



# Turning Data into Knowledge: A case study of researcher- administrator collaboration using multi-university student data

AAIR FORUM – November 2009

Dr Gigi Foster

Mr Graeme Poole

Mr Peter Antony

# Workshop Overview

- Introduction to the project and its requirements (Gigi)
- UniSA systems and protocols (Graeme)
- UTS systems and protocols (Peter)
- Conclusion (Gigi, Graeme, Peter)
- Please interrupt us with questions or comments at any time!

# Academic background

- Who I am and what I do
- Research niche in undergraduate education research in Australia: no large **panel** data sets
- Advantages of panel administrative data:
  - Observation of actual change over time – rather than guessing about change based on cross-sections
  - Students observed in classes, programs, semesters...can isolate the importance of each to explaining outcomes, while removing the student's own idiosyncratic tendencies
  - No non-response (though surveys can also be matched in)

# Policy background

- How do students of different backgrounds fare, in terms of grades and attrition, in one of the largest undergraduate degree areas (business education)?
- Expanded access to higher education: implications for the academic progress of different sorts of students?
- Dearth of reliable, comprehensive information about student welfare to inform universities' strategic planning

# Inception of the project

- Thought up the idea on the bike coming home from work in about December 2006
- Cold-called ATN universities, asking for willingness to provide data, if the project were to get funded (two 'yes' responses)
- Wrote the application and applied for funding under the ARC Discovery Project scheme in March 2007
- Project approval in October 2007

# Setting up project protocols

- Nov-Dec 2007: Met with data extractors at both universities; applied for Ethics Approval of the entire project (surveys approved individually)
- Data extraction timeline
- Survey timeline
- Data items required

# UniSA data extraction

- Multiple student systems with different fields in each and different levels of analysis
- Data obtained from the Division of Business office plus Student Systems (two points of contact)
- SATAC IDs versus Student IDs
- The first extracts were not easy!

# UTS data extraction

- More user-friendly system...only minor teething problems
- One point of contact

# Post-extraction

- Student and course-level data from different extracts merged by student ID for each university
- Key analysis variables constructed (e.g., attrition, GPA, dropping and adding courses), plus program/course dummy variables, student background dummy variables, etc
- Survey data is merged in by student ID-semester-uni
- Analysis data sets constructed at different levels
- Data sets from the two universities are stacked
- All data massage performed in SAS

Questions so far?

# Concluding comments

- What we have learned so far:
  - Massive amounts of data are there to be used!
  - Differences across institutions in ease of access; terminology; field codings; student processing (e.g., credits and graduation data)
  - Current data storage and handling protocols used by HRECs are antiquated: a central secure repository would be ideal
  - Necessary inputs for this type of project: Excellent logical programming skills; willingness to make judgment calls when all else fails; time and energy; patience

# Future of the project

- Our efforts will produce the first universe-level panel of Australian undergraduate student data at this level of detail
- Several academic papers are in the pipeline
- Beyond 2010, if we find funding elsewhere, we can continue building the panel
- Key focus for the next 18 months is finding and communicating effectively with education policymakers

# Questions and Comments

[gigi.foster@unsw.edu.au](mailto:gigi.foster@unsw.edu.au)

[graeme.poole@unisa.edu.au](mailto:graeme.poole@unisa.edu.au)

[peter.antony@uts.edu.au](mailto:peter.antony@uts.edu.au)



# Turning Data into Knowledge: A case study of researcher-administrator collaboration using multi-university student data

AAIR FORUM – November, 2009  
Presentation by Graeme Poole  
Coordinator: Information Services and Reporting,  
University of South Australia

# Covering ....

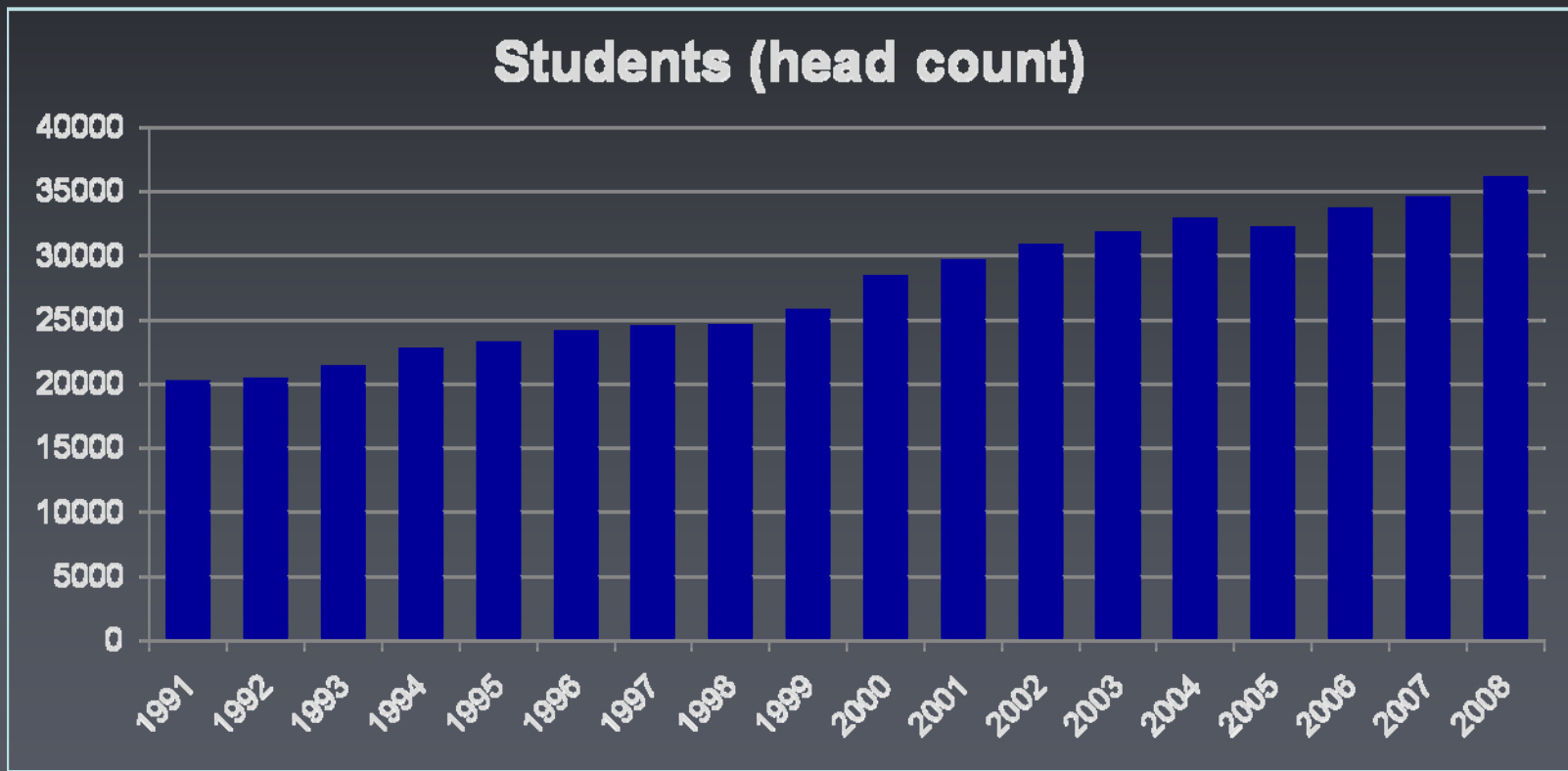
- Institution profile
- Data requirements
- Limitations
- Methodology
- Problems Encountered

# Institution Profile

- UniSA was founded in January 1991 through the amalgamation of the South Australian Institute of Technology and the Magill, Salisbury and Underdale campuses of the South Australian College of Advanced Education

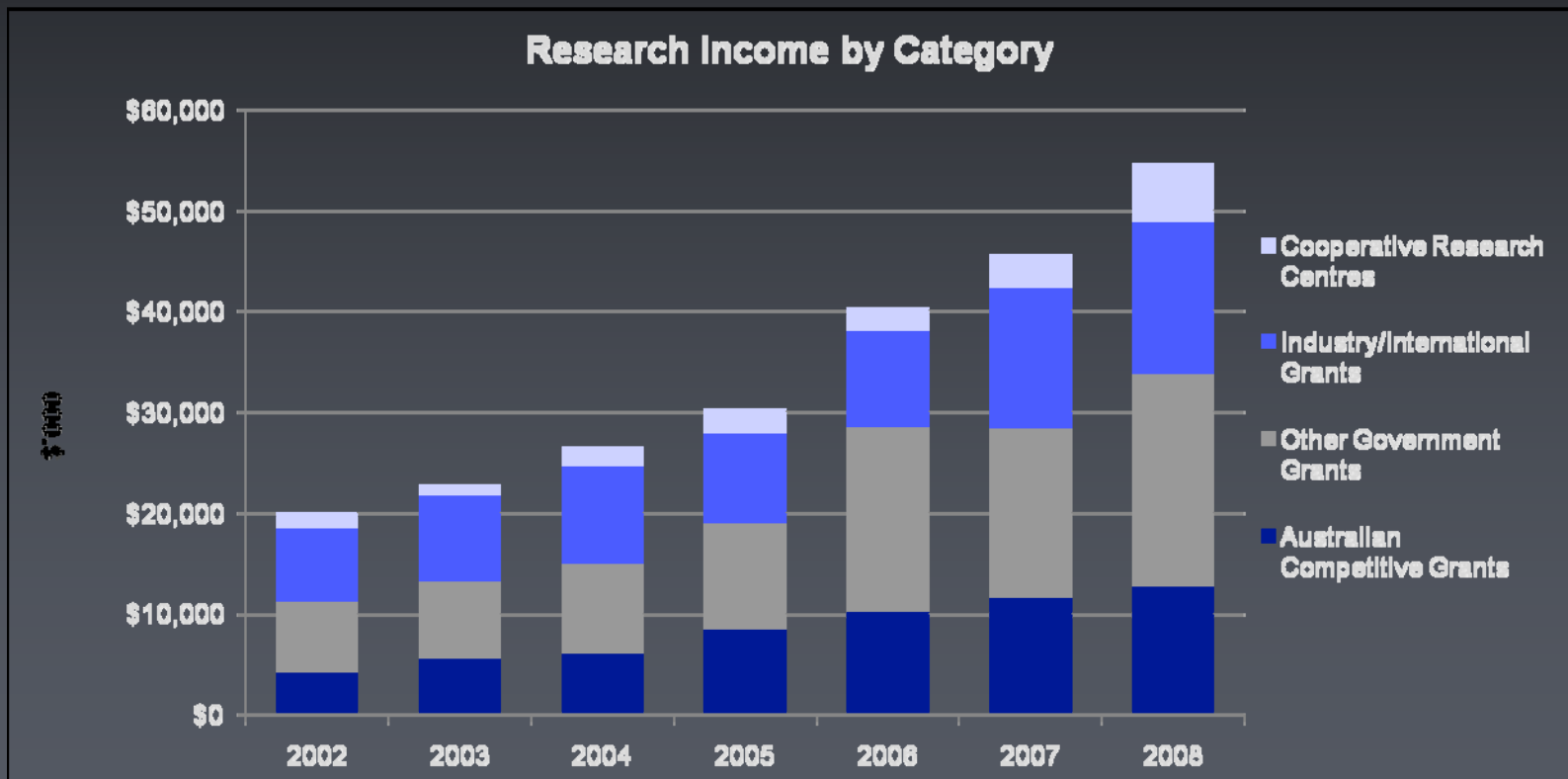
# Institution Profile

- ATN Member 36,000 students



# Institution Profile

- Research income \$54.7 million



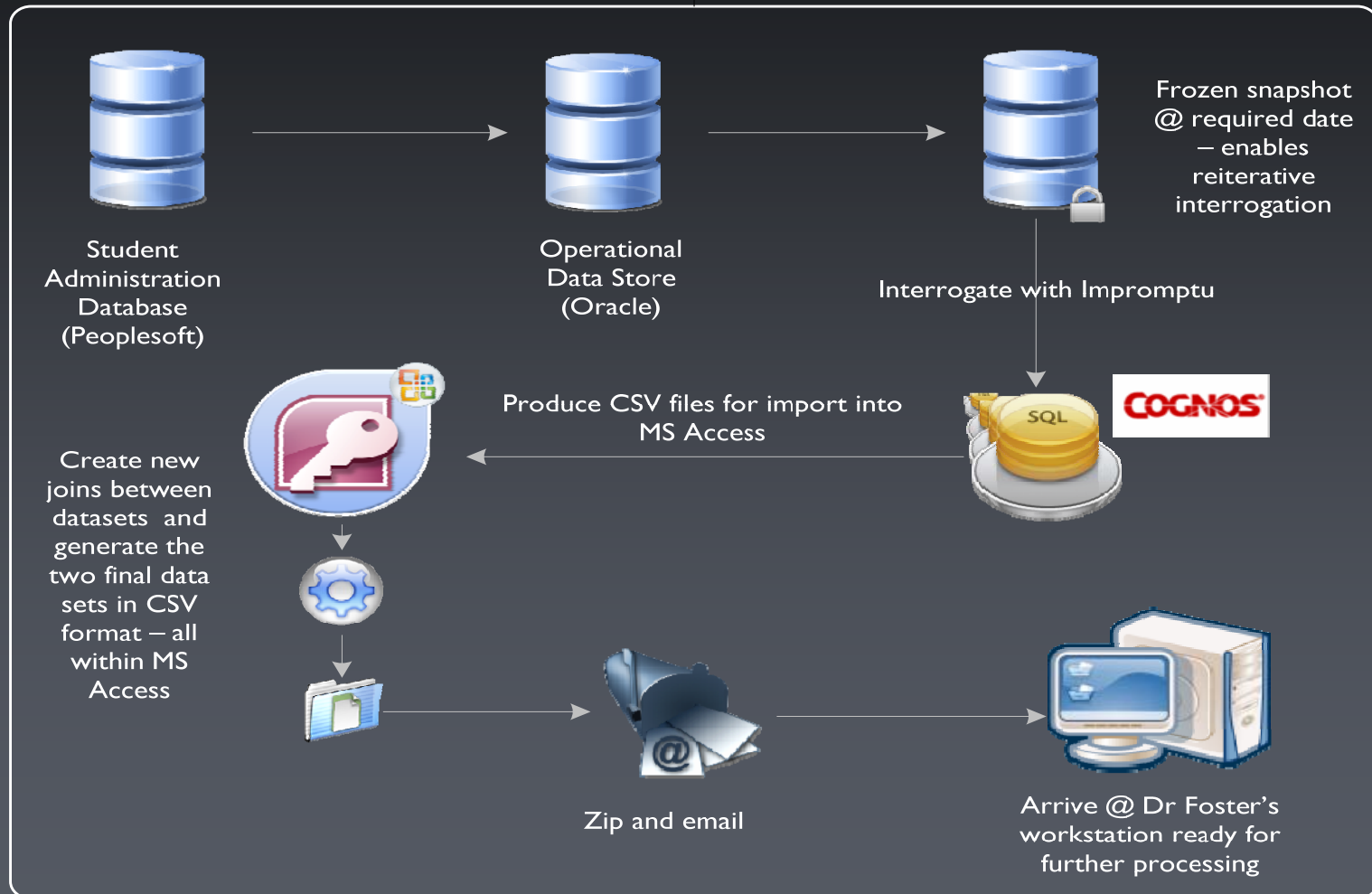
# Data Requirements

- require Ethics Committee approval
- compromising data wish list with 'what is available'
- compare with available UTS datasets
- agree final data scope
- ability to query snapshot multiple times

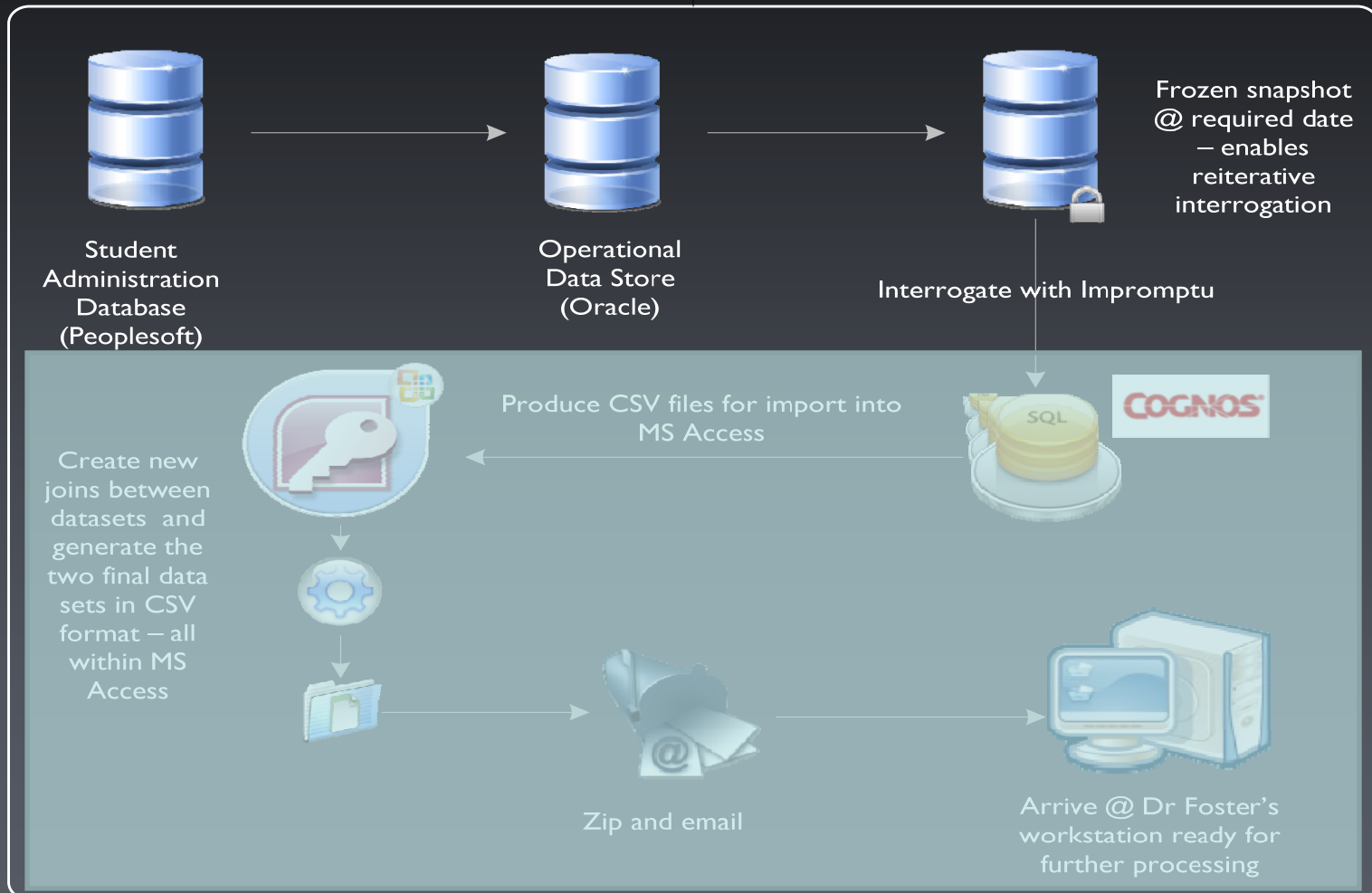
# Limitations

- timeliness
- scope creep
- database structures
- reporting tools and access

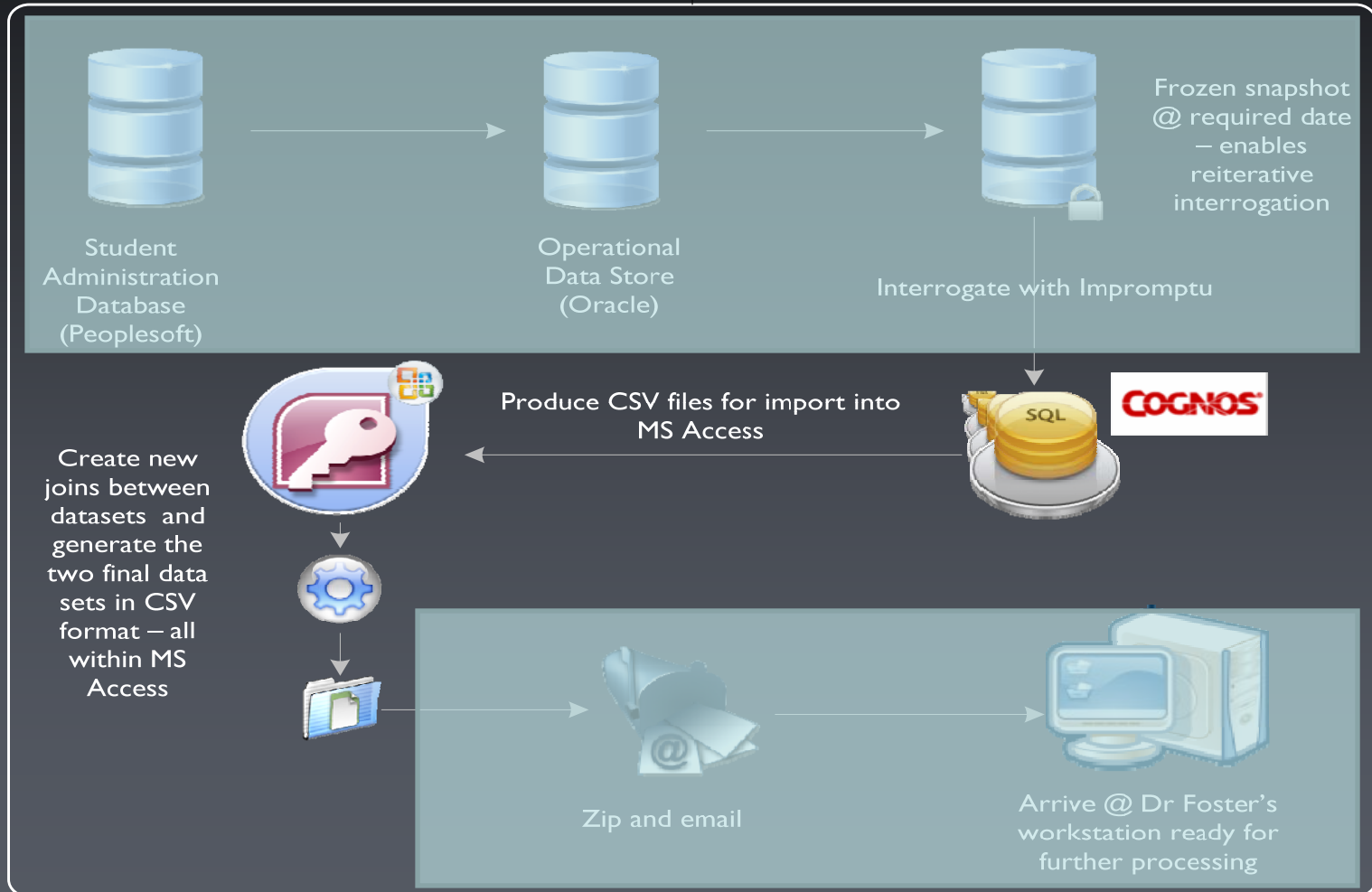
# Methodology



# Methodology



# Methodology



# Methodology

COGNOS

COGNOS

- Report 1 – Set student universe
- Report 2 – SATAC data

COGNOS

- Report 3 – Student demographics
- Report 4 – Student program data

COGNOS

- Report 5 – Student admission basis
- Report 6 – Student course data

COGNOS

- Report 7 – Student class data
- Report 8 – Staff class data



# Methodology

Filter definition:

Student Id + Program Code in [DATASET] and Program Action Code = 'ADMT' ...

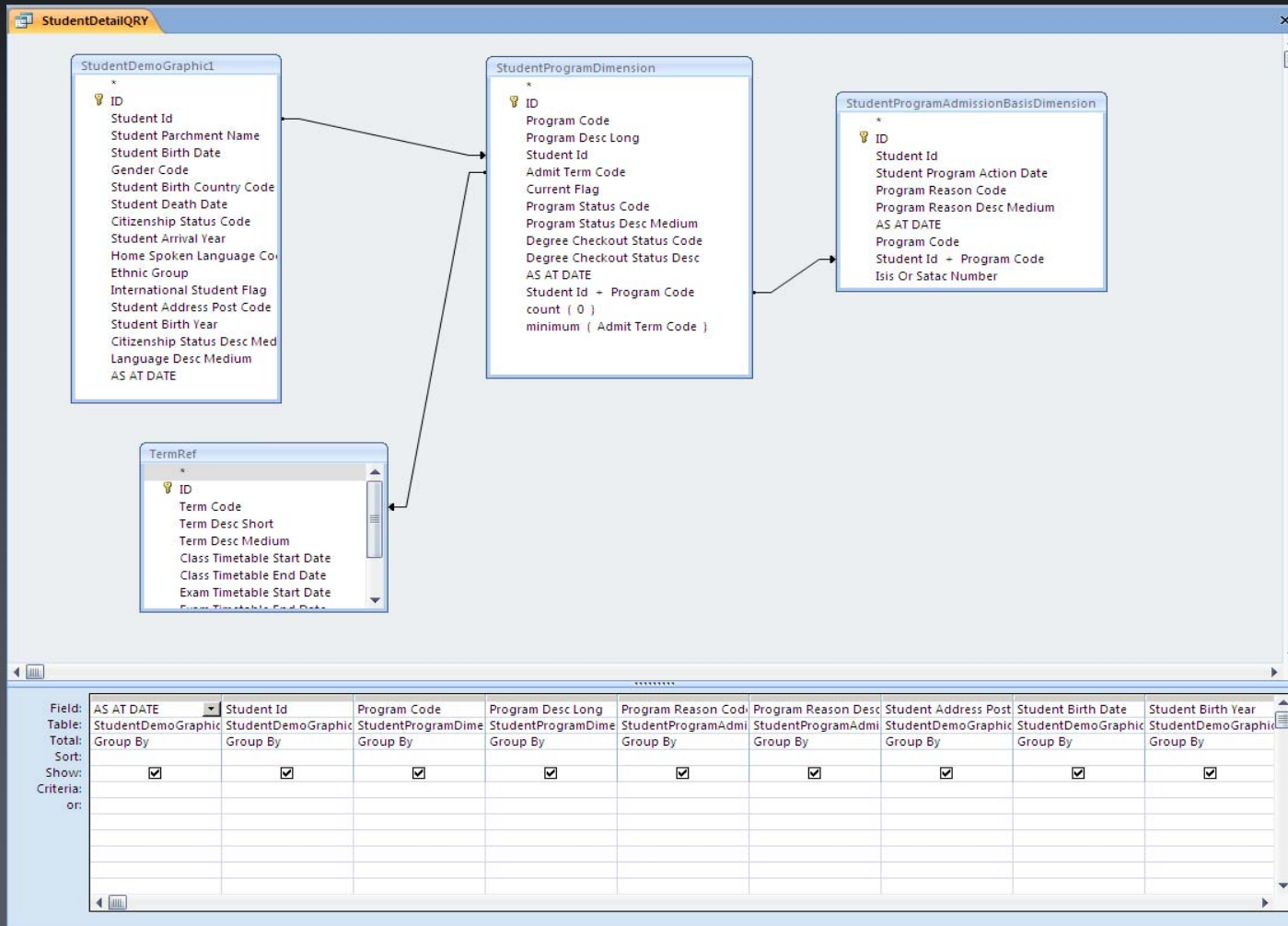
## PROGRAM\_REASON\_DESC\_MEDIUM

Not a Commencing Student  
 Other Basis  
 Higher Education Course  
 Secondary Education  
 TAFE (Non Secondary)  
 Mature Age Special Entry  
 Professional Qualification

SQL :

```
select T1."STUDENT_ID" "c1", T1."STUDENT_PROGRAM_ACTION_DATE" "c2", T1."PROGRAM_REASON_CODE" "c3",
      T1."PROGRAM_REASON_DESC_MEDIUM" "c4", to_date('2009-08-05', 'YYYY-MM-DD') "c5",
      T1."PROGRAM_CODE" "c6", (T1."STUDENT_ID" || T1."PROGRAM_CODE") "c7", T2."ISIS_OR_SATAC_NUMBER" "c8"
from "DWREP"."C_STUDENT_PROGRAM_237" T1, "DWREP"."V_STUDENT_APPLICATION" T2
where T1."STUDENT_ID"=T2."STUDENT_ID" and T1."CAREER_CODE"=T2."CAREER_CODE"
and T1."STUDENT_CAREER_NUMBER"=T2."STUDENT_CAREER_NUMBER"
and (T1."STUDENT_ID" || T1."PROGRAM_CODE") in
      (select distinct (T1."STUDENT_ID" || T1."PROGRAM_CODE") "c1"
      from "DWREP"."C_STUDENT_PROGRAM_237" T1
      where T1."STUDENT_ID" in
            (select distinct T1."STUDENT_ID" "c1"
            from "DWREP"."C_STUDENT_CLASS_235" T1
            where T1."CAREER_CODE"='UGRD' and T1."PROGRAM_ORG_UNIT_1_CODE"='BUE'
            and T1."TERM_CODE" like ('09' || '%')
            and T1."PROGRAM_CAMPUS_CODE" <> 'OFF'
            and T1."PROGRAM_CODE" not in ('BUEUN', 'BUEUX', 'WTAC', 'WTBV'))
            and T1."CURRENT_FLAG"='Y'
            and T1."PROGRAM_STATUS_CODE" in ('AC', 'LA', 'CN', 'CM')
            and T1."PROGRAM_CODE" not in ('BUEUN', 'BUEUX', 'WTAC', 'WTBV'))
            and T1."PROGRAM_ACTION_CODE"='ADMT'
      order by "c1" asc
```

# Methodology



Elements:

- Date of snapshot
- Student ID
- Program code
- Program name
- Basis of admission
- Student post code
- Date of birth
- Year of birth
- Gender
- Country of birth
- Admit term
- Program status (active or on leave)
- International flag
- Citizenship code
- Ethnic group
- Language spoken at home

# Methodology

The screenshot displays a data modeling interface with three dimension tables:

- StudentClassDimension:**
  - ID
  - AS AT DATE
  - Student Id
  - Program Code
  - Course Id
  - Subject Area Code
  - Catalog Number
  - Class Number
  - Class Component Code
  - Class Component Desc Medium
  - Class Instruction Code
  - Class Instruction Desc Medium
  - Class Location Code
  - Class Location Desc Medium
  - Facility Code
  - Facility Desc Medium
  - Term Code
  - Class Schedule Start Date
  - Class Schedule End Date
  - Class Schedule Start Time
  - Class Schedule End Time
  - Class Schedule Mon Flag
  - Class Schedule Tues Flag
  - Class Schedule Wed Flag
  - Class Schedule Thurs Flag
  - Class Schedule Fri Flag
  - Class Schedule Sat Flag
  - Class Schedule Sun Flag
  - Stud\_Id - Prog\_Code - Course\_Id
  - TermCode - ClassNum
- StudentCourseDimension:**
  - ID
  - Student Id
  - Program Code
  - Course Id
  - Course Units Minimum
  - Efts Class Load
  - Term Code
  - Subject Area Code
  - Catalog Number
  - Course Desc Long
  - Course Grade Code
  - Mark Course Grade Code
  - AS AT DATE
  - Student Id - Program Code
  - Student Id - Program Code -
- StaffClassDimension:**
  - ID
  - Course Id
  - Term Code
  - Facility Code
  - Associated Class Code
  - Session Code
  - Class Section
  - Employee Id
  - Instructor Role Code
  - Org Unit 1 Code
  - Class Number

Relationships are shown with arrows: StudentClassDimension and StudentCourseDimension both link to StaffClassDimension. Additionally, StudentCourseDimension has a self-linking arrow.

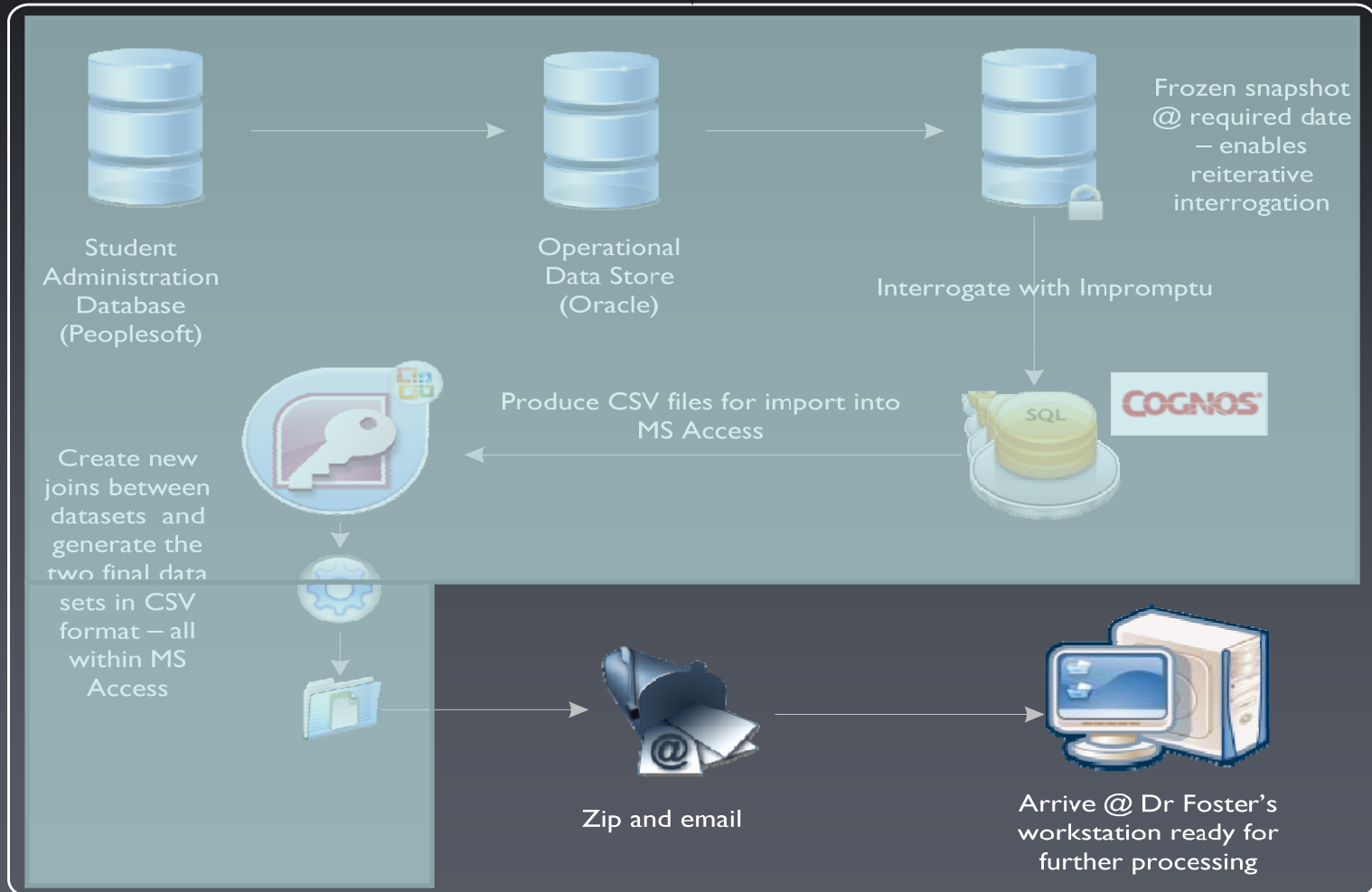
Below the tables is a table grid for query configuration:

Field:	AS AT DATE	Student Id	Program Code	Course Id	Course Units Minimum	Term Code	Subject Area Code	Catalog Number	Course Grade Code	M
Table:	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	StudentCourseDimen	St
Total:	Group By	Group By	Group By	Group By	Group By	Group By	Group By	Group By	Group By	St
Sort:										
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Criteria:										
or:										

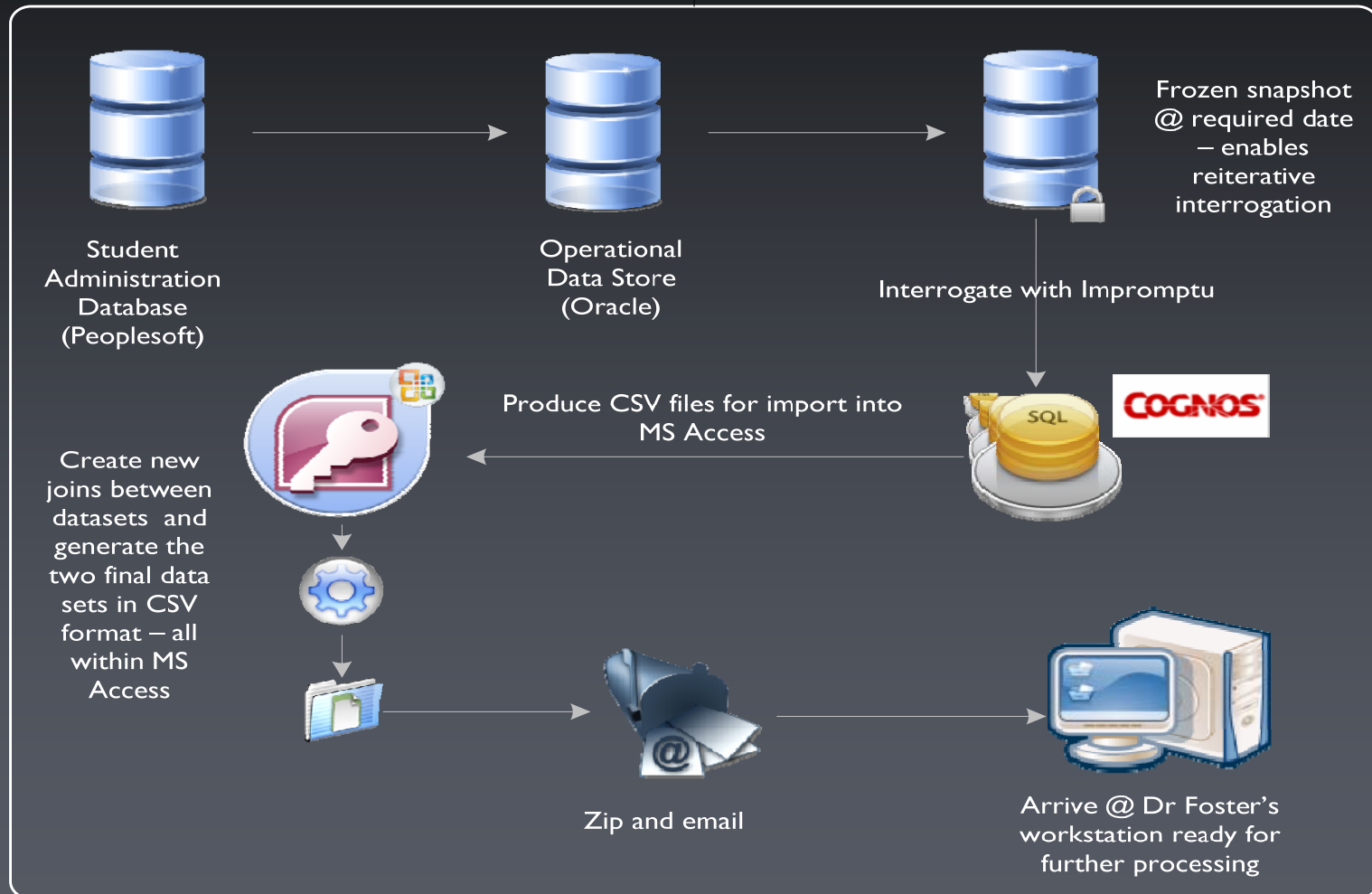
Elements:

- Date of snapshot
- Student ID
- Program Code
- Course ID
- Course Units
- Term code
- Grade
- Mark
- Class ID
- Class Component
- Class Instruction
- Building ID
- Room ID
- Campus ID
- Day of Week
- Class Start Time
- Class End Time
- Class Length
- Staff ID
- Staff Name
- Staff Role
- Class Owner (division and school)

# Methodology



# Methodology



# Problems Encountered

- source system data constraints
- de-normalising the data
- transferring final datasets
- MS Access database size limitations
- need cooperation of IT unit to obtain initial snapshots



[graeme.poole@unisa.edu.au](mailto:graeme.poole@unisa.edu.au)

The background of the slide features a grayscale image of water ripples, creating a sense of movement and depth. The ripples are concentric and spread across the entire frame, with varying shades of gray highlighting the peaks and troughs of the waves.

# Turning Data into Knowledge: A case study of researcher-administrator collaboration using multi-university student data

AAIR FORUM – November, 2009

Presentation by Peter Antony

Solutions Team Leader, University of Technology Sydney

# Agenda

- Institution profile
- Data requirements
- A Data Delivery Framework
- Benefits
- Lessons learnt

# Institution Profile

- New South Wales Institute of Technology
- Australian Technology Network member
- 32,000 students
- Research grants \$27 million

# A Data Delivery Framework

## Data Requirements

- Meet academic and policy goals
- Cross-institutional
- Data quality attributes

# A Data Delivery Framework

## Overview



# A Data Delivery Framework

## Attributes

- Accuracy
- Timeliness
- Consistency
- Confidentiality and security

# A Data Delivery Framework

## Solution Design

- Data-source identification
- Technology landscape
- Testing and Impact analysis

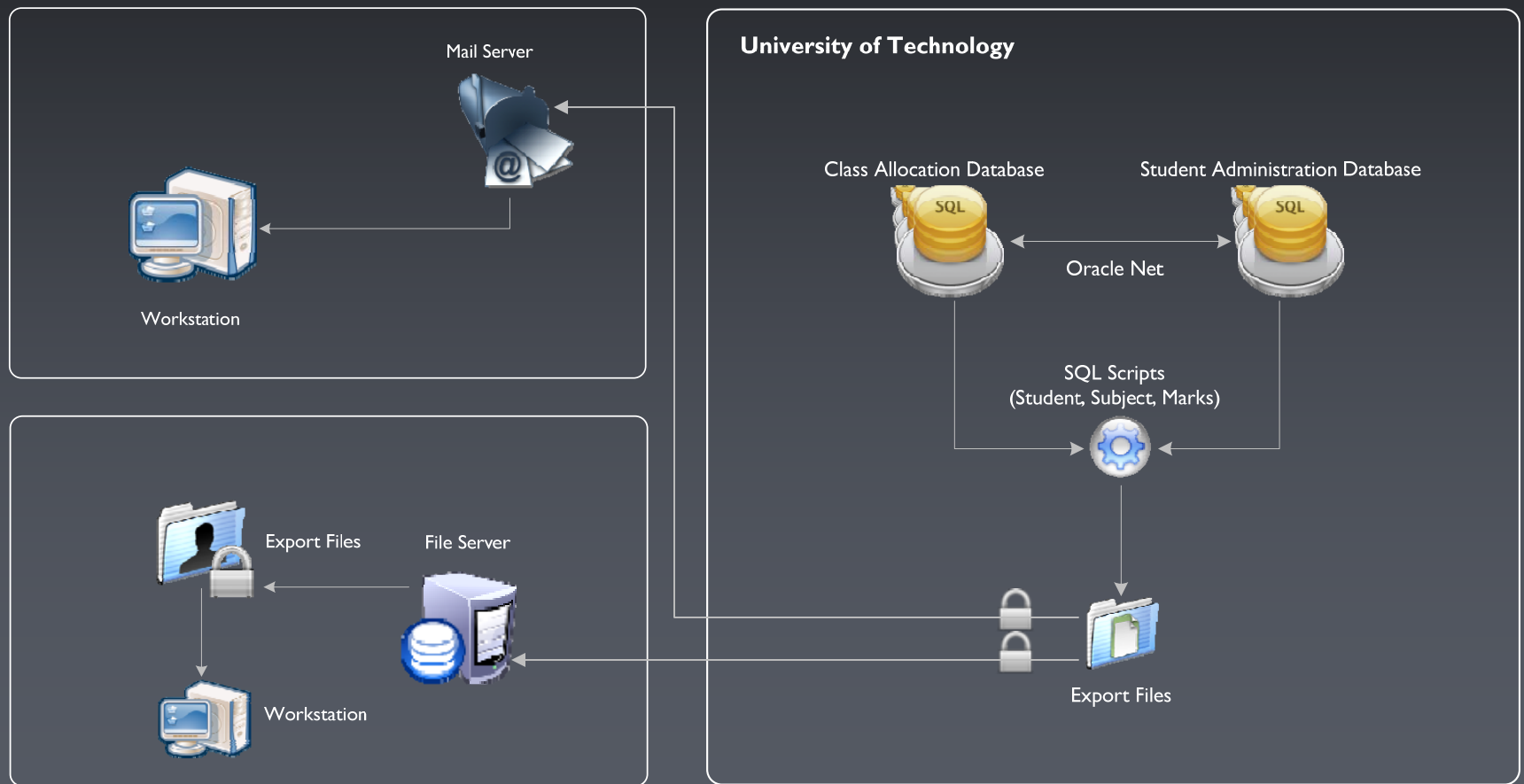
# A Data Delivery Framework

## Implementation

- Student One system
- Allocate+ class allocation system
- Linked via Oracle Net

# A Data Delivery Framework

## Implementation



# A Data Delivery Framework

## Implementation

- Student details
- Subjects
- Marks

# A Data Delivery Framework

## Review

- Identify inefficiencies
- Secure transmission and location

# A Data Delivery Framework

## Challenges

- Data and infrastructure availability
- Data presentation
- Cross-institutional goals

# Benefits

- Consistent
- Flexible
- Sustainable
- Scalable



# Lessons / Conclusions

- Data, data everywhere
- Create consistent process
- Collaboration to meet goals



[peter.antony@uts.edu.au](mailto:peter.antony@uts.edu.au)