



ARE DATA CURATORS EVER DATA SCIENTISTS?

**LOUISE
CORTI**

ASSOCIATE DIRECTOR
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

IDCC
AMSTERDAM,, 15-16 JANUARY 2013



Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

- challenges of big data, open data and linked data
- skills to manipulate the data and extract insightful patterns...in multiple petabytes of data

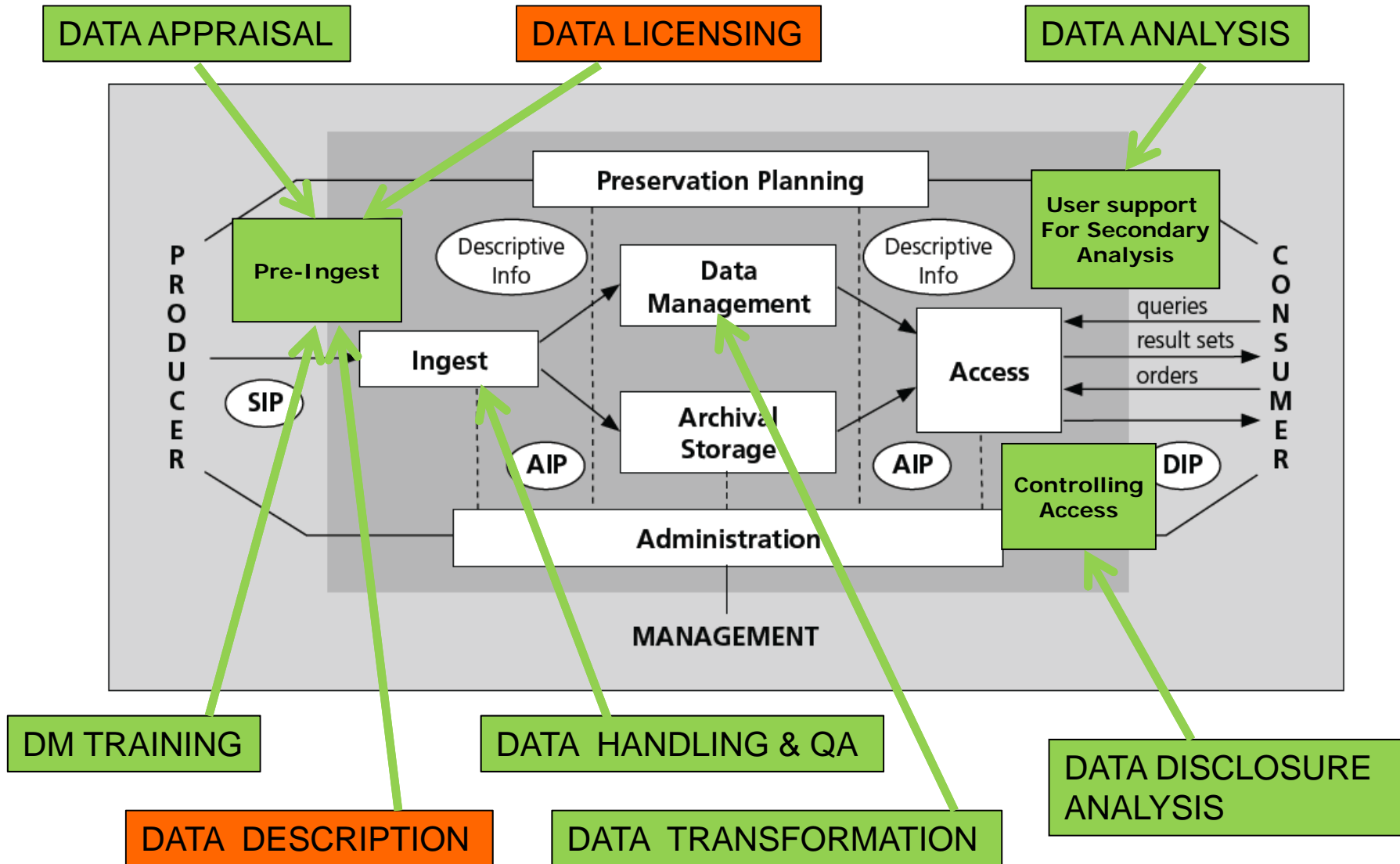
UK Data Archive:

- approximately 6,336 studies
- 1,632 GB
- 1,255,814 files
- 113,832 directories
- Average file size 1.29MBytes
- Grows by 120GB per year



UK DATA ARCHIVE

DATA SKILLS & AND OUR OAIS MODEL



UK DATA ARCHIVE

DATA SKILLS AND DOMAIN EXPERTISE

- Roles discreet
- No one person does them
- Require data skills/ domain expertise
- Most who interact with data have post grad qualifications in social science
- In our organisation, 'data scientist' role only appropriate to those who:
 - transform data - use scripting to help, analyse integrity
 - manipulate, link and merge data sources - harmonised and added value products, e.g. European variables on education, or historic census data

WHO ARE OUR DATA SCIENTISTS?

- Only ever one of them at any one time
- They are definitely data ‘geeks ‘and are proud of that term
- They are highly intelligent social science postgrad /ECRs with experience in analysing socio-demographic data

They are database/ programming-curious and have picked up these skills along the way. Extract-Transform-Load skills

They have picked up curation skills within the organisation, e.g. metadata and preservation requirements

We run an annual ‘How to run a Data Service which lets people see inside our ‘data factory’

'DATA SCIENCE' TRAINING – DOMAIN ISSUE?

The Little List of Free Data Science Courses #datascience #opendata

Posted: December 7th, 2012 | Author: | Filed under: [Data Science Courses](#), [Data Science Workshop](#) | No Comm

As part of our org mission, @ds_idn is dedicated to the free, open, dissemination of #datascience #opendata #opensource

This is a list of free data science courses:

[Introduction to Data Science by Jeff Hammerbacher at University California, Berkeley](#)

[Introduction to Data Science @coursera](#)

[Learning from Data at California Institute of Technology](#)

[Data Science and Analytics at University California, Berkeley School of Information](#)



- Increased data management / preservation training in academia
- We train a wide range – HE staff, students, research support staff
- Why? More/better data, more users
- Other social science data archives/ data libraries starting to provide this

To draw great insights from data, you have to know the data, know the business, and know the contextual relationships that are built into the business (Pierson, Smart Data Collective)

UK DATA ARCHIVE