oiioiioii **Oxford Internet Institute**
oiioiioii **University of Oxford**
oiioiioii

UNIVERSITY OF
OXFORD

ALFRED P. SLOAN FOUNDATION
S
1934

Big Data: What's New?

Ralph Schroeder, Professor
Eric T. Meyer, Research Fellow (@etmeyer)

October 9, 2012

Is the (big data) tail wagging the (research) dog?

Is the (big data) tail wagging the (research) dog?

Big data are data that are *unprecedented in scale and scope in relation to a given phenomenon*. They are often streams of data (rather than fixed datasets), accumulating large volumes, often at high velocity.

Is the tail of the availability of big data and computational methods wagging the dog of good research questions and advancing social science?

If not, how do big data advance research?
What are the opportunities and challenges?

Is the (big data) tail wagging the (research) dog?

Commercial/Governmental versus Social Science Research: Diverging Aims, with Overlap

Manipulation of Behaviour: For aims limited to research in social science.

The 'threat' of social science knowledge, and of commercial/governmental knowledge and control of the natural environment.
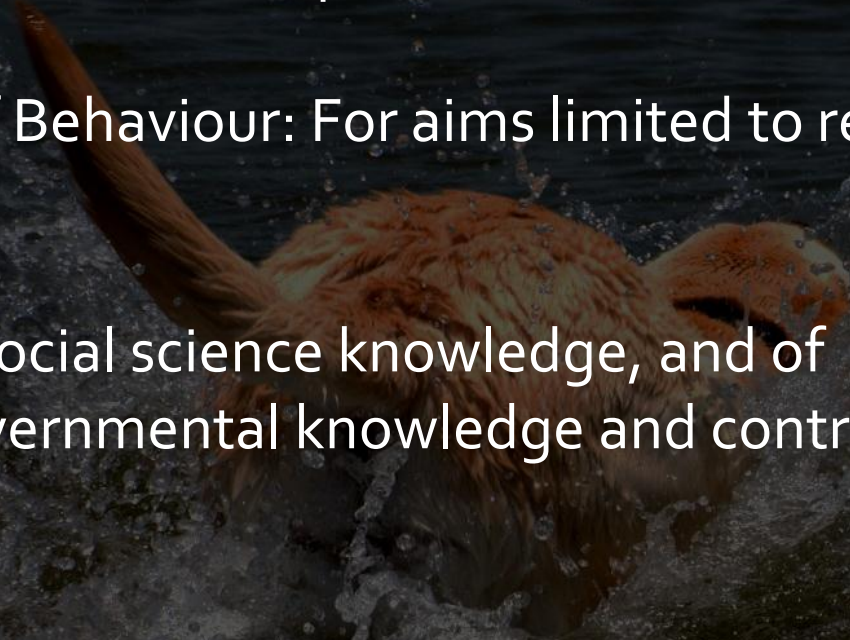
Computational Manipulability?

- 'the distinctiveness of the network of mathematical practitioners is that they focus their attention on the pure, contentless form of human communicative operations: on the gestures of marking items as equivalent and of ordering them in series, and on the higher-order operations which reflexively investigate the combinations of such operations'

- 'mathematical rapid-discovery science…the lineage of techniques for manipulating formal symbols representing classes of communicative operations'

Case 1: Search engine behaviour

Waller's analysis of Australian Google Users

Key findings:
- Mainly leisure
- > 2% contemporary issues
- No perceptible 'class' differences

Novel advance:
- Unprecedented insight into what people search for

Challenge:
- Replicability
- Securing access to commercial data

Case 2: Large-scale text analysis   Google books

Michel et al. 'culturomic' analysis of 5 Million Digitized Google Books and Heuser & Le-Khac of 2779 19th Century British Novels

Key findings:
- Patterns of key terms
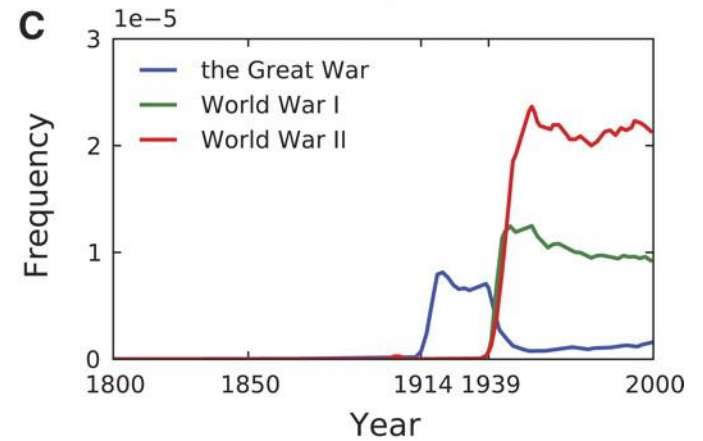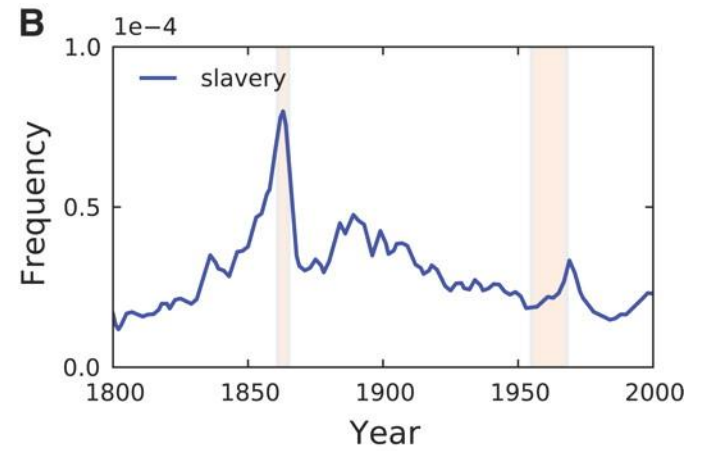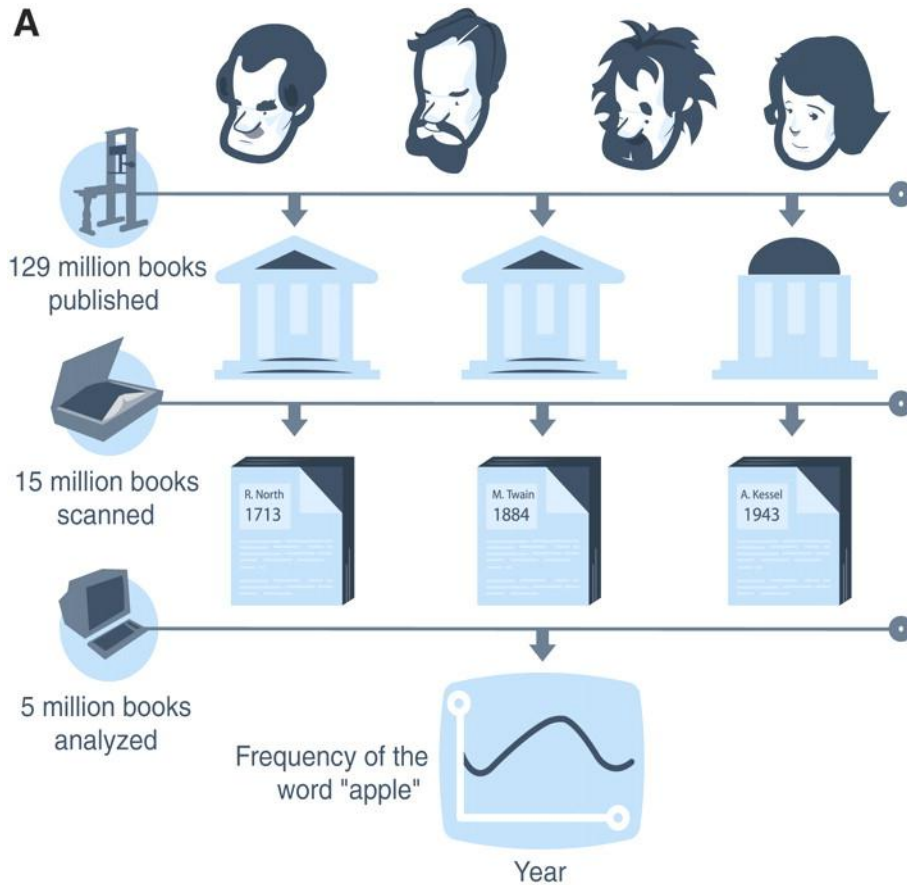- Industrialization tied to shift from abstract to concrete words

Novel advance:
- Replicability, extension to other areas, systematic analysis of cultural materials
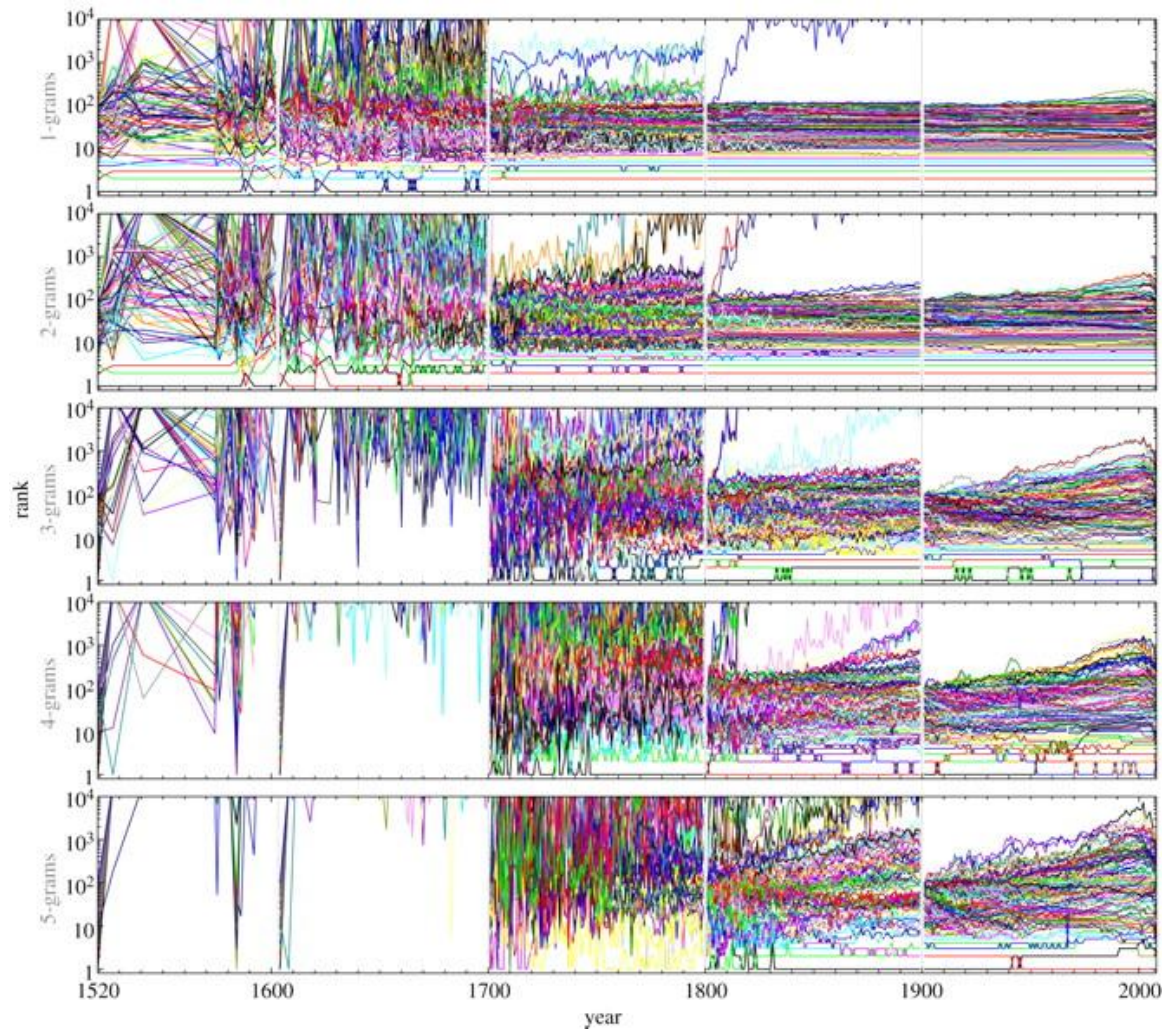
Challenge:
- Data quality

Fig. 1 Culturomic analyses study millions of books at once.

J Michel et al. Science 2011;331:176-182

Evolution of popularity of the top 100 n-grams over the past five centuries.

JOURNAL
OF
THE ROYAL
SOCIETY
Interface

Case 3: Social network or news? twitter

Kwak et al.'s analysis of Twitter

Key findings:
- 1.47 billion social relations
- 2/3 of users are not followers or not followed by any of their followings
- Celebrities, politicians and news are among top 20 being followed

Novel advance:
- Volume of relations and topics

Challenge:
- News or social network needs to be contextualized in media ecology
- Securing access to commercial data

Conclusions

Savage and Burrows?, who ask are commercial data outpacing social science?
Boyd and Crawford?, who ask if big data raise epistemological conundrums?

… No …

The connection between research technologies and the advance of knowledge

The threats and opportunities represented by unprecended windows into people's minds and thoughts

Does this lead to more 'scientific' (i.e. cumulative) social sciences and humanities?

Big Data

Accessing and Using Big Data to Advance Social Science Knowledge

See http://www.oii.ox.ac.uk/research/projects/?id=98
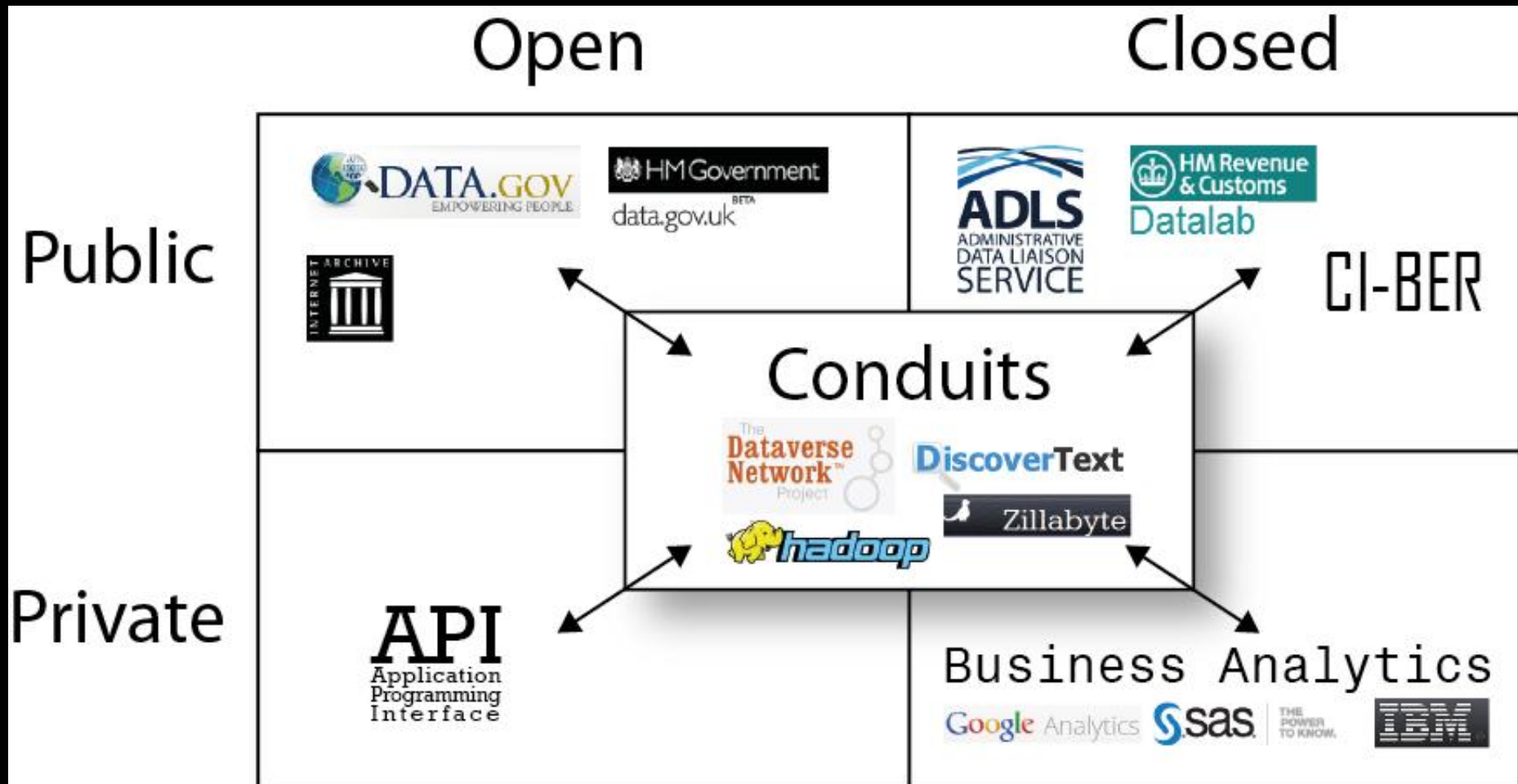
Oxford Internet Institute

Ralph Schroeder
ralph.schroeder@oii.ox.ac.uk
http://www.oii.ox.ac.uk/people/?id=120

Eric T. Meyer
eric.meyer@oii.ox.ac.uk
http://www.oii.ox.ac.uk/people/?id=120
@etmeyer