

Service Integration to Enhance RDM: RSpace electronic laboratory notebook (ELN) case study

Stuart Macdonald (RDM Services Coordinator,
University of Edinburgh)

stuart.macdonald@ed.ac.uk

Rory Macneil (CEO, Research Space)

rmacneil@researchspace.com



University of Edinburgh RDM Policy

- University of Edinburgh is one of the first Universities in UK to adopt a policy for managing research data:
<http://www.ed.ac.uk/is/research-data-policy>
- The policy was approved by the University Court on 16 May 2011.
- It's acknowledged that this is an aspirational policy and that implementation will take some years.

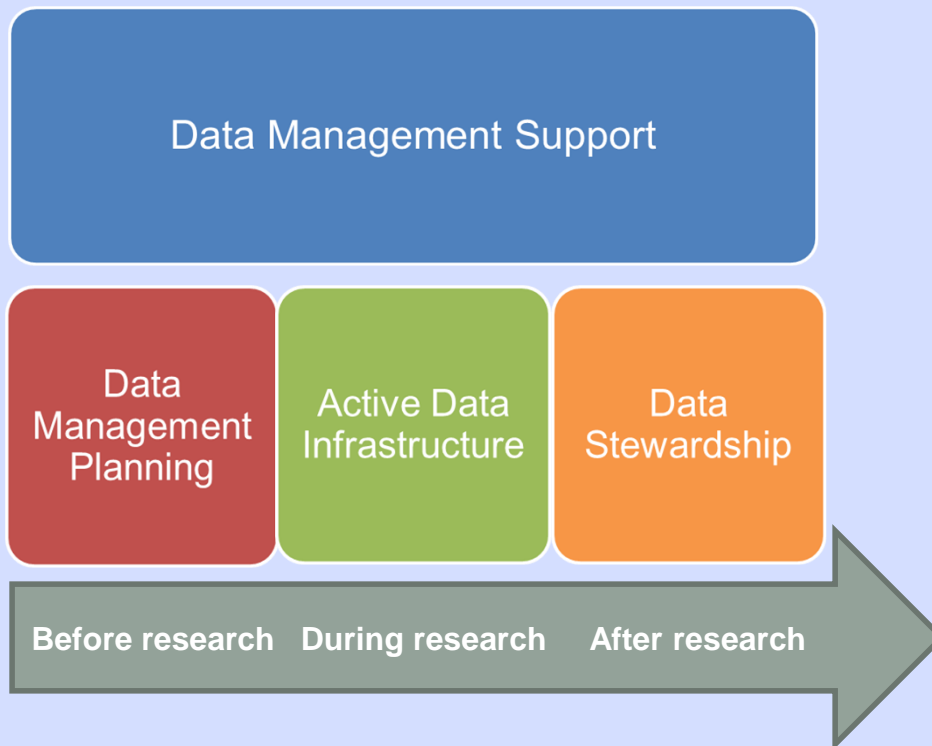
Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.
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.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

Policy implementation: RDM Roadmap



<http://edin.ac/1u3sKqy>

Colleagues in IS developed a roadmap to deliver a suite of services that will meet:

- RDM policy objectives
- the needs of our researchers

Cross-divisional collaboration

- 3 Phases (Aug. 2012 – May 2015)

Services already in place:

- Data management planning
- Active working file space = DataStore
- Data publication repository = DataShare

Services in development:

- Long term data archive = DataVault
- Data Asset Register (DAR)

RDM support: Awareness raising, training & consultancy

Research Data Management Planning

Customised instance of DCC's DMPonline toolkit for Univ. of Edinburgh use:

- A range of funders and local (non-funder) DMP templates
- Institutional guidance (storage, services, support)

Tailored DMP assistance for researchers submitting research proposals (F-2-F)

DataStore

- NAS facility to store data that are actively used in current research activities
- 0.5 TB (500GB) allocated per researcher (incl. PGRs)
- Up to 0.25TB can be used for “shared” group storage
- Extra storage costs: £200 per TB per year incl. back-up and DR copies
- Infrastructure in place. Allocation of space devolved to School IT departments overseen by Heads of IT from each College.

DataShare

- Edinburgh DataShare is the University's OA multi-disciplinary data repository hosted by the Data Library : <http://datashare.is.ed.ac.uk>
- Assists researchers who want to share their data, get credit for data publication, and preserve their data for the long-term (DOI, licence, citation)

Data Vault

- Safe, private and secure long-term data archive
- Current focus on front-end application requirements (authorisation, retention & deletion, file structure, file transfer, integration)

Data Asset Register (DAR)

- Using PURE as the catalogue of data assets produced by Edinburgh researchers for discovery, access, and re-use as appropriate.

Interoperation

- Systems are more likely to be used if some or all of the components are integrated and developed to minimise 'duplication' of effort

RDM Support

- RDM team work with Research Administrators , Academic Support Librarians and IT staff in each of the 22 Schools.
- Queries can be sent to the IS Helpline who will direct them to appropriate RDM staff via CMS.

Training & Awareness Raising

- Introductory sessions on local RDM services and support for researchers and research administration staff in Schools / Institutes
- RDM website: <http://www.ed.ac.uk/is/data-management>
- RDM blog: <http://datablog.is.ed.ac.uk>
- RDM wiki:
<https://www.wiki.ed.ac.uk/display/RDM/Research+Data+Management+Wiki>

Training: Tailored Courses

Formal and bespoke training in the form of workshops, seminars and drop in sessions to help researchers with RDM issues.

- Creating a data management plan for your grant application
- Handling data using SPSS
- Managing your research data: why it is important and what should you do? **NEW**
- Publishing and sharing sensitive data (pilot) **NEW**

MANTRA - <http://datalib.edina.ac.uk/mantra>

- An internationally recognized free online RDM training course for researchers - developed by the Data Library
- Software-specific data handling exercises
- CC License & embed units in VLE's e.g. Moodle

Service Integration examples

- DataShare is a customised DSpace instance with OAI-PMH compliant DCMI metadata fields for data discovery through Google and other search engines
 - Records are harvested by Thomson-Reuters Data Citation Index
- SWORD API utilised for batch deposit of large and/or many files from remote computers ('Push using http')
- Internal batch ingest of many/large files to circumvent 2.1GB limit via the web interface ('Pull via command line interface')
 - checksums determine that delivered object mirrors deposited object
- DSpace GITHUB plugin* - allows software used in research to be archived from GitHub (or similar) source code repository into DataShare, which can then be assigned a DOI to facilitate citation - using the SWORD deposit protocol

DataSync – a secure dropbox-like facility for synchronising data on DataStore with desktop and mobile machines:

- uses open source ‘ownCloud’ technology

Refresh of **ECDF Computing Cluster** (‘Eddie’) complete with ‘Data Centric Computing’ business model – integrate Eddie storage & HPC, parallel and cloud computing layers with DataStore for data sharing i.e. data transferred from DataStore for analysis run on Eddie and then data ported back to DataStore (DataVault)

Linking of **SDA** toolkit with numeric ASCII data held in DataShare for the purposes of analysis (**re-use**)

Facility to embargo variables within numeric files (in statistical analysis package formats) for subsequent open deposit into DataShare of de-sensitised version

Research data deposit directly from **RSpace** Electronic Lab Notebook (ELN) interface into DataShare and Datastore (& Data Vault) using SWORD protocol

Who and what is driving demand for ELNs?

- **Researchers**

- Utility and convenience of paper lab book + online capabilities
- On multiple devices
- File management/integration

- **Groups/PIs**

- Controlled sharing
- Collaboration
- Group management
- File management/integration

- **Institutions: data librarians, research admins, IT, commercialisation offices**

- Enterprise features: Scalable deployment, Single Sign On
- IP protection: audit trail, signing
- Publishing
- Archiving
- Repository integration
- File management/integration

RSpace



RSpace
First electronic notebook for research
institutions

Researchers

Sketching
Chemistry
Notebook
Templating
Snippets
PDF export
Export to html
File gallery
Tablet friendly
Performance
Round trip editing
Offline access
File store integration
Box, Dropbox integration

Labs

Sharing
Messaging
Lab set up enabled
Group management
Inter-group collaboration
File store integration
Box, Dropbox integration

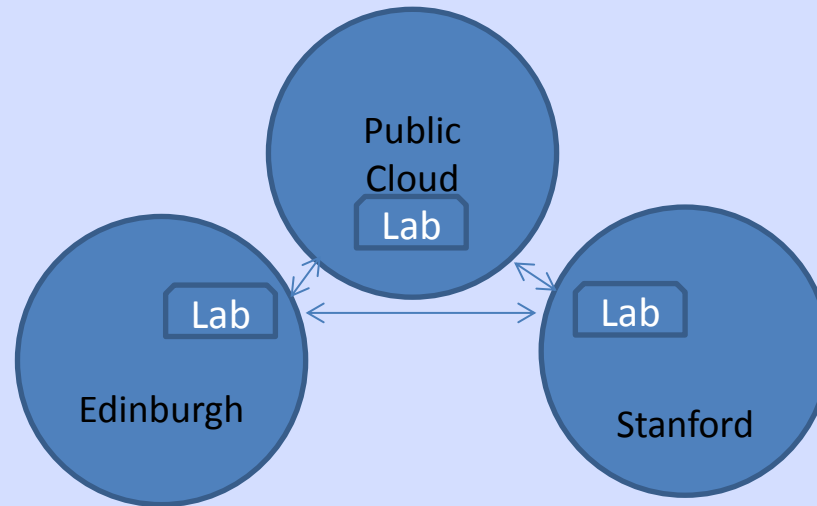
Institutions

Single sign on
Tiered admin
Communities
IP support
Repository integration
Archive integration
File store integration

Export and reimport of data
Between RSpace instances

Business Model

- Free public cloud for labs and individuals
- Institutional deployments @\$100/user/year
- Seamless movement of groups and data between different RSpaces



Researchers

Convenience
Productivity
Portability

Institutions

Control
Compliance
Data mining

Funders

Data mining

Value

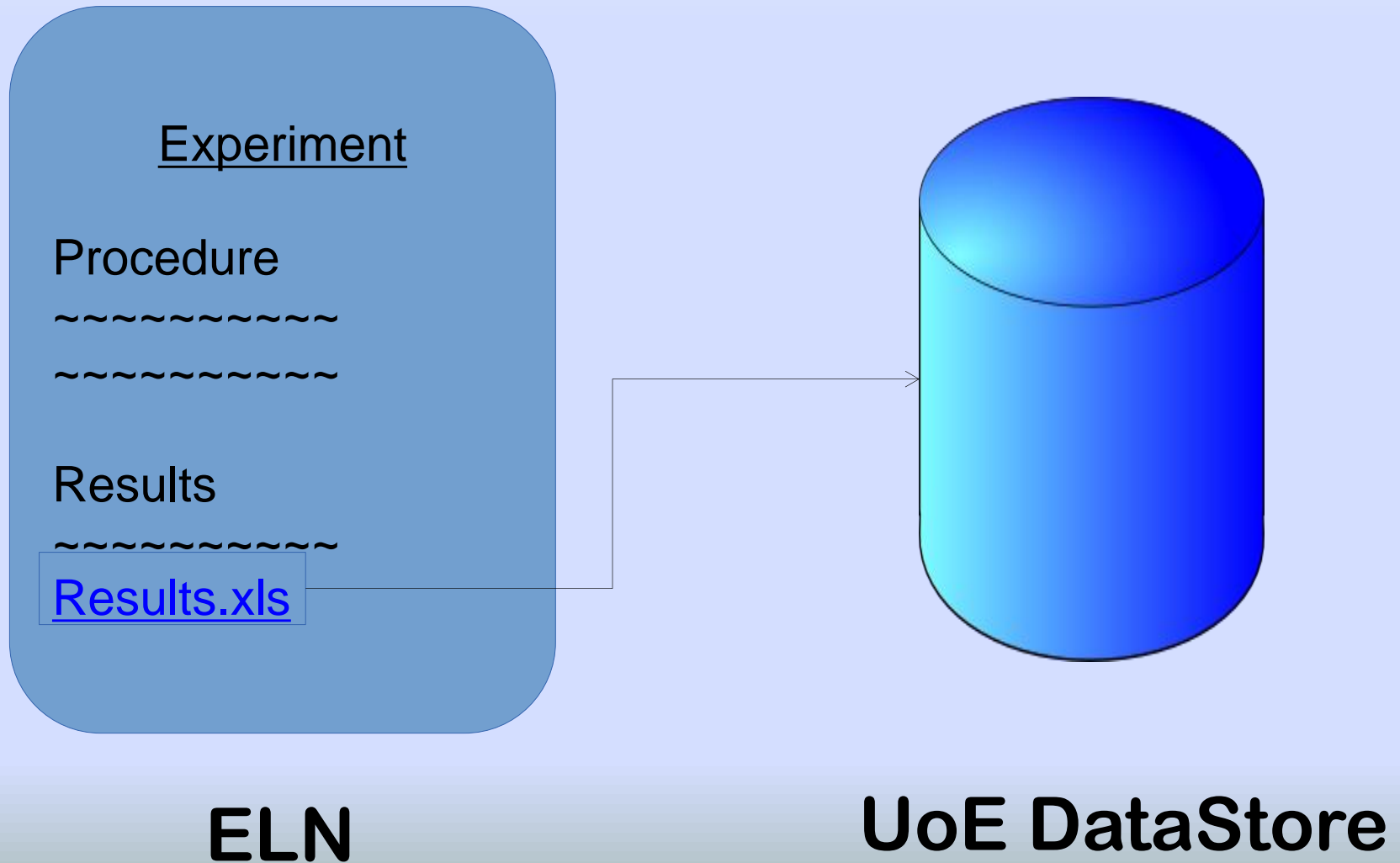
RSpace at Edinburgh

- Linking to files in Edinburgh DataStore**
- Depositing content in Edinburgh DataShare**
- Archiving in Edinburgh DataVault**

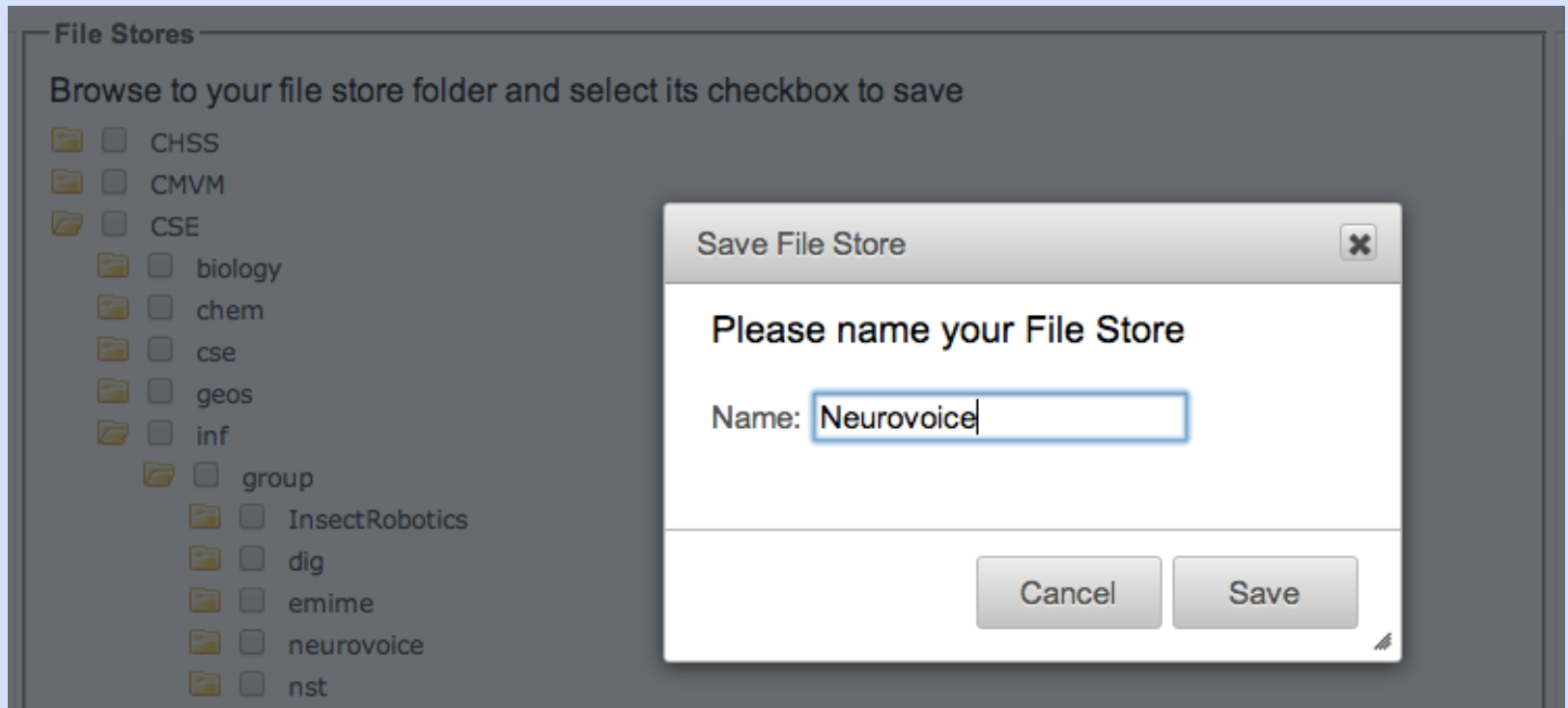
Linking to DataStore

“My plan for workflow would be generally to deposit my data in DataStore either from the wet lab instruments (gel photos, elisa data, etc, and also possibly directly from an iPad) or from in silico data analysis I’ve been doing, and then link to it from within RSpace.”

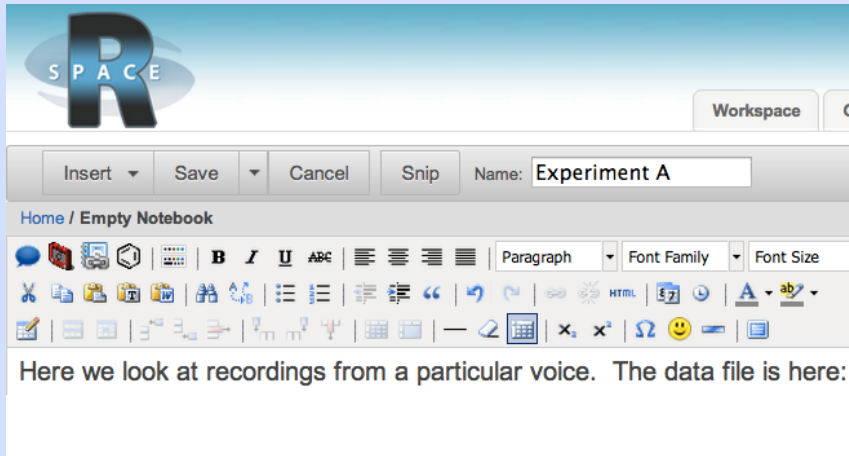
Linking to DataStore



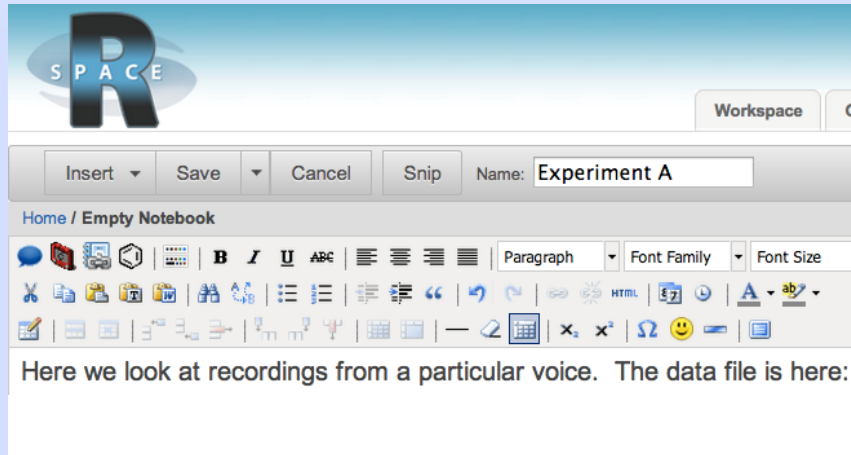
Exposing DataStore File Roots



Linking to a DataStore File

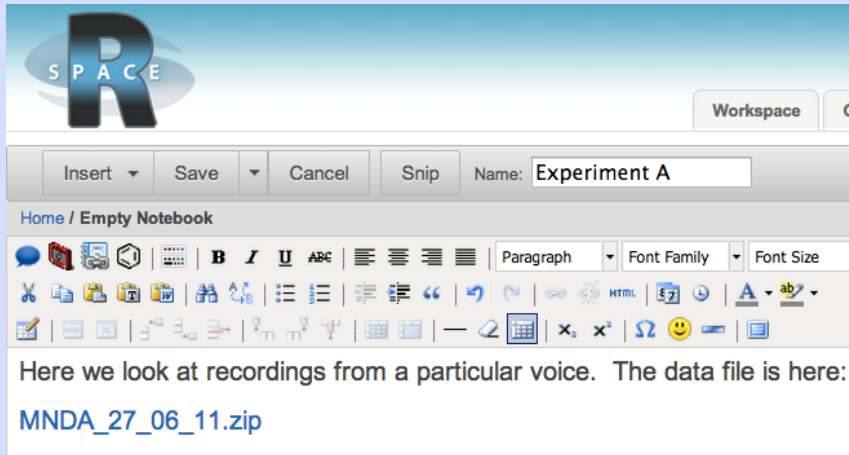


Linking to a DataStore File



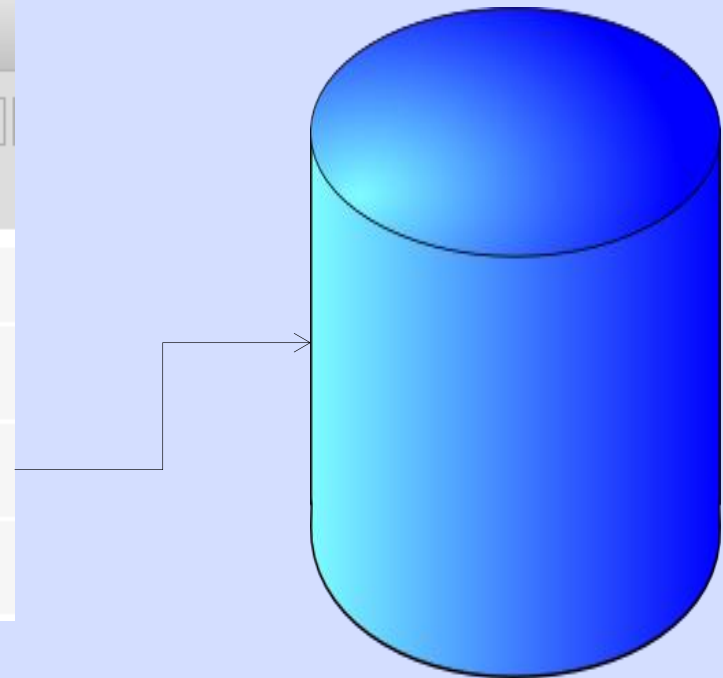
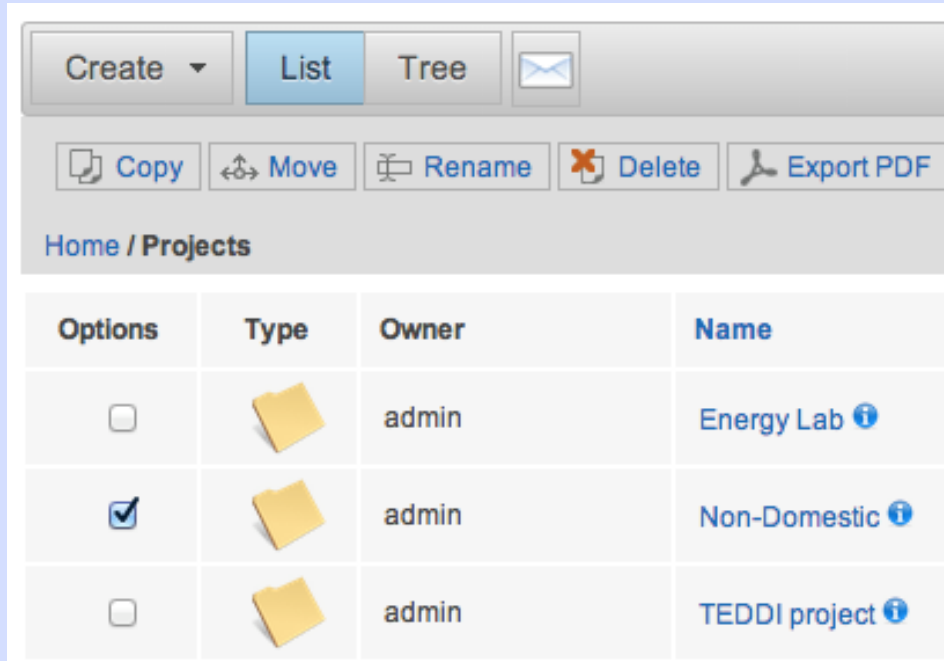
File Stores		
File Store: Neurovoice		
.TemporaryItems		
.Trash-628401		
quota.txt	69	2014-08-24
recordings		
database		
quests		
MND_A_27_06_11.zip	203469205	2011-06-27

Linking to a DataStore File



File Stores		
File Store: Neurovoice		
📁 .TemporaryItems		
📁 .Trash-628401		
📄 quota.txt	69	2014-08-24
📁 recordings		
📁 database		
📁 guests		
📄 MND_A_27_06_11.zip	203469205	2011-06-27

Exporting to DataShare






RSpace

UoE DataShare

Adding metadata

CopyMoveRenameDeleteExport PDFExport Zip ArchiveExport to Repository

Home / Projects

Options	Type	Owner	Name
<input type="checkbox"/>		admin	Energy Lab ⓘ
<input checked="" type="checkbox"/>		admin	Non-Domestic ⓘ
<input type="checkbox"/>		admin	TEDDI project ⓘ

Records per page: 10 ▾

Required Information for Repository Deposit

Enter your name. Put any forename(s) in the first box and your last name in the second

Enter the publishing institution, followed by any school or department, eg University of Edinburgh, School of GeoSciences, Institute of Geography

Enter a title for the deposit

Enter the names of any contributors who created the contents of the deposit, with first and then surname. Use 'Add' to add contributors and 'X' to remove

Add

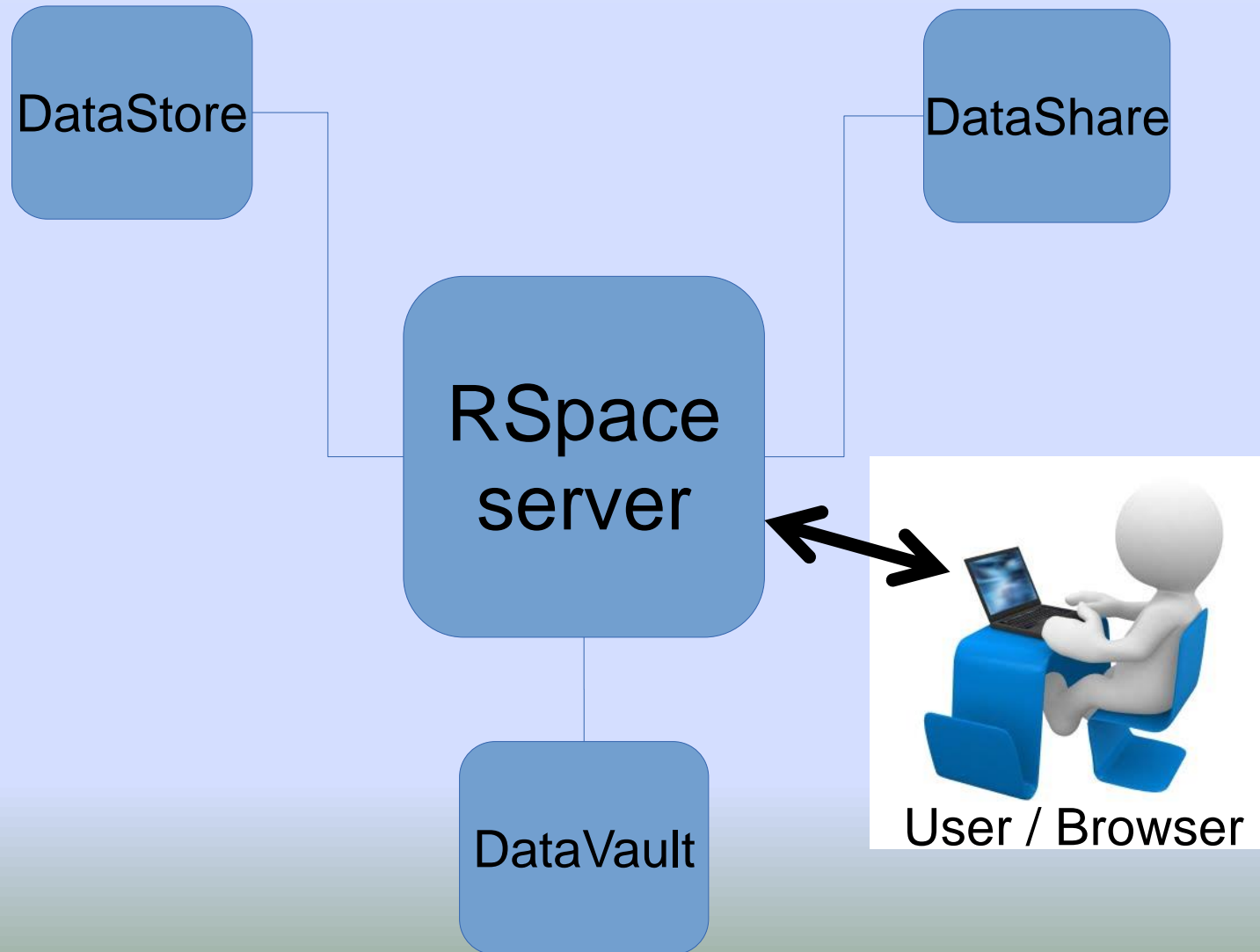
Cancel

Deposit

Archiving in Edinburgh DataVault

- DataVault functionality/API not yet specified
- Anticipate use of XML zip archive
- Many requirements to be determined
 - e.g., searching, restoration

RSpace and Edinburgh RDM



Thanks!

Acknowledgements:

Sunny Yang, Richard Adams, Nigel
Goddard Robin Rice, Kevin Tomlinson,
George Hamilton, Orlando Richards