Data: institutional & data centre roles
The DCC experience

Kevin Ashley
Digital Curation Centre
www.dcc.ac.uk
@kevingashley
Kevin.ashley@ed.ac.uk

Reusable with attribution: CC-BY

The DCC is supported by Jisc

DCC because good research needs good data

A summary

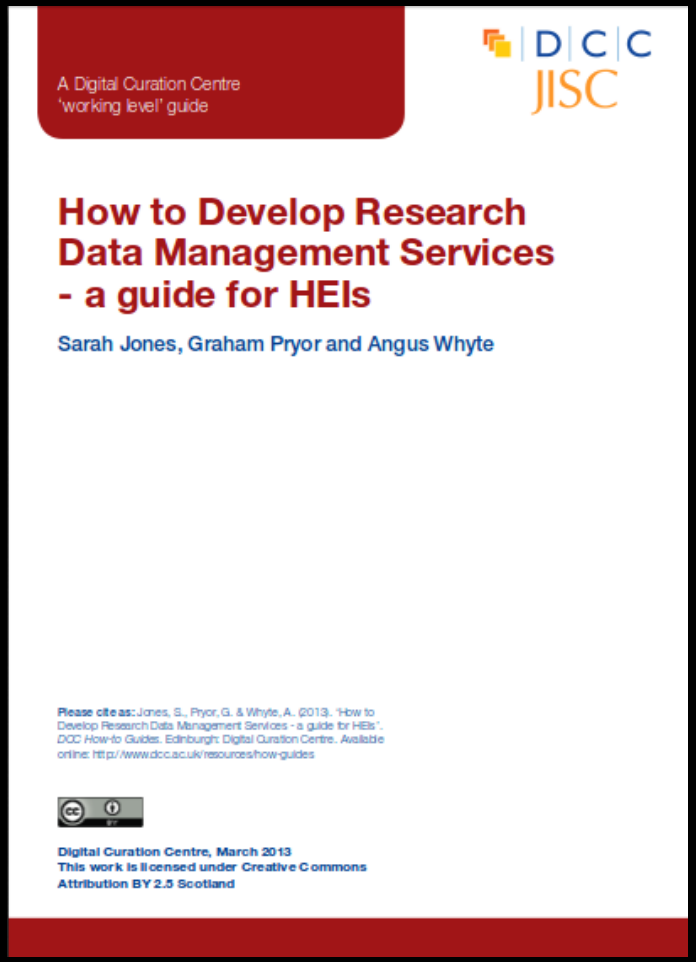
- Why data reuse ?
- What stops us ?
- How data management helps
- Barriers & costs
- The case for reuse - again

My home – the DCC

because good research needs good data

- increase capability and capacity for research data services in UK institutions
- Not just a UK problem – an international one
- Training, shared services, guidance, policy, standards, futures

DCC guidance



A Digital Curation Centre
'working level' guide

DCC JISC

How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

Please cite as: Jones, S., Pryor, G. & Whyte, A. (2013). 'How to Develop Research Data Management Services - a guide for HEIs'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides

Digital Curation Centre, March 2013
This work is licensed under Creative Commons Attribution BY 2.5 Scotland

CASE STUDY

A Digital Curation Centre Case Study
March 2013

DCC JISC

RDMRose

Section of the How to guide that this supports
Guidance, training and support

RDM Training for Librarians

Marieke Guy, Digital Curation Centre

Introduction

This case study looks at the approaches taken by three Jisc Managing Research Data Projects (2011 – 2013) and one institution to providing effective training for librarians and information services professionals in Research Data Management (RDM).

Background context

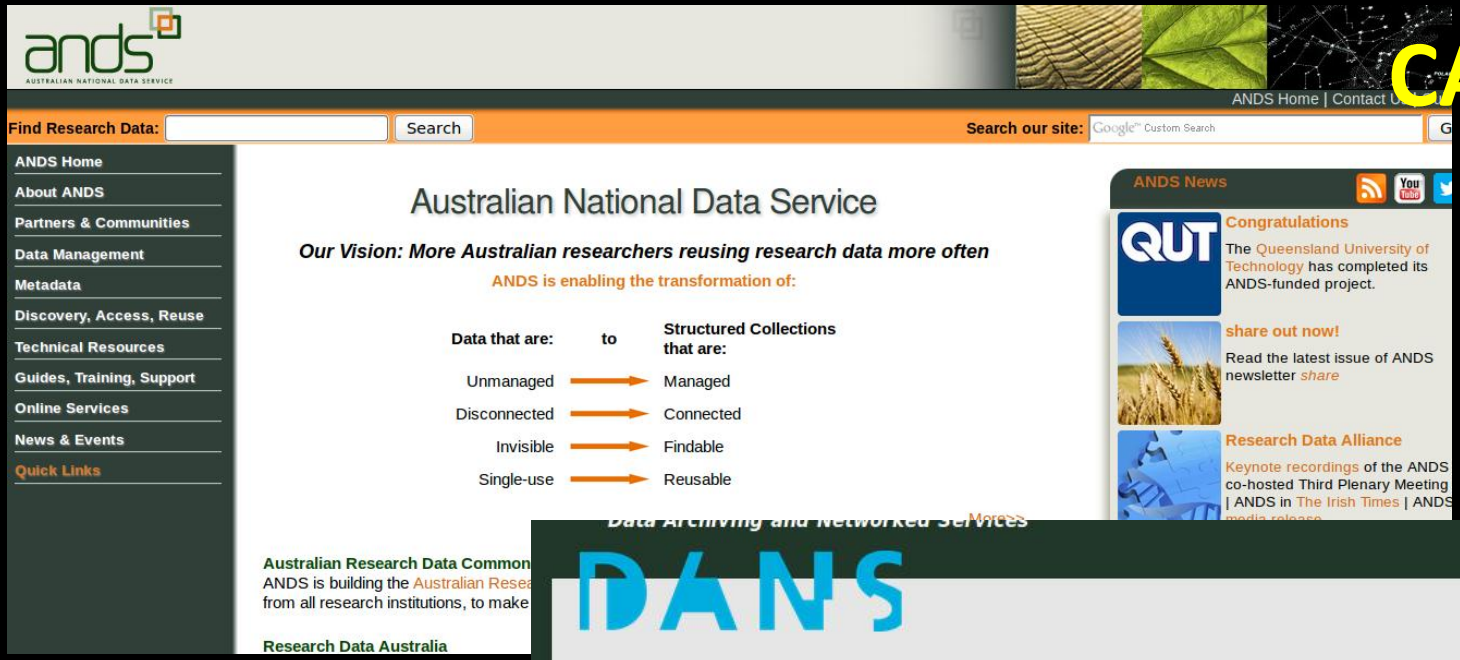
Through its institutional engagement programme the Digital Curation Centre (DCC) has seen many institutional (RDM) initiatives emanating from libraries. Librarians are carving out a new role for themselves in promoting and embedding good RDM practices. They are well placed for this role having information science skills in areas such as metadata, open access, institutional repository use - key constituents for RDM. Librarians also tend to have good working relationships with other service departments and researchers. This results in requirements for them to sit on institutional working parties and steering groups redesigning institutional strategies and infrastructure necessary to meet operational and regulatory requirements. Despite these factors various reports indicate that researchers do not immediately turn to the library, even when librarians could help a lot.

The 2012 RLUK report on *Re-skilling for Research* looked at the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. It states that it "is clear that as the nature of research within our institutions changes, so must the role of the library in supporting research". More recently the Association published a paper: *Academic Libraries and Research Data Services: Current practices and plans for the future*. The paper found that Libraries tend to rely on external conferences or workshops to provide research data services training and that there is a lack of internal institutional support in this area. As the landscape changes individual institutions will need to take measures to ensure that their library and information services staff are effectively trained in aspects of RDM. This case study looks at initiatives in this area.

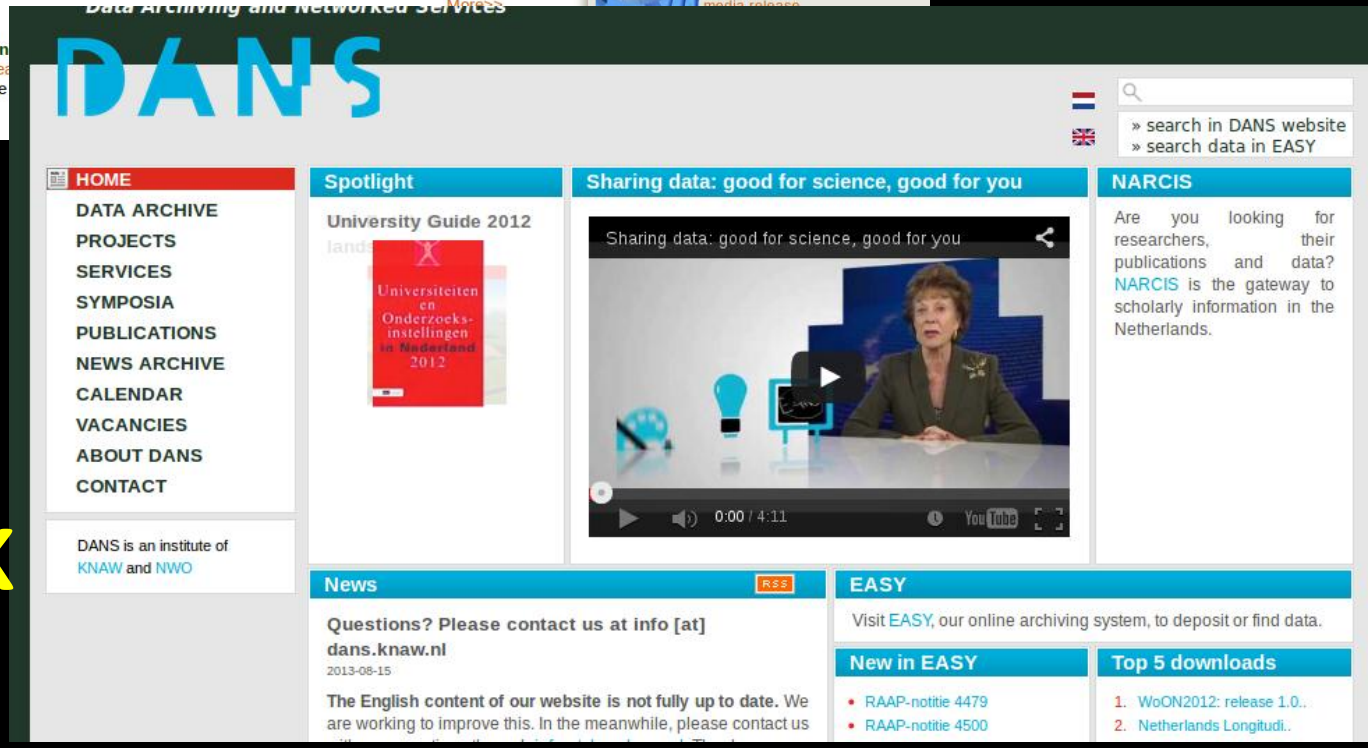
Overview: RDMRose

RDMRose is a Jisc-funded project to produce taught and continuing professional development (CPD) learning materials in RDM tailored for information professionals. It is looking in particular at the specific needs of liaison librarians in university libraries, and deliverables include OER materials suitable for learning in multiple modes, including face-to-face and self-directed learning. All materials can be reused by other library and information service educators and a version for self-supported Continuing Professional Development is available.

The RDMRose project brings together the University of Sheffield iSchool with a practitioner community based on the White Rose University Consortium's libraries at the Universities of Leeds, Sheffield and York. Development of content and teaching was iterative, based on a highly participative curriculum development process and with a strong strand of student evaluation of learning materials and activities. Version 1 of the training materials was



CANADA



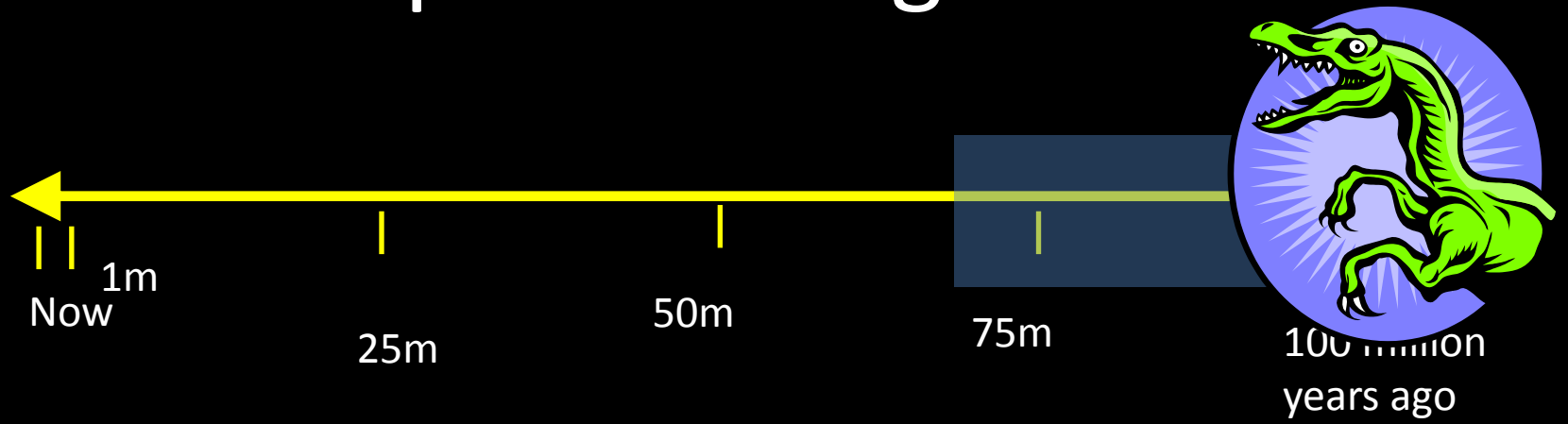
SWEDEN

DENMARK

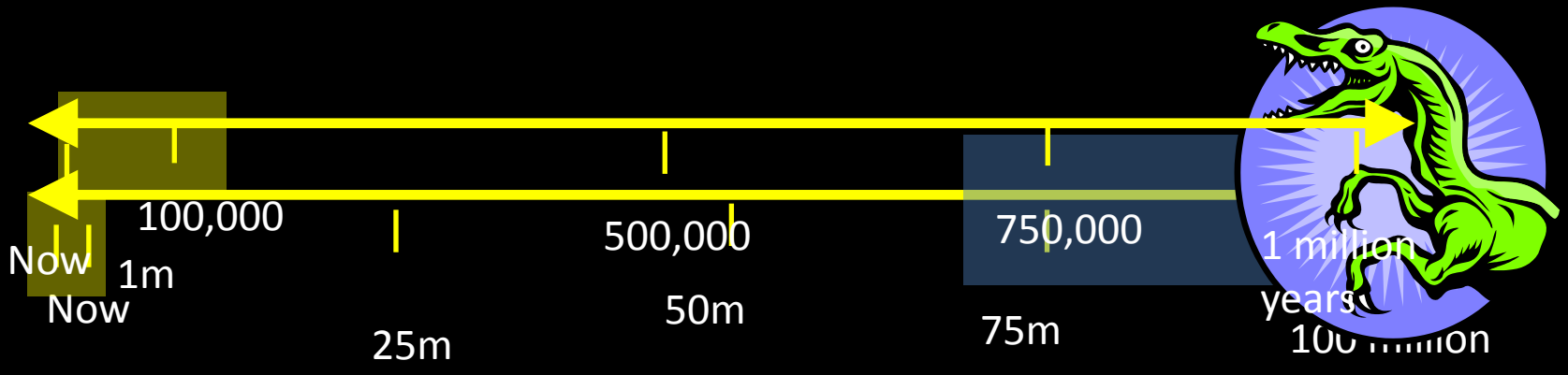
Data reuse stories

- The palaeontologist who saved years of work with archaeological data

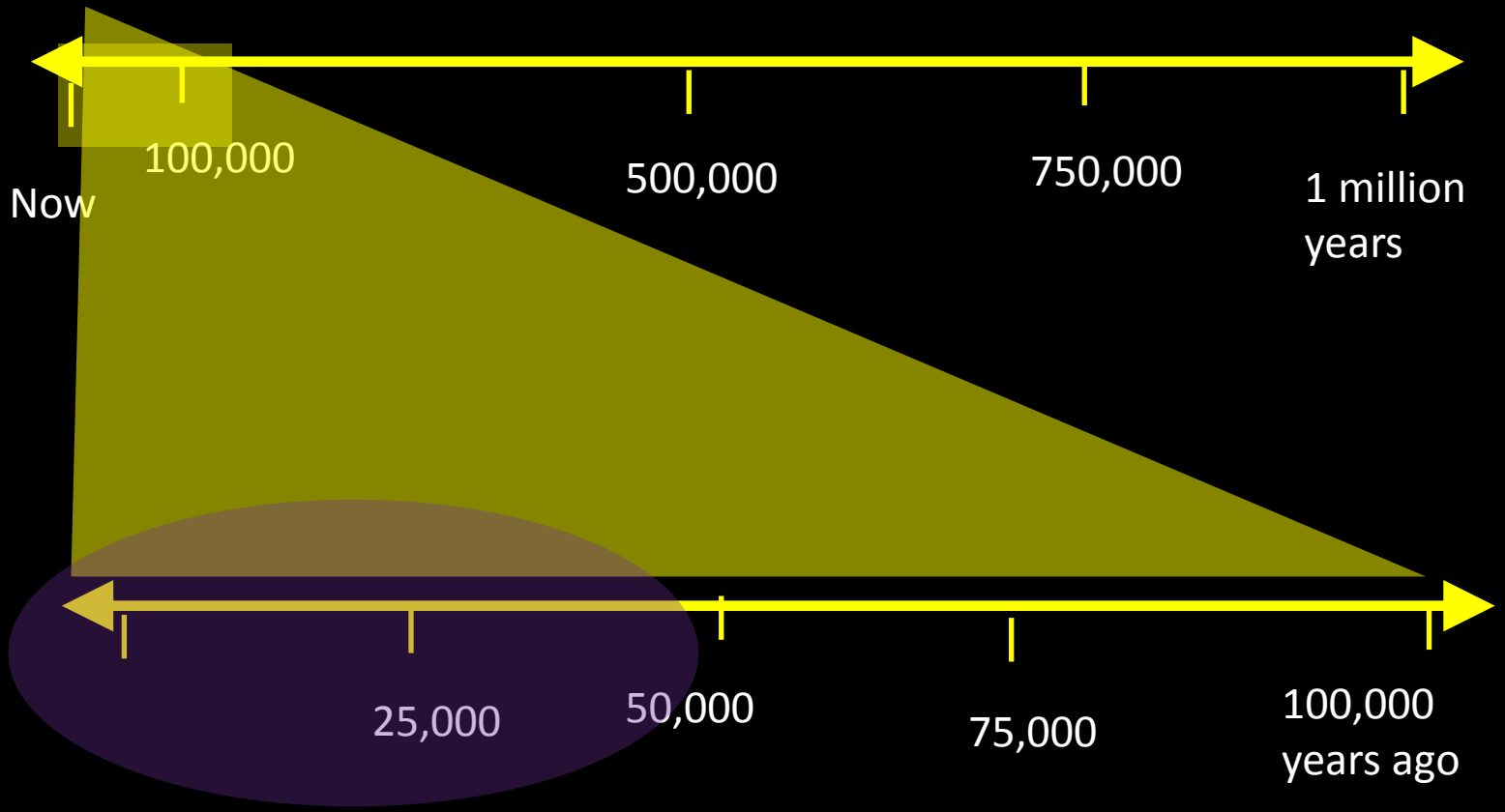
What a paleontologist looks at



What looks at



What an archaeologist looks at

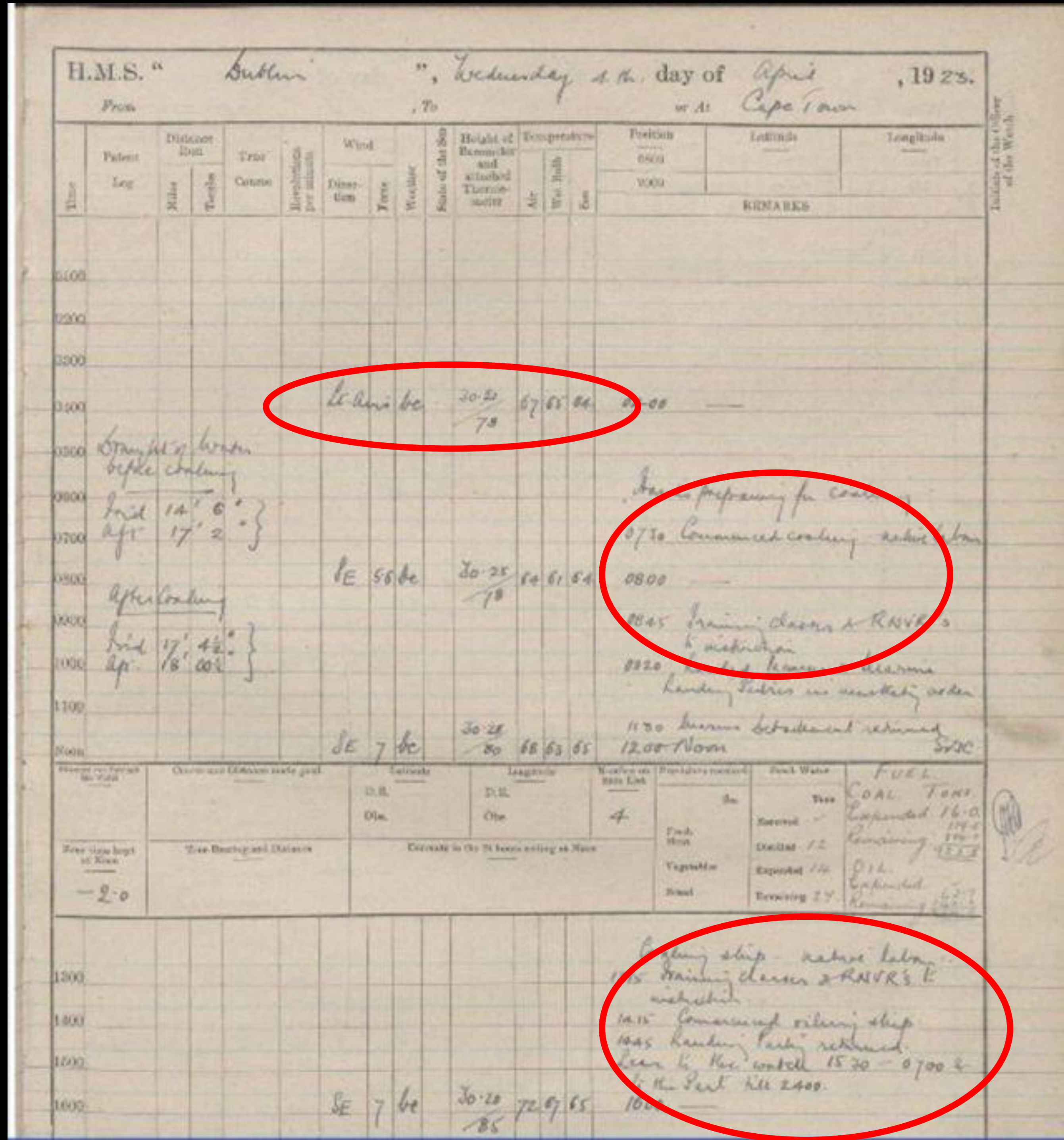


Kevin Ashley –e-IRG2014 - CC-BY

Data reuse stories

- The palaeontologist who saved years of work with archaeological data
- The 19th-century ships logs that help us model climate change

The Old weather project



Data for research, not from research

2014-06-09

Data reuse stories

- The palaeontologist who saved years of work with archaeological data
- The 19th-century ships logs that help us model climate change
- The ‘noise’ from research radar that mapped dust from Eyjafjallajökull

Data reuse - messages

Often your data tells stories that your publications do not

Discipline-bounded data discovery doesn't give us all we need or want

Not all data comes from other researchers

One person's noise is another person's signal

Why care?

- Data is expensive – an investment
- Reuse:
  - More research
  - Teaching & Learning
  - Planning
- Impact – with or without publication
- Accountability
- Legal & regulatory requirements

Why does this matter?

- Research quality
  - How close can we get to the truth?
- Research speed
  - How quickly can we get to the truth?
- Research finance
  - How much does the truth cost?

- Improving one or more of these is of interest to all actors:
- Researchers as data creators
- Researchers as data reusers
- Research institutions
- Funders – hence government and society

G8UK - Endorses OA Open Data Charter Policy Paper 18 June 2013

G8UK - Billigt offenen Zugang Eine offene Daten Charter Strategiepapier.

Cabinet Office

Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

Contents

1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5: Releasing Data for Innovation
6. Technical annex

Funder requirements

- UK

http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx

- USA – NSF, NEH, NIH
- Europe

- Most place burden on researcher – some on the institution

RCUK policy - The 1-minute version

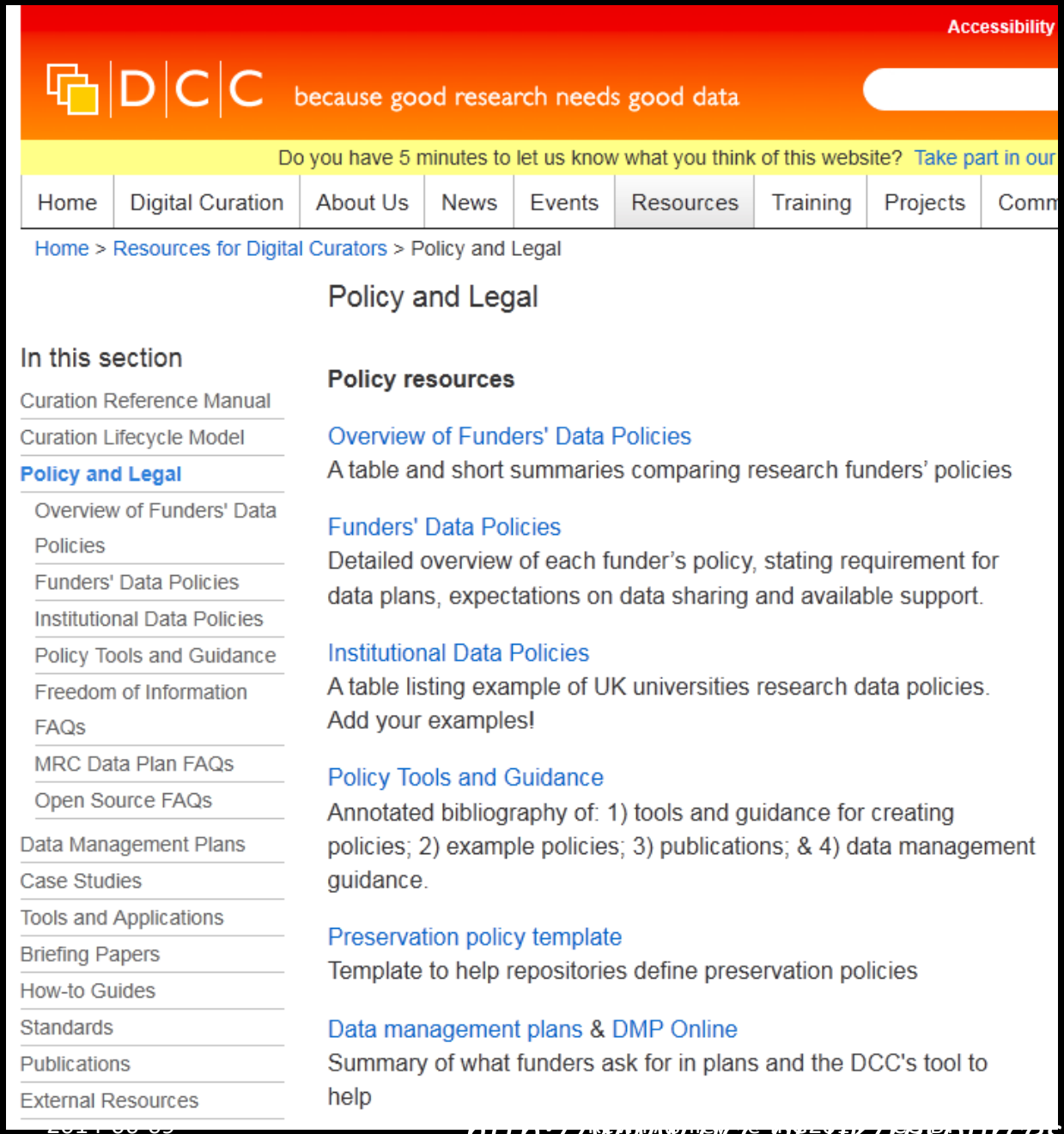
- Research data are a public good – make openly available in timely & responsible way
- Have policies & plans. Data with long-term value should be preserved & usable
- Metadata for discovery & reuse. Link publications & data
- Sometimes law, ethics get in the way. We understand.
- Limited embargos OK. Recognition is important – always cite data sources
- OK to use public money to do this. Do it efficiently.

EPSRC policy points

- Awareness of regulatory environment
- Data access statement
- Policies and processes
- Data storage
- Structured metadata descriptions
- DOIs for data
- Securely preserved for a minimum of 10 years **from last use**

Compliance expected by 2016

Kevin Ashley –e-IRG2014 - CC-BY



DCC Policy Summary

http://www.dcc.ac.uk/resources/policy-and-legal

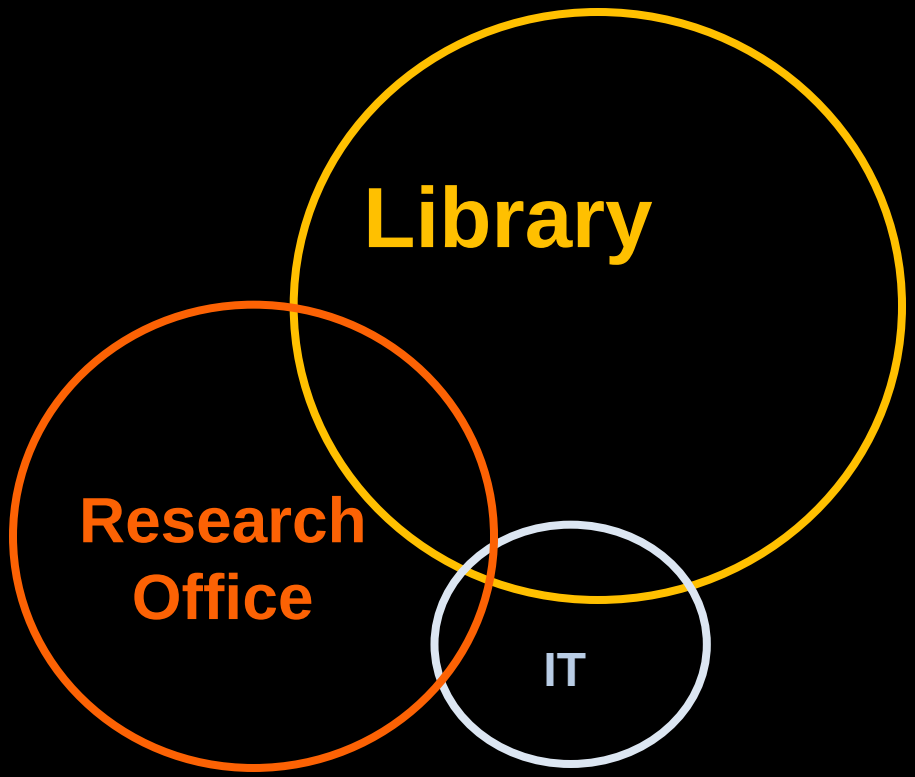
Compliance

Benefits

Some institutional roles

- Leadership – coordinate action
- Audit – who has what, where does it go?
- Advice on access – data, wherever it is
- Preservation – permanence
- Citability
- Data/publication linking
- Promoting data in teaching
- Selection
- Education – early career researchers

Who (in the UK) is addressing RDM ?



Kevin Ashley –e-IRG2014 - CC-BY

How?

- Create policy – collaborate with others
- Develop existing digital services
- Learn about audit tools (DCC & others)
- Learn about data & sources
- Reskill subject librarians
- Learn about your own data
- Bridge between publishers & researchers

Understanding Data Requirements

Data Asset Framework

CARDIO
collaborative assessment of research data infrastructure and objectives

If research data lies at the heart of your organisation, you need to know that you have adequate infrastructure, staff skills and resources, and senior management support in place to ensure that your data is effectively managed for validation, reuse and evidential purposes.

CARDIO enables you to:

- collaboratively assess data management requirements, activity, and capacity at your institution
- build consensus between data creators, information managers and service providers
- identify practical goals for improvement in data management provision and support;
- identify operational inefficiencies and opportunities for cost saving;
- make a compelling case to senior managers for investment in data management support

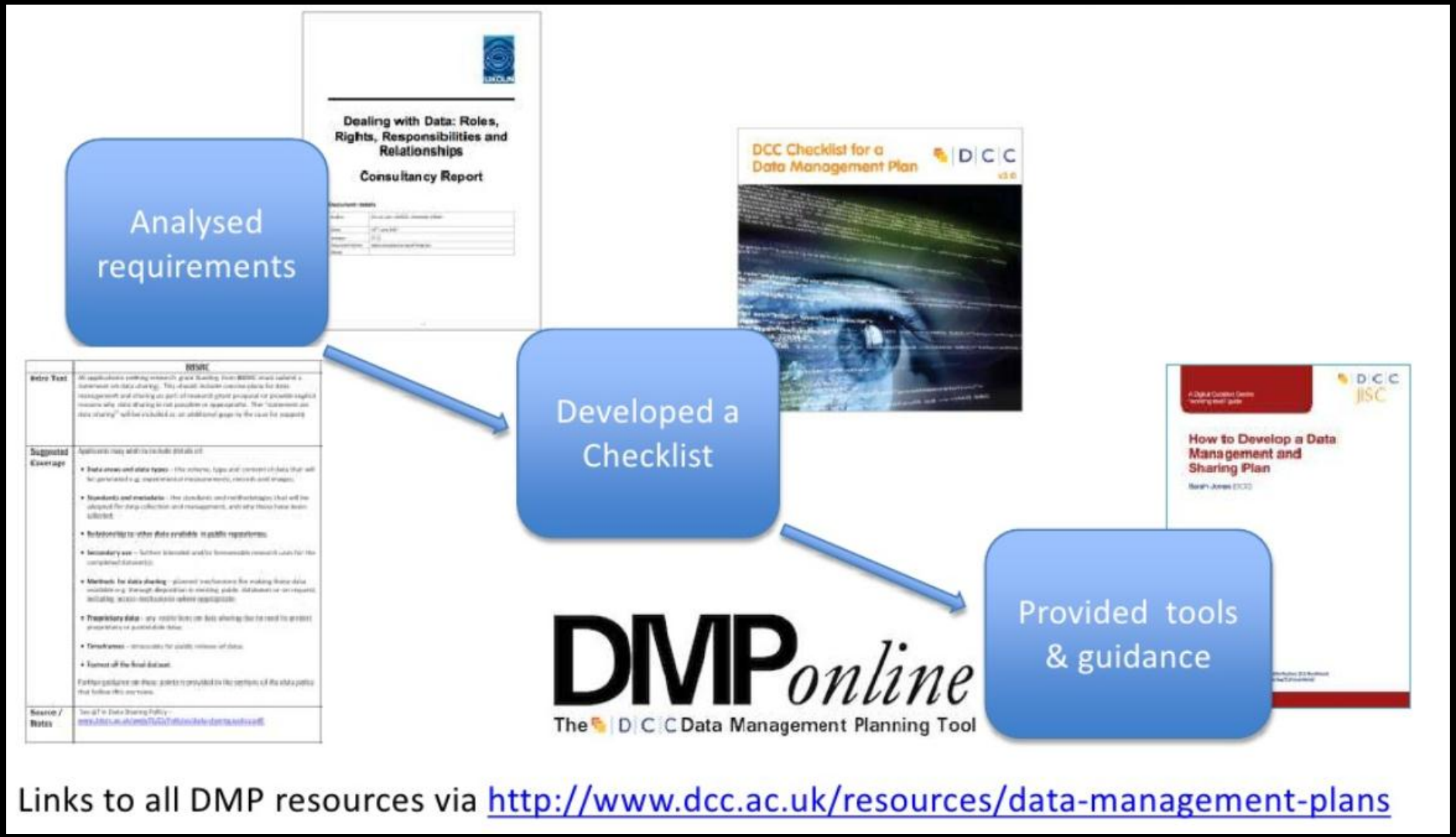
What stops data reuse

- Loss
- Destruction
- Pride
- Gluttony
- Ineptitude
- Concealment
- Bureaucracy
- Complexity
- Procrastination
- Lack of potential

http://www.flickr.com/photos/mattimattila/3003324844/

"Departments don't have guidelines or norms for personal back-up and researcher procedure, knowledge and diligence varies tremendously. Many have experienced moderate to catastrophic data loss"

Incremental Project Report, June 2010



Links to all DMP resources via http://www.dcc.ac.uk/resources/data-management-plans

What data to keep

Roles and Responsibilities

A Digital Curation Centre and Australian National Data Service 'working level' guide

DCC

ands

How to Appraise & Select Research Data for Curation

Angus Whyte (DCC) and Andrew Wilson (ANDS)

Researcher ('data creator')

- Provide enough information for others to assess the research data's scientific and scholarly quality and compliance with disciplinary or ethical norms.
- Provide relevant information for the repository to identify who will use the data and how i.e. the 'designated community', and any specific access requirements or constraints.
- Provide the research data in formats recommended by the data repository.
- Provide the metadata requested by the repository.

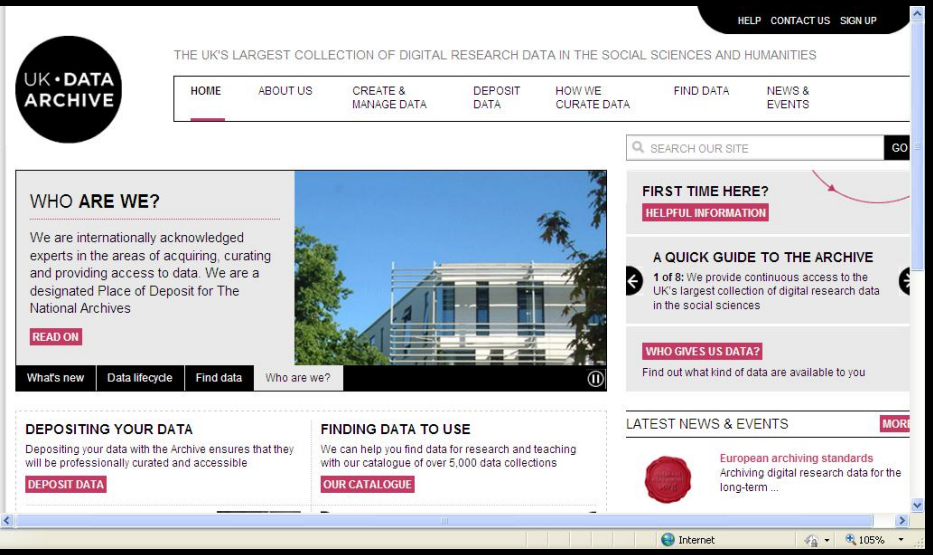
Data centre or repository

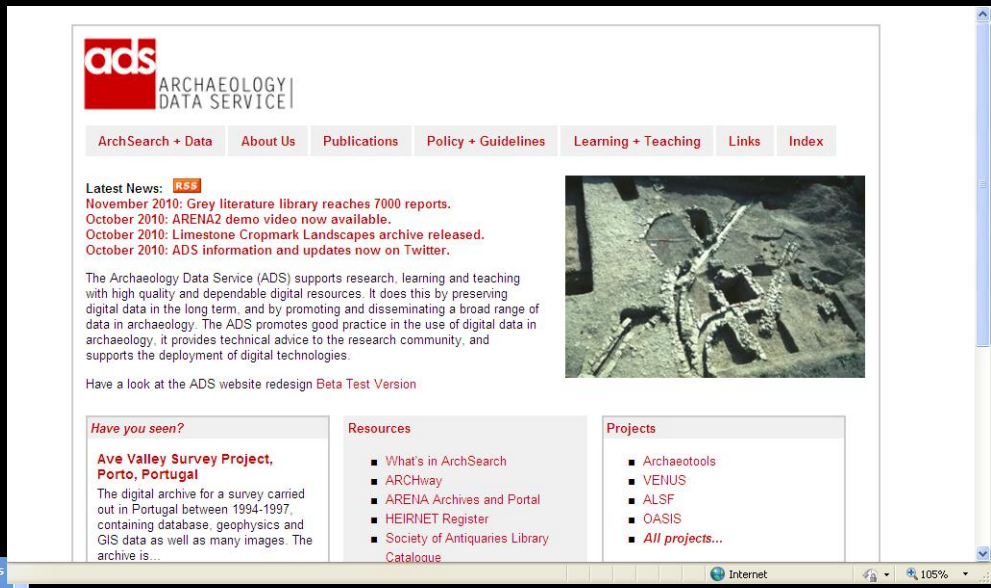
- Make explicit its mission in the area of digital archiving, and its selection policy for digital objects.
- Ensure compliance with legal regulations and contracts.
- Ensure the authenticity and integrity of the digital objects and the metadata.
- Assume responsibility from the data producer for ensuring the digital objects are accessible and available to a defined 'designated community'.
- Plan for long-term preservation of the digital assets.

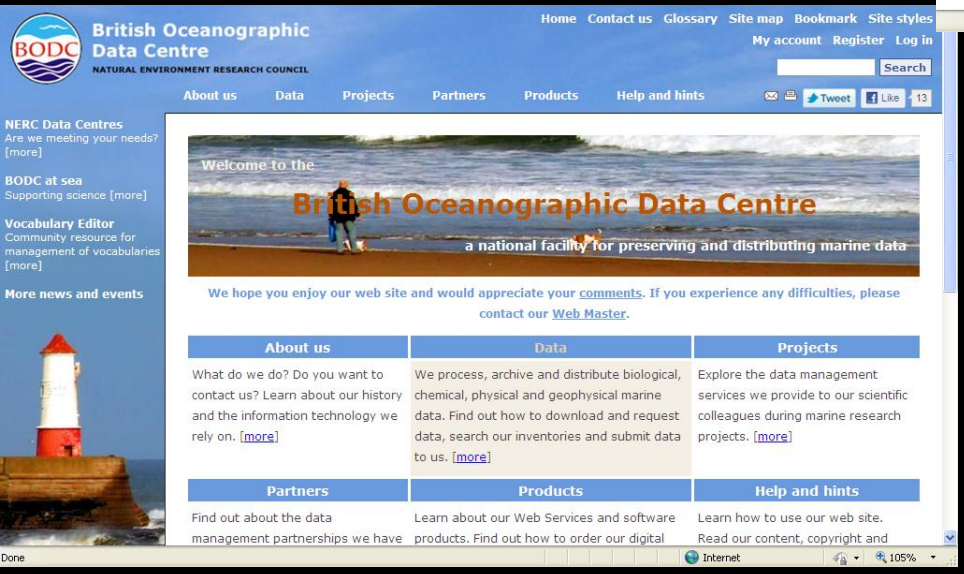
Excuses – and responses

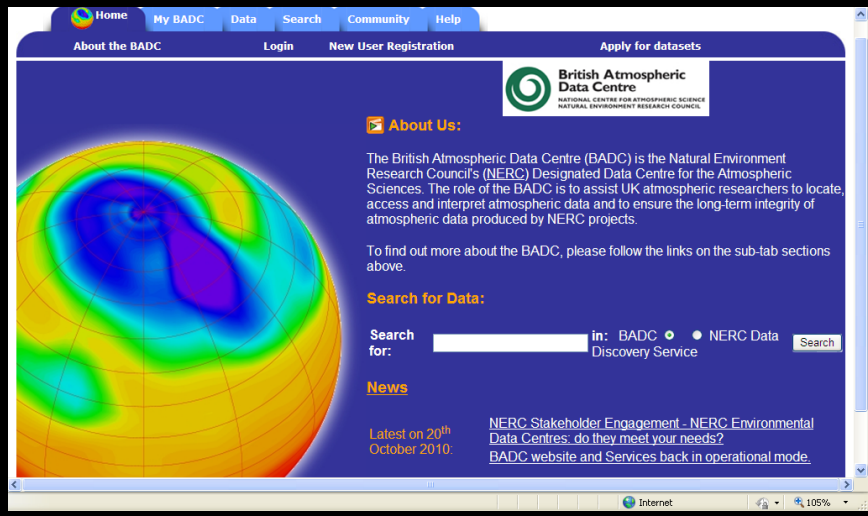
- "People will ask questions"
  - So use a data centre or repository
- "It will be misinterpreted"
  - Stuff happens. Also, openness encourages correction
- "It's not interesting"
  - Let others be the judge – your noise is my signal
- "I might get another paper out of it"
  - Up to a point. We might get more research out of it
- "I don't have permission"
  - A real problem. But solvable at senior level
- "It's too bad/complicated" –see above
- "It's not a priority"
  - Unfortunately, funders are making it so. But if you looked at the evidence, it would be your priority as well

See e.g. Carly Strasser's blog:
http://datapub.cdlib.org/2013/04/24/closed-data-excuses-excuses/


HELP CONTACT US SIGN UP
UK • DATA ARCHIVE
THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES
HOME
ABOUT US
CREATE & MANAGE DATA
DEPOSIT DATA
HOW WE CURATE DATA
FIND DATA
NEWS & EVENTS
SEARCH OUR SITE
GO
WHO ARE WE?
We are internationally acknowledged experts in the areas of acquiring, curating and providing access to data. We are a designated Place of Deposit for The National Archives
READ ON
What's new
Data lifecycle
Find data
Who are we?
FIRST TIME HERE?
HELPFUL INFORMATION
A QUICK GUIDE TO THE ARCHIVE
1 of 8: We provide continuous access to the UK's largest collection of digital research data in the social sciences
WHO GIVES US DATA?
Find out what kind of data are available to you
DEPOSITING YOUR DATA
Depositing your data with the Archive ensures that they will be professionally curated and accessible
DEPOSIT DATA
FINDING DATA TO USE
We can help you find data for research and teaching with our catalogue of over 5,000 data collections
OUR CATALOGUE
LATEST NEWS & EVENTS
European archiving standards
Archiving digital research data for the long-term ...
Internet
105%


ads
ARCHAEOLOGY DATA SERVICE
ArchSearch + Data
About Us
Publications
Policy + Guidelines
Learning + Teaching
Links
Index
Latest News:
RSS
November 2010: Grey literature library reaches 7000 reports.
October 2010: ARENA2 demo video now available.
October 2010: Limestone Cropmark Landscapes archive released.
October 2010: ADS information and updates now on Twitter.
The Archaeology Data Service (ADS) supports research, learning and teaching with high quality and dependable digital resources. It does this by preserving digital data in the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, it provides technical advice to the research community, and supports the deployment of digital technologies.
Have a look at the ADS website redesign Beta Test Version
Have you seen?
Ave Valley Survey Project, Porto, Portugal
The digital archive for a survey carried out in Portugal between 1994-1997, containing database, geophysics and GIS data as well as many images. The archive is...
Resources
What's in ArchSearch
ARCHway
ARENA Archives and Portal
HEIRNET Register
Society of Antiquaries Library Catalogue
Projects
Archaeotools
VENUS
ALSF
OASIS
All projects...
Internet
105%


BODC
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL
Home
Contact us
Glossary
Site map
Bookmark
Site styles
My account
Register
Log in
Search
About us
Data
Projects
Partners
Products
Help and hints
Tweet
Like
13
NERC Data Centres
Are we meeting your needs? [more]
BODC at sea
Supporting science [more]
Vocabulary Editor
Community resource for management of vocabularies [more]
More news and events
Welcome to the
British Oceanographic Data Centre
a national facility for preserving and distributing marine data
We hope you enjoy our web site and would appreciate your comments. If you experience any difficulties, please contact our Web Master.
About us
What do we do? Do you want to contact us? Learn about our history and the information technology we rely on. [more]
Data
We process, archive and distribute biological, chemical, physical and geophysical marine data. Find out how to download and request data, search our inventories and submit data to us. [more]
Projects
Explore the data management services we provide to our scientific colleagues during marine research projects. [more]
Partners
Find out about the data management partnerships we have
Products
Learn about our Web Services and software products. Find out how to order our digital
Help and hints
Learn how to use our web site. Read our content, copyright and
Done
Internet
105%


Home
My BADC
Data
Search
Community
Help
About the BADC
Login
New User Registration
Apply for datasets
British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL
About Us:
The British Atmospheric Data Centre (BADC) is the Natural Environment Research Council's (NERC) Designated Data Centre for the Atmospheric Sciences. The role of the BADC is to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by NERC projects.
To find out more about the BADC, please follow the links on the sub-tab sections above.
Search for Data:
Search for:
in: BADC NERC Data Discovery Service
Search
News
Latest on 20th October 2010:
NERC Stakeholder Engagement - NERC Environmental Data Centres: do they meet your needs?
BADC website and Services back in operational mode.
Internet
105%

Data centres are good value!

- See Jisc reports on ADS, BADC, UKDA:
- Returns on investment between 400% and 1200%

What about collaboration?

- Collaborate within the university
- Collaborate with partners
- Collaborate with regional, national services
- Not everything can be done well locally
- Infrastructure needed at research group, institution, national, (discipline) & international level

DCC because good research needs good data

Search

Home | Digital curation | About us | News | Events | Resources | Training | Projects | Community

Earth Science

Palaeontology Geology Genomics Environmental Science Hydrology Hydrogeology Botany Livestock Cartography Fish Farming Climatology Glaciology Ecology Marine Biology Topography Geoscience Biogeography Molecular biology Planetary science Agricultural Economics Hydrography Oceanography Planning (Urban, Rural and Regional) Entomology Geography Meteorology Minerology Multi-disciplinary Marine Science Agricultural Science Biochemistry Genetics Soil Science

Metadata Standards

AgMES - Agricultural Metadata Element Set

A semantic standard for description, resource discovery, interoperability and data exchange for different types of agricultural information resources.

CF (Climate and Forecast) Metadata Conventions

A standard for climate and forecast "use metadata" that aims both to distinguish quantities (such as physical description, units, or prior processing) and to locate the data in space–time.

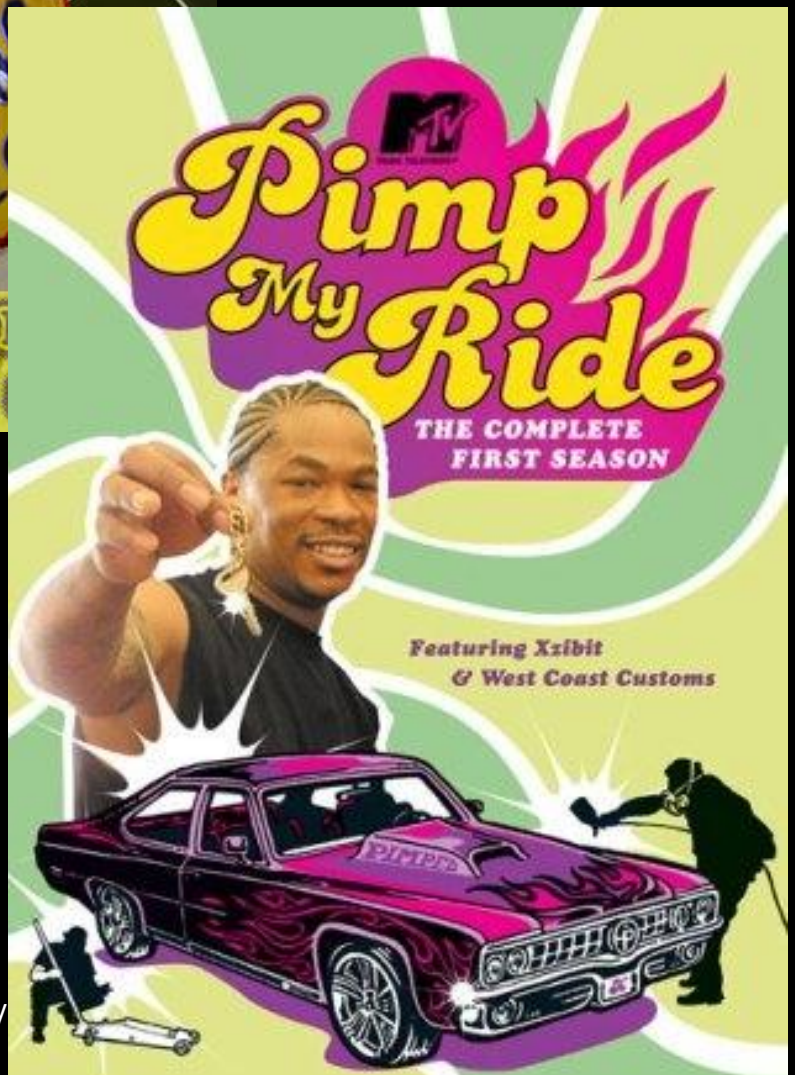
DIF - Directory Interchange Format

Curation training

Looking to develop your data management and curation skills? Learning is easy when you sign up for any of our introductions to digital curation, which cover all those activities you need to consider when planning and implementing new projects.

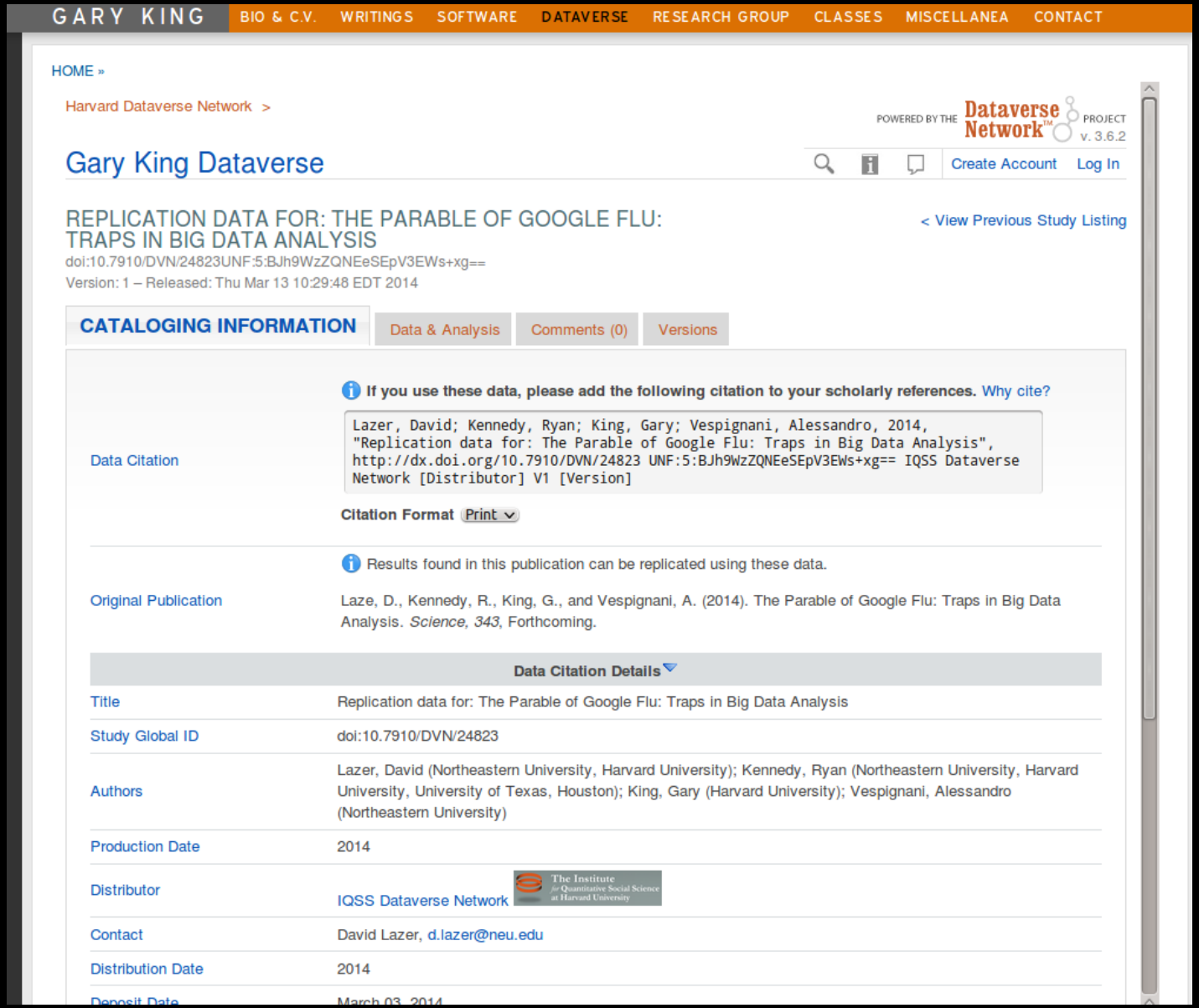
Read more


Cadbury


MTV
Pimp My Ride
THE COMPLETE FIRST SEASON
Featuring Xzibit
& West Coast Customs

Gking.harvard.edu/data

Pimp your data – make it findable & reusable



On costs

- Costs of data curation relatively simple to measure: see work of 4C (4cproject.eu)
- Charging and payment are more complex
- Funder rules can lead to perverse, inefficient payment systems
- Fundamental question is ‘who pays’. This changes the answer to ‘what does it cost’

Commercial services

figshare

creative commons

DIGITAL science

OPEN ACCESS

ARKIVUM
ASSURED ARCHIVING

What we do | Credentials | Partners | News etc | About us

+44 1249 405060
info@arkivum.com

Why Arkivum?

100% data guarantee
We understand data preservation
Territory specific data centres

Learn more »

The UK funding model

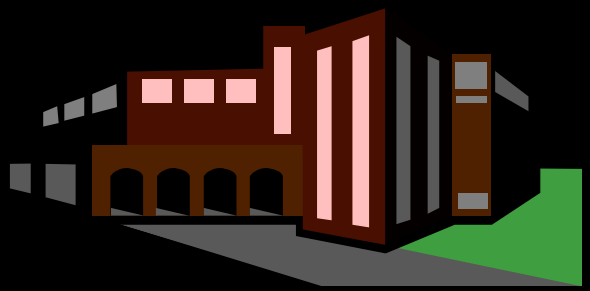
Research Council

Project funding, to PI

Higher Education Funding Council

Annual funding, per institution

The complexities appear in every country, just in different ways

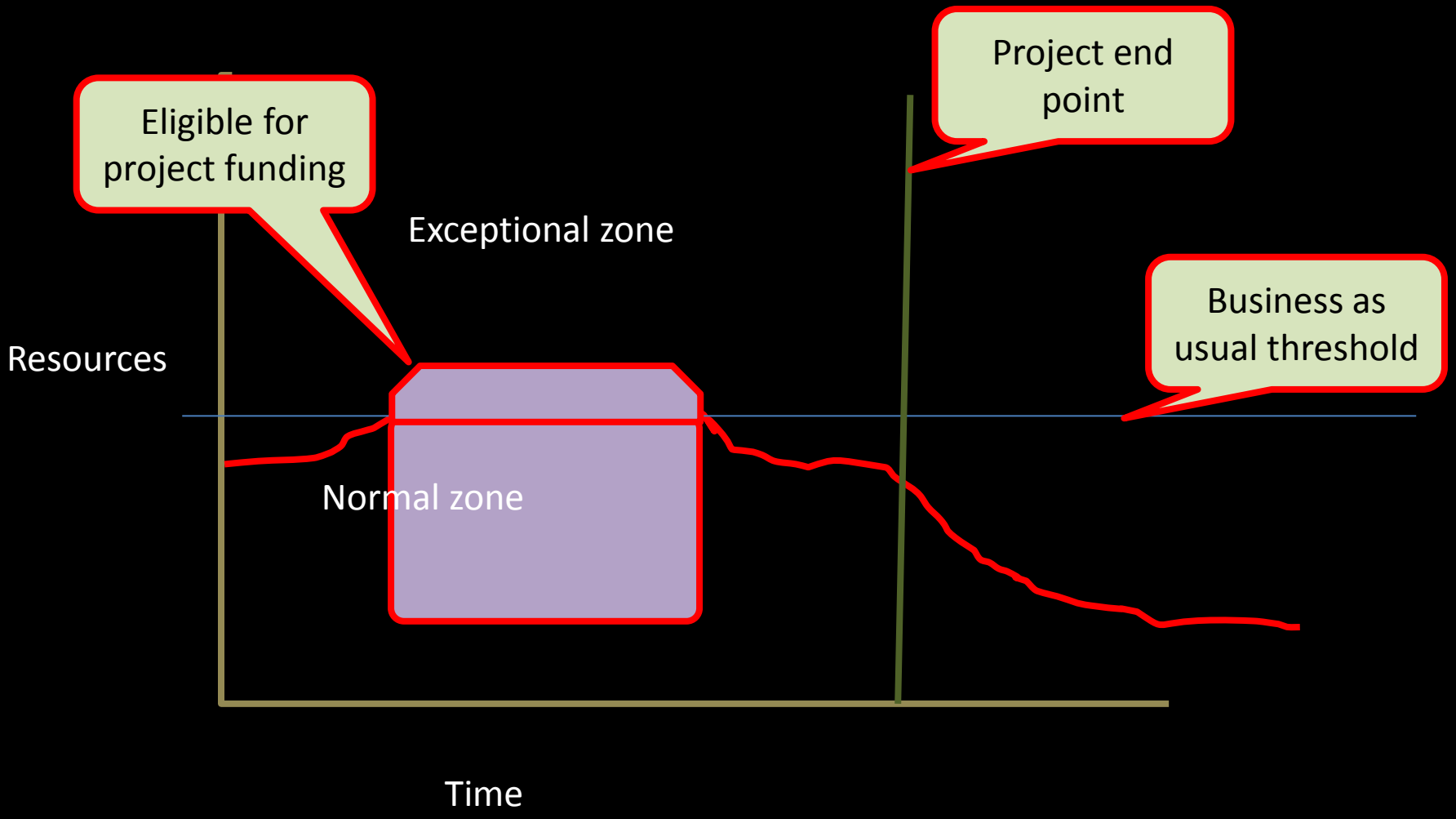


Kevin Ashley –e-IRG2014 - CC-BY

What it means

- Project funding can only be spent during projects on direct project costs
- Project funding comes with overheads, which universities must use for research infrastructure
- Ongoing ('QR') money is continuous, relates to research ranking
- Important to distinguish business-as-usual from exceptional requirements

A research lifecycle



Kevin Ashley –e-IRG2014 - CC-BY

We have money

Funders view

We have requirements

We have rules about how you use money to meet requirements

Over to you!

Being clever with costs

- Ongoing costs beyond project end cannot be charged to a grant, but…
- ‘Pay once, store forever’ charges acceptable.
- Thus, incentive to outsource lon curation
- Yet universities are only acting as last-resort option in any case – discipline data archives preferred

Many of these are run by funders

Closing thoughts

- Library/data centre roles:
  - selecting content
  - protecting it
  - enabling and encouraging reuse
  - Assisting with data management planning
- Library:
  - helping users find the most relevant content – much research data does not come from research
- Data center:
  - setting standards
  - enabling uptake

Infrastructure levels

- Truly international – instruments, standards
- National variation, international core:
  - Training
  - Data management planning
  - Policy
  - ..

My message to researchers

- The credit belongs to you
- The data belongs to all of us
- Share, and we all reap the benefits