Importing datasets into Django

Malcolm Tredinnick
July, 2010
Importing data is a puzzle to solve
Motivation...
“The dataset creator hates you!”
General Principles

- Allow restarting
- Assert for all assumptions
- Use brain and eyes before computer
- Use all available tools, not just Python
  - grep
  - cut
  - head / tail
  - ...

Case study
Collection Data

- 75,814 items in a text file
- Creative Commons license
- Appears “mostly sane”
Worked examples...