



* The Heterogenous Zone: Six use cases for six research data collections in Edinburgh DataShare

Robin Rice

EDINA and Data Library, University of Edinburgh

DSpace User Group, Open Repositories 2013

Charlottetown, PEI, Canada: 11 July, 2013

- * Edinburgh DataShare is a free-at-point-of-use data repository service which allows University researchers to upload, share, and license their data resources for online discovery and re-use by others.
- * Built in DSpace during Jisc-funded DISC-UK DataShare project (Edinburgh, Oxford, Southampton) 2007-2009 as an exemplar of institutional data repositories.

* What is Edinburgh DataShare?



Search Edinburgh DataShare

[Advanced Search](#)

Browse

Edinburgh DataShare
[Research Communities](#)
[Date Issued](#)
[Data Creators](#)
[Titles](#)
[Subjects](#)

My Account

[Login](#)
[Register](#)

RSS Feeds

 [RSS 1.0](#)
 [RSS 2.0](#)
 [Atom](#)

What is Edinburgh DataShare?

Edinburgh DataShare is an online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh, hosted by the [Data Library](#) in Information Services.

Edinburgh University researchers who have produced research data associated with an existing or forthcoming publication, or which has potential use for other researchers, are invited to upload their dataset for sharing and safekeeping. A persistent identifier and suggested citation will be provided.

Deposit Your Data

 [Deposit](#)

- [How to deposit](#)

Research Communities in Edinburgh DataShare

Select a community to browse its collections.

- [Business School](#)
- [Edinburgh College of Art](#)
- [Information Services \(IS\)](#)
- [Moray House School of Education](#)
- [Royal \(Dick\) School of Veterinary Studies](#)
- [School of Biological Sciences](#)
- [School of Biomedical Sciences](#)
- [School of Chemistry](#)
- [School of Clinical Sciences](#)
- [School of Divinity](#)
- [School of Economics](#)
- [School of Engineering](#)

Spotlight



Chromatin Fiber Condensing by drs18 on flickr

New to Edinburgh DataShare - A genome-wide screen in human embryonic stem cells reveals novel sites of allele-specific histone modification associated with known disease loci : School of Molecular, Genetic and Population Health Sciences.

Information for Depositors

- [About Edinburgh DataShare](#)
- [Checklist for deposit](#)
- [Benefits of deposit](#)
- [Service background](#)
- [Our definitions](#)
- [Service policies](#)

Latest Items

- [Wally Referring Expression Corpus \(WREC\) v0.1.0 \(08 Jul 2013\)](#)



Open Data and Institutional Repositories

3 years ago, 23 slides

Edit ▾

Actions ▾



1025

0

0

10



Edinburgh DataShare: Tackling research data in a ...

3 years ago, 23 slides

Edit ▾

Actions ▾



2112

0

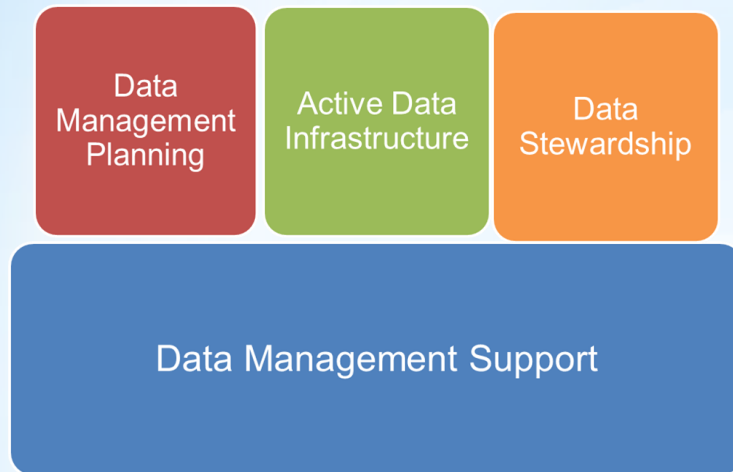
0

25

* On DSpace, data and IRs
over time

“9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.”

***The data repository and
University RDM policy**



Edinburgh DataShare is seen by the RDM Steering Group as one of the key RDM services offered by Information Services, and as such has challenged its staff to meet the requirements of a number of pilot submissions from a range of different types of research communities with special kinds of data.

*DataShare and the University RDM Roadmap



*Dr. Nunno Feirrera,
Teaching Fellow*



Single item deposit,
dataset behind an
article.

Desire to get students
to deposit their data
from theses as norm -
need unambiguous
deposit workflow.

Fieldwork in NHS
means much data is
'sensitive'. Permanent
embargoes?

*Clinical Psychology



*Dr. Bert Remijssen
Chancellor's Fellow*

Dinka Songs of South Sudan collection, 62 items.

Used collection template for metadata; files uploaded by assisted deposit.

User “amazed”. Happy with usage statistics also. Referred a colleague and will distribute leaflets for us.



Village of Fafanlap, Indonesia

*Linguistics and
English Language

* “The Listening Talker” collection identified for deposit, ongoing.

* Very large video files plus software as VM image. Tar files containing millions of files. > 2.1 GB!

* Desires user registration, non-standard licenses and checksums with downloads.



Prof. Simon King



**Centre for Speech
Technology Research**

- * Lots of 'omics data:
local repository vs
subject repositories -
storage cost concerns
- * Interested in push-pull
of metadata to
websites, from CRIS
- * Spearheaded by Data
Manager

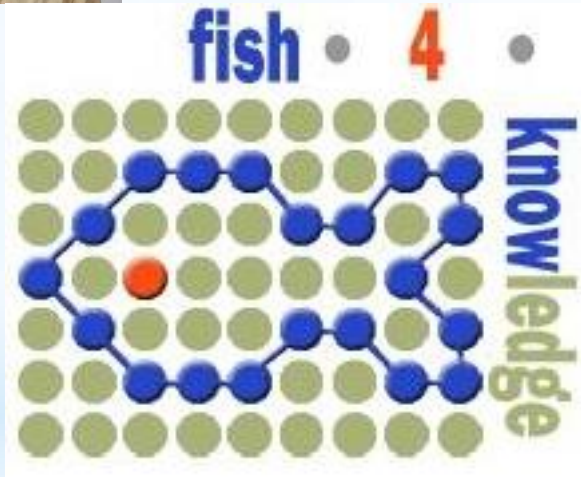


Dolly the Sheep

* Roslin Institute

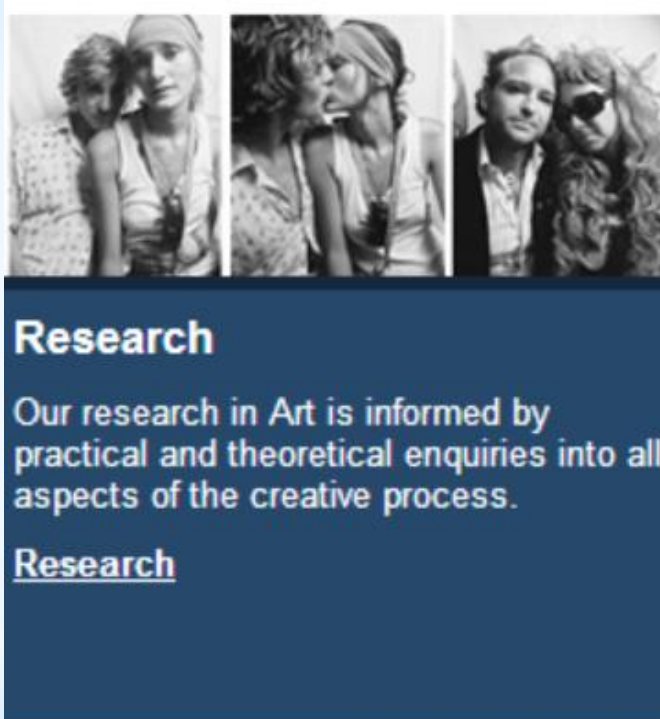


Prof. Bob Fisher



- * Fish4Knowledge EU-funded research project
- * Long-term sustainability issues for observational data
- * Search engine maintained on their website - using METS feed to locate items
- * Testing SWORD implementation, 5% sample >10K files, video + sql rows (3 TB)
- * Efficiency & performance

* **Institute of Perception, Action
& Behaviour (Informatics)**



- * New member of Uni
- * Digital asset mgmt needs
- * Nature of research data in the arts
- * Streaming & display requirements (high quality desired)

* **Edinburgh College of
Art**

- * Better hints, documentation (depositor's checklist, user guide with screenshots)
- * Make use of faceted browse for flexible display (e.g. by collection, not community)
- * Remove initial questions
- * 'Collapse' non-required fields
- * Fine-tune embargos (suppress filenames, allow group)
- * Explore implementing mobile view, usage stat capability, etc. of DSpace 3
- * Wishlist: Figshare-like alt-metrics capability, ratings, etc. for end-users
- * Wishlist: Drag and drop file upload

* Summary: usability issues

- * SWORD implementation: batch ingest solution for technical depositors, what about non-technical?
- * Disable automatic zip on download, & index for large collections (?)
- * Make administrative metadata visible to end-users (e.g. checksums). Raises issues about AIP.
- * Investigate streaming, CSV display, other rich media display options. Raises DIP issues.
- * Determine ingest flow with CRIS (existing system) and Data Asset Registry and active data store.

***Summary: technical issues**