

Course for Doctoral Students

RESEARCH DATA MANAGEMENT AND OPEN DATA

23th July 2015, Social Science Data Arhives, Faculty of Social Sciences, University of Ljubljana

ECPR Summer School 2015

RESEARCH DATA MANAGEMENT PLANNING

Laurence Horton, Library, London School of Economics and Political Science



Contents

- 1. What kinds or archives/repositories are there?
- 2. What to look for in a reputable archive/repository
 - Re3data
 - Data Seal of Approval
 - DOIs
- 3. What do archives/repositories want from researchers?
- 4. Licenses: Defining how your data can be used
- 5. Depositing data in an archive
 - UK Data Service ReShare
 - ORCiD

Types

Multidisciplinary

Zenodo Figshare

Institutional

LSE Research Online Project

European Social Survey

Disciplinary

UK Data Archive Slovenian Social Science Data (ADP)





Global registry of repositories across all disciplines



Selecting a data repository: subject







Selecting a data repository: content type







Selecting a data repository: country/region









Top 5 countries for number of research data repositories









Almost all have some form of Open Access



Small percentage have a certification (Data Seal of Approval)



Not many offer Permanent Identifiers





• 88 (7%) of 1260 repositories in re3data have:



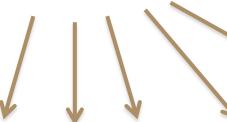
50 come from Humanities and Social Sciences



Data Seal of Approval

- Recognition of being a:
 - trustworthy source of data
 - Long-term preservation
- Self-assessment based on:

• Criteria: 1 2 3 4 5



Guidelines: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16







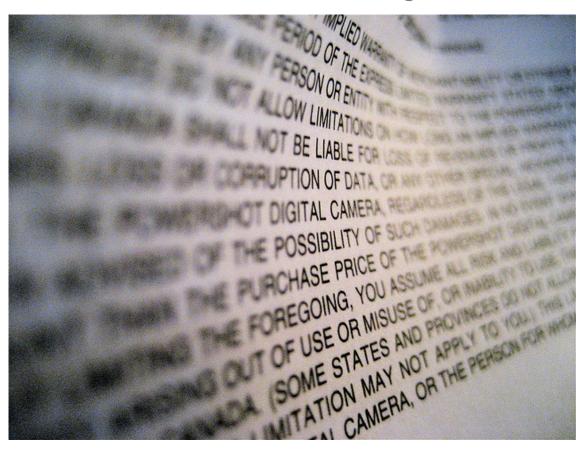
Data can be found on the Internet







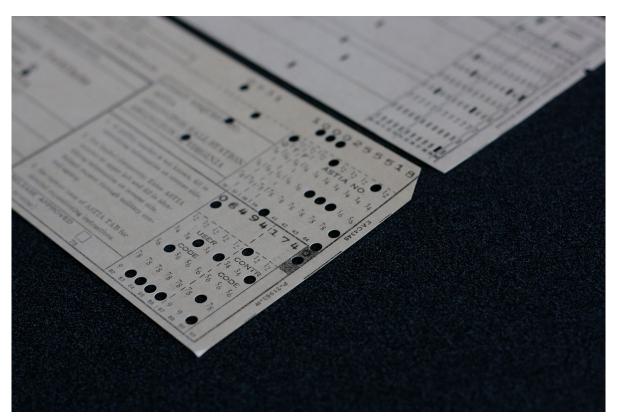
The data are accessible (clear rights and licences)







The data are in a usable format







The data are reliable







The data are identified in a unique and persistent way so they can be cited





DOI: Digital Object Identifier

Permanent identifiers



- A type of identifier based on a fixed and unique reference to an object
- Allows data to be cited
- Not a symbol of data quality
- Obtained from publisher/archive



틀

Criteria for depositing research data

Everything



or

Nothing





Criteria for deposit: Does the data underpin published research findings?





Criteria for deposit: Are people waiting for the data?





Criteria for deposit: Uniqueness





Criteria for deposit: Can data be re-used by others?







Criteria for deposit: Context

Understand context



- . Do the hard work to make it simple . Iterate. Then iterate again.





Licensing data for re-use

Think about how you want others to use it.







License: Clarify Intellectual Property ownership

- You can only archive data you own (or if you have permission)
- If you create it, check you own it

(261) Cup 19

Anno Odavo

Annæ Reginæ.

An Act for the Encouragement of Learning, by Veffing the Copies of Printed Books, in the Authors or Purchafers of fuch Copies, during the Times therein mentioned.



bereas Pinters, Bookellers, and other Perlous babe of late frequerity taken the Liberty of Phinting, Repinting, and Publifting, dyraufing to be Pointto, Repinted, and Publifted Books, and other Edictings, inthout the Conlent of the Authors of Proprietors of into Books, and Edictings, to their bery great Dettinent, and too often to the Ruin-of them and their families: For Persenting, therefore fuch

very great Detruient, and tod otten to the Kuin-of them and their families: For Presenting, therefore furly Pradices to Airby Curry, and to the Kuin-of them and their families: For Presenting, therefore furly Pradices to Airby Curry and the Compose and Airby Compose, and Excellent House, and the Tooks 5 Hope is placed House, since the Airby of English year the Airby and with the Airby of Councils with the Airby and Councils and Crecilent House, and Excellent House, a



License: Clarify funder expectations

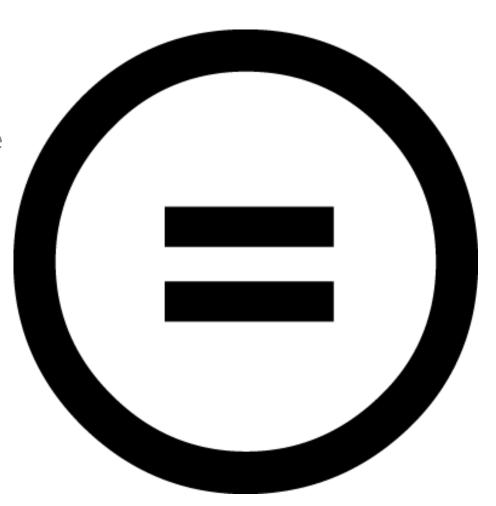
Funder expectations: data to be as open as possible





License: Do you want to allow adaptation and modification?

- Is format and design intrinsic to the work?
- Does the work need to be used in its entirety?





License: Do you want to allow commercial re-use

- Is there a contractual reason for preventing commercial use?
- Will you or your work benefit from or be harmed by commercial usage







Choosing a Creative Commons Licence

CC Licence	Require attribution	Allow sharing	Commercial use	Authorise modification
CC 0 ZERO Public domain	No	Yes	Yes	Yes
Attribution	Yes	Yes	Yes	Yes
Attribution-Share Alike	Yes	Yes	Yes	Yes
Attribution-No Derivatives	Yes	Yes	Yes	No
Attribution-Non Commercial	Yes	Sometimes	No	Yes
Attribution-Non Commercial-Share Alike	Yes	Sometimes	No	Yes
Attribution-Non Commercial-No Derivatives	Yes	Sometimes	No	No





Criteria for deposit: further reading

DCC (2014) 'Five steps to decide what data to keep: a checklist for appraising research data v.1 Edinburgh: Digital Curation Centre









Five steps to decide what data to keep

DCC Checklist for Appraising Research Data

Please cite as: DCC (2014) 'Five steps to decide what data to keep: a checklist for appraising research data v.1 Edinburgh: Digital Curation Centre, Available online: www.dcc.ac.uk/resources/how-guides



Digital Curation Centre, October 2014

This work is licensed under Creative Commons Attribution BY 2.5 Scotland except Section 4, which is adapted under licence CC-BY-NC-SA. from UK Data Archive (2013) Data management costing tool and checklist. Available at: www.data-archive.ac.uk/create-manage/planning-for-sharing/costing

UK Data Service (2014) Collections Development Selection and Appraisal Criteria

CD234-Collections-Appraisal_01_00w.docx

Criteria relating to potential usability and accessibility:	
Able to provide access to the data in ways that make it useful/usable	
Data that have been prepared for access with definable conditions within the access spectrum	
Legal and ethical agreements and protocols used do not prohibit sharing and/or re-use, e.g. consent to share/reuse data, IPR, copyright, DPA	
Data has adequate supporting documentation/ contextual information	
Data are produced in a suitable format for reuse (e.g. no specialised software needed or such software made available) or can be converted to a UK Data Service delivery platform	
The cost of conversion to a UK Data Service delivery platform is not prohibitive, or costs can be recouped	
Data are more openly accessible or where a national licensing model will offer considerable savings	
Third party host: is reliable and reputable	
Third party host: is the data likely to change/be updated over time?	
Third party host: suitable preservation processes are used to manage and curate data	
Third party host: a persistent identifier can be made available (robust access)	



http://ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf

http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

Where can you put data?

- UK Data Service ReShare
- GESIS Leibniz Institute for the Social Sciences datorium
- ADP
- Zenodo
- Figshare
- The Dataverse Project





Logged in as Laurence Horton Logout UK Data Service home Help About FAQ Contact Home Legal Review procedures My data Before you start to deposit a new data collection, check these guidelines on how to prepare your data files and documentation: . group your data files where relevant in zip bundles according to their content or file format, to make upload and download easier, e.g. a zip bundle of interview transcripts, a zip bundle of audio files · for large collections, keep a folder structure for the file · check our recommended file formats · give files meaningful names that reflect the file content, avoiding spaces and special characters · check that data files contain no disclosive information; to anonymise: · remove names or use pseudonyms o remove addresses and detailed location, change postcode to postal district o change date of birth to year remove names and disclosive info from 'file properties', e.g. using MS Word Document Inspector beware of hidden tracked changes in text or table files provide essential documentation for the data: o clear variable and value labels o questionnaire form or data dictionary for surveys o consent form and information sheet o methods description o data list of interviews · PDF of website materials Create a new data collection My ReShare data collections Choose an action

Data Collection



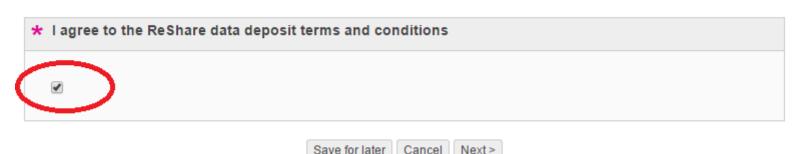


Edit collection: Data Collection #851566



To deposit a data collection, you must accept the ReShare Terms and Conditions.

- I confirm that I am the owner of the copyright and associated intellectual property rights in the whole Data Collection or am otherwise lawfully entitled to grant this licence on behalf of each and every owner;
- I grant a non-exclusive, royalty-free licence to the UK Data Archive (a department of the University of Essex and not a separate
 legal entity) of Wivenhoe Park, Colchester, CO4 3SQ (the "University") to hold, make copies of, and disseminate copies of the
 Data collection, in accordance with the access conditions I will specify when uploading data files: open data accessible to users
 without registration; or safeguarded data accessible to users registered with the data services provided by the UK Data Archive.







Edit collection: Data Collection #851566 Terms and conditions → Award details → People → Data collection → Upload → Deposit Please note: The pink asterisk symbol * indicates a mandatory field Your progress will be automatically saved when you click 'Next' or 'Save'. 'Cancel' will result in the loss of any changes since last If your data collection results from a funded research project step). This is different information from the information you will submit for the data collection. Submit your RCUK grant number (e.g. ES/K000000/1) to retrieve project information from the RCUK Gateway to Research. If grant number is unknown, you can find it by searching the Gateway to Research. Project title; pre-filled if project information is retrieved from the RCUK Gateway to Project title: Research. Project description: The project description or abstract explains the general aims, purpose and background to the research project. This field will be pre-filled if project information is retrieved from the RCUK Gateway to Research. * Research funders: The funding source for the research project; pre-filled if project information is retrieved from the RCUK Gateway to Research. Type N/A if not applicable. 1. Economic and Social Research Council 2.





dit collection	Data Colic	Ction #0515	,000	
erms and conditions	Award details →	People → Data o	collection → Upload	d → Deposit
Creators				
Family name	Given name / Initials	Email	Affiliation	ORCID
1. Horton	Laurence	I.horton@lse.ac.uk	London School of E	0000-0003-2742-6
2.				
Add row				
Copyright holders				
Family name	Given name / Initials	Email	Affiliation	ORCID
1.				~
Add row				
Contact				
Family name	Given name / Initials	Email	Affiliation	ORCID
1.				~
Add row				
Contributors				
Family name	Given name / Initials	Email	Affiliation	ORCID
1.				~
Add row				



ORCID

ORCID

- A unique identifier for individuals (ORCiD iD)
- You can obtain one yourself or maybe through your institution
- Open API code that allows its information to be exported
- Open source, not-for-profit organization
- Increasingly adopted by funders and organizations for capturing publication information



ORCID ID



Connecting Research and Researchers

FOR RESEARCHERS FOR		FOR OR	GANIZATIONS	ABOUT	HELF	SIGN OUT
MY ORCID RECORD	ACCOUN	NT SETTINGS	DEVELOPER TOOLS	LEARN MORE		

1,438,077 ORCID iDs and counting. See more...

Laurence Horton

ORCID ID



D orcid.org/0000-0003-2742-6434

Return to my view

Country

United Kingdom

Keywords

Research Data Management, Digital Preservation

Websites

LinkedIn

Twitter

Biography

Data Librarian at The London School of Economics and Political Science.

▼ Education (I)

IT Sort

University of Essex: Colchester, Essex, United Kingdom

1998-10 to 1999-09 | Master of Arts (Government)

Source: Laurence Horton Created: 2014-01-21

▼ Employment (4)

IT Sort

London School of Economics and Political Science: London, London, United Kingdom

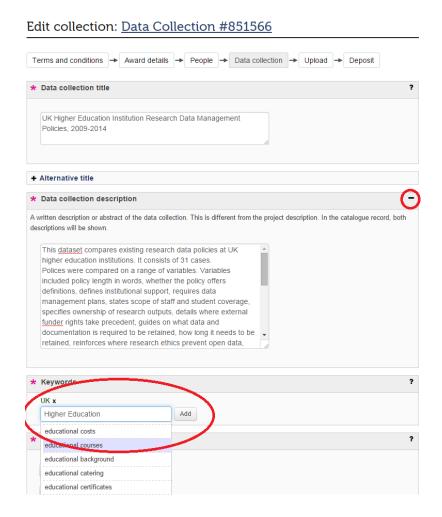
2014-03 to present (Library)

Source: Laurence Horton Created: 2014-01-21

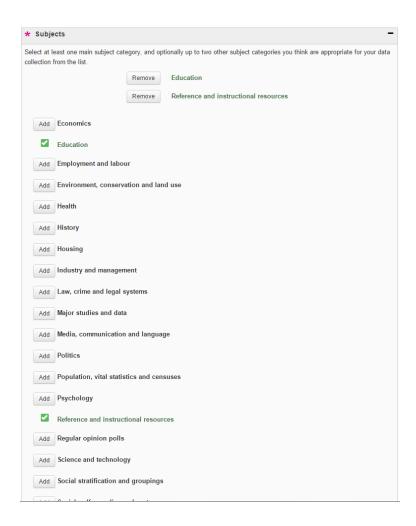
GESIS Leibniz Institut für Sozialwissenschaften Köln: Köln, Nordrhein-











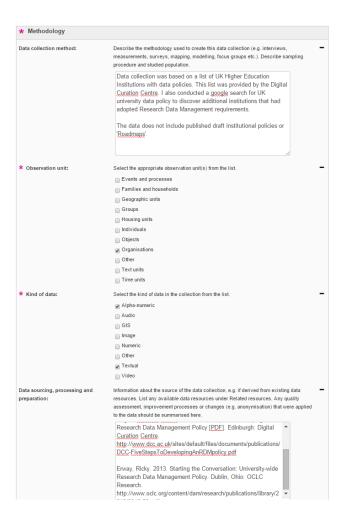


remporar co	verage							
Provide a start a collection.	and end data of either t	he actual data co	llection period, or else the	time period that is	covered by the data			
Collection period:		Start and end date for the data collection period.						
		Date from:	Date to:					
		2014	2014					
Temporal cove	erage:	The time period	(start and end date) for the	nis data collection,	f it is different from th	ne data		
		collection period	d, e.g. for data representin	g a historical perio	d.			
		From	То					
		2009	2014					
* Geograph	ical coverage							
Provide informa	tion about the geograph	í	our data were collected.	ollected.				
Provide informa Geographical area:	tion about the geograph The name(s) of the loc United Kingdom	cation(s) or area(s	s) where your data were co					
Provide informa Geographical area: Bounding	tion about the geograph The name(s) of the loc United Kingdom Define the geographic	cation(s) or area(s	s) where your data were co		de coordinates in deci	mal degrees, as th		
Provide informa	tion about the geograph The name(s) of the loc United Kingdom Define the geographic smallest rectangular sl	cation(s) or area(s al area covered b nape that totally e	s) where your data were co y your data collection by f encloses all of the data.		de coordinates in deci	mal degrees, as th		
Provide informa Geographical area: Bounding	The name(s) of the loc United Kingdom Define the geographic smallest rectangular st Example: 60.85, 2.69,	cation(s) or area(s al area covered b nape that totally e 49.84, -9.23 (bot	s) where your data were co y your data collection by f encloses all of the data. unding area for UK)	our latitude/longitu				
Provide informa Geographical area: Bounding	tion about the geograph The name(s) of the loc United Kingdom Define the geographic smallest rectangular sl	cation(s) or area(s al area covered b nape that totally e 49.84, -9.23 (bot	s) where your data were co y your data collection by f encloses all of the data.			mal degrees, as th Westlongitude		
Provide informa Geographical area: Bounding	The name(s) of the loc United Kingdom Define the geographic smallest rectangular st Example: 60.85, 2.69,	cation(s) or area(s al area covered b nape that totally e 49.84, -9.23 (book	s) where your data were co y your data collection by f encloses all of the data. unding area for UK)	our latitude/longitu South latitu	de	West longitude		

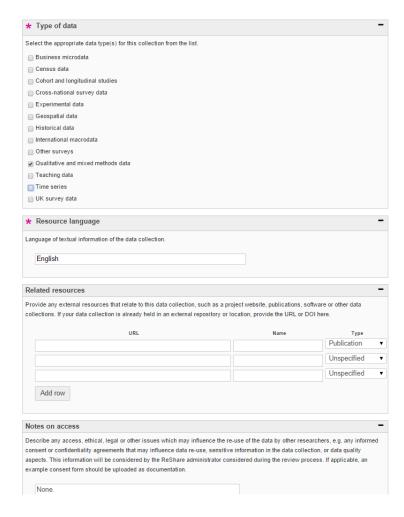


Spatial unit		
Select all spatial units or geographical identifiers for studied individuals. Catego		eference the data collection to a particular location, e.g. the location d to show individual options to select.
	Remove	Counties
	Remove	Countries
	Remove	Districts (Metropolitan)
	Remove	London Boroughs
	Remove	Unitary Authorities (England)
	Remove	Unitary Authorities (Wales)
Add No Spatial Unit + Grid + Postal		
— Alministrative		
Add Communities (Scotland)		
Add Communities (Wales)		
Add Council Areas		
Counties		
Add Counties (Metropolitan)		
Countries		
Add District Council Areas		
☑ Districts (Metropolitan)		
Add Districts (Non-Metropolitan)	
Add Greater London		
Add Local Authority Districts		
London Boroughs		
Add Parishes (Civil)		





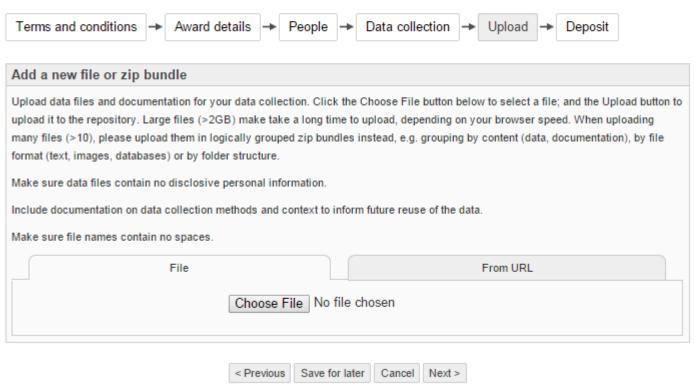






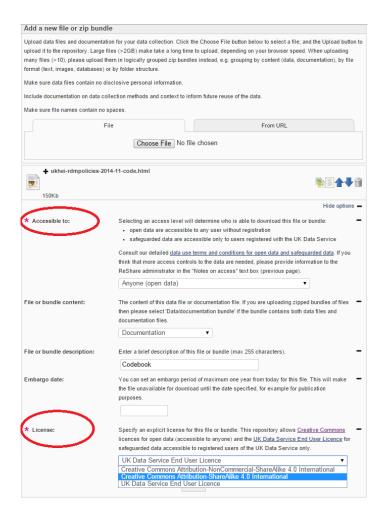


Edit collection: <u>UK Higher Education Institution Research Data</u> Management Policies, 2009-2014











Deposit item: <u>UK Higher Education Institution Research Data</u> <u>Management Policies, 2009-2014</u>



By submitting this data collection I confirm that:

- . I understand and agree to the conditions under which data users will have access to my data collection (link to legal info)
- · data files have been anonymised where needed and contain no disclosive or personal information
- sufficient documentation is uploaded to make this data collection usable for further research
- I understand that a submitted Data Collection will not be visible on the live repository until the UK Data Service will have reviewed it
 for disclosure risk, copyright breaches, validity of file formats and documentation (all numerical data files and at least a 10% random
 sample of textual data files will be reviewed), communicated any concerns to me, and approved it for publication.

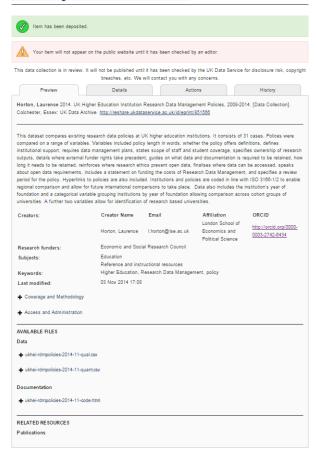
After depositing this data collection you can no longer edit it or upload more data files.

Deposit collection now Save for later





View data collection: <u>UK Higher Education Institution Research</u>
Data Management Policies, 2009-2014







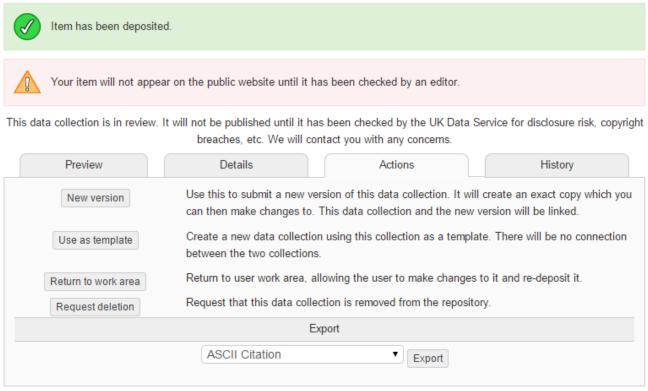
Preview	Details		Actions	History			
Terms and conditions							
I agree to the ReShare data de terms and conditions:	posit Yes						
Award details							
Research funders:	Economic ar	nd Social Research	Council				
Unspecified fields:	Project title,	Project title, Project description, Grant reference, Grant holders, Project dates					
People							
Creators:	Creator Name	Email	Affiliation	ORCID			
	Horton, Laurence	l.horton@lse.ac.ul	London School of Economic and Political Science	http://orcid.org/0000- 0003-2742-6434			
Copyright holders:	Name	Email	Affiliation	ORCID			
	Horton, Laurence	I.horton@lse.ac.ul	London School of Economic and Political Science	http://orcid.org/0000- 0003-2742-8434			
Contact:	Name	Email	Affiliation	ORCID			
	Horton, Laurence	I.horton@lse.ac.ul	London School of Economic and Political Science	http://orcid.org/0000- 0003-2742-8434			
Unspecified fields:	Contributors	Contributors					
Data collection							
Data collection title:	UK Higher E	ducation Institution	Research Data Management	t Policies, 2009-2014			
Data collection description:	This dataset	This dataset compares existing research data policies at UK higher education institutions.					
		It consists of 31 cases. Polices were compared on a range of variables. Variables include					
			the policy offers definitions, o				
			s, states scope of staff and s details where external funder	• • •			
			n is required to be retained, h				
	retained, rei	nforces where resea	arch ethics prevent open data	, finalises where data can b			
			lata requirements, includes a	•			
			ement, and specifies a review				
	Hyperlinks to policies are also included. Institutions and policies are coded in line with ISC 3168-1/2 to enable regional comparison and allow for future international comparisons to						
	3105-1/2 to enable regional comparison and allow for future international comparisons to take place. Data also includes the institution's year of foundation and a categorical						
	variable grouping institutions by year of foundation allowing comparison across cohort						
	groups of universities. A further two variables allow for identification of research based universities.						
		Higher Education, Research Data Management, policy					
Keywords:	Higher Educ	ation, Research Da					
Keywords: Subjects:	Education						
Keywords: Subjects: Type of data:	Education Reference a	ation, Research Da Ind instructional resoluted mixed methods	purces				
Subjects:	Education Reference a	nd instructional reso	purces				





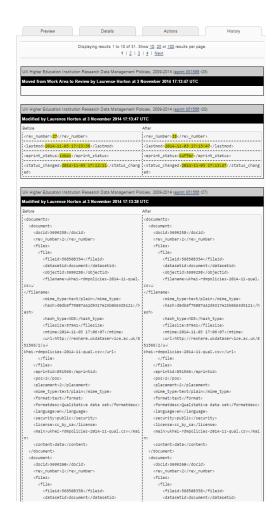
View data collection: <u>UK Higher Education Institution Research</u>

<u>Data Management Policies, 2009-2014</u>











UK Higher Education Institution Research Data Management Policies, 2009-2014

Download

DDI XML

TITLE DETAILS

SN: 851566

Title: UK Higher Education Institution Research Data Management Policies, 2009-2014

Persistent identifier: 10.5255/UKDA-SN-851566

Depositor: Laurence Horton, London School of Economics and Political Science

Principal investigator(s): Laurence Horton, London School of Economics and Political Science

Sponsor(s): Economic and Social Research Council

Grant number: Unknown

SUBJECT CATEGORIES

Education

Reference and instructional resources



Lessons

- There are different types of archives and repositories
- Re3data can help you find a reputable, suitable one
- Data Seal of Approval is an example of a trustworthy archive
- Archives want data with reuse potential: good quality data with meaningful contextual information
- Pick a licence that's appropriate to the data
- Self-archiving platforms exist that allow you to archive your data
- Get an ORCiD



Images

- 5: Logopedia, "Bullseye" http://logos.wikia.com/wiki/Bullseye Content available under CC-BY-SA.
- 6: By Austinmurphy at en.wikipedia [Public domain], from Wikimedia Commons https://commons.wikimedia.org/wiki/File%3AAssorted_video_tapes.JPG
- 7: By National_archives_2007_02_03.JPG:Nick Cooper at en.wikipedia derivative work: Harumphy (National_archives_2007_02_03.JPG) [CC BY-SA 2.5 (http://creativecommons.org/licenses/by-sa/2.5)], from Wikimedia Commons https://commons.wikimedia.org/wiki/File%3ANational_Archives_2007_02_03.JPG
- 12: "Apple3". Licensed under CC BY-SA 3.0 via Wikimedia Commons https://commons.wikimedia.org/wiki/File:Apple3.jpg#/media/File:Apple3.jpg
- 13: CJ Sorg "Fine Print". Licenced under CC-BY-SA 2.0 https://www.flickr.com/photos/cjsorg/366631413
- 14: "Remington Rand punched card.mw" by Marcin Wichary Flickr: Punched cards. Licensed under CC BY 2.0 via Wikimedia Commons
 - https://commons.wikimedia.org/wiki/File:Remington_Rand_punched_card.mw.jpg#/media/File:Remington_Rand_punched_card.mw.jpg
- 15: "Tacoma-narrows-bridge-collapse". Licensed under Public Domain via Wikimedia Commons https://commons.wikimedia.org/wiki/File:Tacoma-narrows-bridge-collapse.jpg#/media/File:Tacoma-narrows-bridge-collapse.jpg
- 16: "Individuality" by VaughnSaball Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons -
- https://commons.wikimedia.org/wiki/File:Individuality.jpg#/media/File:Individuality.jpg



Images

17: "Box.agr". Licensed under CC BY-SA 3.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Box.agr.jpg//media/File:Box.agr.jpg

17: Quinn Dombrowski "Attendee lists" licenced under CC BY-SA 2.0 https://www.flickr.com/photos/quinnanya/3183174451

18: BoneDaddy.P7 "There are 2 things missing in this photo" licenced under CC BY-SA 2.0 https://flic.kr/p/5iuQeU

19: "Bus Queue" by Jimmcgovern15 (talk) - I created this work entirely by myself.. Licensed under CC BY-SA 3.0 via Wikipedia -

https://en.wikipedia.org/wiki/File:Bus_Queue.jpg#/media/File:Bus_Queue.jpg

20: "Snow Flakes" by Dakota Lynch - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Snow_Flakes.jpg/

21: "Reduce, Re-Use, Recyle, (3032678489)" by Marcus Quigmire from Florida, USA - Reduce, Re-Use, Recyle, Uploaded by Princess Mérida. Licensed under CC BY-SA 2.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Reduce,_Re-

Use,_Recyle,_(3032678489).jpg#/media/File:Reduce,_Re-Use,_Recyle,_(3032678489).jpg

22: Paul Downey "Understand context" licensed under CC BY 2.0 https://www.flickr.com/photos/psd/9116635297

23: "SF PERMIT PARKING". Licensed under Public Domain via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:SF_PERMIT_PARKING.svg#/media/File:SF_PERMIT_PARKING.svg



Images

24: "Statute of anne" by British Government - Statute of anne.jpg on the English Wikipedia; original source: The History of Copyright: A Critical Overview With Source Texts in Five Languages. Licensed under Public Domain via Wikimedia Commons -

https://commons.wikimedia.org/wiki/File:Statute_of_anne.jpg#/media/File:Statute_of_anne.jpg

25: Aisletwentytwo "Alec - Mr. Money Bags" CC-BY 2.0 https://flic.kr/p/eZmHq7

