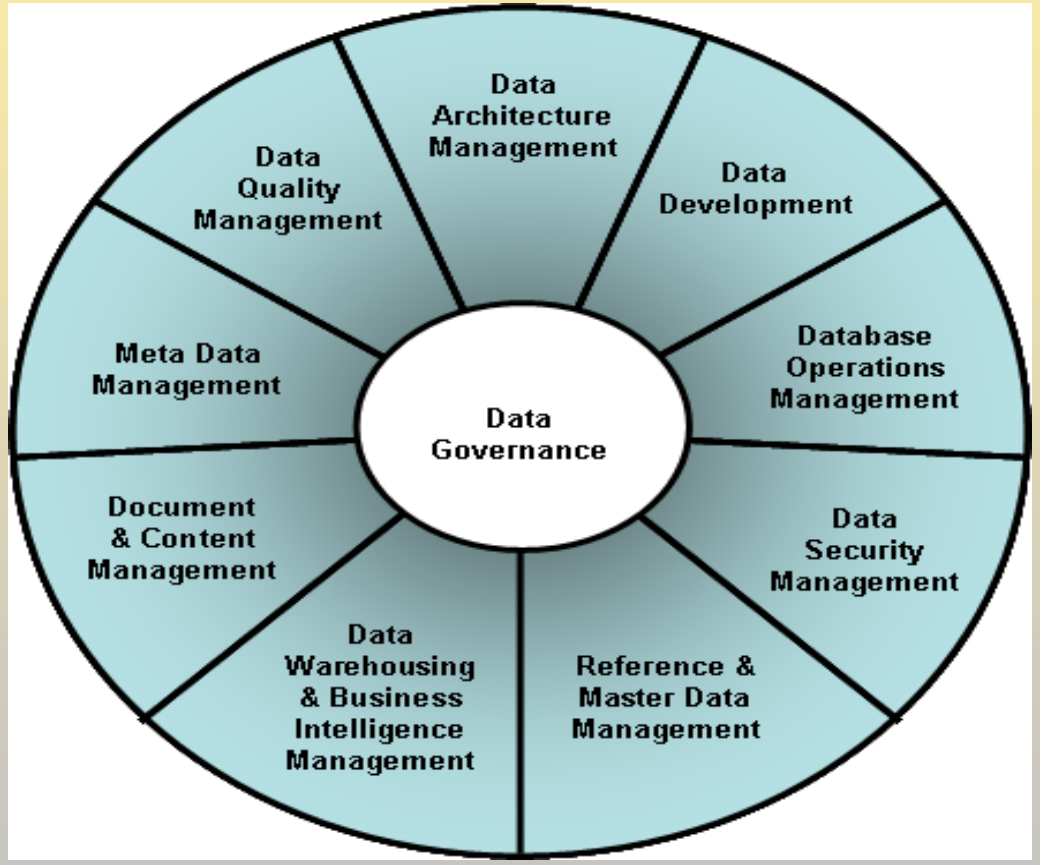
- When is a data warehouse actually a “midden”?

Does anyone care?

- YES
  - The business/agency cares
  - Auditor
  - Researchers
  - Legal counsel
  - Data/enterprise architects
  - Core government agencies that exchange data
  - Private enterprise, unlike government, could go out of business
- The government agency has to care
  - Archives NZ won't be receiving base level data …

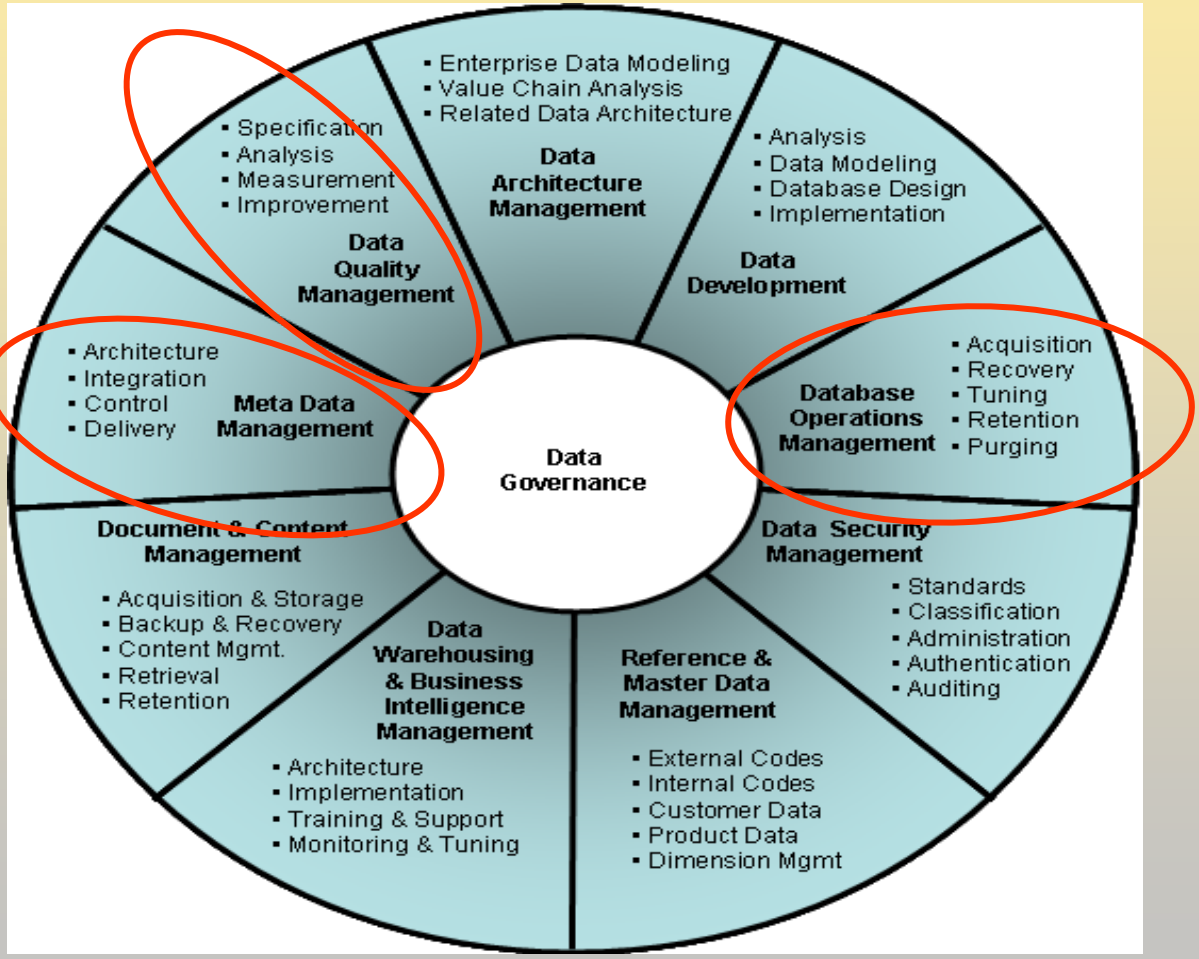
Suggesting solutions

What is Data Management ?



Body of Knowledge

DAMA-DMBOK Framework V.3

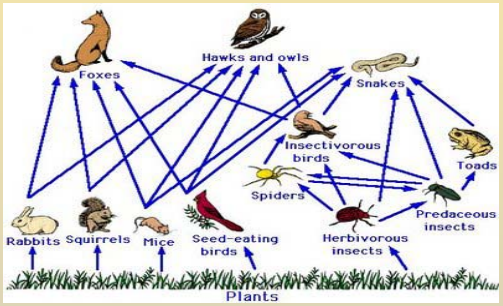


Improve data quality – The People

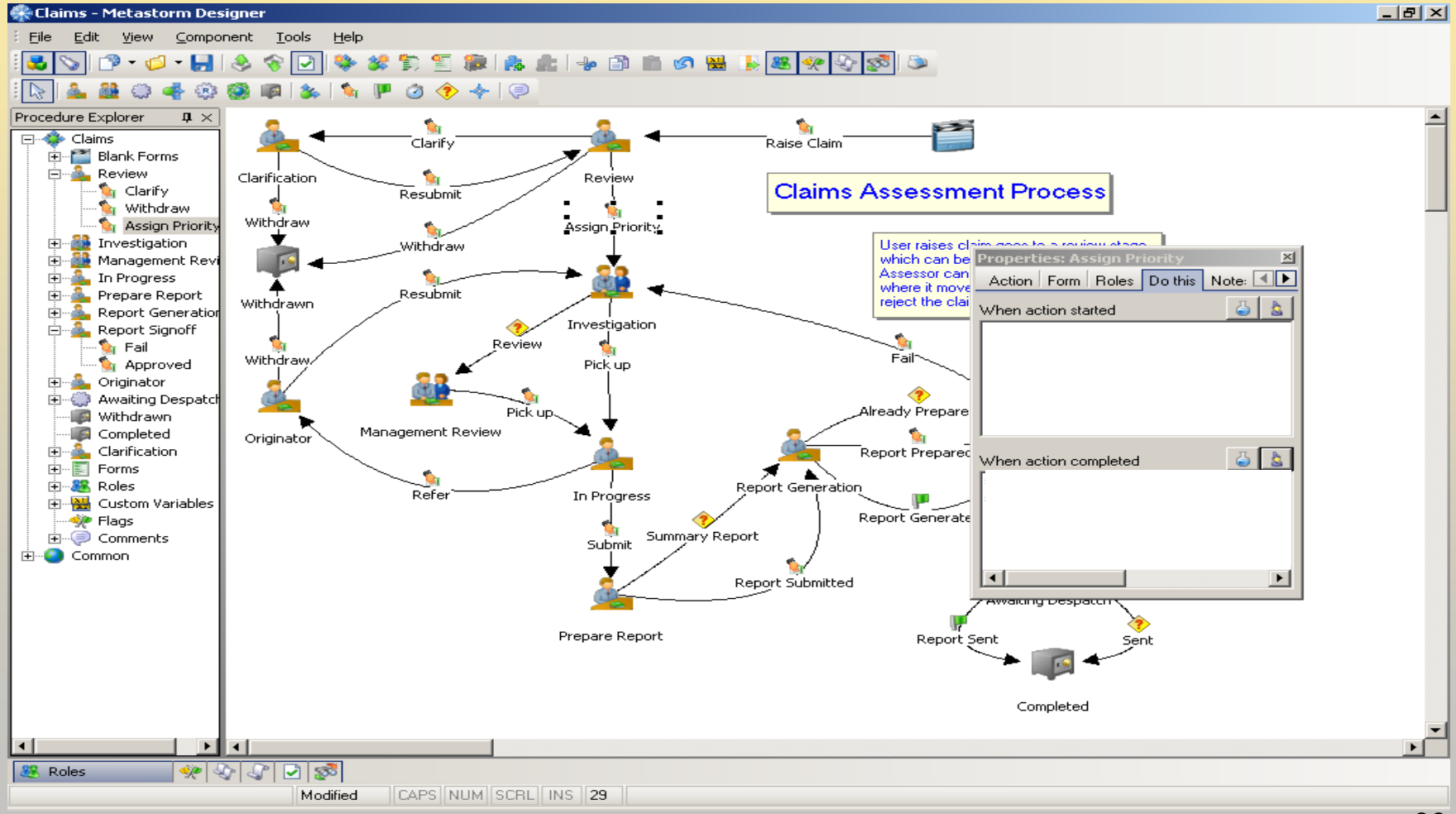
- Alert senior management to the risks
  - What is the “cost” of ignoring poor data?
  - Do managers understand the implications?
- Give authority to the right people who
  - Understand structured data issues
  - Are able to influence Business Managers
  - Are able to work with IT

Improve data quality – The Process

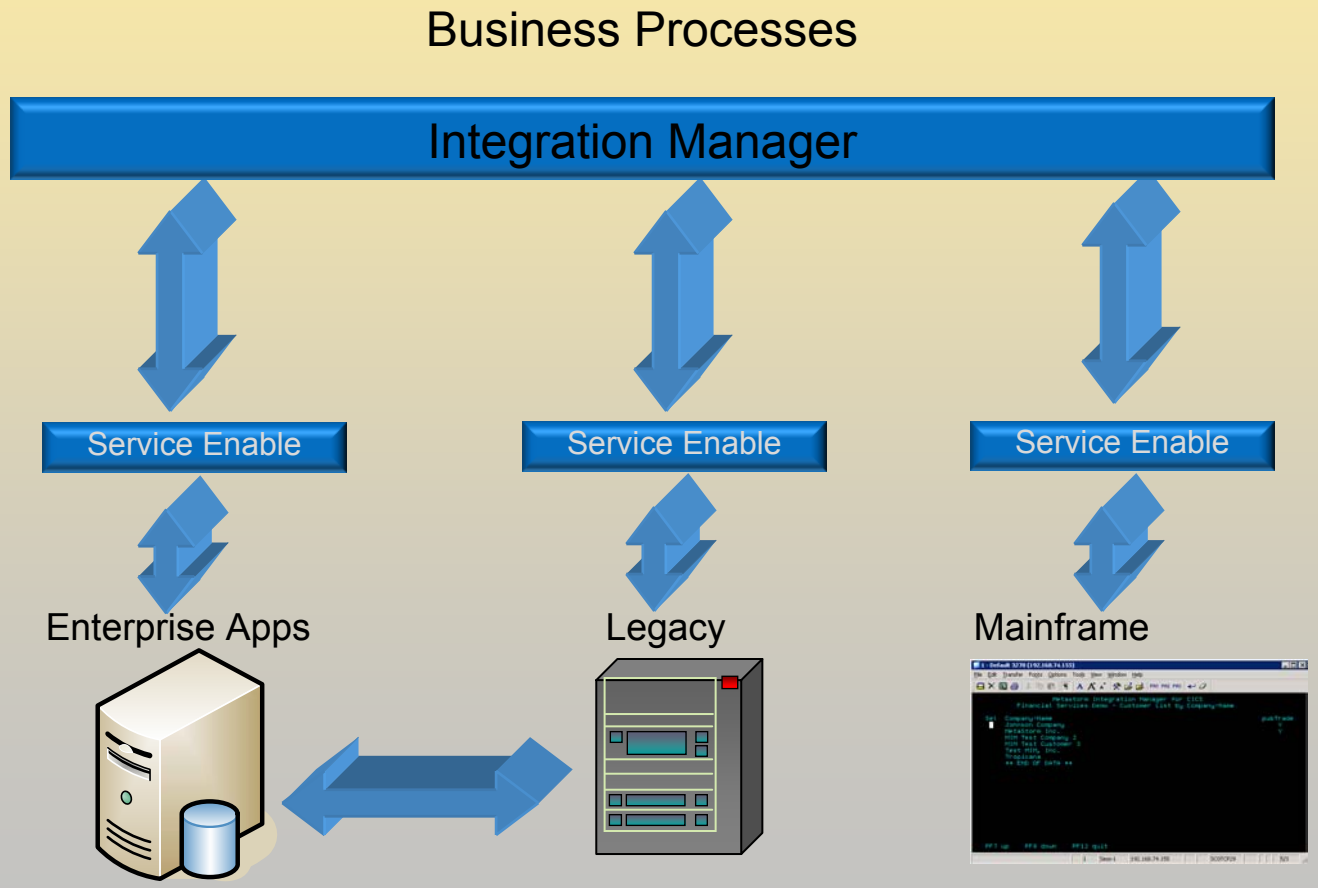
- Introduce Business Process Management
  - Enforces
    - Correct business rules to be applied
    - Consistency
  - Enables end-to-end process
  - Enables data management
  - Produces
    - Accurate, reliable information
    - Audit trail
  - BPM software actually improves staff morale if done right!



Use BPM to engage users
Draw a picture of every procedure



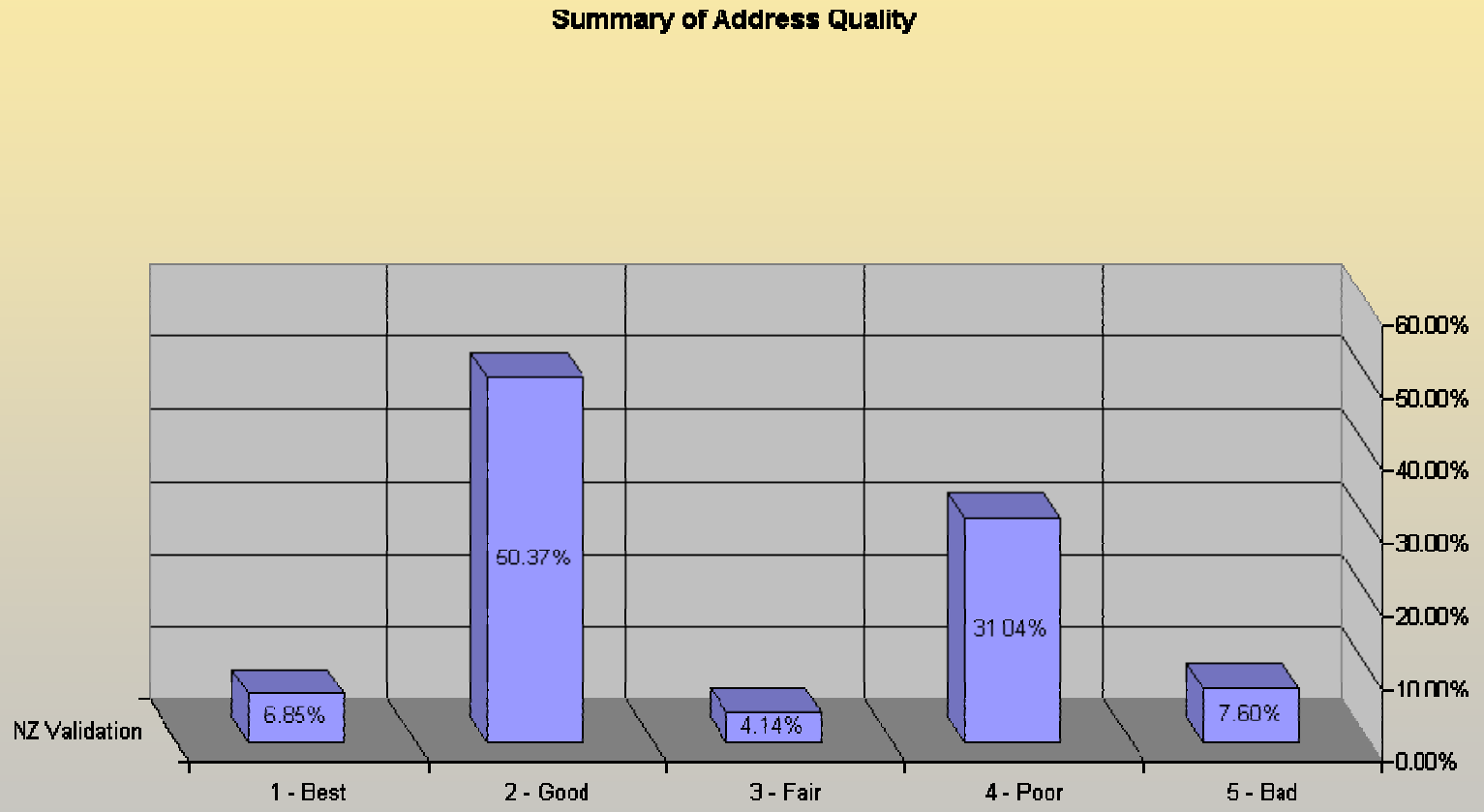
BPM records the process of capturing data



Improve data quality - Use software

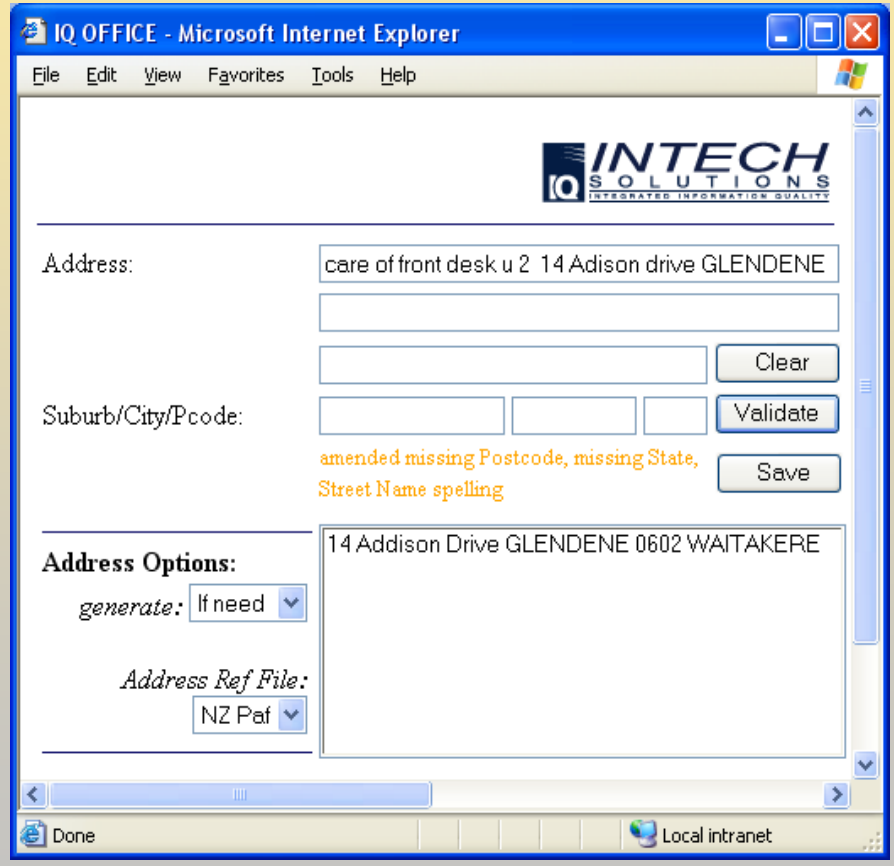
- Validate and Clean the legacy data
  - Data parsing and cleansing can be rapid
  - Free text data becomes structured
- Do intelligent matching
  - Create one source of truth
  - Use proven probability matching techniques
  - Remember that search engines can pick up gems

Measure data quality - Use software



Improve data quality - at Point of Entry

- Validate and verify at point of entry
  - Classic validation tests,
    - e.g. date format, range
  - Verify against existing data
    - Known Addresses, Reference Lists
  - Auto-prompt using Pick Lists
  - Begin with correct data!

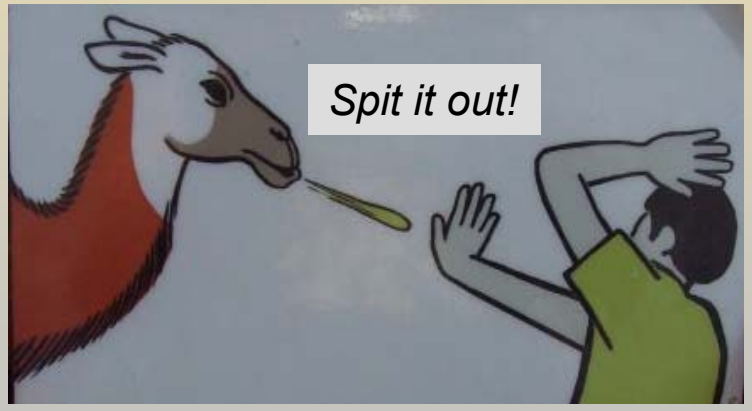


Improve data quality – Test

- Do Usability Testing
  - Enforces consistent user interface
  - Reveals aggravations e.g. too many prompts at the wrong time
- Remember
  - Unhappy Users will 'break' the system or walk away

Where would be a safe place for a complete record?

- If it's not safe, move it
  - Out of record creating systems
- To where?
  - Into record keeping systems
- So, the generating systems don't have to keep records



Spit it into what?

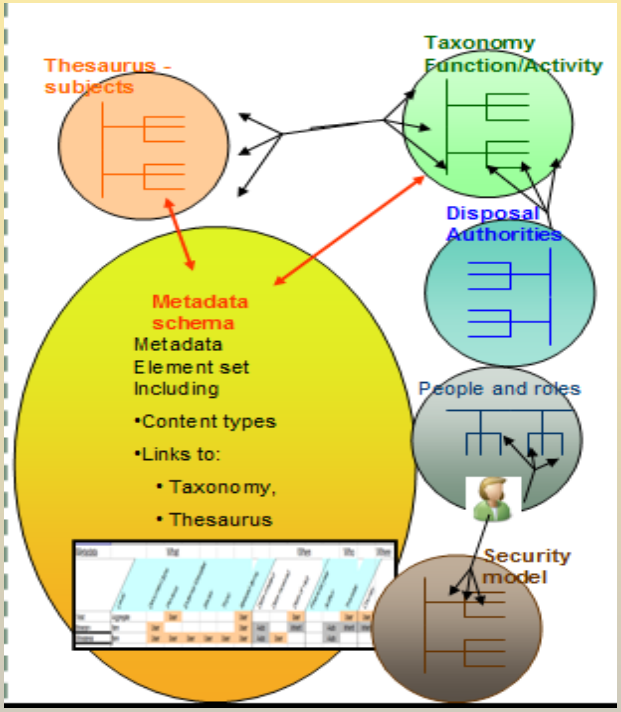
How about a data warehouse?

What would enable a data warehouse to be a recordkeeping system?

- If it is
  - Comprehensive
  - Authentic
  - Reliable, have integrity and be tamperproof
    - Contain only static data
    - Any transformation automatically creates a new dataset
  - Useable, accessible and retrievable
  - Complete
    - All relevant data in a self-contained packet
    - Includes relevant metadata
  - Enabling
    - retention and disposal actions to be applied systematically

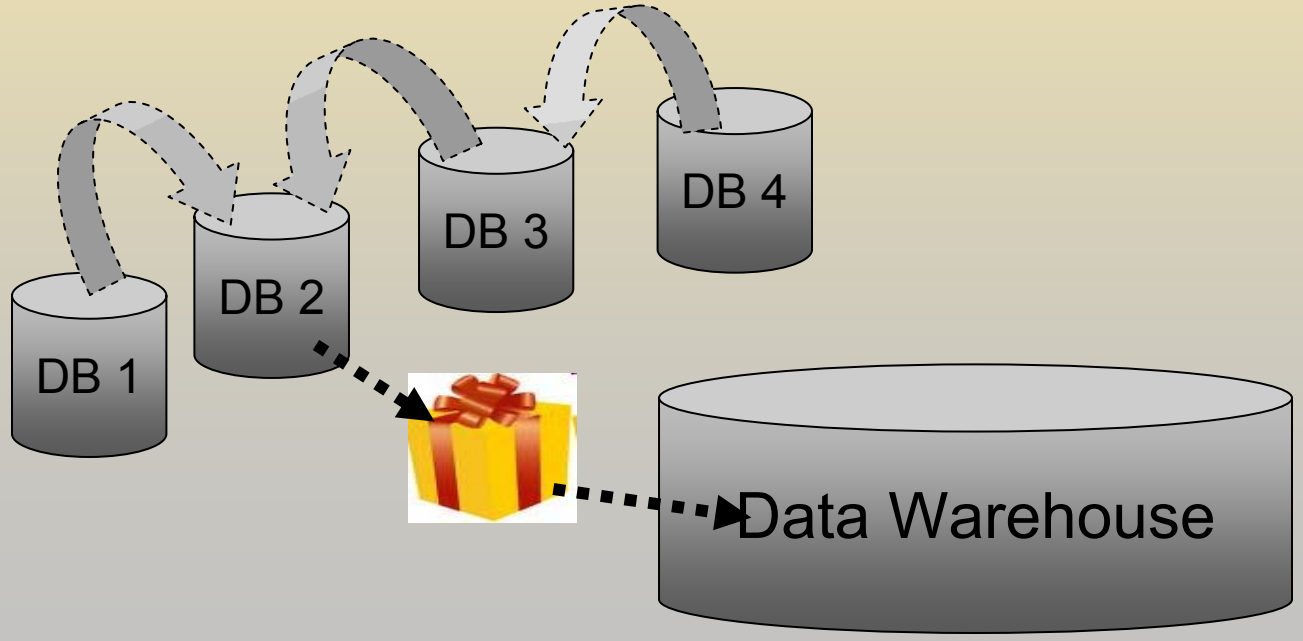
Create a metadata framework for data

- Defines
  - Functions/activities
  - Business rules for retention disposal, security
- Lives independently
  - Outside the EDRMS
  - Applies to all data, structured and unstructured

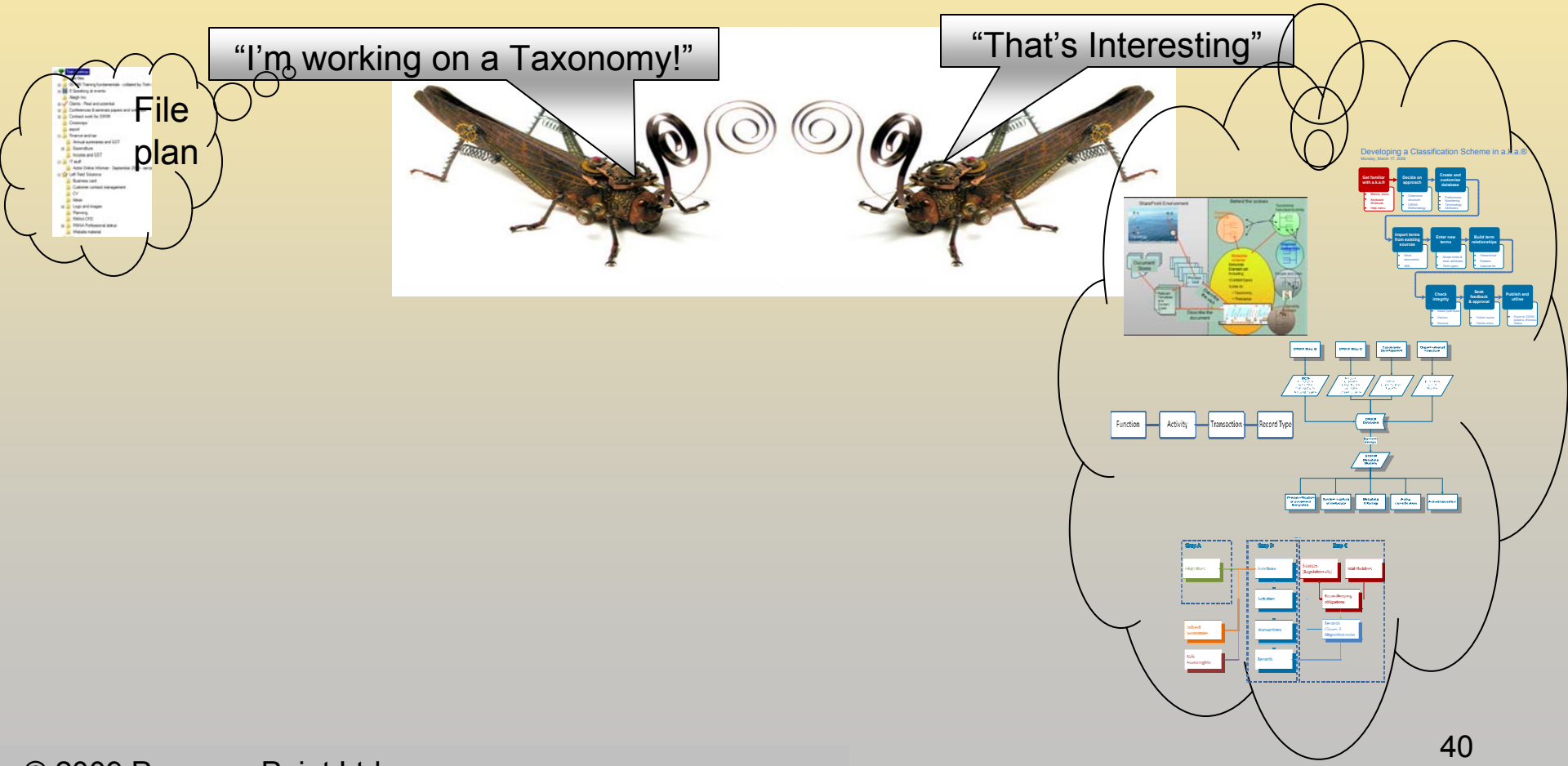


Manage destruction of data

- Complex linked databases are not manageable
- Migrate transaction records into data warehouse
  - Easiest at time of creation
- Decommission transactional databases at will?
  - Only if data already in data warehouse

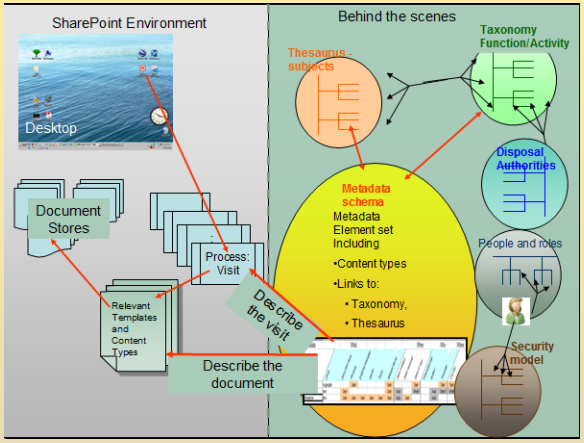


IM Managers should talk to Enterprise Architects



Use Pictures

- Diagrams assist common understanding
  - They don't have to be right
- Develop Scenarios
  - Clark Kent, DBA, maintains tables of staff and their roles
    - He needs to update the job titles of staff in the Krypton Group
  - Lois Lane, Manager, sends regular reports to Privacy Commissioner on data exchanges
    - She needs to …..



- Refer to DRAMBORA Audit Methodology
  - Digital Repository Audit Method Based on Risk Assessment
  - Jointly discuss the detailed appendices

Three Positive Statements

1. Data can become trustworthy
2. We can create system to keep reliable records
3. Destruction of data can be planned

Key learning points

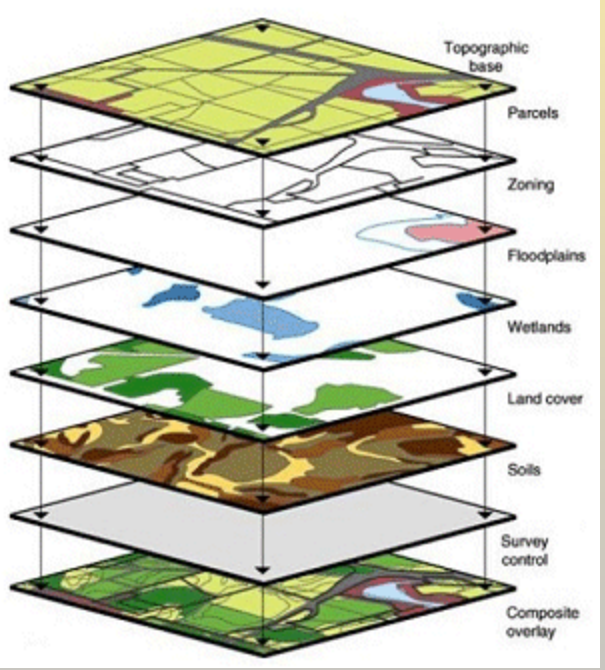
1. Assume that data is inherently untrustworthy
   - Not always accurate
   - Often ephemeral/irretrievable
   - Let software tools structure and clean the legacy data
   - Consider BPM tools to improve process plus data quality

2. Don't assume the system that created the records is suitable for keeping and managing them
   - Business Process Management can help
   - Data Warehouses have potential as recordkeeping systems
   - Defining the context and disposal rules are the key differentiators

3. Data must be tightly managed
   - Principle-based and well understood
   - Events must be controlled by approved business rules

Get a Grip on Your data

Before it is too late !



Thank You!
Here are some references

- Trish O'kane
  - tokane@actrix.co.nz
- Elizabeth Eastwood
  - elizabethe@intechiq.com
- Data Management International
  - www.dama.org
- Business Process Management Institute
  - www.bpminstitute.org
- NZ Government Data Management Policies and Standards
  - www.e.govt.nz/standards/e-gif/data-management
- DRAMBORA Audit methodology
  - www.repositoryaudit.eu
- Team Digital Preservation!
  - www.youtube.com/watch?v=pbBa6Oam7-w