A (Data) Tale Of Two Faculties

Conference or Workshop Item

How to cite:

King, David and Benatti, Francesca (2018). A (Data) Tale Of Two Faculties. In: Data Conversations, 26 Nov 2018, The Open University, UK.

For guidance on citations see FAQs.

© 2018 The Open University

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
A (DATA) TALE OF TWO FACULTIES

David King (STEM) and Francesca Benatti (FASS)
Francesca

Currently Research Fellow in Digital Humanities

Studied Book History and Irish Studies for my PhD

Created my first digital research data for it (an Access database of 1,000 periodical articles)

But far more analogue data
  - handwritten notes
  - annotated photocopies of microfilms
  - marginal notes in books

How can I preserve, organise and use all of this data?

Can I even call it data?
INTRODUCTIONS

David

Currently Visiting Fellow in Computing & Communications
Member of OU’s Digital Humanities Steering Group
Studied International History and Politics for my first degree
…which meant travelling to libraries to read their microfilm
Worked in industry, through computing to management consultancy
…and specialised in document/knowledge management and workflow,
hence extensive experience sharing data, and responsibilities
Mature student at the OU for my PhD in Computing, and stayed…
…thesis in ORO, downloaded more than 5,700 times
Attelabus nitens
Working in biodiversity, shared data

Biodiversity Information Standards (TDWG)

We are a non-profit organization and community dedicated to developing biodiversity information standards.

Historically known as the Taxonomic Databases Working Group, today's Biodiversity Information Standards (TDWG) is a not-for-profit, scientific and educational association formed to establish international collaboration among the creators, managers and users of biodiversity information and to promote the wider and more effective dissemination and sharing of knowledge about the world's heritage of biological organisms.

To achieve its goals, TDWG:

- develops, adopts and promotes standards and guidelines for the recording and exchange of data about organisms
- promotes the use of these standards and guidelines through the most appropriate and effective means
- acts as a community forum for discussion through meetings, online discussions, and publications.

See the About section to learn how TDWG operates and how you can participate. Active Interest Groups are described under Community.
DATA IN SCIENCE

At large - metadata as well as domain specific interest groups
Main resource in biodiversity literature
Some hiccups because of funding
Virtual Biodiversity Research and Access Network for Taxonomy (ViBRANT) was an European Union FP7 funded project starting in December 2010 and completed in November 2013.

ViBRANT supported the development of virtual research communities involved in biodiversity science. Our goal was to provide a more integrated and effective framework for those managing biodiversity data on the Web.

ViBRANT paved the way towards more interoperable systems and integrated services in biodiversity research. We are already looking into preparing a successor project to support data, processes and communities towards the long term vision of predictive modeling of the biosphere.
EU FP7 Infrastructure projects – Awkward because of commercial aspects
Many data stores – For slides
Many data stores – For code

The following command on a *nix like system (Linux, OSX, etc) will create a clone of our Git repository, which will enable you to browse our code and develop against it.

```
git clone https://github.com/NaturalHistoryMuseum/scratchpads2.git
```
Many data stores – Github
Many data stores – Contribute, but deleted own github
DATA IN SCIENCE
Data rot – NHM gone, as Drupal migrated
DATA IN SCIENCE

Data rot – OU inaccessible

... Sorry - you are not authorised to perform this operation

Please use 'Back' button and try again.
DATA IN SCIENCE

Data rot – OU gone totally

RefBank

RefBank nodes of the world:

- Plazi - Boston, USA
- ZooBank - Hawaii, USA
- Pensoft - Sofia, Bulgaria (1 100 500)

References stored in RefBank:

1 100 500
That’s how things were…

Delighted (genuinely!) to state that things are different now

ORO, ORDO and Office 365

Less Google Docs and other ad hoc data sharing needed now too

Do these experiences strike a chord?
Four main types of data:

- **observational** (e.g. weather measurements);
- **computational** (e.g. computer modelling or simulation);
- **experimental** (e.g. laboratory studies);
- **records** (e.g. newspapers, photographs, artworks, letters, diaries, books, articles; records of birth, death, marriage; maps)

Scientists generally **create** their own data

Humanities scholars generally **use data created by others**

Plus: “Any record of human experience can be a data source to a humanities scholar”

Minus: records are often constrained by intellectual property and other restrictions
Let’s revisit:

    Humanities scholars generally **use data created by others**

    I would argue it depends how you define ‘data’

Francesca can expand on a humanists view on data in a later *Data Conversation*
THANK YOU!

Download our corpus from ORDO (search for RSVP)
doi: 10.21954/ou.rd.6850865

Follow us on
http://www.open.ac.uk/blogs/styleproject/
@rhymesontheroad
@dh_ou