"I'm leaving you my data!" - Practical research data sharing within your institution and the wider community

Dr Jonathan Tedds – University of Leicester - @jtedds
Health And Research Data Informatics, Department of Health Sciences
jat26@le.ac.uk

Elizabeth Newbold – The British Library - @scienceBL datasets@bl.uk

Disclaimer

- No data was lost creating this presentation
- * No researcher or librarian was harmed
- * There may sweeping statements and provocative views presented (that we may or may not agree with!)

Researchers, data and repositories – what's all the fuss about?

 Research has always used and been built on data so why are we "worrying" about it now – surely it's nothing new?

The Recent Total Eclipse of the Sun

J. NORMAN LOCKYER Nature 1869 1(1) 14 doi:10.1038/001014a0

sun there is therefore none of the glare which we usually | forward, both in the observatory and in the laboratory, was see-a glare caused by the reflection of the sun's light by against any such extensive atmosphere as the corona had

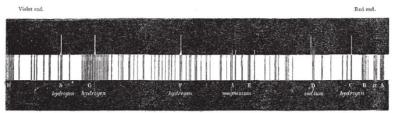


Fig. 1.—Showing the solar spectrum, with the principal Fraunhofer lines, and above it the bright-line spectrum of a prominence containing magnesium, sodium, and iron vapour at its base.

the sun ordinarily hidden from us by this glare, we ought | conclusion to the Royal Society. Since that time, I conto see it during eclipses.

In point of fact, strange things are seen. There is a strange halo of pearly light visible, called the corona, and there are strange red things, which have been called red stronger and stronger; but there was always this considerflames or red prominences, visible nearer the edge of the ation to be borne in mind, namely, that as the spectromoon-or of the sun which lies behind it.

these things revealed to us during eclipses if they be- as the glare might defeat the spectroscope with an unlonged to the sun, it does not follow that they belong to eclipsed sun in the coronal regions, where the temperature the sun because we see them. Halley, a century and a and pressure are lower than in the red-flame region. half ago, was, I believe, the first person to insist that they were appearances due to the moon's atmosphere, and it is only within the last decade that modern science has shown to everybody's satisfaction-by photo- both before and after the eclipse, and observing the prographing them, and showing that they were eclipsed as | minences with the telescope merely during the eclipse, we

our atmosphere. If, then, there were anything surrounding | been imagined to indicate; and we communicated our fess, the conviction that the corona is nothing else than an effect due to the passage of sunlight through our own atmosphere near the moon's place has been growing scopic evidence depends mainly upon the brilliancy of the Now, although we might, as I have pointed out, have lines, that evidence was in a certain sense negative only,

> The great point to be settled then, 'in America, was, What is the corona? and there were many less ones. For instance, by sweeping round the sun with the spectroscope,

A Revision of the Atomic Weight of Cadmium Gregory Paul Baxter and Miner Louis Hartmann PNAS 1915 1 (1) 26-29

ANALYSIS NUMBER	SAMPLE	GRAMS OF CdCl ₂	GRAMS OF Cd	RATIO Cd:Cl2	ATOMIC WEIGHT
1	A	6.08570	3.73181	1.58538	112.426
2	A	4.20489	2.57863	1.58562	112.443
3	A	5.36203	3.28817	1.58553	112 436
4	В	7.50512	4.60221	1.58538	112.426
5	в	6.71591	4.11839	1.58551	112.435
6	В	5.91556	3.62763	1.58555	112.438
7	c	5.49323	3.36805	1.58483	112.387
8	В	3.14416	1.92791	1.58513	112.408
9	c	7.58705	4.65173	1.58474	112.381
10	в	2.26738	1.39006	1.58444	112.359
11	В	5.93501	3.63962	1.58562	112.443
12	F	6.04122	3.70490	1.58578	112.454
13	D	4.07400	2.49821	1.58537	112.425
14	E	9.00004	5.51879	1.58529	112.419
15	G	6.56891	4.02808	1.58534	112.423
16	G	7.12956	4.37174	1.58522	112.414
17	E	8.57291	5.25679	1.58522	112.415
18	G	7.76294	4.76011	1.58521	112.413
Average				1.58529	112.419
Average of analyses 14 to 18				1.58526	112.417

It is a great pleasure to express our gratitude to the Carnegie Institution of Washington and to the Elizabeth Thompson Science Fund for very generous assistance in providing indispensable apparatus.

What's driving the change?

- * Policy pressures
 - * International
 - * National
 - * Institutional
 - * Other?
- * Expectations
 - * Peers
 - * Society



Policy pressure

Guidelines on Data Management in Horizon 2020

Data Management Plans (DMPs) are introduced in the Horizon 2020 Work Programme for 2014-15:

* "A further new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research."

Guidelines on Open Access to Scientific Publications and Research Data In Horizon 2020

* "Open access to research data refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user."

National

Home / Research / 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.

RESEARCH INNOVATION SKILLS NEWS, EVENTS AND

Home About us Service standards and policies

EPSRC policy framework on research data

EPSRC POLICY FRAMEWORK ON RESEARCH DATA

This policy framework sets out EPSRC's expectations \mathscr{O} concerning the management and provision of access to EPSRC-funded research data. EPSRC recognises that a range of institutional policies and practices can satisfy these expectations, and encourages research organisations to develop specific approaches which, while aligned with EPSRC's expectations, are appropriate to their own structures and cultures.

The expectations arise from seven core principles \mathscr{O} which align with the core RCUK principles on data sharing. Two of the principles are of particular importance: firstly, that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data.

The framework was endorsed by the <u>EPSRC</u> Council in March 2011 and implemented from 01 May 2011. It was developed with the benefit of advice from university administrators, from academics, and from research collaborators based in industry.

The policy reflects the principal UK legal provisions intended to assure public access to publicly held information, the most relevant of which to EPSRC-funded research data are contained in the Freedom of Information Act (2000) e^0 and the Freedom of Information (Scotland) Act (2002) e^0 (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Regulations 2004). These Acts allow any person to ask any public authority (including Universities) for any information they believe to be held by that authority, and require the authority to respond in writing stating whether or not they hold the information sought and, if so, to supply that information unless certain exemptions apply. The exemptions, which may be absolute or qualified, generally relate to considerations such as national security, law enforcement,

Institution policies

From the **University of Southampton**: http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html

"The purpose of this policy is to: (i) create model research data management practices for all Researchers (defined below) at the University of Southampton (the University); (ii) foster responsibility for research data management through the development of research data management plans; and (iii) ensure that research data which is significant (as determined by the Researcher with guidance from the Faculty) is stored, retained, accessible and disposed of securely in accordance with all legal, statutory, ethical, contractual and funding requirements.

A robust **Research Data Management Policy** is required to demonstrate and ensure: good research practice and procedures;

- protection of intellectual property rights (IPR);
- * proper recording, maintenance, storage and security of Research Data;
- * compliance with relevant legislation and regulations; and
- * appropriate access to Research Data is maintained.

This policy aims to address the dual objectives of making research data accessible to the public or wider academic community where possible whilst recognising that there are circumstances when access to research data must be restricted, either to achieve the highest standards for secure research data management or to fulfil commercial objectives."

Expectations

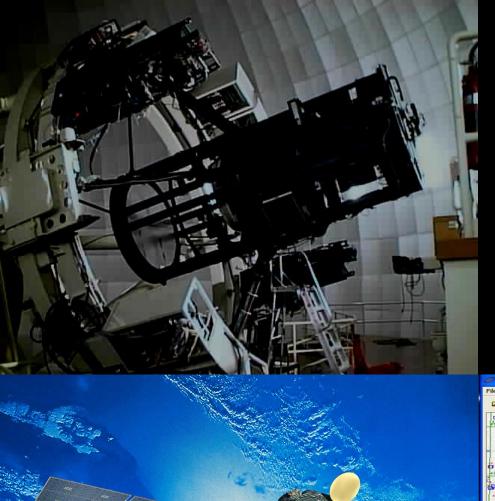
* Royal Society – Science as an Open Enterprise

"The Science as an open enterprise report highlights the need to grapple with the huge deluge of data created by modern technologies in order to preserve the principle of openness and to exploit data in ways that have the potential to create a second open science revolution"

* Open Data: unleashing the potential, June 2012, Cabinet Office

"Data is the 21st century's new raw material. Its value is in holding governments to account; in driving choice and improvements in public services; and in inspiring innovation and enterprise that spurs social and economic growth.

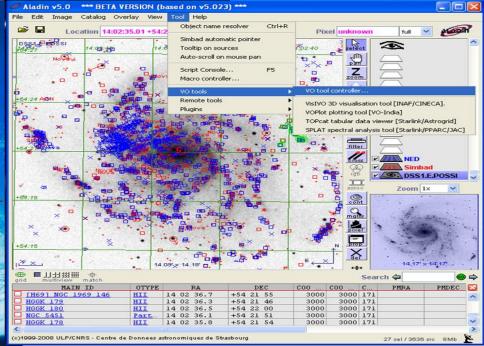
In the last 20 years the world has opened up and citizens across the globe are proclaiming their right to data; this White Paper sets out how we intend to use that energy to unlock the potential of Open Data and for the first time the technology exists to make the demand for greater openness irresistible. We are at the start of a global movement towards transparency – and the UK is leading the world in making data more freely available."











* "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives."

* "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives." - e.g. discipline specific

- * "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives." e.g. discipline specific
- * "Any such scheme needs to allow for a wide variety in the scale and nature of data-generating research projects, and the resources (especially staff) available, i.e. one-size doesn't fit all."

- * "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives." e.g. discipline specific
- * "Any such scheme needs to allow for a wide variety in the scale and nature of data-generating research projects, and the resources (especially staff) available, i.e. one-size doesn't fit all." existing staff...

- * "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives." e.g. discipline specific
- * "Any such scheme needs to allow for a wide variety in the scale and nature of data-generating research projects, and the resources (especially staff) available, i.e. one-size doesn't fit all." existing staff...
- * "The resources which might be put into RDM at the University level need to be carefully weighed in light of the current resource pressures on researchers, research groups and departments."

- * "The main problem to me is that as HEFCE have not yet fully defined what they want I'm unclear how the University can plan adequately....I'm sure we will not have to duplicate existing data archives." e.g. discipline specific
- * "Any such scheme needs to allow for a wide variety in the scale and nature of data-generating research projects, and the resources (especially staff) available, i.e. one-size doesn't fit all." existing staff...
- * "The resources which might be put into RDM at the University level need to be carefully weighed in light of the current resource pressures on researchers, research groups and departments." – hands off our hard won research wonga!

* "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives."

* "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives." – my institutional service doesn't understand me!

- * "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives." my institutional service doesn't understand me!
- * "While the idea of storing data centrally has a number of perceived advantages, the "data" is not well defined, which is bound to bring a lot of confusion.

- * "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives." my institutional service doesn't understand me!
- * "While the idea of storing data centrally has a number of perceived advantages, the "data" is not well defined, which is bound to bring a lot of confusion. I think I know my data best

- * "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives." my institutional service doesn't understand me!
- * "While the idea of storing data centrally has a number of perceived advantages, the "data" is not well defined, which is bound to bring a lot of confusion. I know my data best
- * The data management rules proposed are going to complicate starting new research projects rather than encourage them."

- * "On the positive side, it seems that this is in response to requirements on everyone, everywhere from RCUK etc., and so it has to be dealt with; whatever IT/library can do to encourage some system that is easy to use and painless would be important. I'd strongly like the footprint of this process to be at the acceptable minimum, as I can't see it ever adding value to existing facility archives." my institutional service doesn't understand me!
- * "While the idea of storing data centrally has a number of perceived advantages, the "data" is not well defined, which is bound to bring a lot of confusion. I know my data best
- * The data management rules proposed are going to complicate starting new research projects rather than encourage them." you have no idea how busy my job is as it is

* Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit.

* Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit. – I'm a maverick, it's a jungle out there!

- * Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit. I'm a maverick, it's a jungle out there!
- * The bureaucracy associated with enforcement of the RDM policy is a drain on researchers time and, in the end of the day, research funds. At the time when we are already being stretched to the limit by a declining funding and an ever increasing teaching load this should be avoided at all costs.

- * Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit. I'm a maverick, it's a jungle out there!
- * The bureaucracy associated with enforcement of the RDM policy is a drain on researchers time and, in the end of the day, research funds. At the time when we are already being stretched to the limit by a declining funding and an ever increasing teaching load this should be avoided at all costs. red tape

- * Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit. I'm a maverick, it's a jungle out there!
- * The bureaucracy associated with enforcement of the RDM policy is a drain on researchers time and, in the end of the day, research funds. At the time when we are already being stretched to the limit by a declining funding and an ever increasing teaching load this should be avoided at all costs. red tape
- * Conclusion: A well meaning proposal which is unfortunately likely to have a strong negative impact on research at the University."

- * Some of the "data" must remain the property of the researcher/Department, such as the computer codes developed by the researchers, otherwise we would be giving away years of our work to our potential competitors in the UK and abroad. One may argue as much as they wish otherwise, but I personally am not going to put my cutting edge research effort out there for my competitors to milk, and neither I am going to spend my time in trying to organise/maintain the codes for their advantage. Let us first see them do it for our benefit. I'm a maverick, it's a jungle out there!
- * The bureaucracy associated with enforcement of the RDM policy is a drain on researchers time and, in the end of the day, research funds. At the time when we are already being stretched to the limit by a declining funding and an ever increasing teaching load this should be avoided at all costs. red tape
- Conclusion: A well meaning proposal which is unfortunately likely to have a strong negative impact on research at the University." – I suppose you mean well...

* It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned.

* It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned. – why don't you try looking? We've thought of everything...

- * It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned. why don't you try looking?
- * "It seems to me that the RDM policy has the same problem it did when I first heard about it about 18 months ago it doesn't clearly define what "data" is, and (crucially) what we're required to store and make publicly accessible. Those definitions affect everything for many of our projects it's the difference between a few Mb (the data that actually end up in papers) and tens to hundreds of Tb (raw data from observations/simulations). It's also not clear what we have to do regarding, for example, simulation code or data reduction scripts. Making everything public (and documenting it) would be an absolutely enormous undertaking."

- * It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned. why don't you try looking?
- * "It seems to me that the RDM policy has the same problem it did when I first heard about it about 18 months ago it doesn't clearly define what "data" is, and (crucially) what we're required to store and make publicly accessible. Those definitions affect everything for many of our projects it's the difference between a few Mb (the data that actually end up in papers) and tens to hundreds of Tb (raw data from observations/simulations). It's also not clear what we have to do regarding, for example, simulation code or data reduction scripts. Making everything public (and documenting it) would be an absolutely enormous undertaking." what do you want, blood?!

- * It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned. why don't you try looking?
- * "It seems to me that the RDM policy has the same problem it did when I first heard about it about 18 months ago it doesn't clearly define what "data" is, and (crucially) what we're required to store and make publicly accessible. Those definitions affect everything for many of our projects it's the difference between a few Mb (the data that actually end up in papers) and tens to hundreds of Tb (raw data from observations/simulations). It's also not clear what we have to do regarding, for example, simulation code or data reduction scripts. Making everything public (and documenting it) would be an absolutely enormous undertaking." what do you want, blood?!
- * "Handling of material developed between the telescope archive and the journal seems to be a major issue in terms of making things public. If they want data products to be archived, then it's not clear how that would be done, and I see it potentially being a huge and unquantifiable burden."

- * It is unclear how simulation data created locally, nationally, or internationally have to be dealt with. Arguably, only the codes (including all previous versions using svn or similar), simulation parameters, and any real data (not generated) should be stored. In fact, I do that already: parameters are published and the code is subversioned. why don't you try looking?
- * "It seems to me that the RDM policy has the same problem it did when I first heard about it about 18 months ago it doesn't clearly define what "data" is, and (crucially) what we're required to store and make publicly accessible. Those definitions affect everything for many of our projects it's the difference between a few Mb (the data that actually end up in papers) and tens to hundreds of Tb (raw data from observations/simulations). It's also not clear what we have to do regarding, for example, simulation code or data reduction scripts. Making everything public (and documenting it) would be an absolutely enormous undertaking." what do you want, blood?!
- * "Handling of material developed between the telescope archive and the journal seems to be a major issue in terms of making things public. If they want data products to be archived, then it's not clear how that would be done, and I see it potentially being a huge and unquantifiable burden." I'll walk...

So what can we do to help?

- * Researchers and institutions coming together to create solutions and systems
- Research Data Alliance working group on workflows for publishing data
- * Identifiers and citation DataCite and ODIN

Research Data Alliance – workflows for data publishing

- * What routes are there for publishing data that will help researchers address the pressures they are facing to make data available for reuse?
- * Having information on "established" or "trusted" workflows and services can help:
 - * researchers
 - and the people/organisations supporting them
- The working group propose to study and tests workflows that allow:
 - * efficient and reliable reuse of research data
 - * Enhance the possibilities for greater discoverability across disciplines and stakeholder groups

Workflows and services

- Looking at established and emerging workflows across sectors and subjects
- * Data journals
- * Data centres
 - * International
 - * National
- * Institutional repositories

http://tinyurl.com/mvtbrek

RDA-W	VDS-F	11 work	flow analys	sis 🖈										
File Edi	it View	Insert	Format Data	a Too	ls Add-ons I	lelp	Last	edit v	as made 6	days a	ago by anonymou	IS		
ē r	~ 7	\$ %	.0	Aria	I - 9	· ·	В	I	<u>5 A</u> -	<u>₩</u> , -	₩ + 199 - 1	<u>∓ - ↓ - </u>	주 🔟 囯 ෙ	- Σ -
Some note	es before	hand:												
Α		В	С		D		E		F		G	Н	1	J
ome notes	beforeha	nd:												
please feel	free to ac	dd workflow	s you think shou	ıld be pa	art of the analysis	by expa	anding	the ta	able to the rig	ght				
if you think,	, there are	categories	s missing (or nee	ed to be	changed), please	add a r	ow and	d dro	an email to	the list	to highlight the ch	ange that affects the	ne other ongoing an	alysis - but pleas
please prov	vide as ma	any links to	documentations	s, charts	as possible to en									
	w	orkflow na	STF0		NSIDC Data centre	Infras ervic	ENVRI ence m structu e prov (data e/repo y?)	nodel ure/S /ider			INSPIRE Digital library	NPG (PubChem & Scientific Data) Publisher	UK Data Archive/Service Service provider/ data centre	PREPARDE (NCAR CISL Data centres data journals
		stakeholder charge [ins publisher, disciplinan epository,	t, /								disciplinary, international data repository	ı		
	A	vailable onl		prelida.	ftp://sidads.colora	to http://	conflue	ence.	http://proje harvard.ed vn & https://doc e.com/doc d/1T-i2a4s 3DCIYyAL h2hLdJJM VVXhc/edi https://doc e.com/file/s LxEN77UZ LWXpodm	u/ojs-d s.googl ument/ ynXlhe 8VYg md6K t# & s.googl d/0Bze oYnB		http://www.dcc.ad	ODIN Deliverable [link?]	http://www.le.ac uk/projects/pre rde. http://proj.badc ac.uk/preparde tachment/wiki/I iverablesList/D 1_D2_2_PREP RDE_Workflow combined_draf pdf

Categories we look at

- * Discipline
- * Function of workflow
- Includes PID assignment to data set
- * PID type e.g DOI, ARK etc
- * Peer review of data, (e.g. by researcher & editorial review)
- * Curatorial review of metadata (e.g. by institutional or subject repository?)
- * Technical review & checks (e.g. for data integrity at repository/data centre on ingest)
- * Discoverability: Indexing of the data -if yes, where?
- Formats covered
- * Persons/Roles involved, e.g. editor, publisher, data repository manager etc. [please list everyone you know]
- * Is there a link to paper or is this standalone data
- * Links to grants, usage of author PIDs
- * Data Citation Facilitated
- Refers to a data life cycle

Identifiers and data citation

- * How can services help with incentives for researchers?
- * Force 11 Data Citation Principles
- * Persistent identifiers
 - * DataCite
 - * ODIN

Force 11 Data Citation Principles

The Data Citation Principles cover purpose, function and attributes of citations. These principles recognize the dual necessity of creating citation practices that are both human understandable and machine-action

* Importance

- * Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications[1].
- * Credit and Attribution
- * Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data[2].
- * Evidence
- * In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].
- Unique Identification
- * A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community[4].
- * Access
- * Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data[5].
- * Persistence
- * Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe[6].
- * Specificity and Verifiability
- * Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verfiying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited [2].
- Interoperability and Flexibility
- * Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities[8].
- * able.

DataCite

How can DataCite help facilitate this:

- * Provide unique persistent identifiers (DOIs) for data
- Developed a metadata schema to support data citation
- * Work at an institutional level to integrate with institutional infrastructures
 - * E.g plugins for EPrints and DSpace

Linking data and publications

Home + Recent Actions Browse Search	My Settings Alerts Help					
Quick Search All fields	Author					
? search tips Journal/book title	Volume Issue	Page	Clear (X) Go	Advanced Search		
Marine Chemistry Volume 95, Issues 1-2, 2 May 2005, Pages 51-63						Font Size: 🚍 🚹
▶ Article Figures/Tables References	PDF (931 K)					Thumbnails Full-Size Images
doi:10.1016/j.marchem.2004.06.040 ② Cite or Link Using DOI					▶ Article Toolbox	
Copyright © 2004 Elsevier B.V. All rights reserved.					Download PDF ○ E-mail Article	Export Citation Add to my Quick Links
Changes in the concentration of iro	Cited By Save as Citation Alert	Add to my Guick Links Add to collab Permissions & Reprints				
Jun Nishioka ^{a,} , ⊠, Shigenobu Takeda ^b , Hein 、	J.W. de Baar ^C , Peter L. Croot ^{C, d} , Marid	e Boye ^C , Patrick La	nan ^c and Klaas R. Tim	mermans ^c	M Citation Feed	Cited By in Scopus (16)
^a Central Research Institute of Electric Power Indus	stry, 1646 Abiko, Abiko-shi, Chiba 270-1	194, Japan			Related Articles in Science	Direct
^b Department of Aquatic Bioscience, University of T	Major deviations of iron complexation during 22 days of					
^c Royal Netherlands Institute for Sea Research, 179 ^d Marine Biogeochemie, Institute für Meereskunde	Marine Chemistry Iron cycling and nutrient-limitation patterns in surfac Deep Sea Research Part II: Topical Studies in Oceanogra					
-marine Biogeochemie, institute für meereskunde	Net and gross O2 production in the southern ocean from					
Received 2 September 2003; revised 17 March 20	04; accepted 16 June 2004. Available	online 30 Decembe	er 2004.		Deep Sea Research Par Papers	t I: Oceanographic Research
Abstract An in situ iron enrichment experiment was carried	out in the Couthern Ocean Relay Frontel	Tone and familized	a notab of motor mithire	on addu aftha Antaratia	We	plankton bloom in the Scotia-
Circumpolar Current (EisenEx, Nov. 2000). During in iron distribution and size-fractionations, includin development of the phytoplankton bloom.	Journal of Marine Systems Insights into Southern Ocean carbon export from the [de Deep Sea Research Part II: Topical Studies in Oceanogra					
Prior to the first infusion of iron, dissolved (<0.2 µm being a minor fraction. For the iron addition, an acid	dified FeSO4 solution was released thr	ee times over a 23-d	day period to the eddy.	High levels of dissolved iron	▶ View More Related Arti	cles
concentrations (2.0±1.1 nM) were measured in the were observed before storm events, there was a si	Supplementary Data View Record in Scopus					
Fe]+0.0519, R ² =0.9346). These results indicate the events caused a significant decrease in iron conce colloidal and labile particle). The changes in each	entrations (<0.61 nM in dissolved iron) a	nd changed the pro	portions of the three in	on size-fractions (soluble,	The research colla	boration tool ②
particulate and soluble fractions. Therefore, particle and naturally dissolved organic ligands. Our data s	e and soluble iron efficiently remain in t	he mixed layer, prob	oably due to the prese	nce of suspended particles	No user tags yet	rrating
the oceanic iron cycle. Keywords: Southern ocean; Iron enrichment exper	iment; Size-fractionated iron; Iron speci	ation			This article has not	yet been bookmarked
Article Outline					No comments on the	
1. Introduction					Be the first to add this article	
The state of the s						

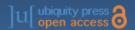


&upmetajournals

open health data

Contact

Repositories





Tweets

Jul 15th

JOPHD

@up jophd

/OL9uo4plyv

Jul 15th

mike galsworthy @mikegalsworthy

#Opendata Hooray! New

data journal just launched

Just launched: the Journal of Open Public Health Data

New report out on realising

an #OpenData culture from

is now live! http://t.co

Ubiquity Press Ju[@ubiquitypress

(Journal of Open Public

Health Data) http://t.co /vdmdINwPaM @up iophd @openscience @UKODI

The Journal of Open Public Health Data (JOPHD) features peer reviewed data papers describing public health datasets with high reuse potential. We are working with a number of specialist and institutional data repositories to

Editorial Board

ensure that the associated data are professionally archived, preserved, and openly available. Equally importantly, the data and the papers are citable, and reuse is tracked.



Articles

Archive

Ads Open Access in developing countries



Home

About

Alexander, Wint Projected Population Proximity Indices (30km) for 2005, 2030 & 2050



Wint, Morley, Alexander Four Rodent and Vole Biodiversity Models for Europe 11 Jul 2013



Knowles, Olafsdottir Data: 'Initial Clinical Referral Standards after Newborn Screening... 09 Jul 2013



Tadic, Knowles, Rahi Data from the Multiprofessional

Most popular articles

FAQ



Alexander, Wint Projected Population Proximity Indices (30km) for 2005, 2030 & 2050 11 Jul 2013



Wint, Morley, Alexander Four Rodent and Vole Biodiversity Models for Europe 11 Jul 2013



Tadic, Knowles, Rahi Data from the Multiprofessional Workshop: 'Paediatric... 09 Jul 2013



Knowles, Olafsdottir Data: 'Initial Clinical Referral Standards

Most recent articles



ODIN project http://odin-project.eu/

- * To support development and stimulate adoption of interoperable identifiers for:
 - * Researchers
 - Inputs (cited work, data)
 - Outputs (publications and data)
- * Facilitate information flow within and between research communities, for greater re-use of data and exploitation of the knowledge created.
- * ODIN proposed to explore the opportunities, highlights gaps and roadmaps, and nurture interoperability solutions, globally, and for specific disciplines and beyond
- * Identify and compare workflows for identifier integration between Humanities and Social Science and High Energy Physics



Scenario 1: sharing and linking data

- I have written my paper and submitted it to a journal
- I have produced data that can be used to reproduce my analysis
- I would like to share my data

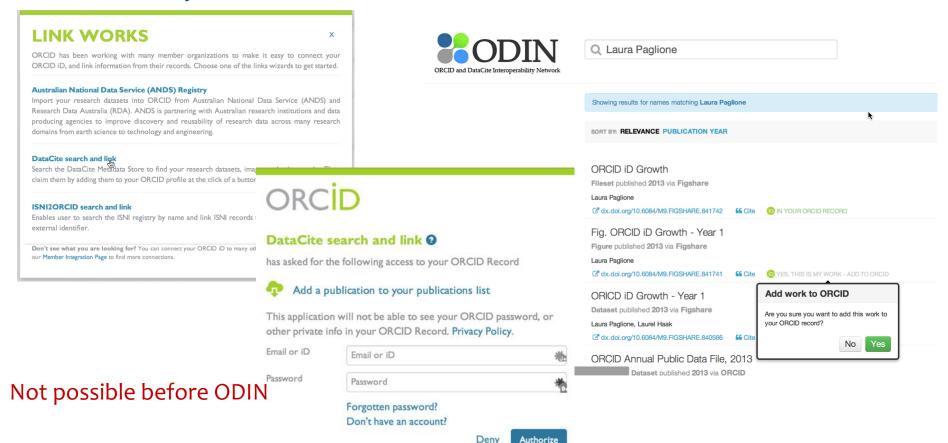


Project metadata for the Archaeology Data Service

Please complete this form as fully as possible with details of your project. This data will form the basis of an entry about your dataset in the ADS Catalogue, and underpins the computerised searching process that allows users to discover and retrieve information.

- 1. Title please indicate the title (and any alternatives) for the dataset.
- Description please provide a brief summary (max. 200-300 words) of the main aims and objectives of the project and the content of the dataset.

- I can link my paper AND my data to my ORCID profile
- There may be thousands of authors named like me



Scenario 2: Sharing and linking thesis data

- I sign up for a PhD programme
- I write my doctoral thesis and deposit it, along with my data
- I see the thesis in EThOS

E-Thesis submission check lis

What	to	do

Ensure that your thesis is produced in accordate University's Research Degree Regulations content, formatting and word count

Clear all rights and obtain all permissions for I from third party copyright holders as you gath resources and write up your thesis



ETHOS

ELECTRONIC THESES ONLINE SERVICE

Opening access to UK theses

You are logged in as Mr Sergio Ruiz

Title:

Presynaptic nicotinic modulation of neurotransmitter release in the cell
Author:

Kaiser, Sergio.

Awarding Body:

University of Bath

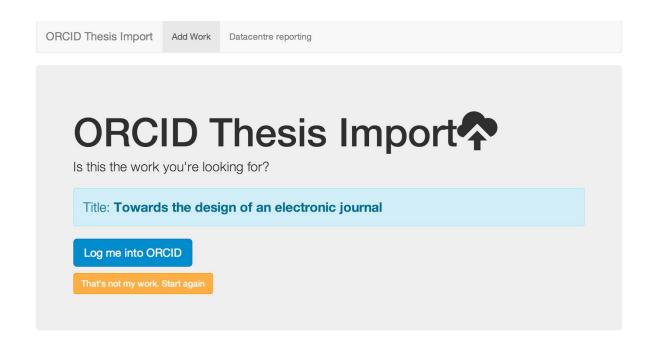
Current Institution:

University of Bath

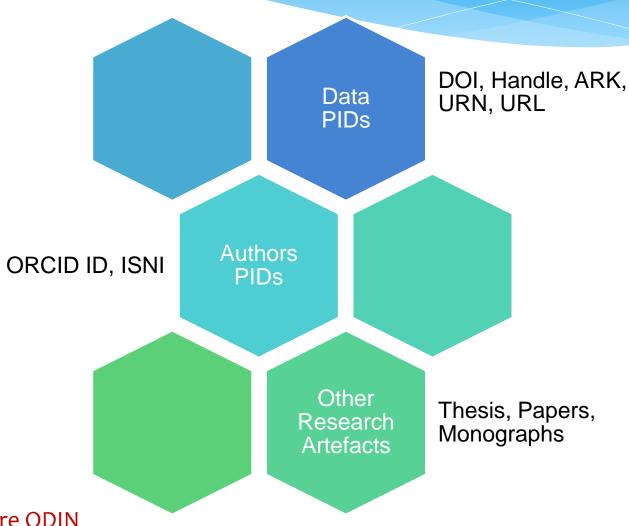
Date of Award:

1999

I can link my thesis AND the data behind to my ORCID profile



Connecting PIDs



Not possible before ODIN

CHALLENGES

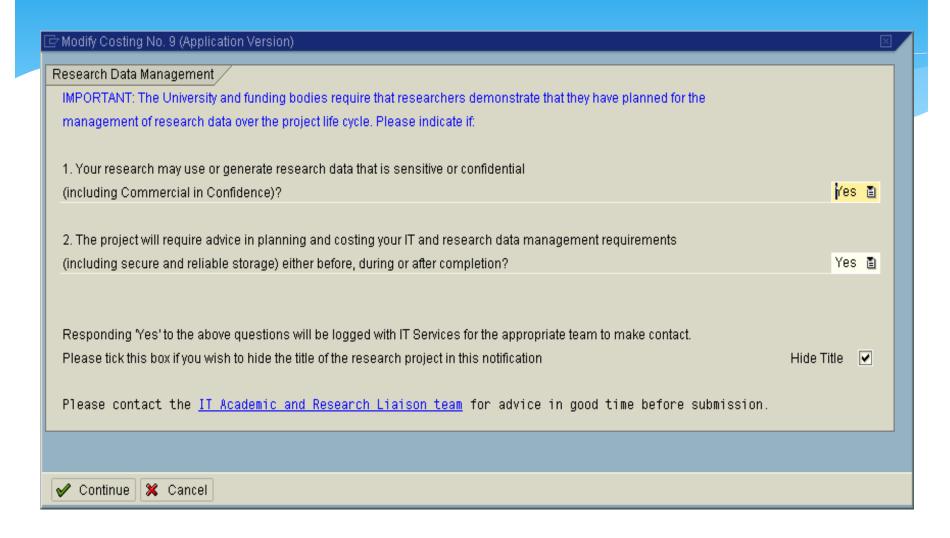
interdisciplinary research database

- * ingest each input dataset in form such that sufficient information is carried forward to enable interoperation
- * Cultural differences
- versioning & provenance for input datasets
- * which software tools, infrastructure, Query interface?
 - suitable for multi disciplinary researchers
- * Requirements upon the institution for sustaining the research assets & skills
- Requirements upon the researchers
 - * Annotating
 - * Refreshing
 - * Maintainence of datasets

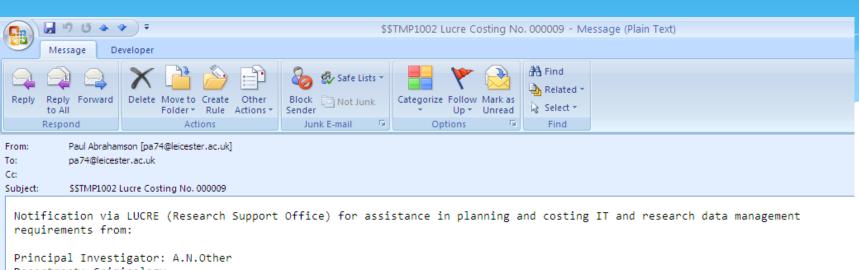
Top Tip: how to get researchers' attention?

- Research grant pre-award costing (LUCRE SAP based)
 - * Dominates researchers' minds!
 - * Enable PIs to build grant application using actual costs of staff, overheads, and the right rules for funder
 - * Trigger involvement of IT Research Liaison and wider institutional expertise via flags
 - * sensitive research data
 - costing/planning support including curation and preservation over research lifecycle
 - * Track institution wide needs via IT Service Desk

Research Costing as a Trigger



Research Costing as a Trigger



Department: Criminology

Email: No email address for PI in Lucre

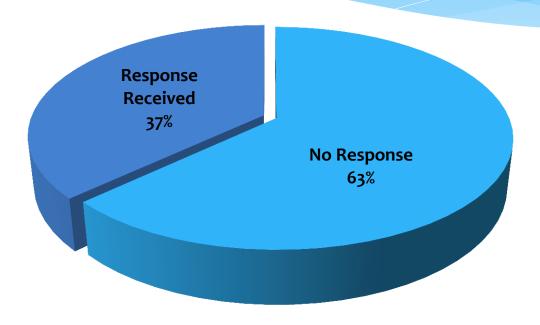
regarding research grant proposal:

Title: David Ad Hoc Lucre Costing No. 000009 Start Date: 01.02.2009 End Date : 30.11.2009

To be forwarded for attention of Jonathan Tedds and the IT Research Computing Support Team

Research costing – only part of the answer

Researcher Responses to Contacts Made



ORGANISATIONAL CHALLENGES AND SOLUTIONS

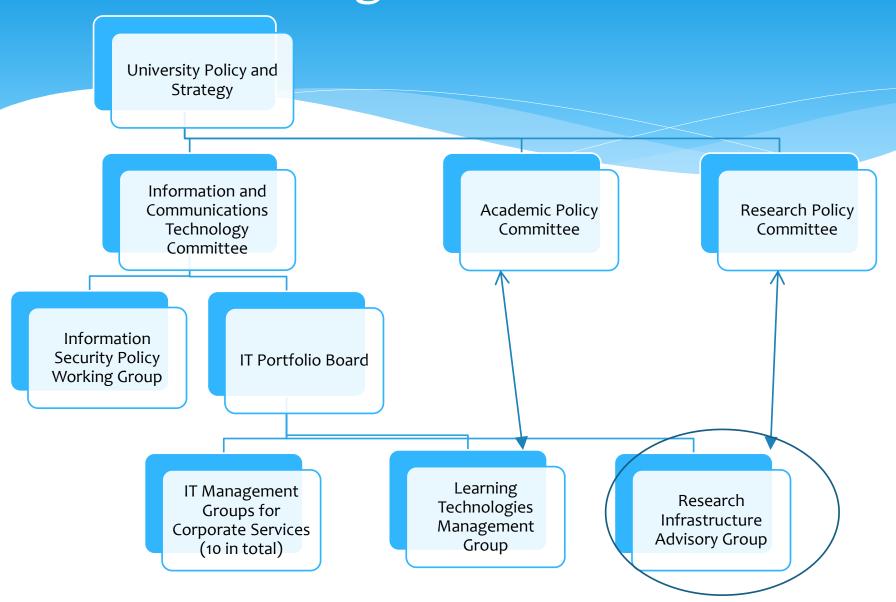
Cultural differences

- * Recognise different cultures and mind sets
 - * research community and IT specialists in central services
 - * different professional language, expectations and working practises
 - management of a research project usually requires a different, iterative methodology than an IT infrastructure project having a more clearly pre determined end point
- Research Liaison Role
 - * strong research background helps!
 - enables effective ways of liaising with research community
 - bridging gaps in understanding
- Leveraging expertise within and external to the organisation
 - * coordinate 'specialists'

Active Data Storage: Identifying the Holy Grail?

- "what is needed is a tool to transparently sync local and network storage" (Marieke Guy, JISCMRD2 Nottingham)
 - CKAN (Orbital, Lincoln)?
 - Research Data Toolkit (Herts) hybrid solutions
- DropBox-like functionality a must
 - Usability
 - Technical interoperability
- aim to help create databases for research data so
 - facilitates collaboration and data sharing
 - enables the subsequent publication of datasets
- challenge is to ensure data are
 - documented
 - Preserved
 - service is sustainable

Add governance....



Find & Share



Do you know how to...

find existing information resources related to your research?

Where can you find research data that you can repurpose or combine with your own to produce new research?

share data with your collaborators securely and effectively?

Whether building a collaborative proposal, generating results for others to comment on or sharing the final outputs of your research – how will you link with your colleagues ahead of wider sharing?

deposit your research data and outputs in an open repository?

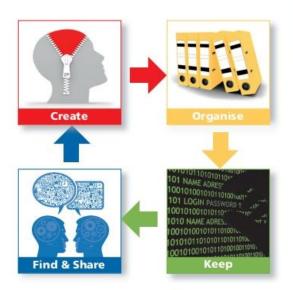
Is there an appropriate disciplinary or institutional repository and what do you need to do to deposit your research output? Plan ahead to avoid refactoring...

publish your research, and get it cited as well?

Institutions and data centres must make research data reusable to others while providing credit to the researchers who did the work. Your future career could depend on it!

Chances are you could use some helpful pointers in all of these!

To find information, support, advice and training, as well as links to external resources, go to www.le.ac.uk/researchdata



email: researchdata@le.ac.uk





What would you do if you lost your research data tomorrow?

Take the research data health check... and find help to secure, share and exploit your valuable research.



www.le.ac.uk/researchdata

www.le.ac.uk/researchdata

Create

Organise

Keep







Have you...

fully understood your research funders' data management requirements?

Government and funders require that publicly funded research is made available for reuse – are you up to date with their latest policies? Your future funding might depend on it!

written a data management plan?

Your funder may already require this but build it in from the proposal stage to avoid headaches in the future.

gained ethics approval/consent?

Writing a data management plan will aid planning and help you to navigate ethics and governance requirements.

protected your intellectual property?

Leaving intellectual property considerations for a rainy day could lose you appropriate credit, damaging career prospects and perhaps your financial future health!

Are your research files and data...

clearly described, in terms of content (using standard metadata)?

Ask yourself honestly: is there a danger of data being lost? Will you be able to remember how you generated your data, and will you or anyone else be able to find it in the future when you wish to reuse and share?

clearly labelled with versions and dates?

How will you remember which was **the** definitive version and which dataset was used in producing a given research outcome?

logically structured and named?

Once you've remembered **how** you generated data, can you still **find** the relevant files?

future-proofed against broken links, using persistent identifiers?

The persistent identification of digital resources can play a vital role in enabling their accessibility and re-usability over time using recommended data standards.

Do you know...

how to restrict access to your research data to the right people?

Have you consulted with university or data centre experts so that only the right people have access to your research?

which data to keep and which data to discard?

Managing research data effectively means being selective: **which** data to discard and **when** as well as **what** to keep and for **how long**?

how securely your data is stored?

What happens if your storage media fail? How resilient is it? Could it get left on the train and could somebody else misuse it?

how your data is backed-up?

Have you made use of university and/or external resources to back up data so that you have multiple copies in case of loss or theft?

Suggested timeline for implementing institutional research data management

Near term (up to a year)

Get key players from across institution together: Identify objectives for Research Data Management (RDM)

Raise awareness and make use of existing general training materials e.g. DCC101

Identify and benchmark existing central provision for RDM, including research information systems, grant costing & staff support services

Identify, disseminate and develop current exemplars to conduct basic audit of RDM. Seek quick wins with high profile academic champions

Draw on external expertise, build links with potential partners

Medium term (1-3 years)

Establish a formal University wide Committee to develop & ratify policies on RDM; disseminate benefits & cost savings

Design and pilot institution-wide and discipline-specific training drawing upon local support and guidance

Appraise RDM provision & conduct gap analysis to develop business case. Benchmark levels of maturity of RDM infrastructure & support

> Develop portfolio of projects to reuse infrastructure, build RDM capacity in further key groups

Formalise links to external partners & stakeholders where necessary

Long term (>3 years)

RDM is embedded across faculties & central services, subject to periodic review, ensuring information flow and support is co-ordinated

Roll out institution-wide and discipline-specific training to all new staff & students. Periodic review and update of materials

RDM service in place and involving all stakeholders.
Periodic benchmarking and target setting

RDM and Policies developed at faculty level; Key academic groups are engaged; central services in widespread use across the institution

> RDM effectively integrated with shared UK and international services; quality assurance in place

From Whyte & Tedds (2011), DCC Briefing

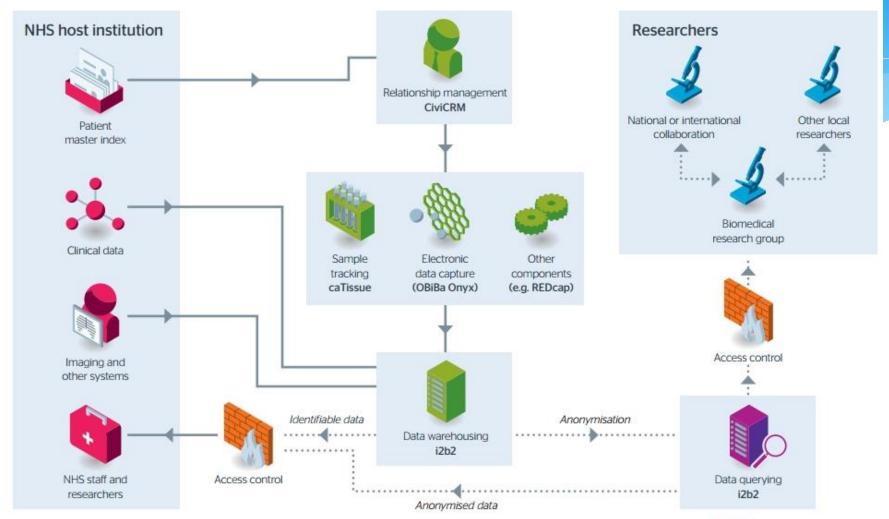
http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm

But that's not all...

- * What about the software underpinning data driven research?
- * If we're going to publish as open data:
 - * How do we help researchers to store, annotate and discover the datasets they create?
 - * How do you sustain and reuse that?



Biomedical research software as a service BRISSKit A Jisc funded project



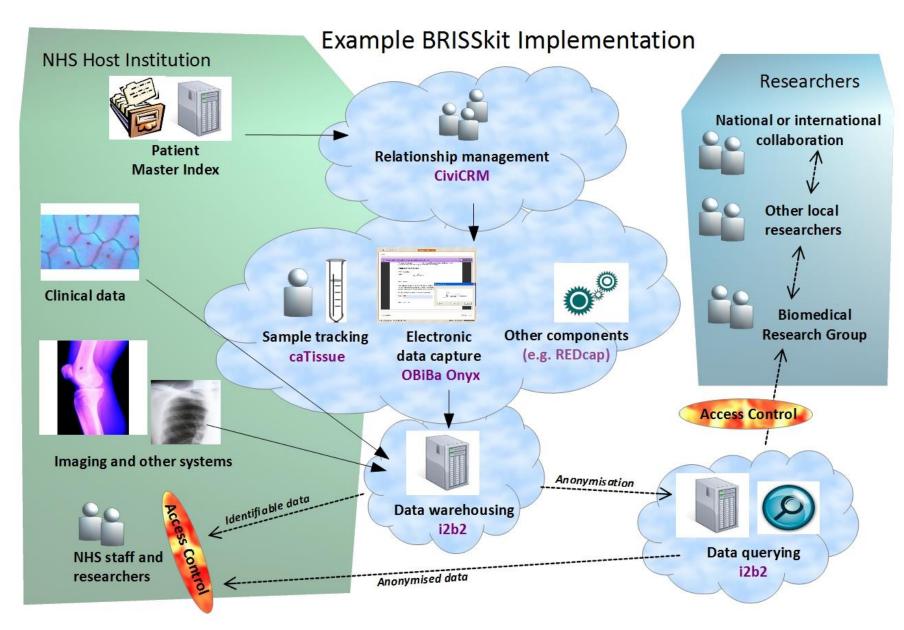






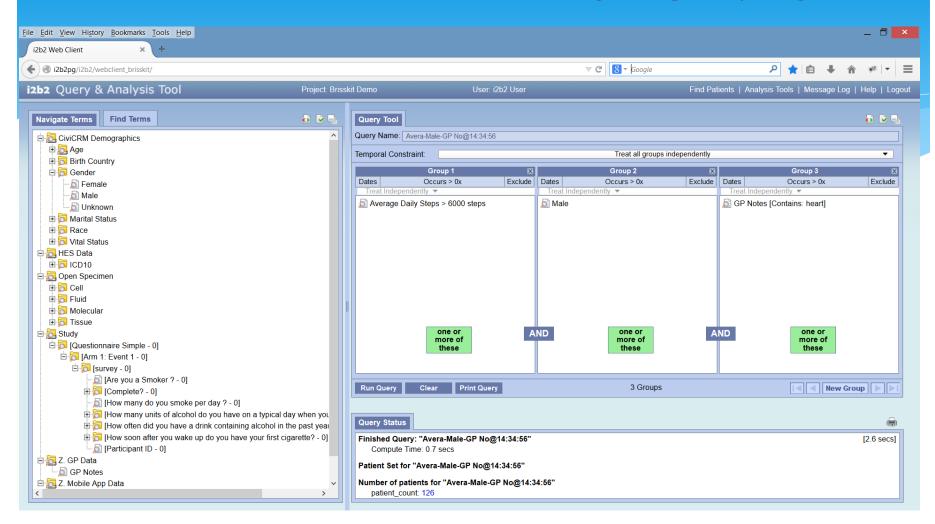






http://www.brisskit.le.ac.uk

BRISSKit i2b2: data warehousing & querying



Data from multiple data sources combined into multiple ontologies for flexible and sophisticated searching, cohort discovery and research

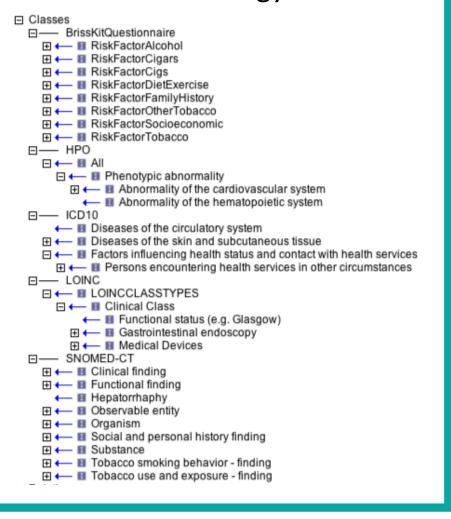
Research: the semantic bridge

Questionn

Records participal consent, question data and primary specimen IDs

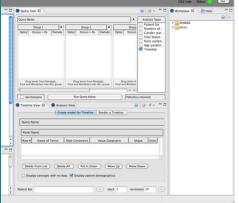


Bio-ontology!



7

rt selection and querying

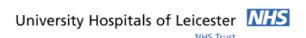


March 2015: i2b2 via public cloud (anon data)

- Users can now get free servers from public clouds, e.g. Amazon, HP, Azure etc
- In March 2015 BRISSKit users will be able to
 - deploy their own i2b2 virtual app onto their own cloud server
 - upload their data through .csv files default nominal ontology created
 - modify/align this ontology to standardised BioPortal codesets – e.g. SNOMED
 - perform queries on their data using the revised ontology through i2b2







Challenge for institutions

- Rise to scientific and research challenge
 - Not just a management challenge
 - Responsibility for the knowledge they create
- * Library
 - Doing the wrong things through the wrong people?
 - New library needs to enable
 - curation of data and publications
 - active support from data scientists
 - from centralised to dispersed support
 - Expert centres such as D-Lab essential!
- * IT Service
 - Provide research data platforms for researchers:
 - Active storage
 - * Enable collaboration
 - Connect to preservation services through Library

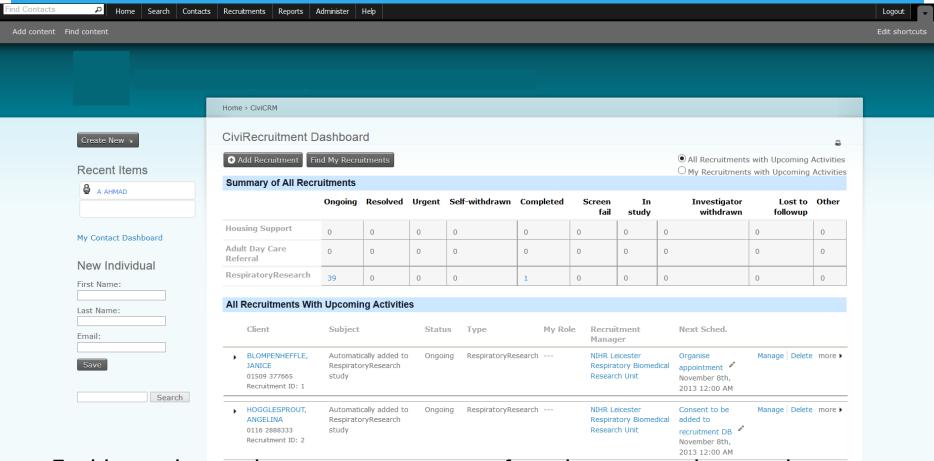
Thanks

Thanks to collaborators and partners

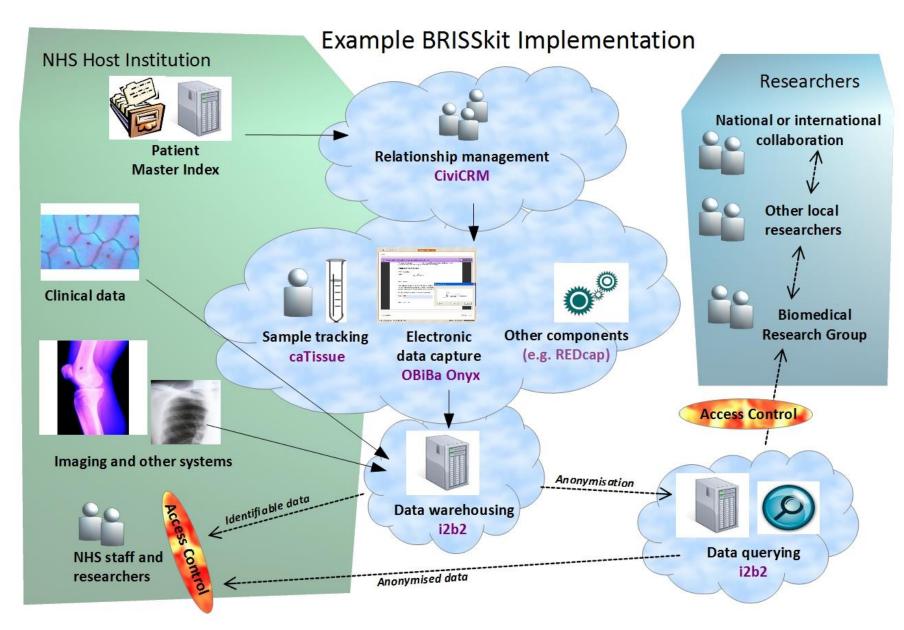
- * ODIN partners
- * RDA-WDS Interest Group & Working Group members
- British Library
- * BRISSKit project colleagues, Jisc & partners
- * PREPARDE project partners
- * University of Leicester



BRISSKit CiviCRM: patient cohort management

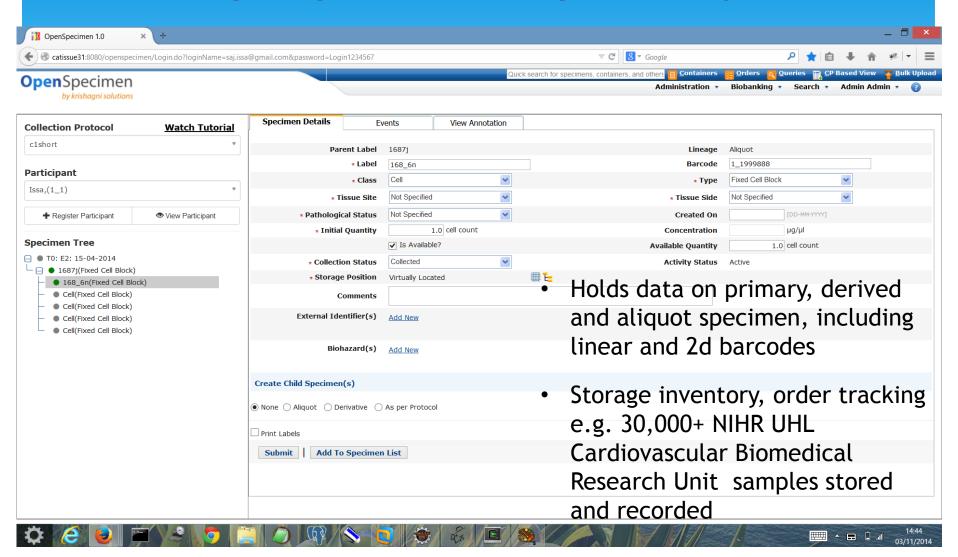


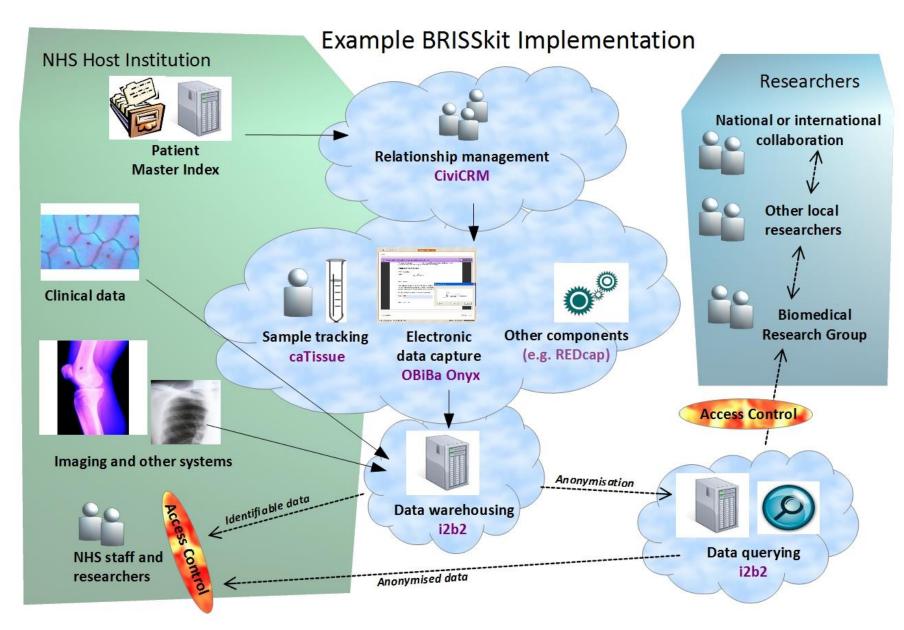
- Enables end-to-end contact management for volunteers and research participants
- track approaches, contact, responses, recruitment, exclusions
- · object model that reflects community building and non-profit relationships



http://www.brisskit.le.ac.uk

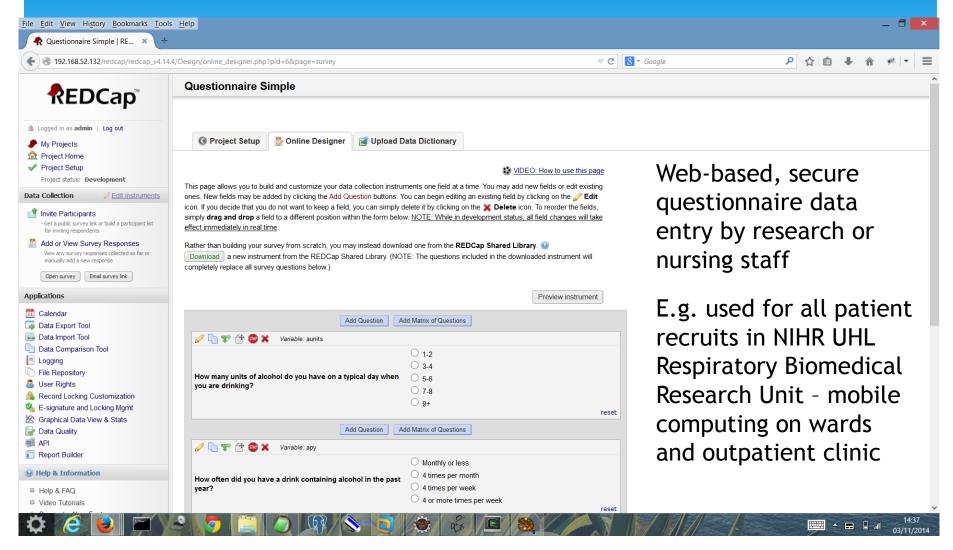
BRISSKit OpenSpecimen: sample management

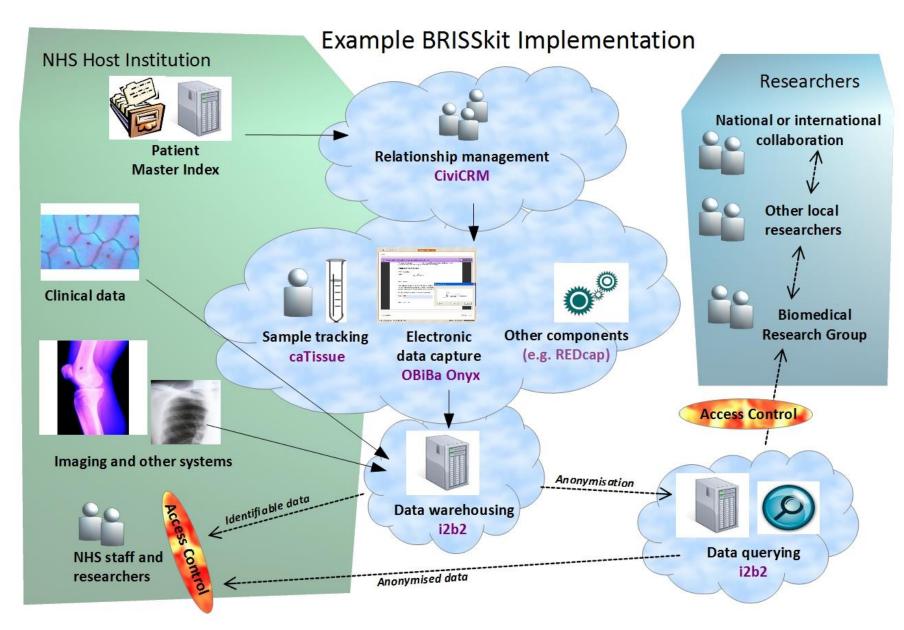




http://www.brisskit.le.ac.uk

BRISSKit RedCap: survey management





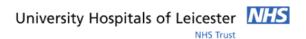
http://www.brisskit.le.ac.uk

BRISSKit USPs

- Integrated support for core research processes
- Well-established mature open source applications as protoyped in e.g. Cardiovascular: fully UK customised
- A platform for seamless management and integration between applications
- An API allows integration with existing clinical systems
- Easy set up, use and administration through browser (including on mobile devices)
- Capability of being hosted in any compliant cloud provider including UHL (NHS information governance)
- Direct secure links through Jisc via Janet network





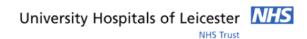


BRISSKit Funding & Partners

- New HEFCE/Jisc investment approved for 2014 2016
 - Jisc endorsed service
 - Co-design with reorganised Jisc
 - Key Janet Framework partners Farr, Crick
- University of Leicester Cancer Biobank
 - Tissue sample management built on caTissue, OpenSpecimen
- NIHR Respiratory Biomedical Research Unit solutions: University Hospitals Leicester NHS Trust
 - linked to UoL Health Sciences Exceed Study
 - Links to Loughborough-Leicester Lifestyle BRU





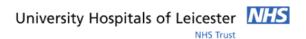


BRISSKit highlighted collaborations

- University of Bristol
 - ALSPAC Birth Cohort Studies
 - DataShield: simultaneous remote, secure access to multiple large international cohorts
 - SAIL-Farr secure NHS data hosting
- University Hospitals Leicester NHS Trust
 - UoL Health Sciences Exceed Study
 - NIHR BRUs: Cardiovascular, Respiratory, Lifestyle (Loughborough-Leicester)
 - Leicester Diabetes Centre
- UoL Data to Knowledge for Practice strategic theme
 - UoL Genomics
 - UHL NHS Trust IBM IT Partnership





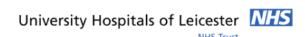


Towards an i2b2 NHS community

- With datasets uploaded into a range of i2b2 instances
- Users will be able to publish their i2b2 datasets
- A community of public cloud i2b2 users will emerge, within which users can publish, exchange and augment data and ontologies
- These merged datasets can then be used to service NHS-wide cohort search, selection and quality management
- Re-identification of cohorts will remain with original sources of i2b2 data







New BRISSKit Community & Hack Events planned for mid Feb 2015







