What’s so different about Arts and Humanities data?

David De Roure
Deluge!!!

Data!!

Scientists
Social Scientists
Funding agencies
Policy makers
Publishers
Humanists
Librarians
Internet architects
What is e-Research?

• Research in every domain is increasingly data- and computationally-intensive, carried out collaboratively over distributed infrastructures

• e-Research is the continuous technological and methodological innovation in digital methods to achieve new research outcomes – using new forms of data and emerging infrastructural capabilities

• The Oxford e-Research Center is a digital methods incubator hosting 40 postdoctoral researchers conducting early-adopter digital research across all disciplines
More people

More machines

Big Data and Computation

Conventional Computation

Social Machines

Social Networking

Cyberinfrastructure
e-infrastructure

Science 2.0
Citizen Science

e-Research

David De Roure
Did you know?

The aperture arrays in the SKA could produce more than 100 times the global internet traffic.

Did you know?

The SKA super computer will perform $10^{18}$ operations per second – equivalent to the number of stars in three million Milky Way galaxies – in order to process all the data that the SKA will produce.

Did you know?

The data collected by the SKA in a single day would take nearly two million years to playback on an ipod.
Economic and Social Research Council
Shaping Society

- Digital Social Research Program
- Social Machines
- Web Observatories
- Responsible Innovation
- Centre for International Social Media Analytics

Digital Social Research
Harnessing technology for social science research
Here is the evidence, now what is the hypothesis?
The complementary roles inductive and hypothesis-driven science in the post-genomic era

Douglas B. Kell

Summary
It is considered in some quarters that methods are the only way to achieve large advances in scientific knowledge. Marginal, irrelevant, in the development of technology with "hypothesis-led" (beyond what might be of value) - must be the hypothetico-deductive paradigm here that data- and technology-driven discovery of knowledge discovery is the alternative to hypothesis-driven discovery. Here, we analyse, which may be generating novel hypotheses, novel hypotheses.


A Revolution That Will Transform How We Live, Work and Think

By Chris Anderson 06.23.08

The End of Theory: The New Scientific Method Obsol

Illustration: Marian Bantjes

BIG DATA

Viktor Mayer-Schönberger and Kenneth Cukier

First Paradigm

Data-Intensive Scientific Discovery

EDITED BY TONY HENRY, STEWART TABLES, AND ADAM TUCKE
Citizen Scientists

Image Classification

Talk Forum

data reduction

Scientists
The challenge is to foster the co-constituted socio-technical system on the right i.e. a computationally-enabled sense-making network of expertise, data, models and narratives.
Web as lens

Web as infrastructure

Web as artefact

The Observatory Quarter

http://www.w3.org/community/webobservatory/
Live Music Archive Linked Data

This server provides access to Linked Data that describes the audio held in the Internet Archive’s Live Music Archive (also sometimes known as "etree"). The metadata from etree has been converted to RDF and is exposed through a SPARQL endpoint along with browsable pages.

The dataset contains information describing over 100,000 performances by 4,000 artists including 1,600,000 individual tracks, each of which may be available in a number of formats.
Digital Music Collections

Student-sourced ground truth

Community Software

Supercomputer

Linked Data Repositories

Music Information Retrieval Community

23,000 hours of recorded music
1. From signal to understanding
2. Working with multiple sources of incomplete and inconsistent data, with new born-digital data
3. Innovation and sharing of new digital methods
4. Challenges of resource discovery and publication of new digital artefacts
5. Importance of provenance
6. Challenges of curation
7. Increasing automation (and risks therein)

Commonalities

1. Specific content types and their relationship to physical artefacts
2. Curated collections, and an “infinite archive” of heterogeneous content, richly interlinked
3. Specialist digital methods
4. Publications are subjects and records of research
5. Emphasis on multiple interpretations and critical thinking
Latest opinions and ideas
Read all about it on our blog: Software and Research

The Software Sustainability Institute
Some Social Machines
1. Digital > Digitised
2. Machines are users too
3. Digital brings new decay but also opportunity for automation... and community engagement
4. Software is part of the problem and part of the solution
5. Disciplinary boundaries are a legacy and transcended by today’s research questions
6. Towards Social Machines for Digital Curation?
david.deroure@oerc.ox.ac.uk
www.oerc.ox.ac.uk/people/dder
www.scilogs.com/eresearch
@dder

Thanks to: Christine Borgman, Ichiro Fujinaga, Stephen Downie, Chris Lintott, Iain Buchan, Nigel Shadbolt
PAUL ET
AVIRGINIE

CÁLÓP

CHARLES D’ALBERT.

LONDON:
CHAPPELL & CO., 58 NEW BOND ST.