





Data Quality - How big is the problem, what are the costs and can it be fixed?

Beth Reichwald
Academic Services
University of Exeter

Sandy Day
Academic Services
University of Exeter



Introductions

- Beth Reichwald
- Sandy Day
- Content
 - What is data quality?
 - Why does it matter?
 - Finding problems
 - Fixing
 - Tips



Why we're here

- Business Intelligence and Business Planning Project
- HESA return
- HEFCE report into "Data Quality and Efficiency into Higher Education"

What is Data Quality?


- Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" (J.M.Juran)
- The term data quality refers to "all features and characteristics of data that bear on its ability to meet the stated or implied needs and expectations of the user". (The Brownfields and Land Revitalization Technology Support Center)
- Data Quality - High quality data is "accurate, timely, meaningful, and complete" (eLiteral.com)




Why does data quality matter?



- [Data Quality Video](#)
- Higher Education context
 - e.g. SSRs

	Maths	Social Studies
Average	16.6	22.5
Min	3.5	2.2
Max	66.4	71.6
1st Quartile	13.2	17.3
3rd Quartile	18.2	26.4






Why does it matter...

- Operational
- Time costs
- Internal impact
- External impact

How do you find the problems and how big are they?

- Warning signs:
 - HESA
 - Data cleaning
 - Local databases





How we found them

Prioritise

- Interviews
- Data profiling
 - Applying logic
 - Business rules
- Data structures
- Data cleaning
- Data review



Fixing Data



Why is it difficult to fix?

- Can you answer?
 - Who is responsible for the data?
 - Do you have the manpower?
 - What's the return on investment?
 - Does your institution have a significant problem?
 - Can you work across business units and departments?



Getting it Right

- Make a plan
- Two major strands:
 - Fixing
 - Preventing





Correcting Errors

- Profiling grid (example available)
 - Identify the issues
 - Give context
 - Prioritise (RAG status)
- Core data master sheet





Prevention

- Multi stream work:
 - Education and Training
 - Ownership
 - Procedures for reviewing
 - Embedding



Education and Training

- Source system training
- What is data used for?
- The impact of errors
- All staff responsibility
- External rules concerning data





Ownership

- Identify data owners
- Data Stewardship


"A steward is a person who is expected to exercise responsible care over an asset that he or she does not own. The data is actually owned by the enterprise. The steward is responsible for caring for that asset"

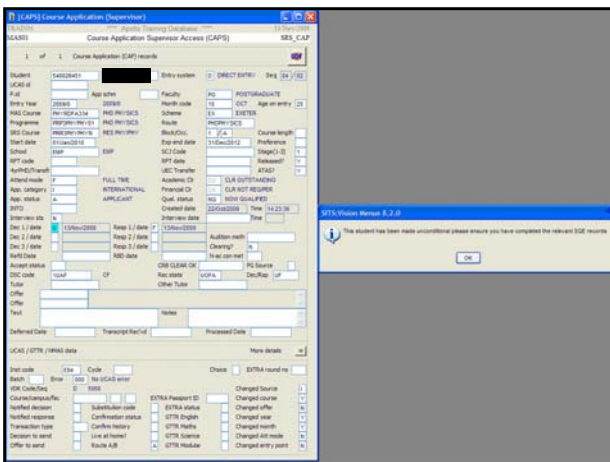
Jonathan G. Gieger, Intelligent Solutions



Reviewing


- Regular review
- Error reporting
- Data quality dash board
- Processes





Embedding

- Cultural change
 - Responsibility of all staff
- Training and education
- Understanding of impact



What have we achieved?




• Originally	• Now
- 80 Schools	- 9 Schools
- 120 Departments	- 20 Departments

- Reduced error reporting
- £30k saving
- Reduction of staff time on HESA
- More to come!



Top Tips

- ⇒ Prioritise
- ⇒ 2 strands – fixing and preventing
- ⇒ Data Stewardship
- ⇒ Staff training
- ⇒ Sanity check data – is it meaningful
- ⇒ Work to add value
- ⇒ When it all goes wrong – KEEP SMILING



Any Questions?

More Information

- Beth Reichwald
B.L.Reichwald@exeter.ac.uk
- Sandy Day
S.F.Day@exeter.ac.uk

