

# Collecting Data Online

Andy Peyton

Department of Health & Ageing

# Agenda

- Background
- Electronic Contact with Participants
- Using Electronic Forms
- Infrastructure Issues
- System Architecture
- Census Period Support Impacts
- Data Management
- Reporting
- Data Cleansing, Enhancement, Rejection
- Analysis

# Terminology

- Providers
  - Organisations that have a legal responsibility
- Outlets
  - Subcomponents of providers that deliver services to clients

# Presentation Structure



# Background

# Background

- Previous paper census in 2002
- Outsourced management to another organisation
- It took about 2 years to get the final census report in 2004
- No data delivered, only the final report
  - So we don't have the data for longitudinal analysis

## Background (2)

- About 1,200 outlets would submit forms
- We ended up with over 1,700 submissions
- Instead of 1 set of forms per outlet, we aimed for 1 set of forms for every person who needed to complete one
- People submitted data on their clients
- 1 form could have data for over 100 people
- We would collect data on about 50,000 people

# Maintaining Contact Electronically

# Electronic Contacts (1)

- We had systems for managing data about organisations that we have dealings with
  - We harvested email addresses from these
- The first broadcast email had massive delivery failures
  - It turned out that none of our systems actually used the email addresses, so they got quite out of date
- Rule of thumb: 10% of email addresses go stale every 3 months

# Electronic Contacts (2)

- We had to build our own contact management system (using Microsoft Access)
- We had to call every Provider and identify a main contact for the census
- We asked each main contact for details of their outlets
- We emailed the providers and told them what they told us
  - We had more delivery failures
- We ended up having to phone each outlet individually and get a main contact with email address

# Electronic Contacts (3)

- **BIG PROBLEM:** outlets used personal email addresses
  - These go stale real quick
  - About 10% per 3 months at least
  - Reflects a large staff turnover
- We ended up sending regular emails to outlets just to maintain the integrity of the email database

# Using Electronic Forms

# Using Electronic Forms (1)

- We had the participants complete and submit a simple form first
  - Non-threatening
  - Showed users how simple it could be
  - Allowed us to see who was able to logon and complete the form and who weren't
- **BIG SURPRISE:** there were no problems with using the electronic forms
- The users ranged from the IT literate to the illiterate
- The time we spent designing the forms paid off
- (show forms)

## Using Electronic Forms (2)

- We were able to change the forms at any time before or during the census if necessary
- We predicted that the business owners would want to change and fine-tune the wording of questions right to the end. And they did.
- Good form-building functionality but not perfect

# Using Electronic Forms (3)

- We shaped the census form to suit the capabilities of the basic tool
  - Only used a single custom plug-in
- Dynamic components had to refresh screen for visually-impaired folks.
  - You can't just hide/unhide blocks of text because most screen-reading software doesn't detect the change

# Presentation Structure



# System Architecture

- Good design
  - Separates online load from analysis load
  - Allowed us to maintain private data entry environment separate from Departmental processing environment
- We don't have access to the data until the user formally submits it

# System Architecture

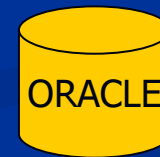
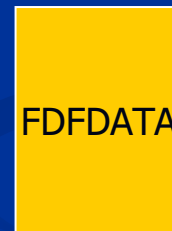
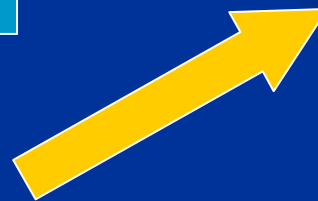
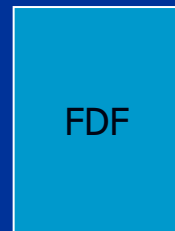


Users interact with Form System

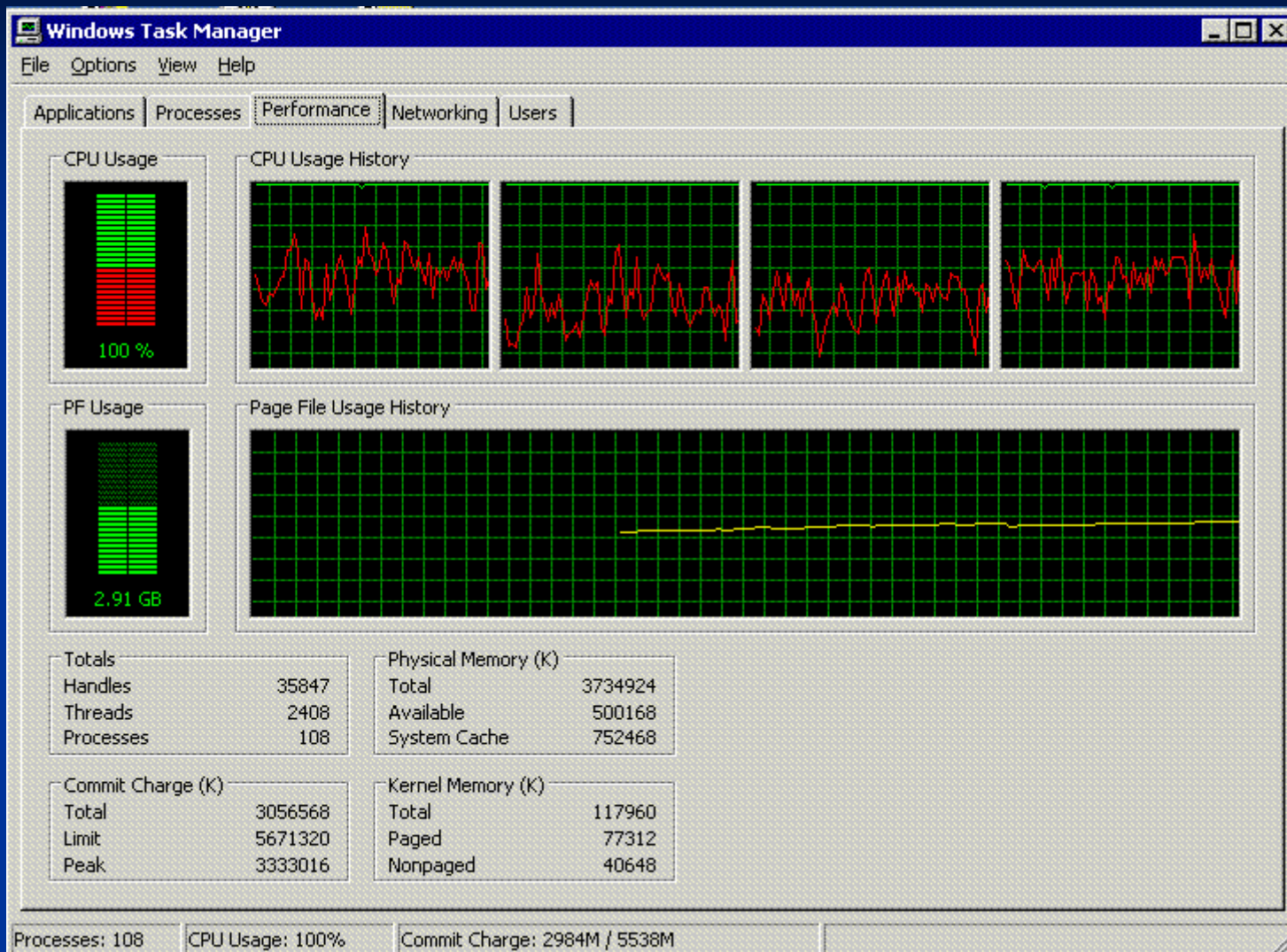
When users "submit" form, data is transferred to database environment

Real-time analysis against database environment

Form system maintains data as XML files



**So what happened when the census  
actually ran?**



# Infrastructure Issues

# Infrastructure Issues

- ColdFusion server issues
- Java issues
- Session timeout:
  - Firstly increased, because of the amount of time people were taking to complete a single form
  - Then decreased, to allow sessions to shutdown and reduce the number of server threads

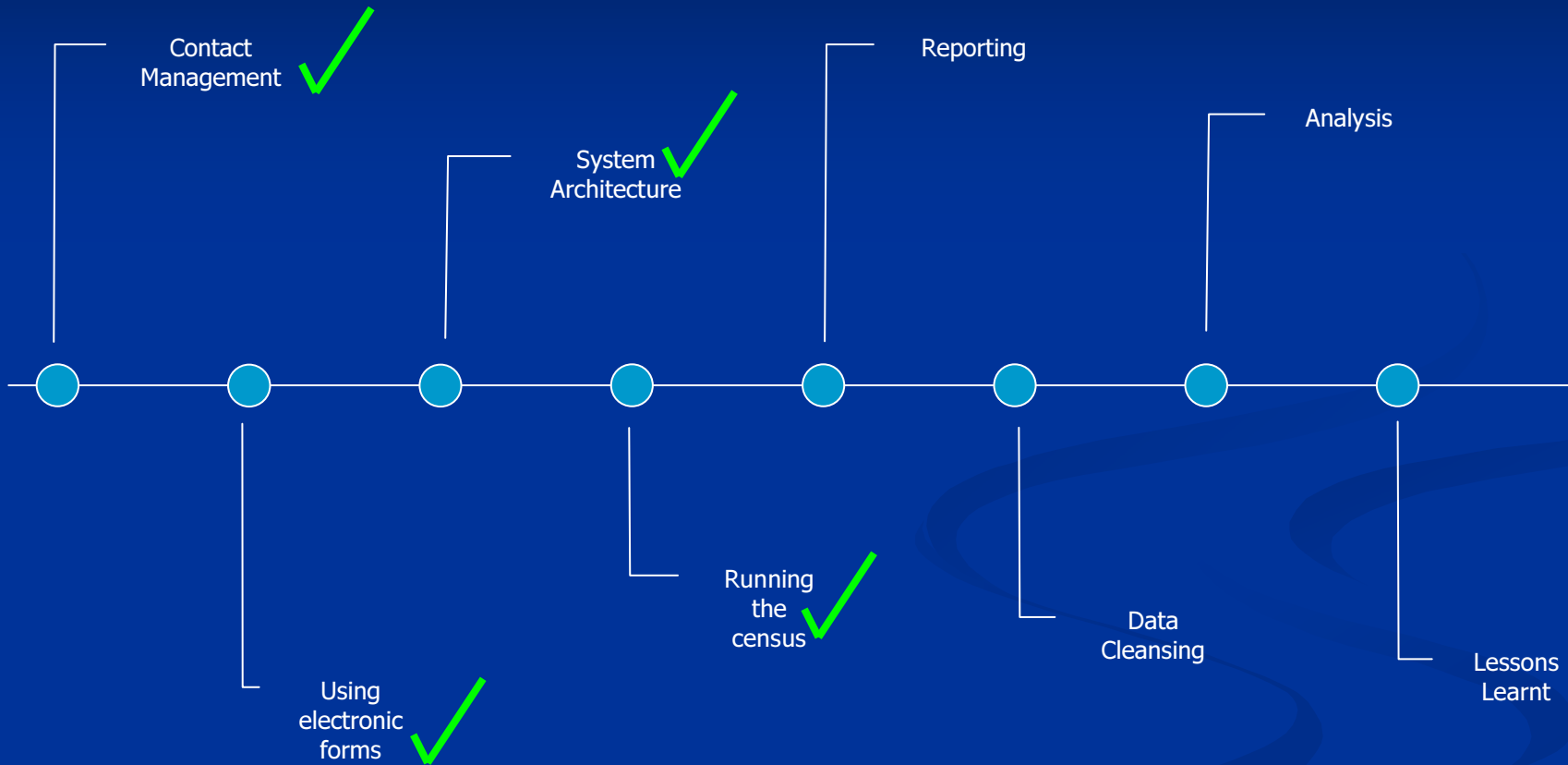
# Census Period

- Help desk
  - Initial inadequacy of help desk only compounded delays
  - People called and emailed
  - People left messages to say that they have already left a message earlier
- Logon problems

## Census Period (2)

- A wonderful feature of our design was that we could see what was going on. We could monitor:
  - Which users had logged on
  - When they last saved
  - When they had submitted their forms
- Monitoring was linked to our contact management system – essential for help desk

# Presentation Structure

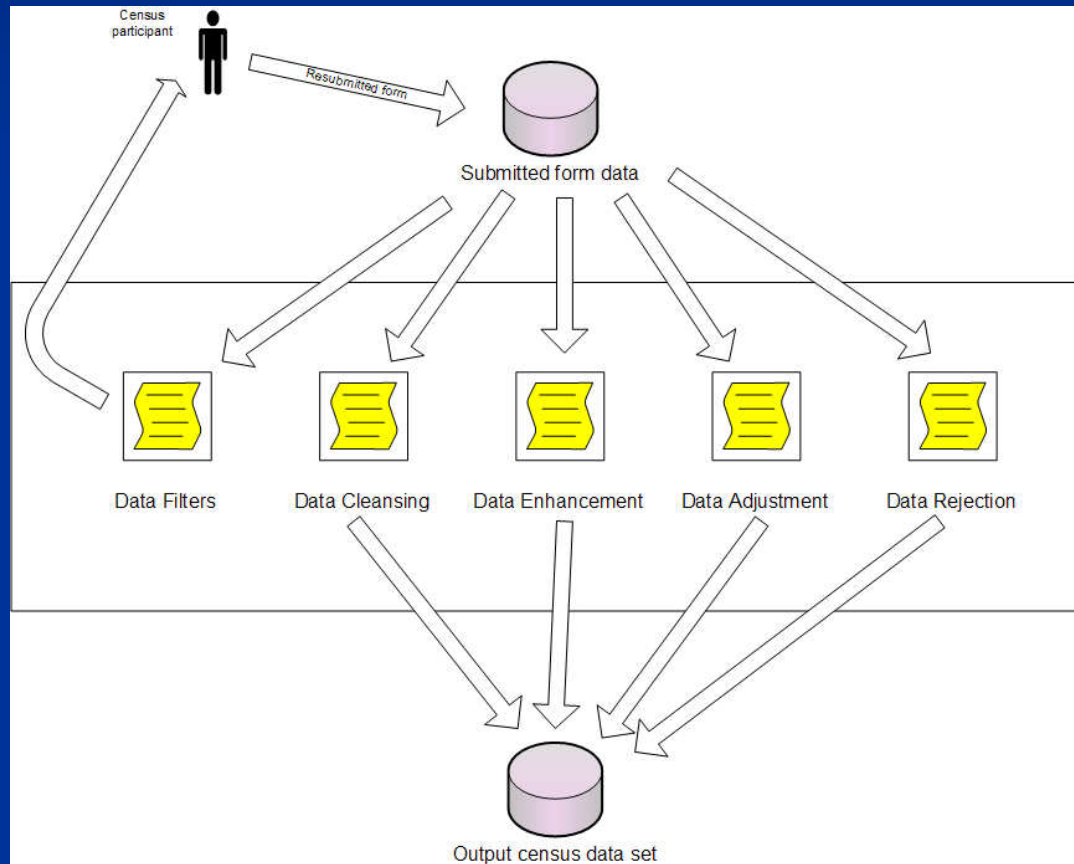


# Reporting

- We were able to report on data as soon as the first form had been submitted
- We could monitor data quality issues in real time
  - We adjusted some field validations as a result
- We could identify systemic issues
  - Eg outlets submitting “default” forms
  - Or outlets submitting bad data that couldn’t be detected by form validations

# Data Cleansing, Transformation, Enhancement, etc

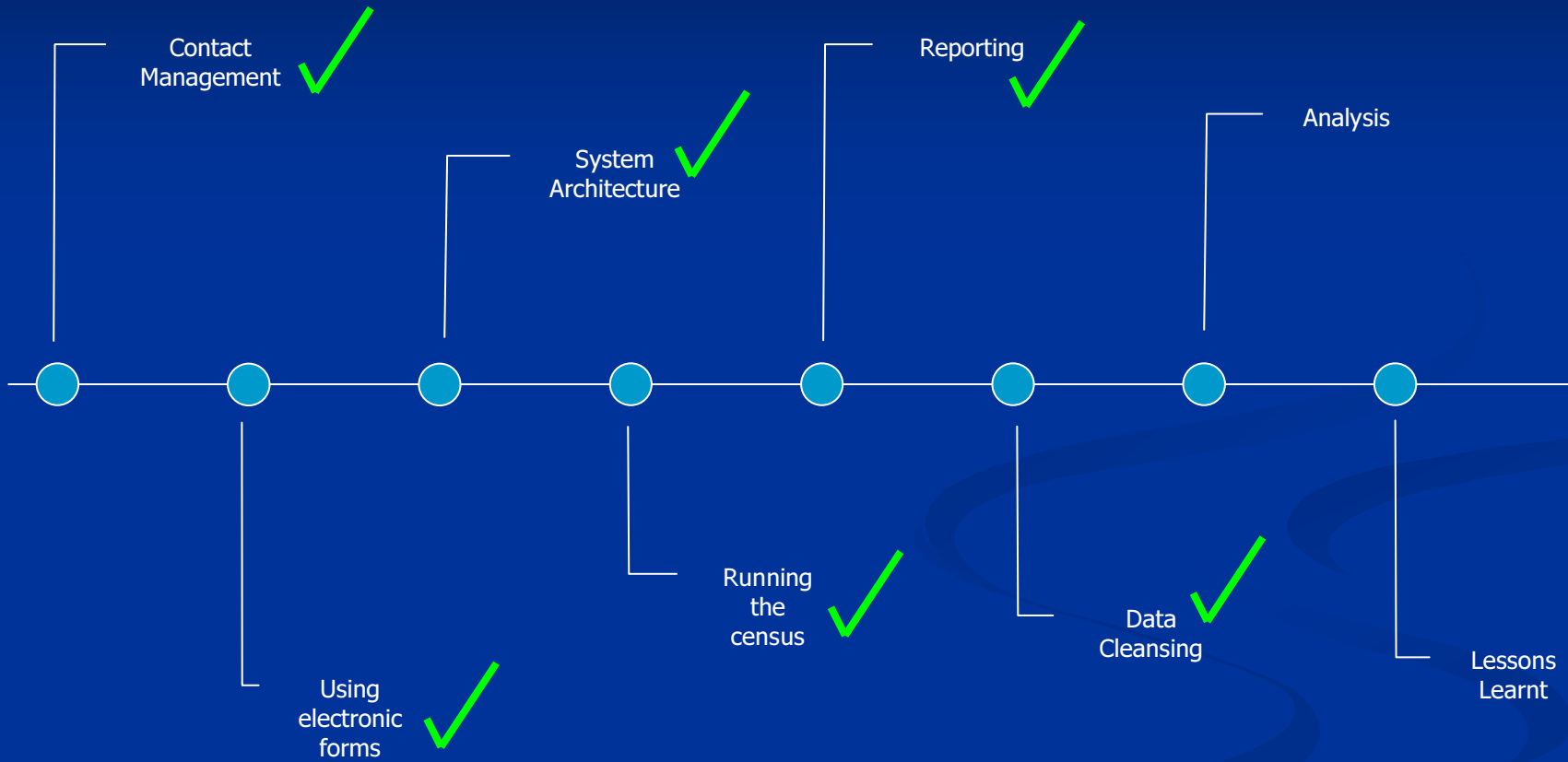
# Data Cleansing, etc (1)



# Data Cleansing, etc

- Key cleansing issue was the use of “unreliability indicators”
  - They indicate what parts of the client records are unreliable
  - Allow good data to still be used
  - Maximise data retention

# Presentation Structure



# Analysis

- We used Excel pivot tables (Microsoft Cubes)
- 27 of them
- They were quick to build, modify, deploy
- We had 250-300 facts about each person
- The largest pivot table had over 100 dimensions
- While there were a lot of standard analyses we had to perform, there were also many unexpected investigations that were necessary

# Lessons Learnt

- You can never have enough server capacity
- People would rather have an e-form that works like a paper form. Didn't like dynamic elements
- Keep the URL & logon process as simple as possible
- Have some spare Help Desk capacity
- Keep on good terms with your infrastructure folks “you want a new server tomorrow???”

## Lessons Learnt (2)

- Make sure you can communicate “hot issues” to users
  - We had a “hot issues” section on the front of the e-Form that we could change at a moment’s notice

*Any Questions?*