



University of St Andrews

600 YEARS
1413 – 2013



University
of Glasgow

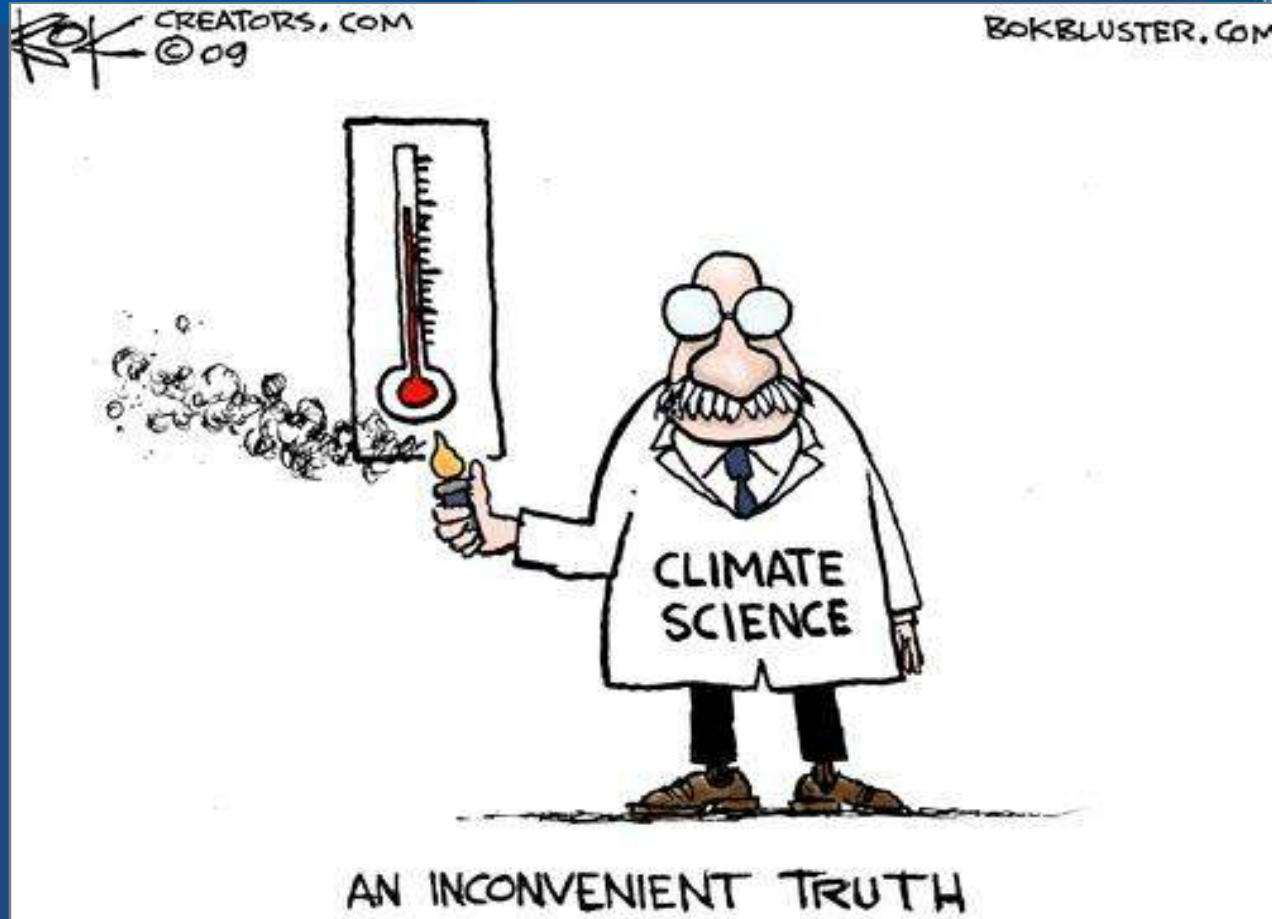
Anna Clements

Head of Research Data & Information Services,
University of St Andrews
akc@st-andrews.ac.uk @annakclements

Valerie McCutcheon

Research Outputs Manager, University of Glasgow
Valerie.McCutcheon@glasgow.ac.uk

Research Data meets Research Information:
Pure CERIF-CRIS & Eprints with CERIF extensions



“ ... that there had been "a consistent pattern of failing to display the **proper degree of openness**". Scientists also failed to appreciate the risk their **lack of transparency** posed to the university and "indeed to the **credibility** of UK climate science"

[The Guardian, Thursday 8 July 2010](#)

RSC | Advancing the Chemical Sciences

Publishing Platform | ChemSpider | MyRSC |

chemistryworld



ONLINE MEDIA AWARDS 2013

FINALIST

HOME NEWS OPINIONS FEATURES REGULARS JOBS PODCASTS

RESEARCH BUSINESS NEWS AND ANALYSIS

Call to arms on data integrity

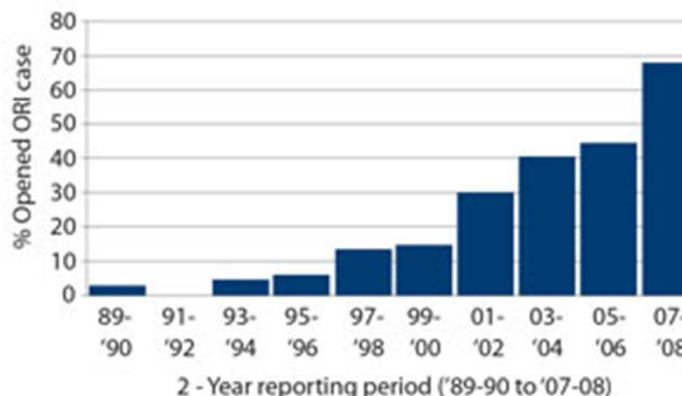
18 July 2013 Patrick Walter

Like 30 Tweet 27 +1 5

Data integrity in the chemical sciences is under the spotlight. In a broadside aimed at shoddy data practices, the editor of *Organic Letters*, [Amos Smith](#), has called on chemists to put their house in order and root out poor practice. But with the focus now on this murky issue the question remains just how widespread manipulation of raw data actually is in the chemical sciences?

Smith declined to be interviewed by *Chemistry World* but in [his editorial](#)¹ he explains that his journal has hired an analyst to scrutinise supplementary data more closely. As a result a number of instances have been discovered where raw data, such as spectra, have been 'cleaned up', in some cases to remove evidence of impurities.

ORI Cases with questioned images



The number of cases of image manipulation brought before the US Office of Research Integrity has been growing steadily



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Volume 483 > Issue 7391 > Comment > Article

NATURE | COMMENT

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Affiliations | Corresponding author

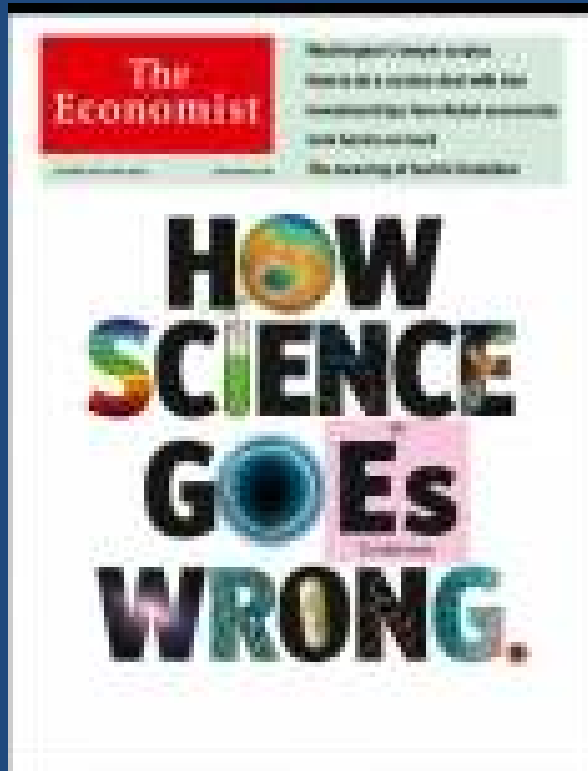
Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

Published online 28 March 2012

[Clarification \(May, 2012\)](#)

“Unquestionably, a significant contributor to failure in oncology trials is the quality of published preclinical data.”

“scientific findings were confirmed in only 6 (11%) cases. Even knowing the limitations of preclinical research, this was a shocking result.”



‘ ... three-quarters of published scientific papers in the field of machine learning are bunk ...’, says Sandy Pentland, a computer scientist at MIT’

Economist, Oct 19th, 2013



“ Publishing data in a reusable form to support findings must be mandatory “



Policy paper

G8 Open Data Charter and Technical

Governance



Can Build New Products

President Barack Obama signed an order Thursday requiring federal agencies to make their publicly accessible data open and machine-readable. Obama's order, which enforces a new Open Data Policy, was issued in hopes entrepreneurs will use government d...

1.5K SHARES



UK Open Access policies

- RCUK policy from 01/04/2013

Papers should also include a statement on how the underlying research materials - such as data, samples or models – can be accessed. It is not necessary that the data itself be openly accessible if there are compelling reasons against this

- REF2020

To be eligible for the next Research Excellence Framework, peer-reviewed manuscripts must be deposited in an institutional or subject repository on acceptance for publication. These requirements apply to all journal articles and conference proceedings accepted for publication after 1st April 2016.

Didn't go as far as saying the data underlying these publications should be open.



Research Policy



Research Careers



Public Engagement



Knowledge Exchange and Impact



International



Press and Media



Publications



About

Home

Research Policy

Eligibility for Research Council funding

How to apply for research funding

Diversity

Areas of Research

Cross-Council Research

Research Infrastructure

Applications across Research Council remits

Principles for Funding Multi-Institutional Collaboration

Efficiency & effectiveness

Changes to the way Research Councils process grants

Peer Review

Research Integrity

Grant Terms and Conditions

Policy on Open Access

Research Outcomes

RCUK Common Principles on Data Policy

Gateway to Research

[Home](#) > [Research Policy](#) > [RCUK Common Principles on Data Policy](#)

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.



 Full Coverage
  Partial Coverage
  No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC											
BBSRC											
CRUK											
EPSRC											
ESRC											
MRC											
NERC											
STFC											
Wellcome Trust											

<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>



EPSRC framework

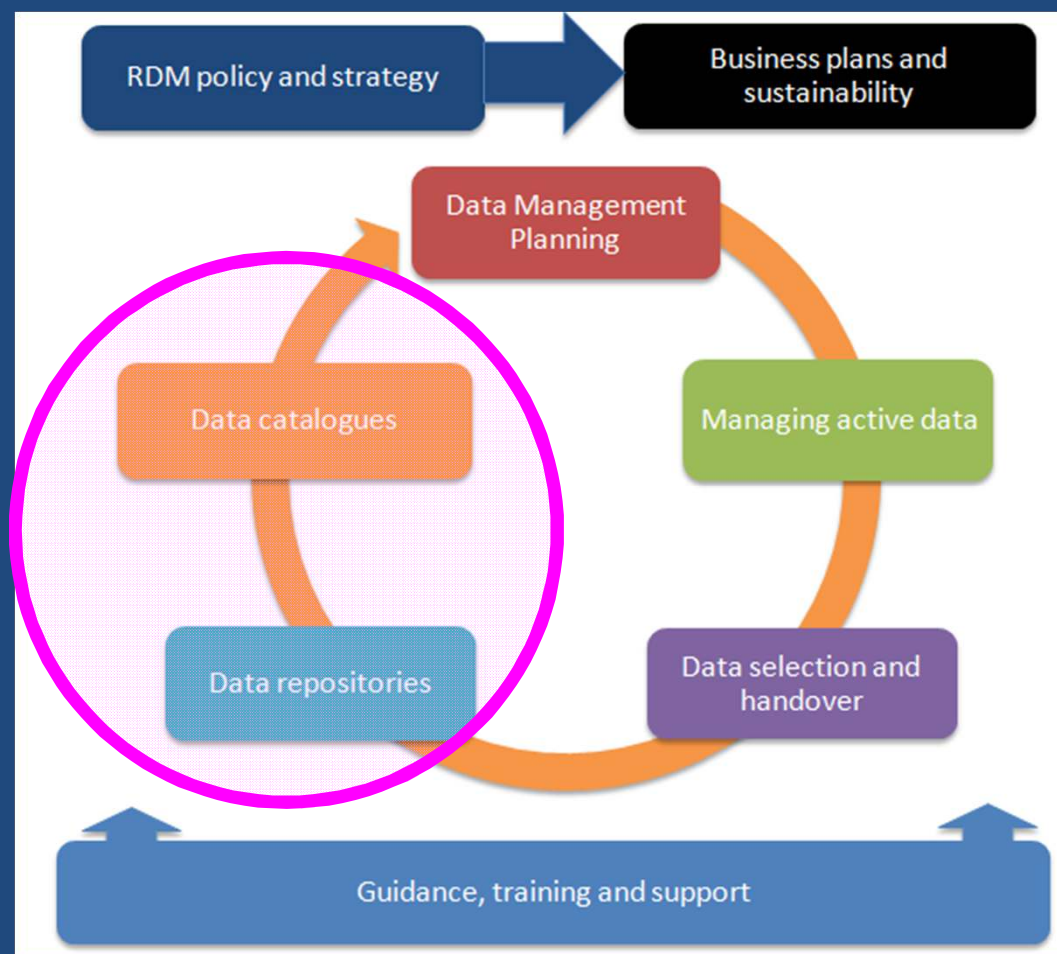
- By **May 2015**, EPSRC has 9 clear *expectations* of **organisations** receiving EPSRC funding, including :
- ‘.. that ***appropriately structured metadata*** describing the research data they hold is published (normally within **12 months** of the data being generated) and made **freely accessible on the internet**; in each case the metadata ***must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it.***’
- Where access to the ***data is restricted*** the published metadata should also give the **reason** and summarise the conditions which must be satisfied for access to be granted. For example ‘commercially confidential’ data, in which a business organisation has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement.
- ‘... that EPSRC-funded research data is ***securely preserved for a minimum of 10-years from*** the date that any researcher ‘privileged access’ period expires or, if others have accessed the data, from ***last date on which access*** to the data was requested by a third party; all reasonable steps will be take to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK



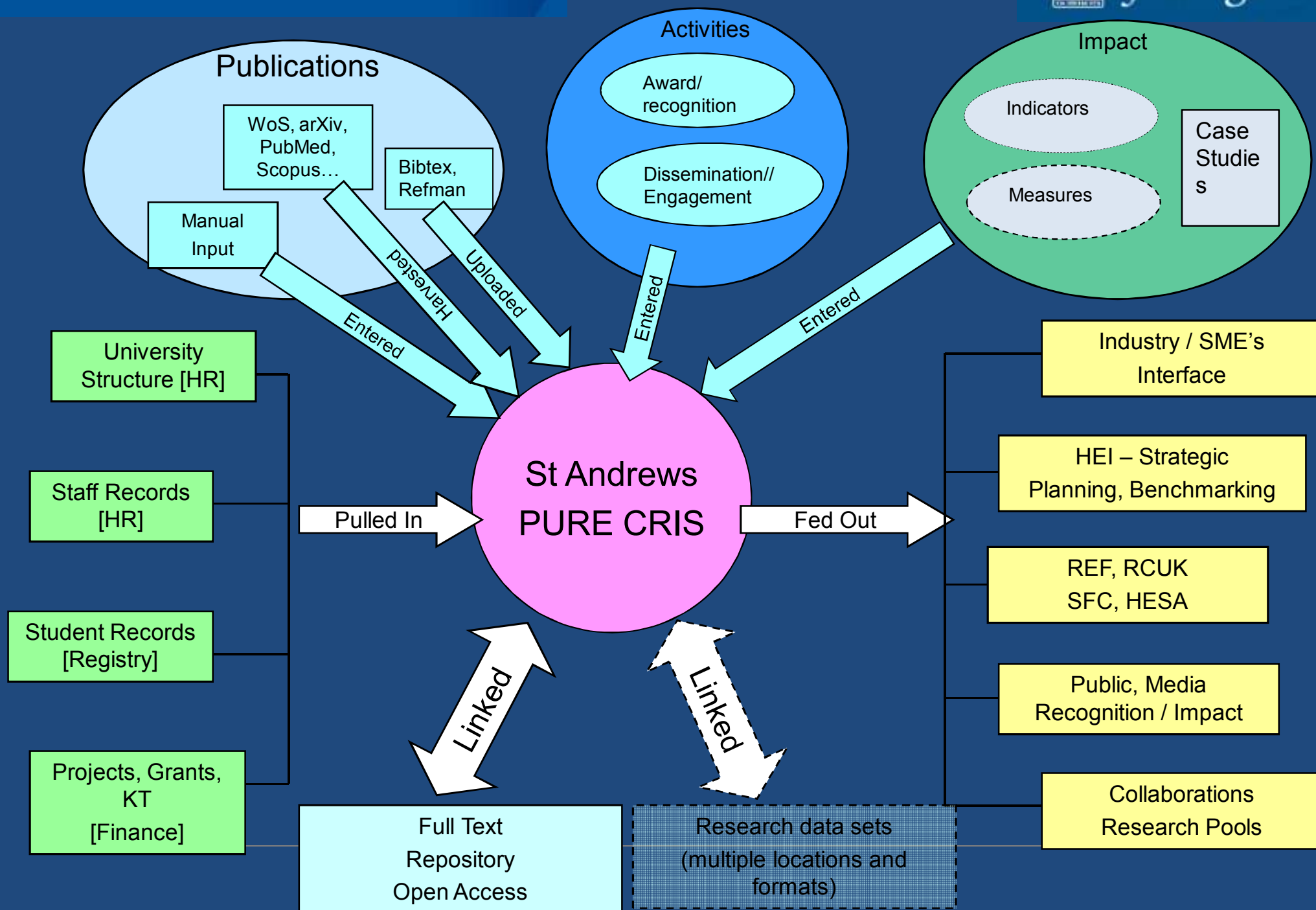
So ... managing research data is important for ...

- Future funding
- Increased impact/citation of research
- Scientific integrity (reproducibility)
- Re-use and sharing of data -> innovation
- Aligned with University strategy
- Future national assessment exercises – beyond REF2020

So what have we been doing?



Components of a data management service : Digital Curation Centre:
<http://www.dcc.ac.uk/resources/how-guides>



Research@StAndrews FullText



Search Repository

Go

[Advanced Search](#)

[Home](#)

Browse

[Communities & Collections](#)

[Issue Date](#)

[Names](#)

[Title](#)

[Subject](#)

[Classification](#)

Sign on to:

[Receive email updates](#)

[My Repository](#)
authorized users

[Help](#)

[About DSpace](#)

[Library Home](#)

[University Home](#)

[Research at St Andrews](#)

[Research@StAndrews:FullText](#) >
[University of St Andrews Research](#) >
[University of St Andrews Research](#) >
[University of St Andrews Research](#) >

Please use this identifier to cite or link to this item: <http://hdl.handle.net/10023/4191>

This item has been viewed **132** times in the last year. [View Statistics](#)

Files in This Item:

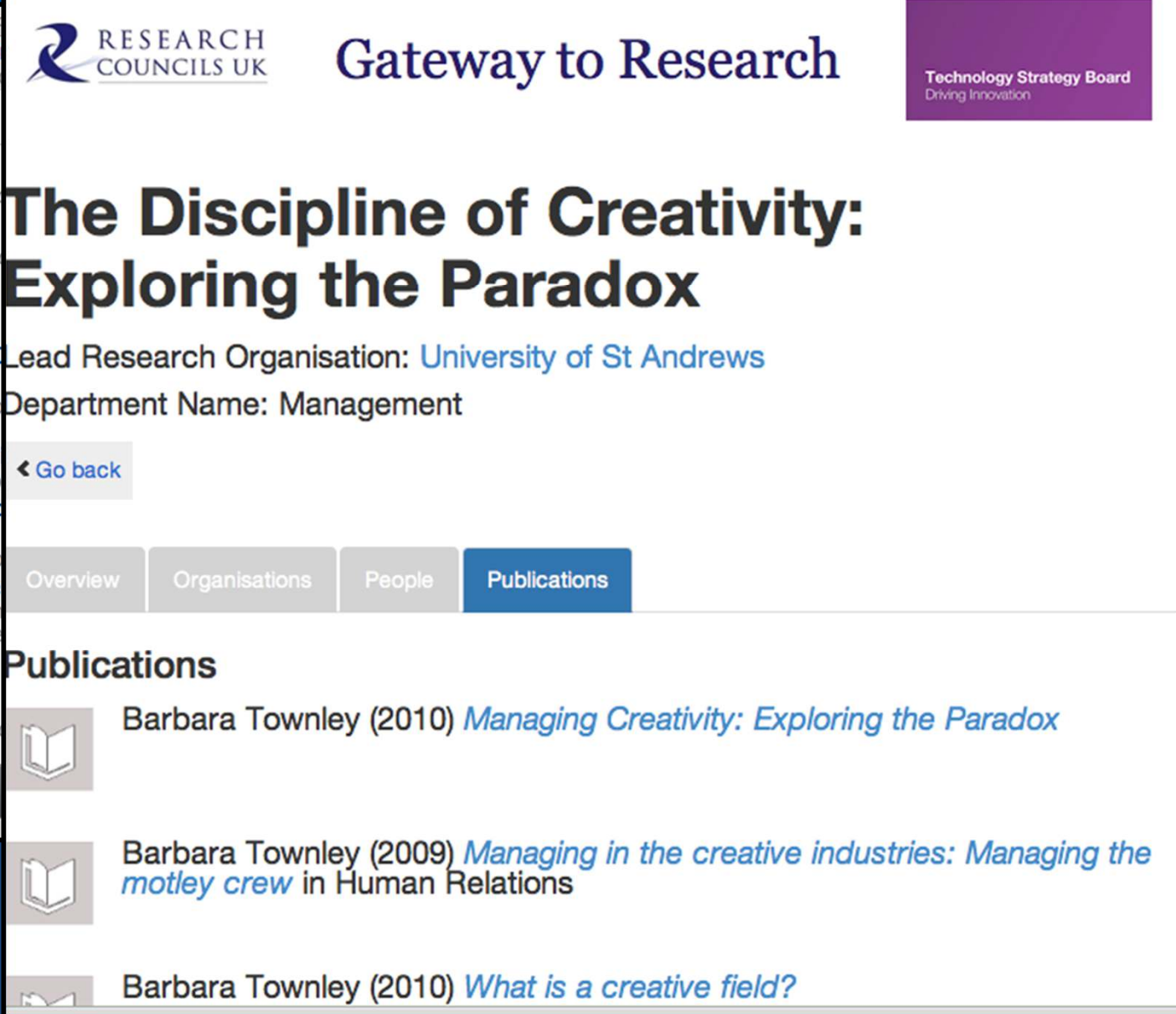
File	Description	Size	Format
digitalheritage2013_submission_536.pdf		611.85 kB	Adobe PDF View/Open

Title: Exploring Canons and Cathedrals with Open Virtual Worlds : The Recreation of St Andrews Cathedral, St Andrews Day, 1318

Authors: [Kennedy, Sarah Elizabeth](#)
[Fawcett, Richard](#)
[Miller, Alan Henry David](#)
[Sweetman, Rebecca Jane](#)
[Dow, Lisa](#)
[Campbell, Anne](#)
[Oliver, Iain Angus](#)
[McCaffery, John Philip](#)
[Allison, Colin](#)

Issue Date: Oct-2013

Citation: Kennedy , S E , Fawcett , R , Miller , A H D , Sweetman , R J , Dow , L , Campbell





EaStCHEM

The Edinburgh and St Andrews Research School of Chemistry



EaStCHEM

The Edinburgh and St Andrews Research School of Chemistry

Home Research Areas Impact Facilities Graduate School About Us People News and Events Links

[Edinburgh Staff](#)

[St Andrews Staff](#)

Robert Alan Aitken



Senior Lecturer

Phone

+44 (0)1334 463865/3846

Email

raa@st-andrews.ac.uk

Research Overview

Synthetic and mechanistic organic chemistry including heterocyclic chemistry, synthetic use of flash vacuum pyrolysis, asymmetric synthesis, reactive intermediates, organophosphorus and organosulfur chemistry

Publications

2014

[The X-ray Structure of N-\(Acetylamino\)phthalimide, an Atypical Triacylhydrazine](#)

Journal of Chemical Crystallography

Kati Marianne Aitken, R Alan Aitken, Alexandra Martha Zoya Slawin

[Formation of unexpected heterocyclic products from pyrolysis of thiocarbonyl stabilised phosphonium ylides](#)

Phone

+44 (0)1334
463865/3846

Email

raa@st-andrews.ac.uk

+44 (0)1334
463779

sema@st-andrews.ac.uk

+44 (0)1334
467236

cjb14@st-andrews.ac.uk

+44 (0)1334
463899

rtb5@st-andrews.ac.uk

beb2@st-andrews.ac.uk

+44 (0)1334
467195

cb2@st-andrews.ac.uk

+44 (0)1334
467232

mb45@st-andrews.ac.uk

+44 (0)1334
467235

mb105@st-andrews.ac.uk

www.eastchem.ac.uk

<http://phyesta.supa.ac.uk>

CRIS2014, 12th May 2014

Enlighten

Enlighten Home

Enlighten Blog

About Enlighten

Latest Additions

Download Statistics

Search

Simple

Advanced

Theses & Other Material

Top 100 Searches

Browse Papers by

Year

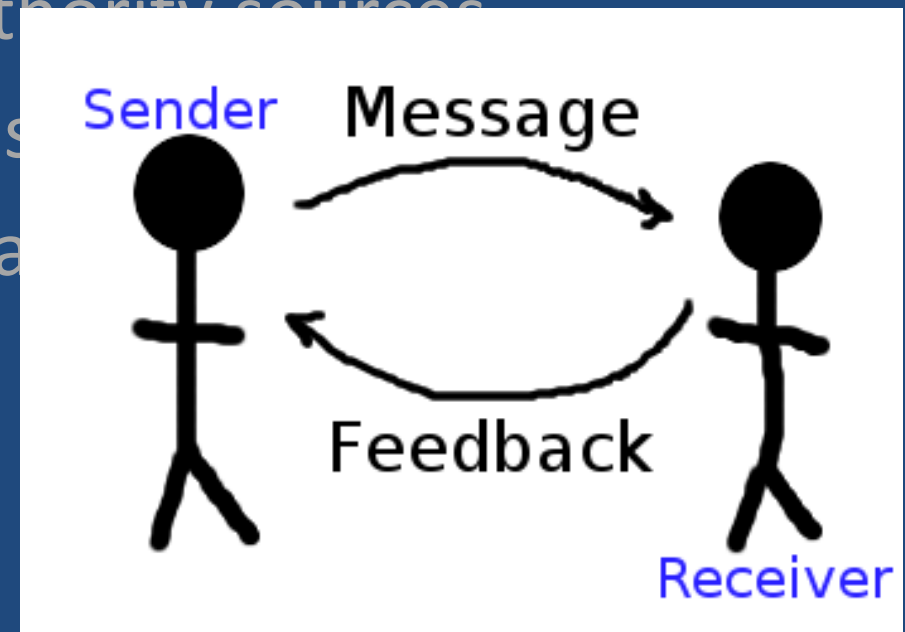
Subject

Funder and Project Information

Project Award Code	Project No	Project Name	Principal Investigator	Funder's Name	Funder Ref	Lead Dept
55242	1	Monolithic Resonant TeraHertz Detectors	David Cumming	Engineering & Physical Sciences Research Council (EPSRC)	EP/I017461/1	ENG - ENGINEERING ELECTRONICS & NANO ENG
58833	1	Triple wavelength superspectral camera focal-plane array (SUPERCAMERA)	David Cumming	Engineering & Physical Sciences Research Council (EPSRC)	EP/J018678/1	ENG - ENGINEERING ELECTRONICS & NANO ENG
58833	2	Triple wavelength superspectral camera focal-plane array (SUPERCAMERA)	David Cumming	Engineering & Physical Sciences Research Council (EPSRC)	EP/J018678/1	ENG - ENGINEERING ELECTRONICS & NANO ENG

Common approach

- Pragmatic
 - Joining up data silos and reusing data
 - Using standards and authority sources
 - Using carrots as well as sticks
 - Demonstrating added value to managers and funders
- Communication is key





Adding datasets : extending our systems

- Institutional dataset register/catalogue
- Metadata harmonisation
 - Cerif for Datasets * Datacite * Pure & EPrints UG
- UK National data registry pilot
 - JISC * DCC * Institutions * Funders * Data Centres
 - CASRAI-UK working group

EPrintsUserGroup_RDMetadataReference ☆

File Edit View Insert Format Data Tools Help All changes saved in Drive

Comments Share

\$ % 123 Arial 10 B I A

	A	B	C	D	E	F	G
5	ReCollect field	Description	Description of the dataset.	There is not consensus whether this is optional or mandatory	Description of the dataset - most likely an abstract	Description of the dataset - most likely an abstract	DataCite Metadata 3.0
6	ReCollect field	Keywords	Keyword or key phrase describing the resource.		Delimited list of keywords or key phrases describing the resource.	Delimited list of keywords or key phrases describing the resource.	Based on DataCite Metadata 3.0
7	ReCollect field	Subject	Subject or classification code describing the resource.	Should we recommend standard ones? Does DC and national catalogue have anything to say about this?	Subject or classification code describing the resource chosen from one or more authoritative sources offered as drop down lists. E.g. HESA, JACS, RCUK	Subject or classification code describing the resource chosen from one or more authoritative sources offered as drop down lists. E.g. HESA, JACS, RCUK	Based on DataCite Metadata 3.0



Id: 63692377

EDIT

Metadata

OVERVIEW

Relations

Display

HISTORY AND COMMENTS

History and comments

Population estimates by single year of age, sex and ethn

Dataset

Managing organisational unit*

Geography & Sustainable Development

Organisational unit: Department

Change organisation...

Date made available

Date made available

Year* Month Day

2009

Publisher

Id: 63692377

Population estimates by single year of age, sex and ethn

Dataset

EDIT

Metadata

OVERVIEW

Relations

Display

HISTORY AND COMMENTS

History and comments

6044userguide

6044userguide.pdf, 179 KB

Added by: jep10, 4 Apr 2014 12:28

Public - No restriction

Show

Edit

+ ↑ -

read6044

read6044.htm, 9 KB

Added by: jep10, 4 Apr 2014 12:28

Public - No restriction

Show

Edit

+ ↑ -

Add document...

Access contact details

Contact persons

• Sabater, Albert

albert.sabater@st-andrews.ac.uk

Geography & Sustainable Development

Person: Academic - Research

Edit

Add contact person...

Relations to other content

Projects

There are no associations

+ ⊞

Equipment

There are no associations

+ ⊞

Publications

Enhancing the Population Census: A Time Series for Sub-National Areas with Age, Sex and Ethnic Group Dimensions in England and Wales, 1991-2001

Research output: Contribution to journal > Article

NOTIFICATIONS

Last saved: 4/04/1...

Save

e Cathie Marsh
yright material is
f HMSO and the

Show Edit

CRIS2014, 12th May 2014



Id: 63692377

EDIT

Metadata

OVERVIEW

Relations

Display

HISTORY AND COMMENTS

History and comments



Population estimates by single year of age, sex and ethn

Dataset

Managing organisational unit*

Geography & Sustainable Development

Organisational unit: Department

[Change organisation...](#)

Date made available ⓘ

Date made available

Year* Month Day

2009

Publisher ⓘ

UK Data Service Discover » Population Estimates by Single Year of Age, Sex and Ethnic Group for Council Areas in Scotland, 1991-2001

discover.ukdataservice.ac.uk/catalogue/?sn=6044&type=Data%20catalogue

Dataset - Pure 4.18.3

UK Data Service Discover » Population Estimates by Single Year of Age, Sex and Et...

Site Search FAQ Help Contact

UK Data Service

About us Get data Use data Manage data Deposit data

Discover > Catalogue

Catalogue

UK Data Service data catalogue record for:

Population Estimates by Single Year of Age, Sex and Ethnic Group for Council Areas in Scotland, 1991-2001

Documentation | Related Studies | Publications | Download/Order | Get full DDI XML

TITLE DETAILS

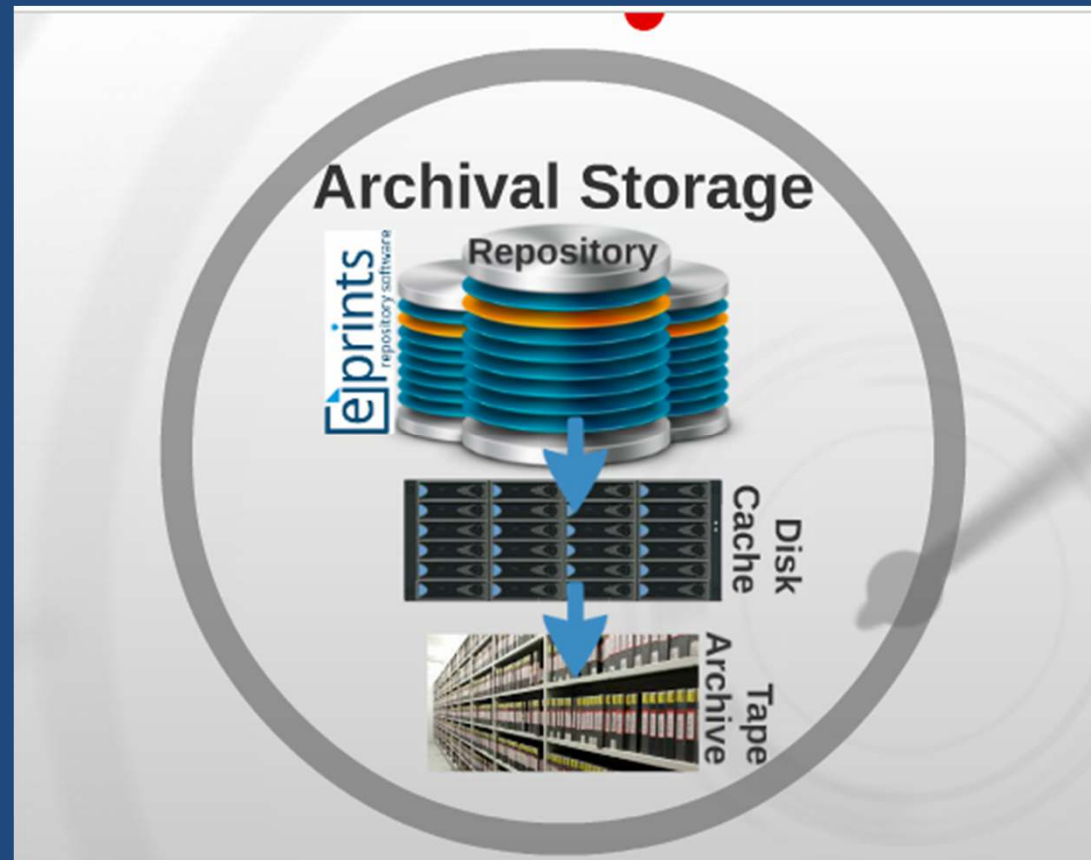
SN:	6044
Title:	Population Estimates by Single Year of Age, Sex and Ethnic Group for Council Areas in Scotland, 1991-2001
Persistent identifier:	10.5255/UKDA-SN-6044-1

e Marsh
material is
and the

w Edit

IS2014, 12th May 2014

Pure as a data repository?



Tim Banks, University of Leeds

<http://prezi.com/3xkzp8x8stfl/university-of-leeds/#>



Search for datasets

Items where Author is "Cresswell, Alan"

[Up a level](#)

Export as

Export



Atom



RSS 1.0



RSS 2.0

Group by: **Dataset Type** | [No Grouping](#)

Jump to: [Dataset](#)

Number of items: **1**.

Dataset

Sanderson, David and Cresswell, Alan and Seitz, Bjoern and Yamaguchi, K. and Takase, T. and Kawatsu, K. and Suzuki, C. and Sasaki, M. (2013) [*Validated radiometric mapping in 2012 of areas in Japan affected by the Fukushima-Daiichi nuclear accident*](#).
[Dataset]

Dataset record

Validated radiometric mapping in 2012 of areas in Japan affected by the Fukushima-Daiichi nuclear accident

Sanderson, David and Cresswell, Alan and Seitz, Bjorn and Yamaguchi, K. and Takase, T. and Kawatsu, K. and Suzuki, C. and Sasaki, M. (2013) *Validated radiometric mapping in 2012 of areas in Japan affected by the Fukushima-Daiichi nuclear accident*. [Dataset]



Archive

ValidatedRadiometricMappingData.zip

[Download \(27Mb\)](#)

Description

On March 11 2011 the north-eastern region of Japan was hit by a magnitude 9 earthquake, which occurred underneath the sea-bed 70 km east of the Oshika peninsula in Tohoku. The north-eastern shore of Honshu was hit by a tsunami resulting from this earthquake.

Edit item: Dataset #72

[Details](#)[DOI](#)[Subjects](#)[Funding](#)[Upload](#)[Deposit](#)[Save and Return](#)[Cancel](#)[Next >](#)

Title



The title of the item. The title should not end with a full stop, but may end with a question mark. There is no way to make italic text, please enter it normally. If you have a subtitle, it should be preceded with a colon [:]. Use capitals only for the first word and for proper nouns.

Example: **A brief history of time**

Example: **Life: an unauthorised biography**

Example: **Mathematics for engineers and scientists. 5th edition**

Example: **Ecosystems of the world. Vol. 26. Estuaries of the world**

Description



Creators



	Family Name	Given Name / Initials	Email	GUID	
1.					
2.					

Dataset Details

**URL of dataset
(if not
uploaded):**

If the resource is available online, provide a web address (URL) that links to the resource.

**Software
Required:**

Software required to open/use dataset.

Media of Output:

The media of output.

Example: **CD-ROM**

Example: **DVD**

Related URLs:

URLs that are related to this item.

URL

URL Type

UNSPECIFIED ▼

More input rows

Publisher:

A person, firm or corporate body responsible for placing an item on the market. Do not use abbreviations, enter in full. This will usually be **University of Glasgow**

**Publication
Date:**

The date when the data was or will be made publicly available. If exact date is not available please fill in the year.

Year: Month: **Unspecified** ▼ Day: **?** ▼

★ Restrictions

★ **Limitations on access:**





Digital object identifier DOI

Edit item: Cerif4Datasets

Details → **DOI** → Subjects → Funding → Upload → Deposit

< Previous Save and Return Cancel Next >

Datacite DOI

DOI from Datacite.org

If you do not know what a DOI is or what Datacite is, please refer to the Datacite DOI section of the About page.

Format to use is doi:xx.xxxx/xxxxxxx

< Previous Save and Return Cancel Next >



Relate dataset to awards

Edit item: test

Details



DOI



Subjects



Funding



Deposit

< Previous

Save and Return

Cancel

Next >

Funding Details

Information on research funding.

Project Code

Award No

Project Name

Principal Investigator

Fu

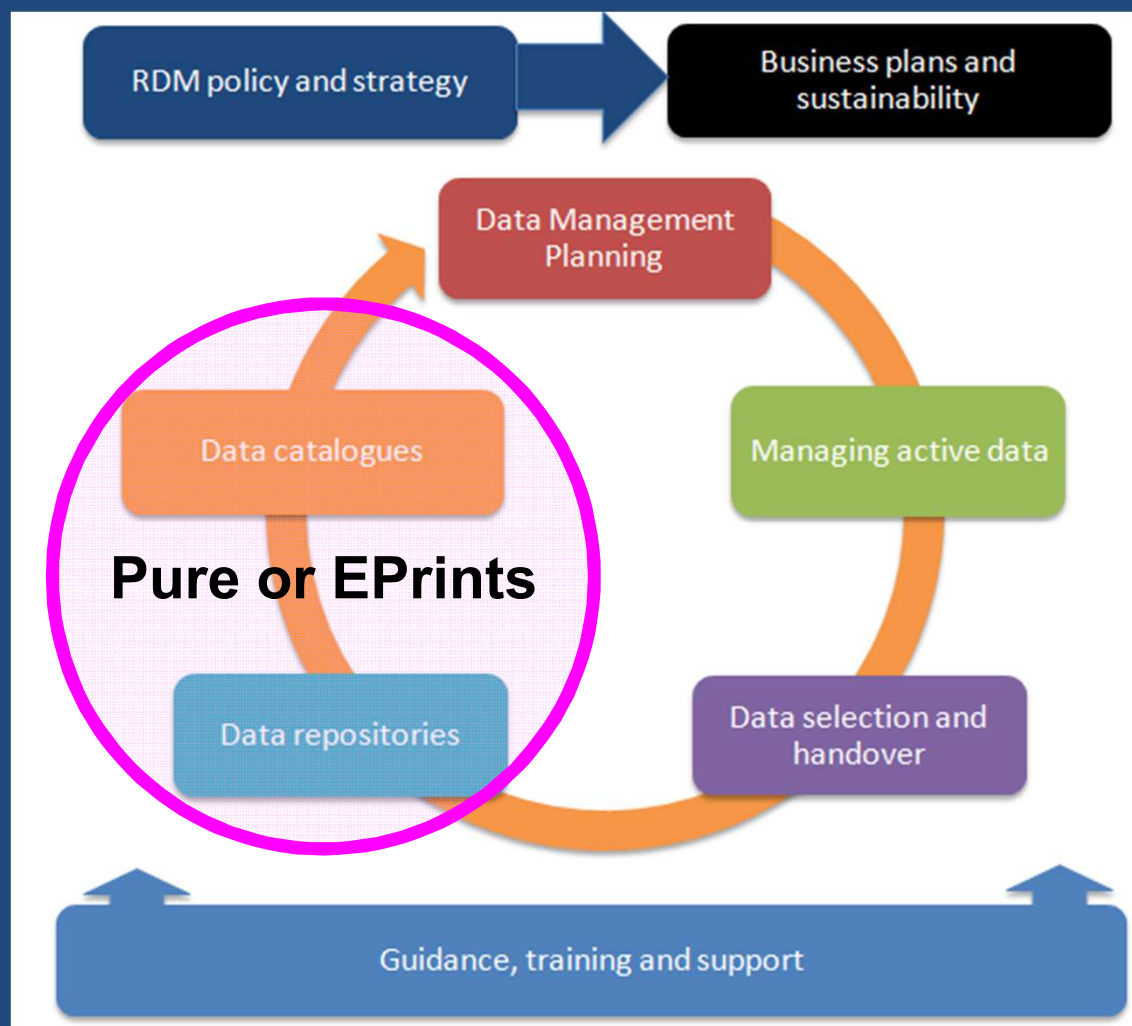
1.

2.

More input rows



So what have we been doing?



Components of a data management service : Digital Curation Centre:
<http://www.dcc.ac.uk/resources/how-guides>



Challenges

- Data is much more heterogeneous than publications
 - How it is manifested
 - How it is created and reused
 - How embedded within the research process
 - Not just picking up a pdf & standard bibliographic metadata at the end of the process
-



Academic-led review

“We need to avoid the idea of a data repository as “a place where data goes to die”.

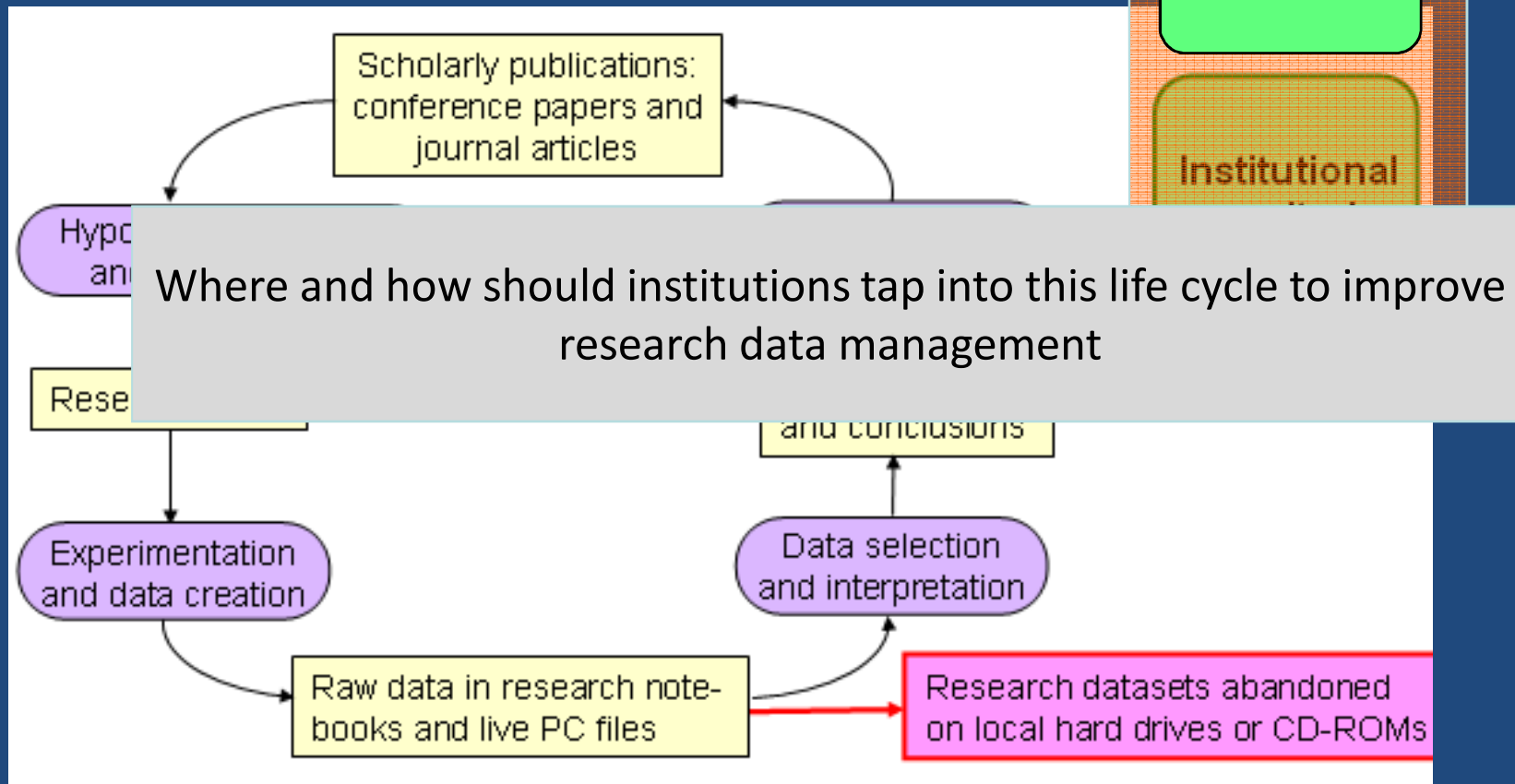
“.. should be designed to act as a hub to encourage and facilitate cross-School interactions data science ... ‘

“ .. broadening accessibility means that datasets must also be suitably annotated so that they can be found easily from simple keyword associations (“tagging”), through to structured metadata such as CERIF, up to full linked open metadata.”

“ Many subject areas already have multiple options for data storage, which need to be integrated rather than competed with.”



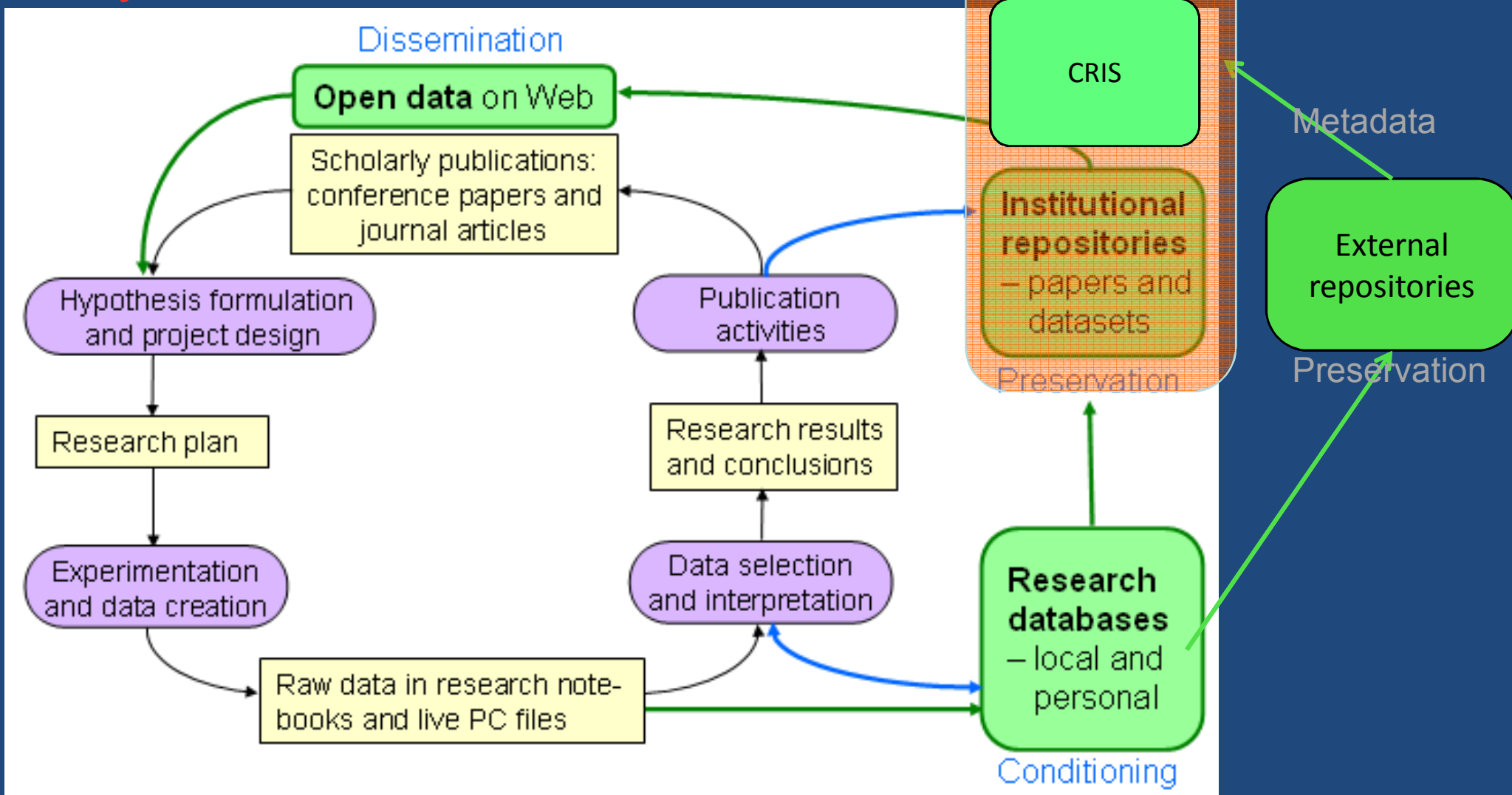
Conventional research life cycle



David Shotton, University of Oxford, 2011, JISC ADMIRAL project
<http://opencitations.wordpress.com/2011/08/04/the-plate-tectonics-of-research-data-publication/>



Enhanced research life cycle



David Shotton, University of Oxford, 2011, JISC ADMIRAL project

<http://opencitations.wordpress.com/2011/08/04/the-plate-tectonics-of-research-data-publication/>



Conclusion

Researchers should be in the driving seat & some already are.

We can provide tools, infrastructure & standards ... & carrots (or sticks) along the way ...

BUT

We can't do the research!

“We need to avoid the idea of a data repository as “a place where data goes to die”.

Research Data Management Infrastructure

Active Data



Managed Storage for Active Data

including:

- HPC Attached
- Generic
- High Performance
- Scratch
- Highly Secure

Archive Data

Rich discipline specific metadata catalogues



Research Data
Repository

Repository Data
Disc Cache

Repository Data
Archive Storage

25 years
Archive as a Service
POSF
Open Plan Storage Framework

Tim Banks, University of Leeds

http://prezi.com/jlrw_nt5igdw/research-data-management-storage/



University of St Andrews

600 YEARS
1413 – 2013



University
of Glasgow

Anna Clements

Head of Research Data & Information Services,
University of St Andrews
akc@st-andrews.ac.uk @annakclements

Valerie McCutcheon

Research Outputs Manager, University of Glasgow
Valerie.McCutcheon@glasgow.ac.uk

Research Data meets Research Information:
Thank you for listening