

WHY MAKE DATA AVAILABLE?



CATH BARNETT

Outline

- Open data – a definition
- Personal experience
- Background
- Why disseminate your data?
- How to disseminate your data - The NERC Environmental Information Data Centre
- Making data effective
- Advantages of data dissemination
- Summary

Research Data – a definition

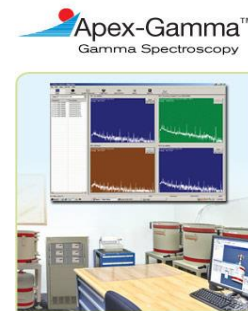
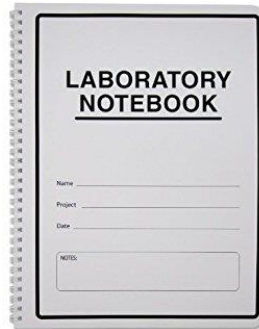
- What is research data?
 - Evidence that underpins the answer to the research question (it has many forms e.g. statistics, surveys, fieldwork observations, maps, methods, models, images, sound recordings and published texts etc.)
- What is Open data?
 - Data that anyone can access, use, modify and share (provided that there is appropriate acknowledgement if required)

Open data **must be accessible to anyone to use** (who agrees to the terms of use), but this **does not mean that it must be cost free** to access as there is often a cost to creating, maintaining and publishing usable data. However, much primary research data are cost-free

Personal experience back in 1989



Many 1000's of samples collected, prepared, analysed.....
results recorded in lab books/databases and data QC'd....



Personal experience in 1989 and beyond

- For only some of it to be summarised into one small table!

186

N. A. Benford, B. J. Howard, C. L. Barnett, N. M. J. Crout

TABLE 2

A Comparison of the Activity Concentration (Bq kg^{-1} DW) of Total and Aged Radiocaesium Determined by Assuming an Original ^{134}Cs : ^{137}Cs Ratio for Chernobyl Fallout of 0.53 (0.17 at the time of sampling in November/December 1989)

Area		Radiocaesium activity concentration		
		Vegetation	0-4 cm soil	Lower soil
Black Combe ^a	^{137}Cs			
	$\bar{x} \pm \text{SE}$	370 ± 45	1180 ± 103	140 ± 36
	Range	147 - 696	669 - 1896	35 - 432
	^{134}Cs			
Corney Fell	$\bar{x} \pm \text{SE}$	50 ± 8	130 ± 15	—
	Range	4 - 96	65 - 245	$< 2 - 35^b$
	^{134}Cs : ^{137}Cs			
	$\bar{x} \pm \text{SE}$	0.12 ± 0.014	0.11 ± 0.008	0.022 ± 0.006
Aged ^{137}Cs	Range	0.02 - 0.15	0.06 - 0.15	0.004 - 0.08
	$\bar{x} \pm \text{SE}$	110 ± 20	420 ± 62	110 ± 23
	Range	30 - 227	118 - 770	30 - 250
Black Combe ^a	^{137}Cs			
	$\bar{x} \pm \text{SE}$	320 ± 44	1020 ± 63	170 ± 23
	Range	44 - 1071	258 - 1775	10 - 498
	^{134}Cs			
Corney Fell	$\bar{x} \pm \text{SE}$	50 ± 7	130 ± 9	—
	Range	3 - 154	35 - 260	$< 2 - 48^b$
	^{134}Cs : ^{137}Cs			
	$\bar{x} \pm \text{SE}$	0.14 ± 0.007	0.13 ± 0.004	0.05 ± 0.007
Aged ^{137}Cs	Range	0.02 - 0.17	0.09 - 0.17	0.004 - 0.15
	$\bar{x} \pm \text{SE}$	60 ± 9	245 ± 29	110 ± 14
	Range	3 - 178	23 - 893	5 - 308

A significant amount of data was just not used in published papers

I know as I collected and prepared a lot of it!

Open data is not just numbers



Sheep Picture by Chris Himsworth (own work)
[CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0>)], WikimediaCommons



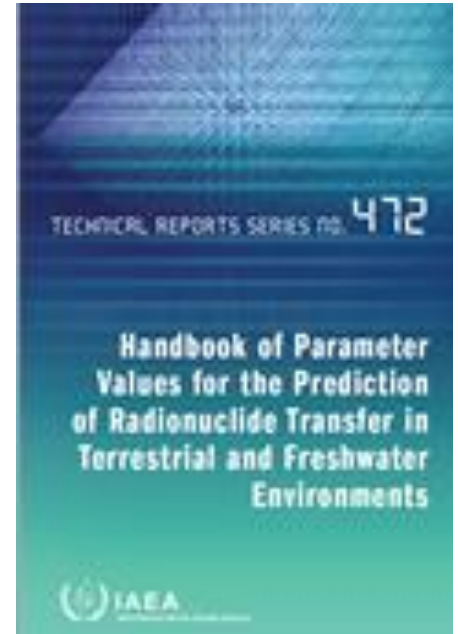
Own work assumed (based on copyright claims)., Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=2496069>



Creative commons, you will see this logo quite a lot in the talk

Personal experience - IAEA TRS 472

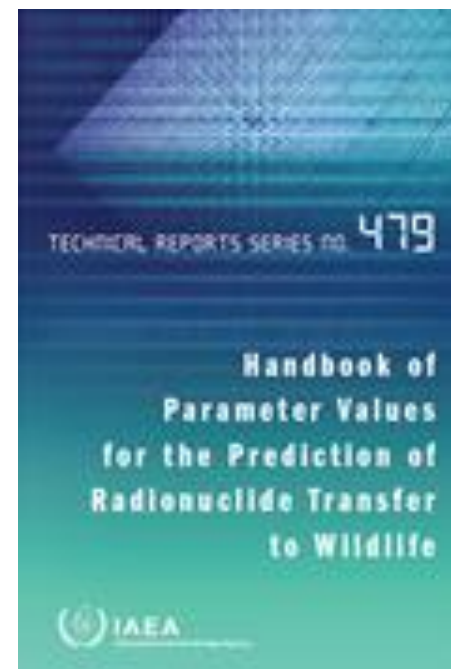
- We conducted review of data for transfer to animals
- Many studies published a long time ago, authors have changed jobs, retired or died
- Much potentially useful information could not be used as unable to clarify some data issues e.g.
 - Unclear if results in fresh or dry weight
 - Number of samples
 - No standard error/deviation reported
- Where a mean \pm error was reported – the individual data would have been more useful



Personal experience - IAEA TRS 479

For the 'wildlife transfer handbook' (TRS 479) things were better

- on-line database created:
[http://www.wildlifetransferdatabas
e.org/](http://www.wildlifetransferdatabas
e.org/)
- Database continues to be updated and used for other things e.g. ERICA Tool update, ICRP



International Atomic Energy Agency & International Union of Radioecologists

Wildlife Transfer Parameter Database

[Help](#)

If you are not registered, [Register Now](#)
If you have forgotten your login details please email us and we will send you a reminder ([Password Reminder](#))

User Name
Password

This database collates data to provide parameter values for use in environmental radiological assessments to estimate the transfer of radioactivity to wildlife. The database has been updated today (12/12/13). Full details of the update will be made available soon. There may be some minor changes with the data over the next few weeks as further evaluation of the changed data takes place. If you have any queries please contact us at the wildlifetransfer@gmail.com address. We will put a note here when the database update is complete.

The database was started to aid both:
(i) the International Atomic Energy Agency (IAEA) in the production of a handbook on wildlife transfer parameters (IAEA Technical Report Series now in press); and
(ii) the derivation of transfer parameter values for the International Commission on Radiological Protection (ICRP) list of Reference Animals and Plants (RAPs) (ICRP Publication 114).

The database has been designed and supported by the following organisations:

- Environment Agency, England and Wales
- Norwegian Radiation Protection Authority, Norway
- Centre for Ecology and Hydrology, NERC, UK
- University of Stirling, UK
- The STAR Radioecology Network of Excellence

Background - where did it all start.....

On June 2013 the Royal Society hosted
**the first ever G8 joint Science Ministers and
national science academies meeting** in London



G8 joint Science Ministers meeting (1)

At that meeting Ministers approved a statement for the G8 to consider **new areas** for collaboration and agreement on:

1. global challenges
2. global research infrastructure
3. **open scientific research data**
4. **increasing access to the peer-reviewed, published results of scientific research**



G8 joint Science Ministers meeting (2)

Within that statement, it states....

3. Open Scientific Research Data

‘Open enquirycan provide society with the necessary information to solve global challenges

We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of scientific research are as widely available as practical, enable transparency in science and engage the public in the scientific process’



G8 joint Science Ministers meeting (3)

4. Expanding access to scientific research results...

‘.... The generation, sharing and exploitation of scientific knowledge are integral to the creation of wealth and the enhancement of our quality of life

We recognise that G8 nations have an important opportunity and responsibility to promote policies that increase access to the results of publicly funded research results to spur scientific discovery, enable better international collaboration and coordination of research, enhance the engagement of society and help support economic prosperity’

G8 Open data Charter



‘Access to data allows individuals and organisations to develop new insights and innovations.... While governments and businesses collect a wide range of data, **they do not always share these data in ways that are easily discoverable, useable, or understandable by the public**’

‘This is a missed opportunity’

Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

G8 governments agreed to follow a set of principles that will be the foundation for access to data. They are:

- Open data by default
- (ensure) Quality and quantity
- Useable by all (e.g. easy to find)
- Releasing data for improved governance
- Releasing data for innovation

Open data in the UK

- So, a year after the G8 meeting, in December 2014, the UK Government announced a “Science, Technology, and Innovation Strategy” - it requested the promotion of open science
- The UK is currently **on course** to make **all taxpayer-funded** research publications available in an open access format by (I think) 2019
- In 2016, an agreement, for the first time, proposes a set of expectations of **best practice** for publishing data



Report produced by
stakeholder group

Agreement principles

The agreement sets out ten principles. By committing to the principles the research community can demonstrate that they:

- safeguard good research practice
- conform to all ethical, legal and professional obligations
- nurture a research environment that makes data open wherever practical and affordable
- use transparent, robust and fair processes to make decisions concerning data openness
- have appropriate mechanisms in place to provide assurances as to the integrity of their research data; and
- recognise the importance of data citation and credit acknowledgement

Why disseminate your data?

To comply with:

- **Legislation and government guidance**
 - In the UK - Freedom of Information act
 - Describes how the public may obtain access to government-held information – full effect in 2005
 - In Europe - INSPIRE directive



- **Funder** requirements
- **Publisher** requirements

What is the INSPIRE Directive?



- Came into force in 2007, full implementation required by 2021
- **Set up because environmental issues** such as flooding, forest fires etc. **do not respect national borders**
- Aims to create an EU spatial data infrastructure - most environmental data, e.g. emission measurements, biodiversity observations, or environmental quality data is spatial
- Data should be collected only once
- Includes data about: species and habitats, population demographics, utility services, energy resources, marine features, metrological features and human health etc.
- It is possible to combine seamless spatial information from different sources across Europe and share it with many users

Aims of INSPIRE directive – an example

Following the eruption of the Eyjafjallajökullin volcano
in Iceland in April/May 2010

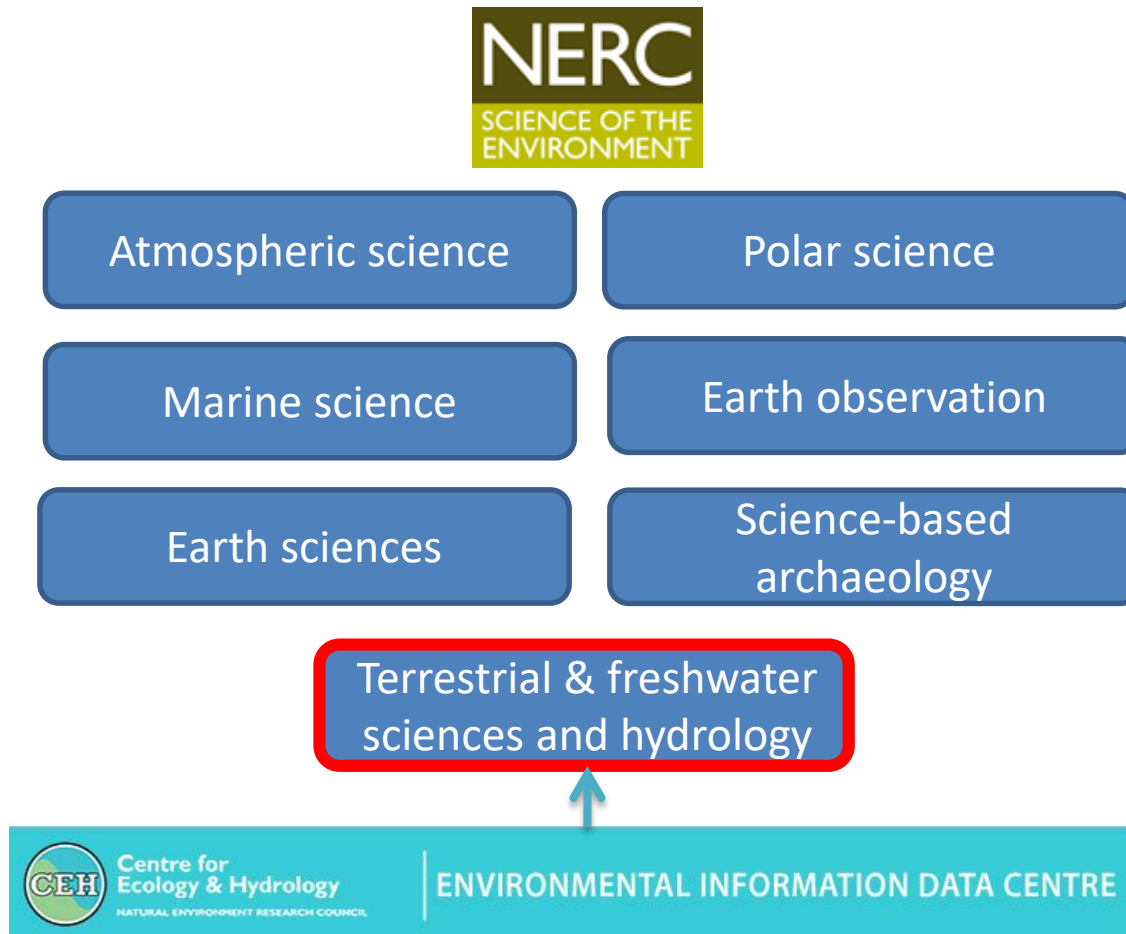


Needed data to be
available cross
country boundaries

data was not always compatible
and that caused problems

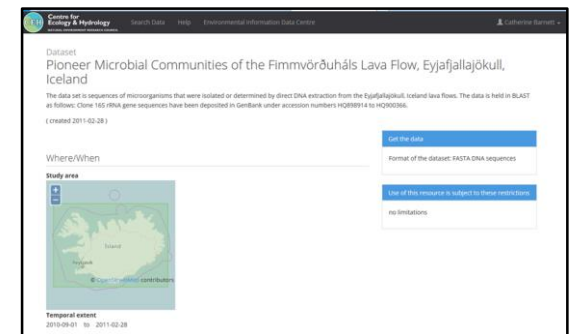
Effort has been made to
improve both the coverage
and compatibility

An example of INSPIRE compliant data centres



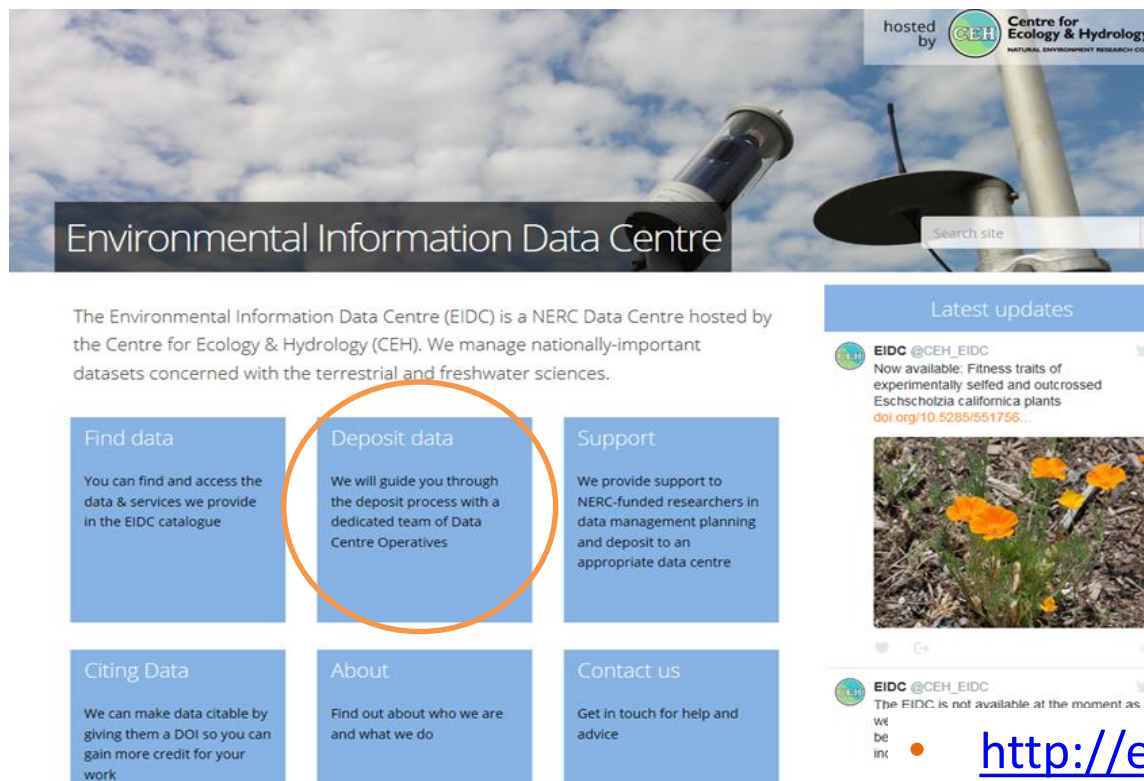
The NERC Data Centres

- They provide a **focal point** for NERC's scientific data and information
- Responsible for maintaining environmental data and **making them available** to all users



NERC Data Centres

- The Environmental Information Data Centre (EIDC) is **hosted by the Centre for Ecology & Hydrology**
- Provides **long-term data curation** of nationally-important datasets concerned with the terrestrial and freshwater sciences
- Gives **support and guidance** to those depositing data



I may refer to this later as an example of a **repository**



<http://eidc.ceh.ac.uk/>

The first time we used the EIDC.. in 2012

Centre for Ecology & Hydrology
Search Data Help Environmental Information Data Centre Catherine B.


Dataset
Observations of Fukushima fallout in Great Britain
Licence terms and conditions apply

Beresford, N.A.
<https://doi.org/10.5285/1a91c7d1-ec44-4858-9af2-98d80f169bbd>

Following 11 samples across countries, it which can be grass sample data and int

Publication date: 2011-12-31 (created 2011-08-03)

Where/When
Study area



Temporal extent

Get the data
Download the data
Supporting documentation
Format of the dataset: text/csv

Use of this resource is subject to these restrictions
Licence terms and conditions apply
You must cite: Beresford, N.A.; Barnett, C.L.; Howard, B.J.; Howard, D.C.; Tyler, A.N.; Bradley, S.; Copplestone, D. (2011). Observations of Fukushima fallout in Great Britain. NERC Environmental Information Data Centre. <https://doi.org/10.5285/1a91c7d1-ec44-4858-9af2-98d80f169bbd>

DOI recognises datasets as valued outputs



Guaranteed persistence

'Self-archiving'
Green
open access



Journal of Environmental Radioactivity

Volume 114, December 2012, Pages 48–53

Environmental Impacts of the Fukushima Accident (PART II)



Observations of Fukushima fallout in Great Britain

N.A. Beresford^a, C.L. Barnett^a, B.J. Howard^a, D.C. Howard^a, C. Wells^a, A.N. Tyler^b, S. Bradley^b, D. Copplestone^b



Within the text it states:

'All data associated with this study are available from the CEH Information Gateway (<https://gateway.ceh.ac.uk/>) and the data have been allocated a digital object identifier (<http://dx.doi.org/10.5285/1a91c7d1-ec44-4858-9af2-98d80f169bbd>)'



Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

Beresford, N.A., Barnett, C.L., Howard, B.J., Howard, D.C., Wells, C., Tyler, A.N., Bradley, S., Copplestone, D. 2012. Observations of Fukushima fallout in Great Britain. J Environ Radioact. 2012 Dec;114:48-53. doi: 10.1016/j.jenvrad.2011.12.008.

DOI - Linking citation to data record

Journals have
now improved
how datasets are
linked to papers

Links to detailed
description, data
access, metadata
(information about
the data)

Dataset
Microsatellite genotype data for five species of bumblebee across an agricultural landscape in Buckinghamshire, UK
This resource is made available under the terms of the Open Government Licence

Dreier, S.; Redhead, J.W.; Warren, I.; Bourke, A.F.G.; Heard, M.S.; Jordan, W.C.; Sumner, S.; Wang, J.; Carvell, C. (2014)
<https://doi.org/10.5285/6a408415-0575-49c6-af69-b568e343266d>

Microsatellite data for five species of common and declining bumblebee (*Bombus terrestris*, *B. lapidarius*, *B. pascuorum*, *B. hortorum* and *B. ruderatus*) collected from the Hillesden Estate, Buckinghamshire, UK, in summer 2011. Worker genotypes were determined from individuals sampled across an agricultural landscape. Genotypes were reconstructed from sampled worker offspring. Data were collected as part of a project led by the Centre for Ecology and Hydrology, funded under the Natural Environment Research Council's Environmental Information Data Centre Initiative.

Publication date: 2014-05-08

Where/When

Study area



Temporal extent

2011-06-20 to 2011-08-05

Online Resources

Supporting information
Supporting information available to assist with re-use of this dataset.

Download the data
Get a copy of this data

Admin

Edit Publish More

Get the data

Download the data

Supporting documentation

Format of the dataset: text/csv

Use of this resource is subject to these restrictions

This resource is made available under the terms of the Open Government Licence

You must cite: Dreier, S.; Redhead, J.W.; Warren, I.; Bourke, A.F.G.; Heard, M.S.; Jordan, W.C.; Sumner, S.; Wang, J.; Carvell, C. (2014). Microsatellite genotype data for five species of bumblebee across an agricultural landscape in Buckinghamshire, UK. NERC Environmental Information Data Centre. <https://doi.org/10.5285/6a408415-0575-49c6-af69-b568e343266d>

BioRx RIS

MOLECULAR ECOLOGY
Molecular Ecology 23, 3384–3395
doi: 10.1111/mec.12823

Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape

STEPHANIE DREIER,*† JOHN W. REDHEAD,‡ IAN A. WARREN,*† ANDREW F. G. BOURKE,§ MATTHEW S. HEARD,‡ WILLIAM C. JORDAN,* SEIRIAN SUMNER,*† JINLIANG WANG* and CLAIRE CARVELL‡

*Instit
Univers
Crouma
Norwich

FINE-SCALE GENETIC STRUCTURE OF BUMBLE BEES 3395

Data accessibility

Microsatellite genotype data for five species of bumblebee across an agricultural landscape in Buckinghamshire, UK. NERC Environmental Information Data Centre CEH:EIDC:1401189757278; doi:10.5285/6a408415-0575-49c6-af69-b568e343266d. Geographic distances between pairs of wild bumblebee colonies across an agricultural landscape in Buckinghamshire, UK: NERC Environmental Information Data Centre CEH:EIDC:1401193864362; doi:10.5285/8b3f4857-9809-43cb-b2aa-a988e64a5449.

agri-environmental habitat quality. *Agriculture, Ecosystems & Environment*, 178, 31–38.
Schmid-Hempel R, Schmid-Hempel P (2000) Female mating frequencies in *Bombus* spp. from Central Europe. *Insectes Sociaux*, 47, 36–41.
Stewart LC, Hale RJ, Hale ML (2010) Species-specific primers for the molecular identification of cryptic *Bombus* species in New Zealand. *Conservation Genetics*, 11, 1207–1209.
Stolle E, Rohde M, Vautrin D *et al.* (2009) Novel microsatellite DNA loci for *Bombus terrestris* (Linnaeus, 1758). *Molecular Ecology Resources*, 9, 1345–1352.
Truett GE, Heeger P, Mynatt RL *et al.* (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *BioTechniques*, 29(52), 54.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P

Important info:
abstract, authors,
Terms & Conditions of
access etc.

Chernobyl data papers – historical data

Dataset

Post Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and Britain

This resource is made available under the terms of the Open Government Licence

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015)

<https://doi.org/10.5285/d0a6a8bf-68f0-4935-8b43-4e597c3bf251>

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. NERC Environmental Information Data Centre. <https://doi.org/10.5285/d0a6a8bf-68f0-4935-8b43-4e597c3bf251>

Publication date: 2015-07-17

- Data from:
 - 'grey literature' reports
 - Previous journal publications
 - Unpublished data
- All combined into multiple 'linked' datasets and given a single doi
- Each dataset described in more detail in a 'data paper' - which in itself also has a doi

Download the data

Get a copy of this data

Supporting information

Supporting information is available to assist in the re-use of this dataset

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

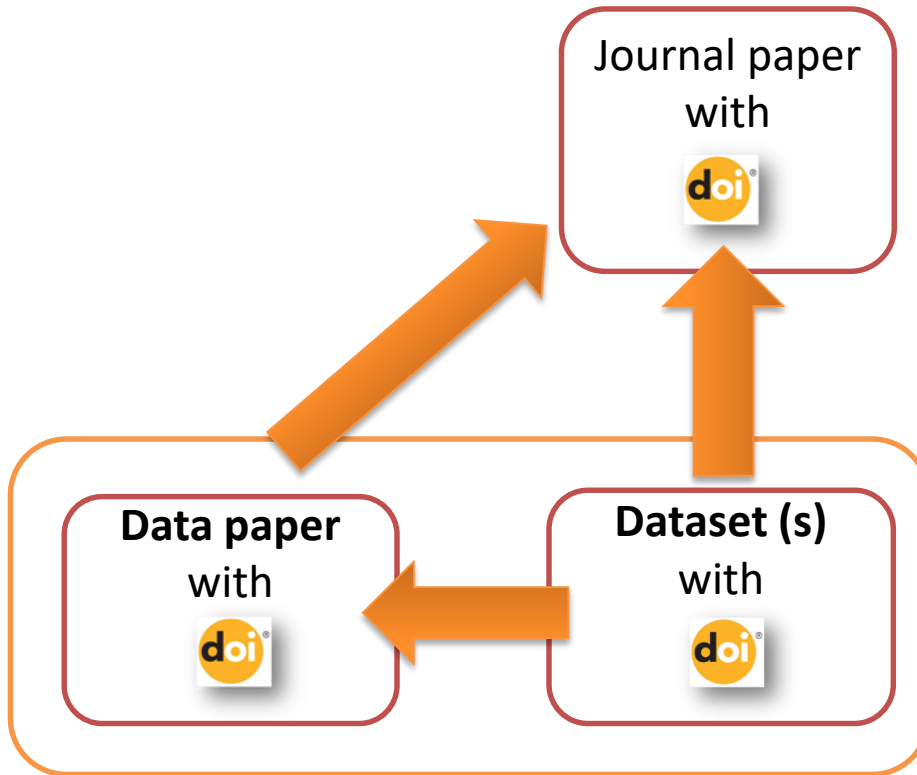
Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.

Journal paper

Chaplow, J.S.; Beresford, N.A.; Barnett, C.L. (2015). *Post-Chernobyl surveys of radiocaesium in soil, vegetation, wildlife and fungi in Great Britain*. *Earth System Science Data*, 7, 215-221. doi:10.5194/essd-7-215-2015, 2015.



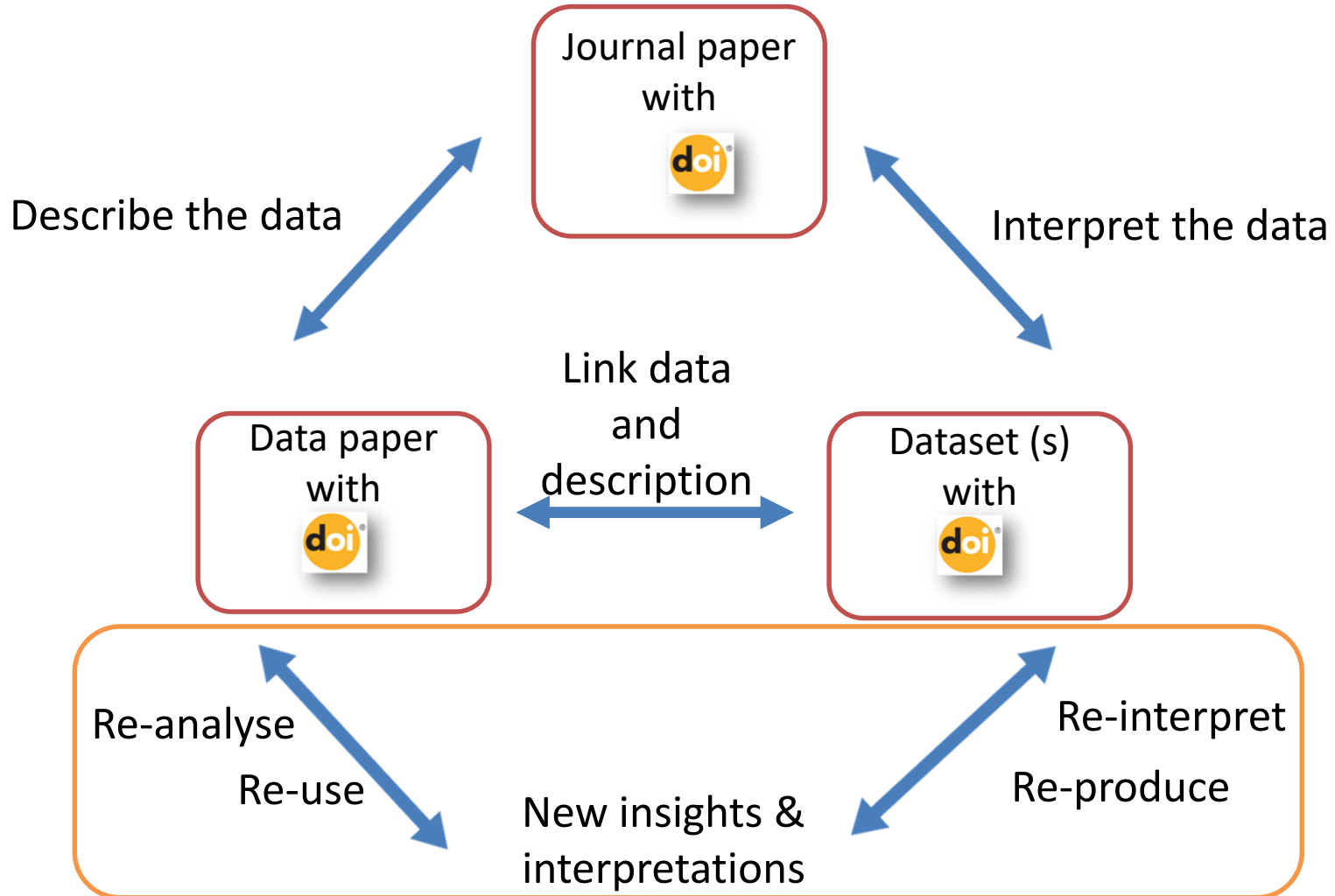
So, the process is.....



Good opportunity
for younger/less
experienced
scientists/analysts
to publish and gain
recognition

So why should we publish our data? (1)

It allows others to develop your work



So why should we publish our data? (2)

It also aids scientific debate

20 August 2010 Last updated at 10:17



Chernobyl species **decline** linked to DNA

By Victoria Gill
Science reporter, BBC News



The scientists have studied the exclusion zone for more than a decade.

Scientists working in Chernobyl have found a way to predict which species there are likely to be most severely damaged by radioactive contamination.

The secret to a species' vulnerability, they say, lies in its DNA.

This discovery could reveal which species are most likely to decline or even become extinct in response to other types of environmental stress.

The researchers published their findings in the Journal of Evolutionary Biology.

Professor Tim Mougeon from the University of South Carolina, US, and

BBC

Mammals decline in
Chernobyl zone

Last Updated: Thursday, 20 April 2006, 05:55 GMT 06:55 UK

E-mail this to a friend

Printable version

Wildlife **defies** Chernobyl radiation

By Stephen Mulvey
BBC News

It contains some of the most contaminated land in the world, yet it has become a haven for wildlife - a nature reserve in all but name.

The exclusion zone around the Chernobyl nuclear power station is teeming with life.

As humans were evacuated from the area 20 years ago, animals moved in. Existing populations multiplied and species not seen for decades, such as the lynx and eagle owl, began to return.

There are even tantalising footprints of a bear, an animal that has not trodden this part



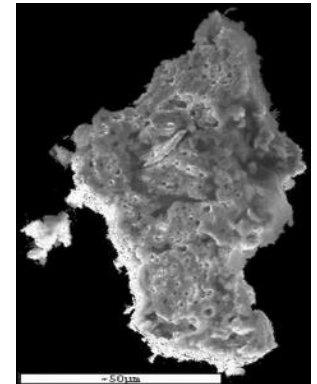
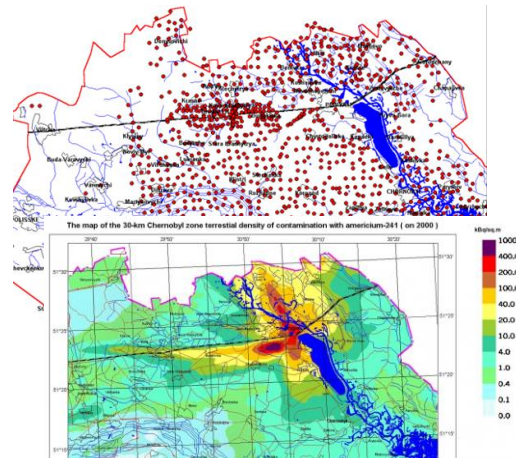
Conflicting articles are published, it leads to confusion
For verification, we need access to the data

Chernobyl data papers – new data

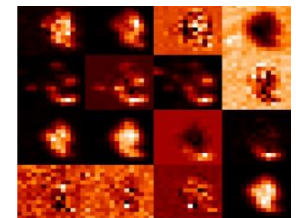
We are currently working with our Ukrainian colleagues to make new data available on wildlife, deposition and particles



Trap cameras



U fuel particle



Data will be available from the NERC Environmental Information Data Centre and published in a data paper soon!

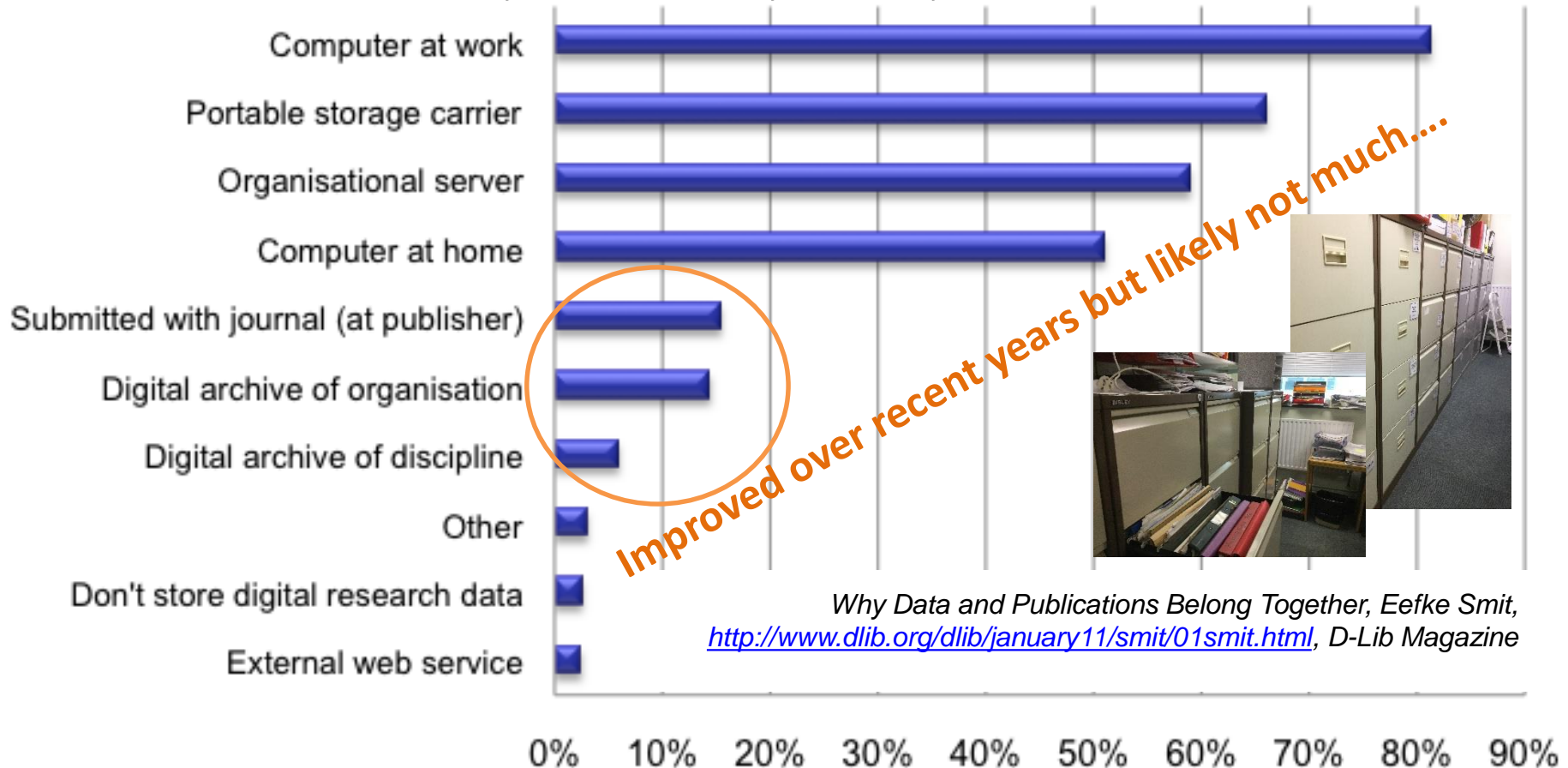
<http://www.radioecology-exchange.org/content/chernobyl-exclusion-zone>

<http://www.ceh.ac.uk/tree>

Data, data everywhere

But not everyone iswe have the opportunity to publish our data but it does not happen often

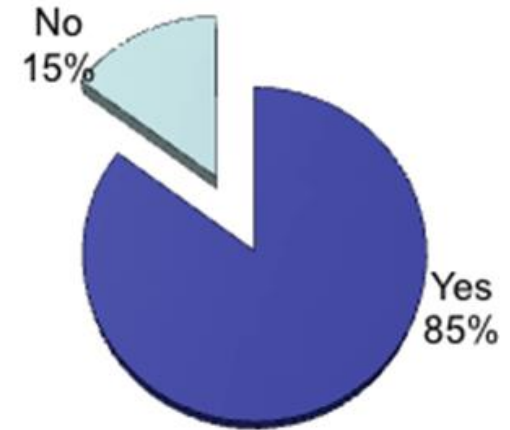
A few years ago 1202 US Researchers were asked
Where do you currently store your research data?



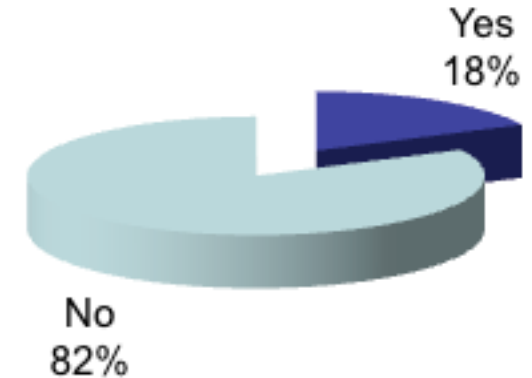
Attitudes towards sharing research data (1)

The same 1202 Researchers were also asked

Do you think it is **useful to** link underlying research data with formal literature (i.e. Journal articles) ?



Do the journals to which you typically submit your work to **require you to** include the data used to create your tables, figures, etc.?



Almost the exact opposite response was received !

Attitudes towards sharing research data (2)

In another study.... Leiden University's Centre for Science and Technology Studies and Elsevier embarked on a project to investigate open data practices

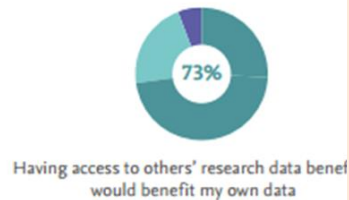
'Open Data: The Researcher Perspective'

'Open data practices facilitate collaboration, drives data analysis, and promotes transparency and reproducibility' but....



Figure 2. Attitudes towards sharing of research data (% , n=1162)

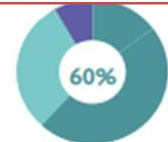
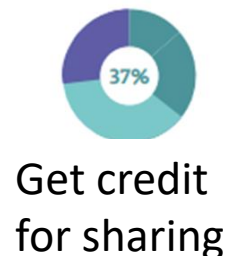
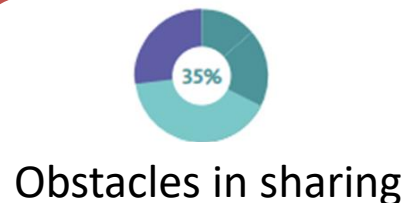
34% of researchers surveyed do not publish data at all



outside of my research team

I provide my research data to publishers so that it can be made accessible with my research article

Research data management specialists need to play a role in research data sharing



Strongly agree/Agree Neither agree nor disagree/Don't know Strongly disagree/Disagree

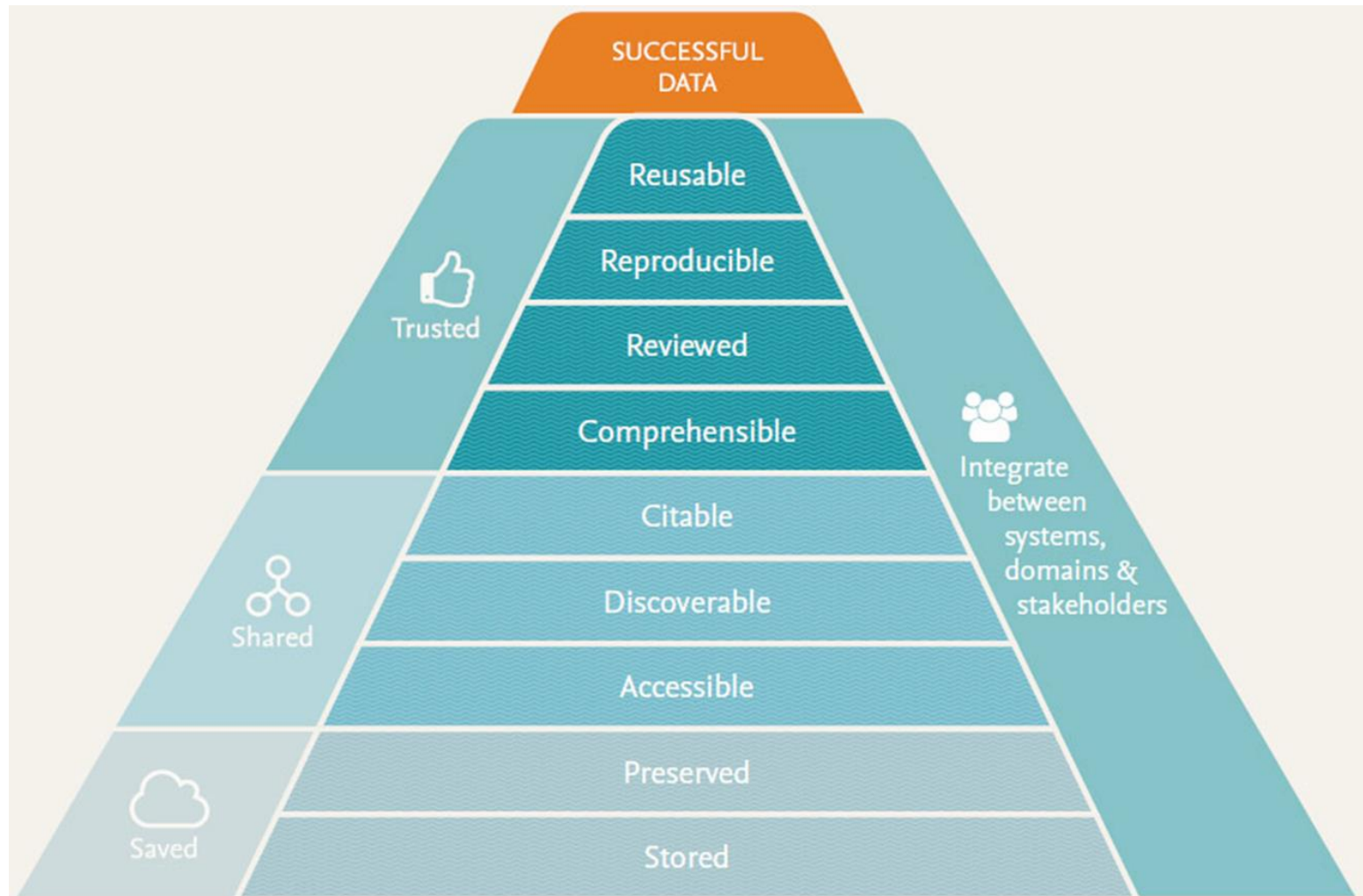
<https://www.universiteitleiden.nl/en/research/research-output/social-and-behavioural-sciences/open-data-the-researcher-perspective>

So what can be done about it?

.....perhaps make it easier

Making data 'effective'

Elsevier have published their interpretation of 'successful data' showing the nine steps involved



Making data effective



Making data effective



Making data effective

Both researchers and machines may want to access data e.g for meta-analysis or re-use. There are a **number of different ways to do this**, via:

- Via **repositories**
- A **data sharing system** e.g. Mendeley Data where users create private data sharing spaces that can be opened up to wider groups
- Publisher systems such as Elsevier's Open data pilot which makes **raw research data** (as submitted with an article) **openly accessible alongside the article** for any web user to access
- Pay publisher for **Gold open access** (an article processing charge)



Making data effective

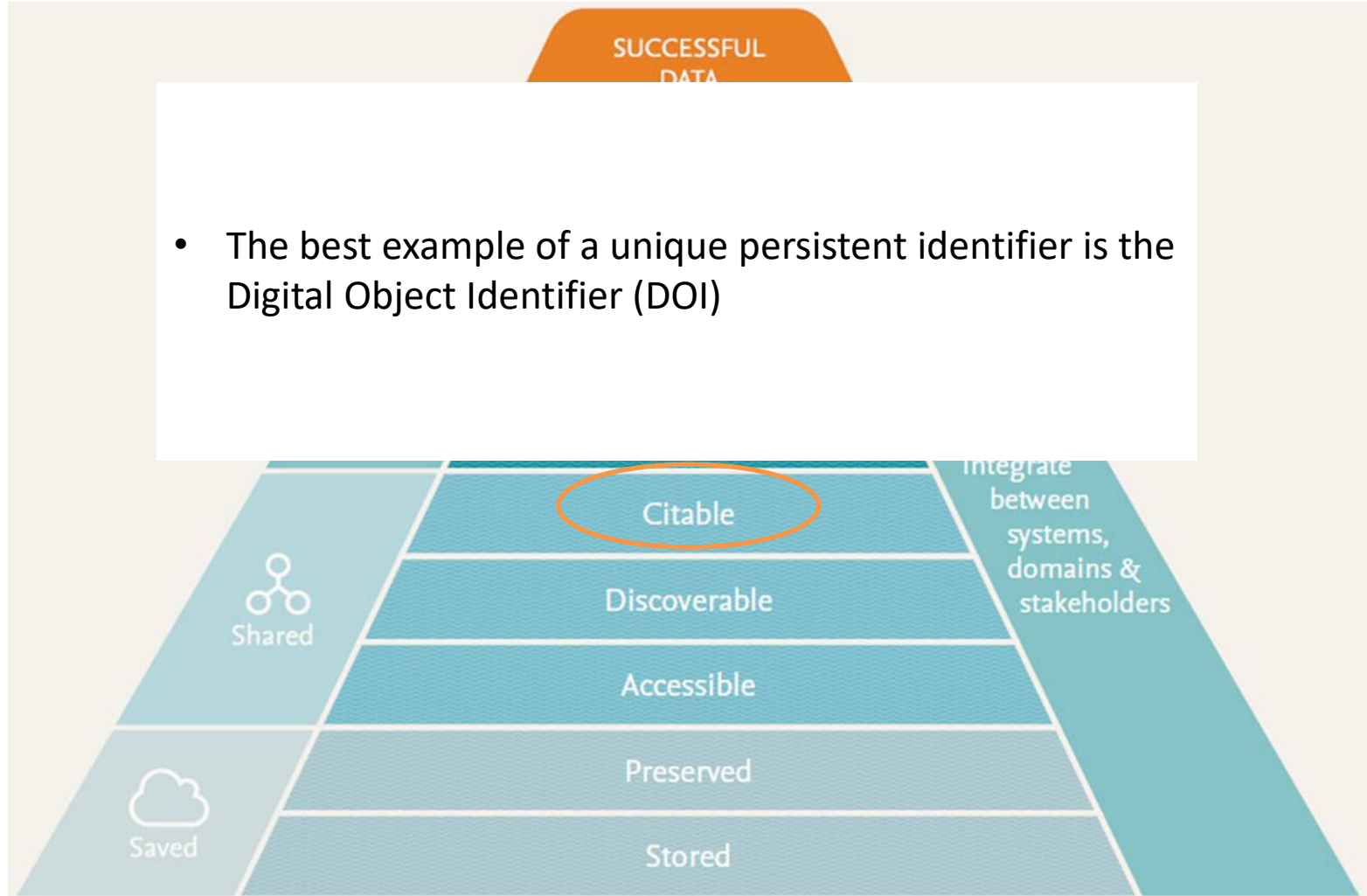
Finding scientific papers is now very straightforward, this is **still not yet the case for research data**

- An easy way to make data more discoverable is to **link articles to the data sets** the articles are based on
- Publishers collaborate with data repositories to ensure data is more easily discoverable
- Data **'indexing'** services exist (i.e. on-line lists)
- Data **search engines** are being developed

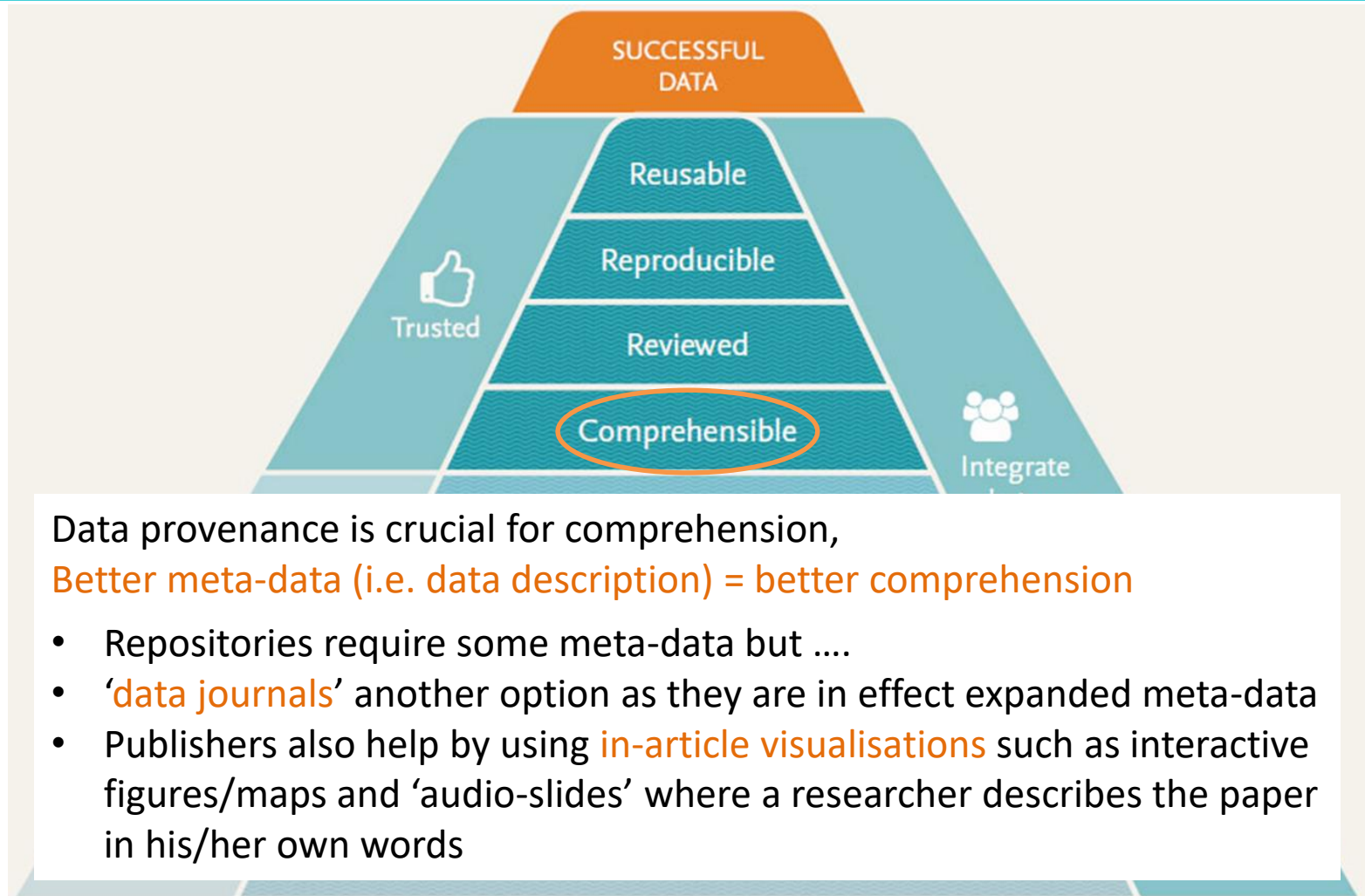


Making data effective

- The best example of a unique persistent identifier is the Digital Object Identifier (DOI)



Making data effective



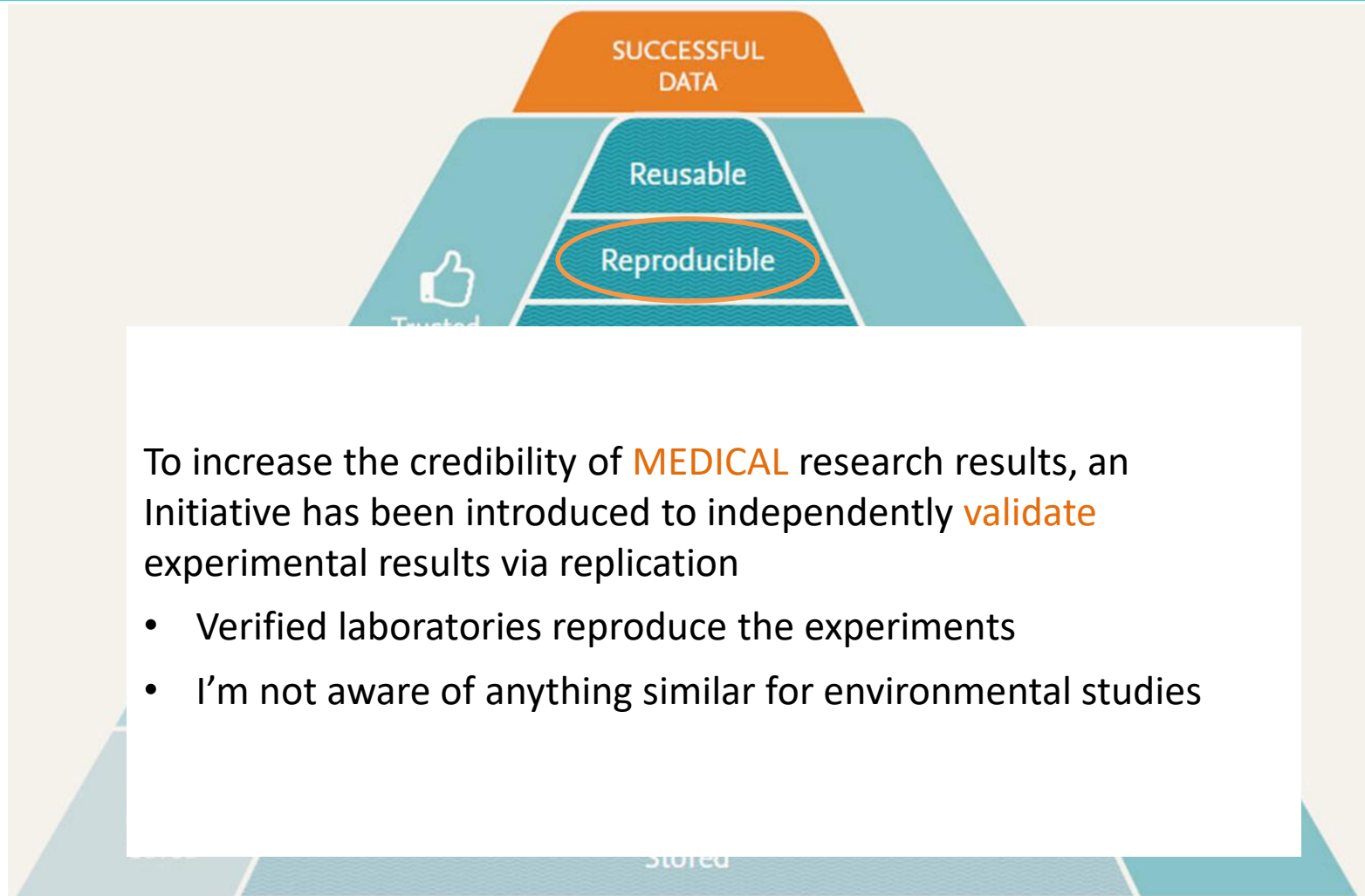
Making data effective



It is common for research articles to be peer reviewed, but it is still quite uncommon for research data to be

- There are different degrees of review between repositories, journals and data journals
- Useful if this was addressed as it is an important step when it comes to quality control and trustworthiness of data

Making data effective



To increase the credibility of **MEDICAL** research results, an Initiative has been introduced to independently **validate** experimental results via replication

- Verified laboratories reproduce the experiments
- I'm not aware of anything similar for environmental studies

Making data effective



The **key benefit** of having shared research data **is the ability to reuse it**

- However, only when it is sufficiently trustworthy and reproducible will other researchers re-use it
- It is recommended to attach a **user license** to datasets so any user can clearly understand what they can and cannot do with the data. It can also help ensure they give researchers and data creators the appropriate credit
- There are a variety of user licenses available with the most common ones being Creative Commons

Database design – a little advice

- **Consult a database expert if you can** - they may suggest things you may not think of
- **Plan as well as you can** - ensure you design to capture all the information you need at the start – talk to all your collaborators
- **Sharing** – do all your collaborators have equal knowledge?
 - perhaps **provide a template** with some examples on how the database should be filled in, it may save time explaining or correcting mistakes
 - Consider the use of **drop down menus** to restrict input into cells
- **Compatibility** - be careful when combining multiple datasets/databases from different sources/collaborators – **people like to add extra information e.g. extra columns!**
- **QC** - **Consider automated routines** they may save you time. e.g. Excel macros or perhaps some of the data validation routines

So how do you prepare a dataset for publication?

- **Documentation** = Have good Meta-data, ensure it is descriptive
- **Filename, formatting and content**
 - Filenames: Keep them short & meaningful. Ensure they have no spaces or special characters (e.g. \$* @%) and reflect the content e.g. FukushimaCs137soil2017.csv
 - Column headings should be unique, meaningful and in the first row only, avoid special characters (e.g. \$* @%)
 - Variables, abbreviations and codes should be unique, meaningful, consistent and either self-explanatory or explained in the metadata
 - Missing values should be handled consistently
 - There should be no unexplained characters or codes e.g. n/d, n/a, x, they should all be explained in the metadata

Format/structure/machine readability all make data more usable

Data papers – how to write one

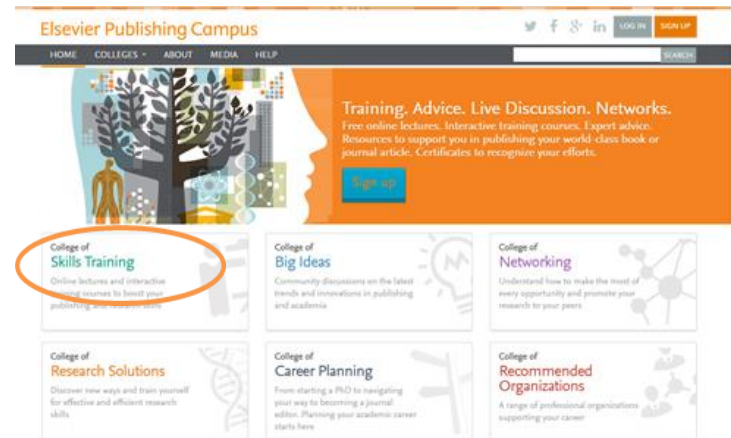
<https://www.publishingcampus.elsevier.com/>

There is help!



Publishing Campus

	Do	Don't
In general	<ul style="list-style-type: none"> The journal Data in Brief publishes a very specific article type that allows you to describe and share data. Follow the Data in Brief template to prepare your article. Review all submission instructions. Use the word 'data' throughout your data article whenever possible. Spell out acronyms and abbreviations. Include enough information so that your data article can be read on its own. 	<ul style="list-style-type: none"> Data articles published by Data in Brief are not microresearch papers or short communications. Don't include a Conclusion, Discussion or Summary section. Don't use words such as 'Wells', 'results', 'background', etc. Don't provide excessive background information (but refer to published articles whenever possible). Data in Brief is primarily a venue to describe original data rather than new methods.
Title	<ul style="list-style-type: none"> The title should be unique and focus on the data you are sharing. If possible use the words 'data', 'dataset', etc. in the title. 	<ul style="list-style-type: none"> Don't use the same title as an existing research paper. Don't use words such as 'effects', 'trends', 'responses', 'implications', 'influence' etc. in the title.
Abstract	<ul style="list-style-type: none"> The abstract should be purely descriptive and describe the contents of the data article. If the data article is related to a research article, include the abbreviated reference in the abstract. 	<ul style="list-style-type: none"> Don't copy the abstract from a research paper. Don't describe any related research article, results, or provide too much background.
Specifications table	<ul style="list-style-type: none"> Fill in the fields in the Specifications table. Remove all instructions from the table, if a field is not applicable, leave it blank (e.g. data source location). Keep the fields in the table brief. If your dataset is deposited in a public repository, include the link in 'Data availability'. If it is not necessary to reproduce the data in the Data in Brief article. 	<ul style="list-style-type: none"> Don't leave the examples provided in the table or they may end up in your final article. Don't include too much information in for example the 'Experimental factors' and 'Experimental features' fields. Instead, move the details to the Materials and Methods in the text. Don't include links to data in repositories that are behind a firewall.
Value of the Data	<ul style="list-style-type: none"> Value of the Data should include 3-5 broad, general, short bullet points. Put into perspective the utility of the data beyond what you used it for in your own research. Explain how/why the data you are sharing in this Data in Brief article could be generally valuable to the scientific community, with an eye towards possibly opening up doors for new collaborations. 	<ul style="list-style-type: none"> Don't offer interpretative statements or conclusions about the data nor state the value of this data relative to any related research study. Don't make conclusions about any data not presented in your data article.



- College of Skills Training is useful
- Some of the 'quick guides' are in Japanese – e.g. how to get published

論文の発表に向けた心得

良い投稿と悪い投稿の分け目

良い投稿は...	悪い投稿は...
<ul style="list-style-type: none"> データが正確で、透明性が高い。 データが公開されている。 データが適切な方法で収集されている。 データが適切な方法で分析されている。 データが適切な方法で報告されている。 データが適切な方法で解釈されている。 データが適切な方法で議論されている。 データが適切な方法で結論づけられている。 	<ul style="list-style-type: none"> データが不正確で、透明性が低い。 データが公開されていない。 データが不適切な方法で収集されている。 データが不適切な方法で分析されている。 データが不適切な方法で報告されている。 データが不適切な方法で解釈されている。 データが不適切な方法で議論されている。 データが不適切な方法で結論づけられている。

Where we have published in 'data journals'



Impact factor: 6.696

Earth System Science indexed | Science Citation Index Expanded, Current Contents/PCE, Scopus, ADS, CLOCKSS, CNKI, DOAJ, EBSCO, Gale/Cengage, GeoRef, GoOA (CAS), Google Scholar, J-Gate, Portico, ProQuest, World Public Library



2-year impact factor: 4.259

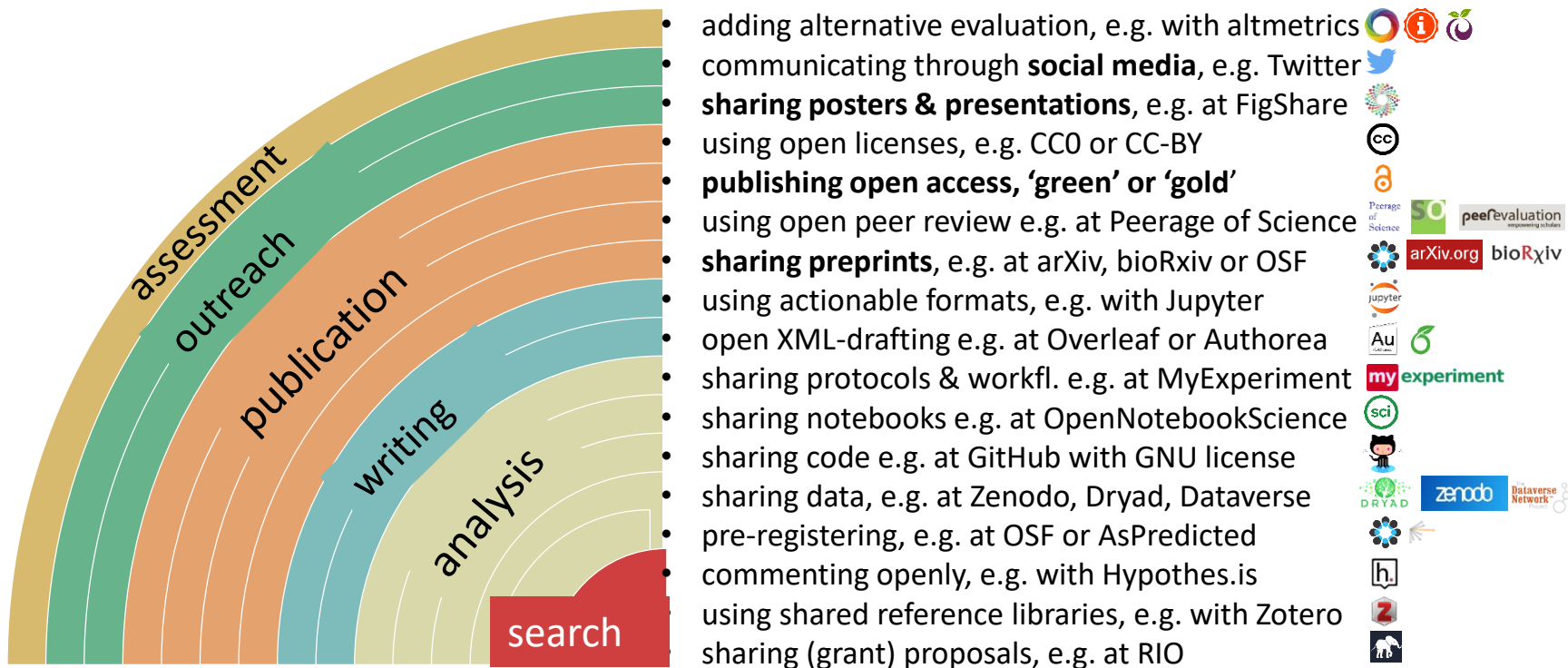
Scientific Reports is indexed in ISI Web of Science, PubMed, PubMed Central, Scopus, Google Scholar, and SAO/NASA ADS.

Note the relatively good impact factors

How to make your work more open...

So, having got your data, how can else can you share it.....

At many stages Admit I'm not familiar with quite a lot of these, but they are options!

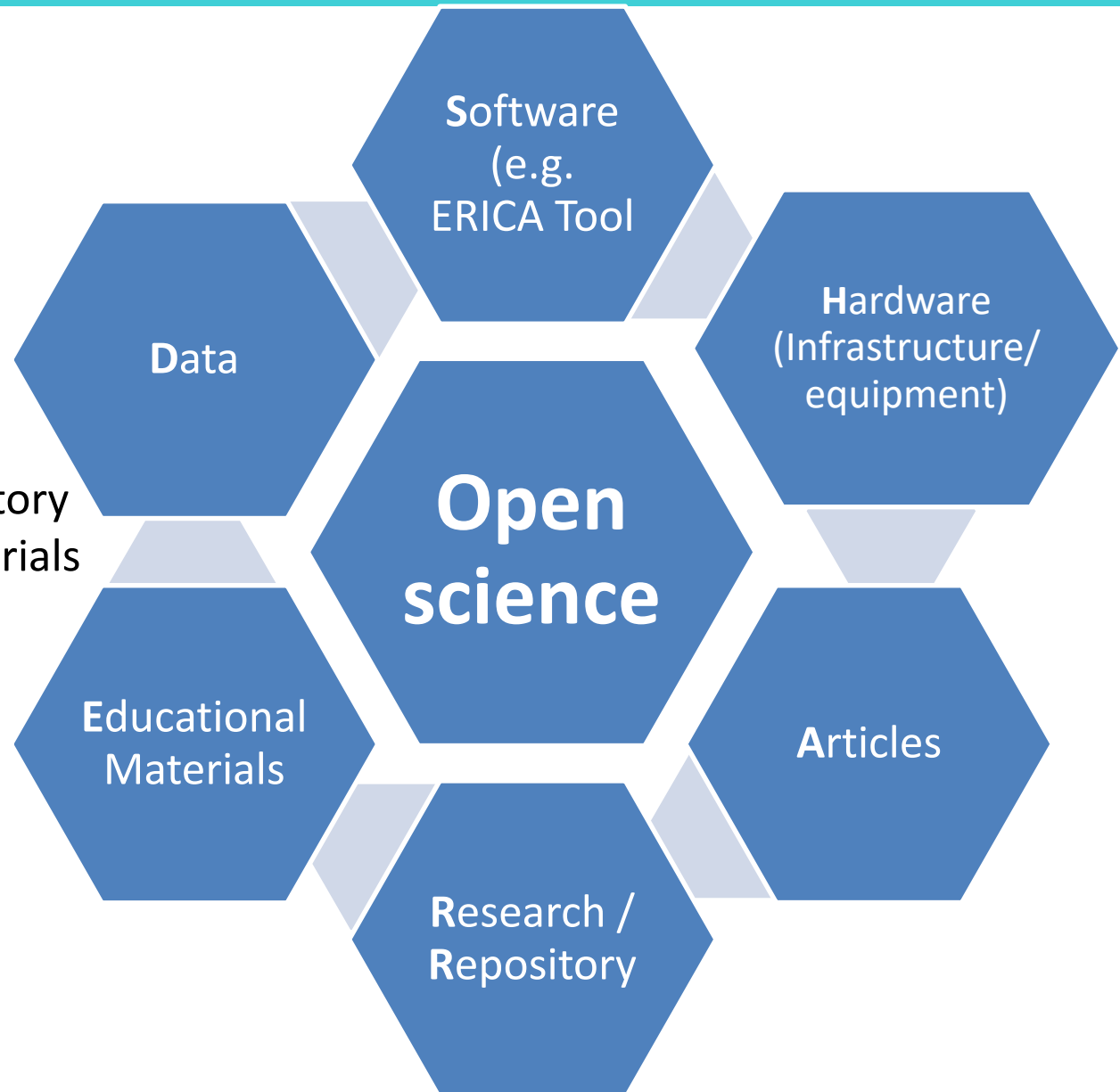


Slide adapted from:
<https://doi.org/10.6084/m9.figshare.5065534>



OPEN science = SHARED science

- Open **S**oftware
- Open **H**ardware
- Open **A**rticles
- Open **R**esearch/**R**epository
- Open **E**ducational materials
- Open **D**ata



Training

There is training available this is an example

The screenshot shows the homepage of the MANTRA (Research Data Management Training) website. The header features the 'MANTRA' logo and the text 'Research Data Management Training' against a background of colorful, abstract patterns. Below the header, a descriptive paragraph states: 'MANTRA is a free online course for those who manage digital data as part of their research project.' To the right of this text are four icons representing different user roles: Research Student, Career Researcher, Senior Academic, and Information Professional. A navigation bar contains links for Home, About, Acknowledgements, DIY Training Kit for Librarians, Feedback, and Contact Us. Below the navigation bar, a section titled 'Learning Units: Select one to start' displays a grid of nine learning unit buttons, each with an icon and an information icon (i). The units are: Research data explained >, File formats & transformation >, Data protection, rights & access >, Data management plans >, Documentation, metadata, citation >, Sharing, preservation & licensing >, Organising data >, Storage & security >, and Data handling tutorials > (highlighted in orange). The 'Organising data >' button has a sub-list: Housekeeping, File versioning, and Naming & re-naming conventions. A star rating system (5 stars) and a 'Rate MANTRA (121 Votes)' button are also visible.

MANTRA
Research Data Management Training

MANTRA is a free online course for those who manage digital data as part of their research project.

Research Student Career Researcher Senior Academic Information Professional

Home About Acknowledgements DIY Training Kit for Librarians Feedback Contact Us

Learning Units: Select one to start ★★★★★ Rate MANTRA (121 Votes)

Research data explained > i

File formats & transformation > i

Data protection, rights & access > i

Data management plans > i

Documentation, metadata, citation > i

Sharing, preservation & licensing > i

Organising data > i

Storage & security > i

Data handling tutorials > i

- Housekeeping
- File versioning
- Naming & re-naming conventions

Available under an open licence for reuse.
Contains videos and lectures.

So what are the advantages of sharing data?



Lots more reasons.....

- No loss when staff/students leave
- Cross disciplinary research possible
- Less duplication
- Less bias
- Cross disciplinary research
- Greater reproducibility
- Further testing of datasets
- Meta-analysis of datasets
- Best information used for R&D
- Opportunities for staff mobility
- Knowledge Exchange



Figure adapted from: Danny Kingsley and Sarah Brown; Other sources of information:
<https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data>

And the challenges.....

...in data sharing

- Most sharing currently occurs among **collaborators**
- Data sharing practices are **not easily standardised**
- **Privacy and ethical issues** hinder the transfer of data

...in data management

- Researchers are **not aware of** data sharing mandates
- Data management plans are **not used consistently**
- **Extra staff** perhaps required for data handling

...in perceptions on sharing and reuse

- Standards for citing another researcher's data are **not universally understood**
- **Global and national differences** need to be addressed
- Licensing issues and formats are **not well understood**
- **Researchers feel they are the drivers of data sharing**

And the opportunities....

...in data sharing

Researchers do recognise the importance of data sharing

- Researchers are already sharing data in ways that can be **optimised**
- Collaborative practices can be used to further **streamline** data sharing
- **Cross-disciplinary** sharing provides big opportunities

...in data management

- Training is required in data management and Creative Commons licencing practices etc.
- need for increasing funding of data management activities

...in perceptions on sharing and reuse

- The scientific credibility system could **reward participation** in open data practices, e.g., through publishing in data journals
- Need for **improving standardisation** and harmonization of processes, e.g. citation practices

Adapted from report: https://www.elsevier.com/data/assets/pdf_file/0004/281920/Open-data-report.pdf

Summary

This presentation has covered

- Background on data sharing in UK and Europe
- How to store and disseminate your data
 - Safely storing data in a repository and getting a DOI from a data centre or via a 'data paper'
- The many advantages, challenges and opportunities of sharing data



Summary

Whilst preparing this presentation I came across this quote
in a report promoting open science in Japan

“Open science may change scientific research.
It will not replace traditional research methods,
but will add new tools to advance science”

I agree!

*Promoting Open Science in Japan, Opening up a new era for the advancement of science,
The Expert Panel on Open Science, based on Global Perspectives Cabinet office, Government of Japan.*

March 30, 2015.

http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_en1.pdf

Thank you for listening

