WILEY

**Enriching the Research Life Cycle:
From Journal Archives to Open Data**
豐富研究循環：由典藏期刊到開放數據

James Murphy, Director, Research & Professional

Enriching the Research Life Cycle:
豐富研究循環：

From Journal Archives to Open Data

- Journal archives

  - The importance of journal archives to research

  - Journal archive usage examples

- Open Data

  - Defining research data

  - The importance and value of data

  - Data publishing

WILEY

The Importance of Journal Archives to Research
典藏期刊對研究的重要性

WILEY

In the year 2000

15%

of papers read by researchers were more than 5-years old

在**2000**年，研究人員閱讀的文章中有**15%**超過**5**年

Tenopir & King, *Learned Publishing* (2002) 15, 259-265

WILEY

In 2011

30%

of papers read by
researchers were more
than 5-years old

2011年，研究人員閱讀的文章中，有**30%**超過**5**年

Tenopir , Volentine,& King, *Learned Publishing* (2012) 25, 279-291

WILEY

And yet only …

70%

of publishers have

90%+ of their backfiles online

然而，70%出版社把自家90%回溯期刊電子化

The STM Report, Fourth Edition

WILEY

20%

of publishers have less than 50% of backfiles digitized; and that's just publishers, not journals!

只有20％出版社擁有不到50％電子回溯期刊，此數字僅代表出版社，而非期刊！
The STM Report, Fourth Edition

WILEY

Wiley Backfile Usage Examples:
Wiley回溯期刊使用率範例：
France and Germany
法國與德國

WILEY

France National Backfile Usage

Full-Text Downloads

To be provided

Source: Matthew Barker, Analyst, Research Insights, Wiley

WILEY

France National Backfile Usage

Access Denied

To be
provided

WILEY

Germany

Potential for future scale

To be provided

Open Data and Data Publication
開放數據和數據出版

Research data underlies all the material that we publish – for every conclusion that is made in a paper, massive amounts of data was collected to reach and confirm that result, both in that paper and in the papers that came before it.

WILEY

Results and statistics are data

結果與統計都是數據

WILEY

```r
source("http://www.r-statistics.com/wp-
content/uploads/2012/01/source_https.r.txt") # Making sure we can
source code from github
source_https("https://raw.github.com/talgalili/R-code-
snippets/master/clustergram.r")

set.seed(250)
Data <- rbind(
                        cbind(rnorm(100,1, sd =
0.3)          d = 0.3),rnorm(100,0, sd = 0.3),rnorm(100,0, sd =
0
                        cbind(rnorm(100,0, sd =
        rnorm(100,0, sd = 0.3),rnorm(100,0, sd =

                        cbind(rnorm(100,0, sd =
        rm(100,1, sd = 0.3),rnorm(100,0, sd =

                        cbind(rnorm(100,0, sd =
        rm(100,0, sd = 0.3),rnorm(100,1, sd =

                        )
```

Code is data

代碼是數據

WILEY

Media is data

媒體是數據

Metadata is data

內設資料是數據

WILEY
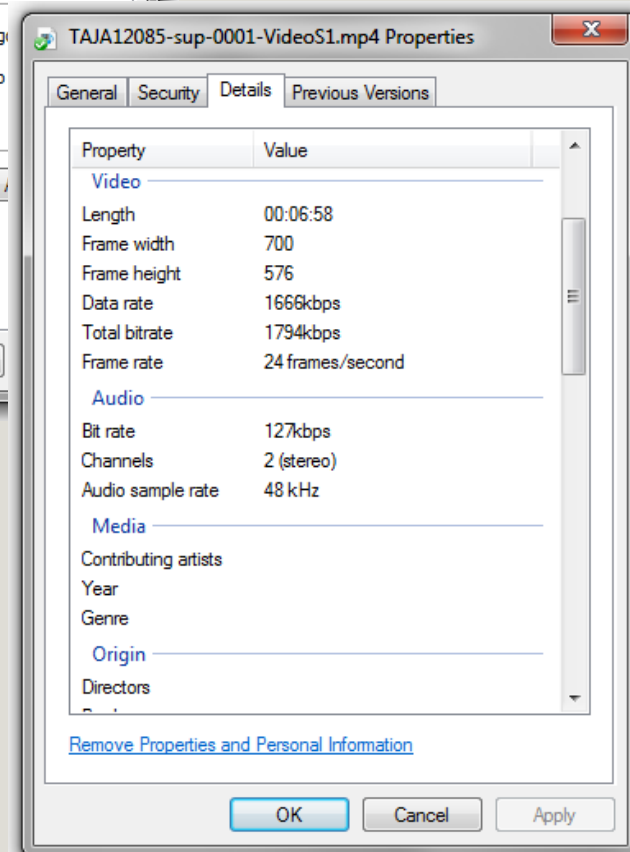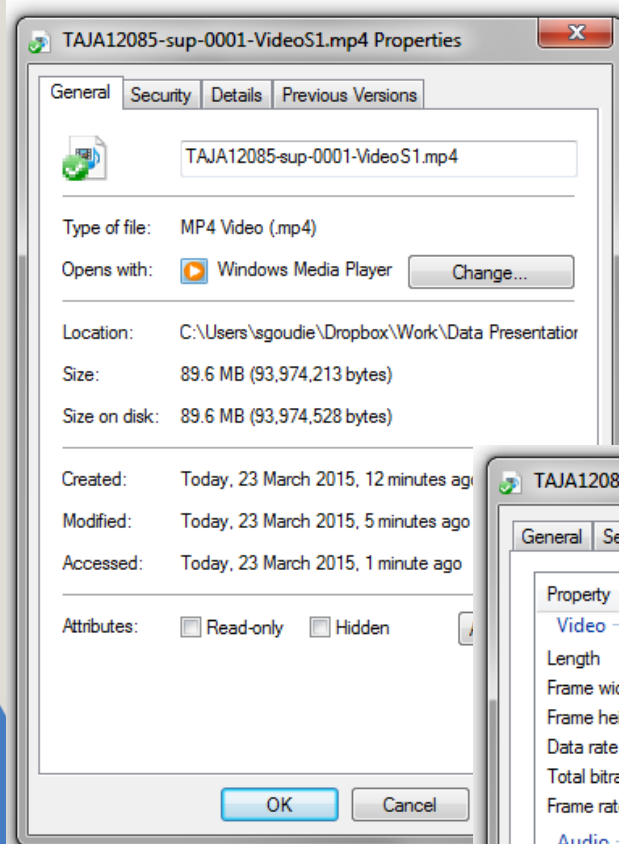
What is research data?
何謂研究數據**?**

- **Research data** means data in the form of **facts**, **observations**, **images**, **computer program results**, **recordings**, **measurements** or **experiences** on which an argument, theory, test or hypothesis, or another **research output is based**.

- Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media.

- **Metadata** means information or facts **about** research data for the purpose of attribution, description, management, verification and discovery.

  (QUT MOPP: http://www.mopp.qut.edu.au/D/D_02_08.jsp)

WILEY

Why is data becoming a major issue?
為何數據成為一個主要議題？

Technology

- No longer restricted to publishing on paper

- Massive amounts of data can be processed, stored, and shared quickly

- Researchers have the ability to discover, combine and analyze data from many sources easily

- Allows access to information sets that would previously have never been reasonable to collect, or to compare many studies together

WILEY

Why is data becoming a major issue?
為何數據成為一個主要議題？

Capturing the value of data

- Research data are the 'output' of research (not articles!)

- Research funders pay for research to occur

- Funders want to get the most value from their investment

- Value ('impact') comes from research data
being made available to influence future research, policies and
practices

WILEY
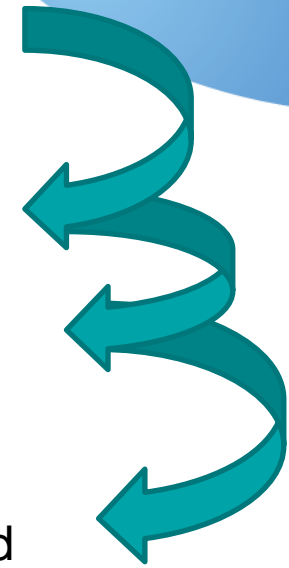
Why is data becoming a major issue?
為何數據成為一個主要議題？

Improving Science

- There has always been the responsibility for research published to be able to be interrogated, replicated and verified

- Easier access to data enables collaboration and much faster innovation in research

WILEY

Why is data becoming a major issue?
為何數據成為一個主要議題？

Funder Mandates

Many funding bodies now mandate that data produced from funded research be made available, similar to OA mandates

The SHERPA/JULIET site can help you review which mandates different funders require:

Funding Organisation (*linked to their JULIET summary*)	Publications Policy		Data Archiving Policy	Country
	Open Access Publishing	Open Access Archiving		
Australian Research Council (ARC)		— — —	— —	Australia
Australian Research Council [2013] (ARC)	✓	✓ ✓ —	— —	Australia
National Health and Medical Research Council (NHMRC)	—	✓ ✓ —		Australia

WILEY

SHERPA/JULIET

Publications Policy

Open Access Archiving

⩗	**Whether to Archive:**	* Requires deposition in Open Access archives
⩗	**What to Archive:**	* Any publications * Publisher's version *and/or* Author's final version * Unspecified format
—	**When to Archive:**	* At the earliest possible opportunity * Acceptable embargo: up to 12 months after publication * Metadata must be deposited at the date of publication
	Where to Archive:	* In appropriate institutional repositories (Required) * In appropriate disciplinary repositories (Alternative) * PubMed Central (Alternative)
	Archiving Conditions:	* Publisher's copyright, licensing & * embargo policies must be respected

Open Access Publishing

—	**Whether to Publish:**	* Encourages publication in Open Access publications
	Where to Publish:	* in a peer-reviewed open access journal (Alternative)
	Publishing Conditions:	* [No information]

General

General Conditions:	* Effective for all new projects from 01-Jul-2012 * Applies to all projects funded totally or partly by the National Health and Medical Research Council * Articles accepted for publication before the 1st of July 2012 but published after this date will be exempt from the revised policy on the dissemination of research findings. * Compliance with the policy is a matter for the Administering Institution to discuss with the NHMRC the NHMRC will not routinely check compliance with individual Chief Investigators (CIs). * NHMRC does not intend to place restrictions on the types of publications that can be included in an institutional repository. The Administering Institution should assist researchers to identify and capture appropriate information. * NHMRC does not intend to place restrictions on the types of publications that can be included in an institutional repository. The Administering Institution should assist researchers to identify and capture appropriate information. * If the copyright transfer/licence agreement does not allow the article (or manuscript) to be made available within twelve months of the date of publication, it needs to be made available as soon as possible after that date. * If the journal never allows the article to be made available, this information must be provided at the time of Final Report submission. * If the print version (journal version) of the article is openly accessible via the publishers website or via a service such as PubMed Central, it is sufficient to just make the article metadata available in the institutional repository and provide a link * If no institutional repository is immediately available to a Chief Investigator, this will need to be recorded in the grant Final Report.
Policy Links:	* Policy guidance - document entitled 'Program Grant Funding Policy' * Dissemination of Research Findings

Data Archiving Policy

Whether to Archive:	* No policy for deposition in Open Access archives
What to Archive:	* Unspecified data
When to Archive:	* Not specified
Where to Archive:	* Not specified
General Conditions:	* [No information]
Policy Links:	* [No information]

RESEARCHER DATA SHARING INSIGHTS

WILEY

- Wiley's Researcher Data Insights Survey was launched earlier this year to understand how and why researchers make their research data publicly available. The study's results, highlighted below, are intended to advance the global conversation about data sharing and help Wiley better meet the needs of our researchers, authors, and partners in the rapidly evolving landscape of scientific research and communications.

- The survey was deployed in March 2014 and received more than 2,250 responses from researchers around the world.

GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.
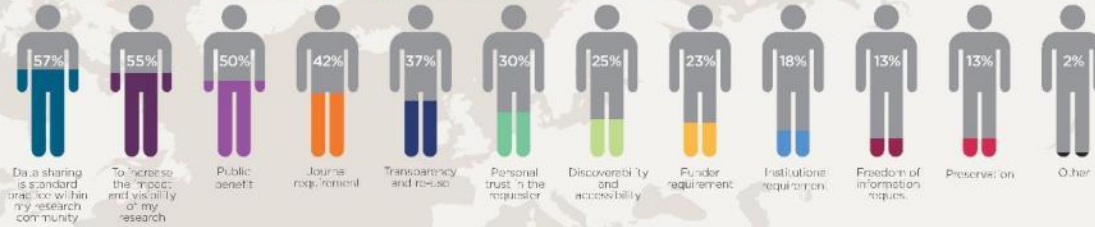
48% DO NOT SHARE DATA
52% SHARE DATA

WAYS DATA IS SHARED

- **67%** As supplementary material in a journal
- **37%** Personal, institutional or project webpage
- **26%** Institutional data repository (i.e. university or institute-sponsored)
- **19%** Discipline-specific data repository
- **6%** General-purpose data repository (e.g. Dryad, figshare)
- **5%** Other

Globally, researchers also report sharing their data in limited and non-permanent ways; 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).

RESEARCHER MOTIVATIONS FOR SHARING DATA

%	Motivation
57%	Data sharing is standard practice within my research community
55%	To increase the impact and visibility of my research
50%	Public benefit
42%	Journal requirement
37%	Transparency and re-use
30%	Personal trust in the requestor
25%	Discoverability and accessibility
23%	Funder requirement
18%	Institutional requirement
13%	Freedom of information issues
13%	Preservation
2%	Other

DATA SHARING TRENDS BY COUNTRY

UNITED STATES — 46% SHARING / 54% NOT SHARING
Among researchers in the US sharing their data publicly, two out of three do so because it is standard practice in the communities and because they believe it benefits the public. Similar to their counterparts in the UK, the majority of US-based researchers also store data to increase the impact or visibility of their research.

UNITED KINGDOM — 43% SHARING / 57% NOT SHARING
While more than 40% of UK researchers are sharing data, only about 14% are using discipline-specific or other public repositories like Dryad and figshare. The two key drivers that motivate UK researchers to share their data are the prospect of gaining increased impact or visibility for their work and to satisfy funder requirements.

JAPAN — 44% SHARING / 56% NOT SHARING
Compared with their counterparts around the world, researchers in Japan cite concerns about being scooped as a reason for not sharing data more frequently. Nearly five out of ten Japanese researchers point to this as a reason for not sharing their data, roughly double the global average.

CHINA — 36% SHARING / 64% NOT SHARING
Nearly five in ten Chinese researchers say they are not sharing data because they are not required to do so by their funders or institutions. They are more likely than their global counterparts to say that they do not see data sharing as a personal responsibility and plan to take direction from funders to guide their data sharing decisions in the future.

BRAZIL — 52% SHARING / 48% NOT SHARING
Two out of three researchers in Brazil say that a guarantee of proper credit or attribution would compel them to share more of their data publicly in the future.

AUSTRALIA — 41% SHARING / 59% NOT SHARING
Researchers in Australia say they would be most incentivized to make their data accessible in the future to ensure preservation as well as transparency and re-use. The majority of researchers also ranked funder requirements among top reasons to share in the future.

GERMANY — 55% SHARING / 45% NOT SHARING
Among German researchers sharing their data publicly, three out of four are driven to share data because they believe it will increase the visibility of their research and want to ensure public transparency and re-use. About 20% of German researchers are making use of general-purpose repositories (like figshare and Dryad), significantly more than their counterparts around the world, including those in the US and UK.

REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

- **42%** Intellectual property or confidentiality issues
- **36%** My funder/institution does not require data sharing
- **26%** I am concerned that my research will be scooped
- **26%** I am concerned about misinterpretation or misuse
- **23%** Ethical concerns
- **22%** I am concerned about being given proper citation credit or attribution
- **21%** I did not know where to share my data
- **20%** Insufficient time and/or resources
- **16%** I did not know how to share my data
- **12%** I don't think it is my responsibility
- **12%** I did not consider the data to be relevant
- **11%** Lack of funding
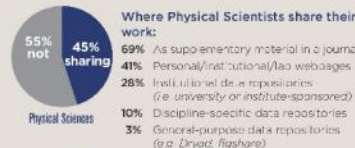- **7%** Other

DATA SHARING BY DISCIPLINE

Data sharing, specifically by way of data repositories, is most prevalent amongst life scientists, particularly those in the earth and environmental and agriculture and food sciences.
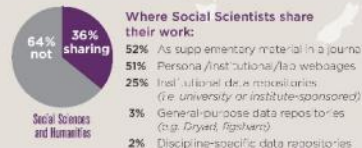
Where Health Scientists share their work:
52% not sharing / 48% sharing
- **68%** As supplementary material in a journal
- **29%** Personal/institutional/lab webpages
- **29%** Institutional data repositories (i.e. university or institute-sponsored)
- **21%** Discipline-specific data repositories
- **5%** General-purpose data repositories (e.g. Dryad, figshare)

A typical *Health Science researcher* says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assuaged.

Where Life Scientists share their work:
34% not / 66% sharing
- **76%** As supplementary material in a journal
- **42%** Discipline-specific data repositories
- **29%** Personal/institutional/lab webpages
- **23%** Institutional data repositories (i.e. university or institute-sponsored)
- **13%** General-purpose data repositories (e.g. Dryad, figshare)

A typical *Life Science researcher* says she would be motivated to share more of her data in the future if she was guaranteed proper credit.

Where Physical Scientists share their work:
55% not / 45% sharing
- **69%** As supplementary material in a journal
- **41%** Personal/institutional/lab webpages
- **28%** Institutional data repositories (i.e. university or institute-sponsored)
- **10%** Discipline-specific data repositories
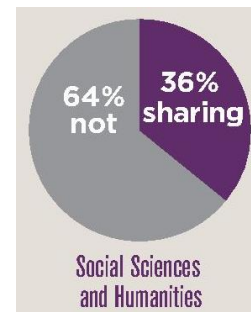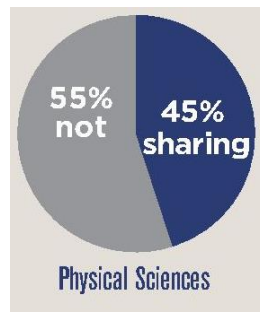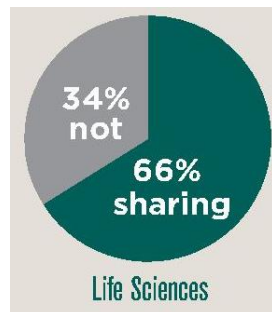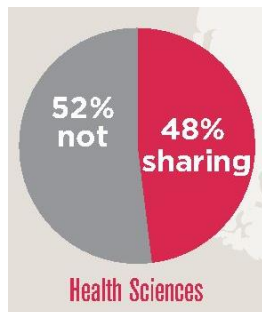- **3%** General-purpose data repositories (e.g. Dryad, figshare)

A typical *Physical Science researcher* says she would be motivated to share her data in the future because it is standard practice within her research community and because it increases the impact and visibility of her work.
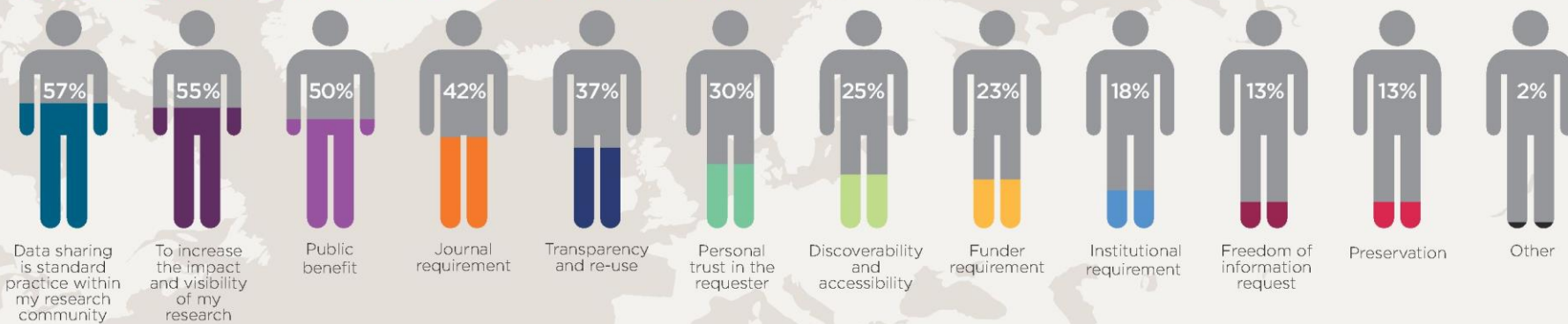
Where Social Scientists share their work:
64% not / 36% sharing
- **52%** As supplementary material in a journal
- **51%** Personal/institutional/lab webpages
- **25%** Institutional data repositories (i.e. university or institute-sponsored)
- **3%** General-purpose data repositories (e.g. Dryad, figshare)
- **2%** Discipline-specific data repositories

A typical *Social Science and Humanities researcher* says she would be motivated to share her data in the future if increased the impact and visibility of her work or if she was required to by her funder.

http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/

WILEY

Health Sciences: 52% not sharing, 48% sharing

Life Sciences: 34% not, 66% sharing

Physical Sciences: 55% not, 45% sharing

Social Sciences and Humanities: 64% not, 36% sharing

RESEARCHER MOTIVATIONS FOR SHARING DATA

57% — Data sharing is standard practice within my research community

55% — To increase the impact and visibility of my research

50% — Public benefit

42% — Journal requirement

37% — Transparency and re-use

30% — Personal trust in the requester

25% — Discoverability and accessibility

23% — Funder requirement

18% — Institutional requirement

13% — Freedom of information request

13% — Preservation

2% — Other

DATA SHARING TRENDS BY COUNTRY

USA: 46% SHARING, 54% NOT SHARING

UK: 43% SHARING, 57% NOT SHARING

Japan: 44% SHARING, 56% NOT SHARING

China: 36% SHARING, 64% NOT SHARING

Brazil: 52% SHARING, 48% NOT SHARING

Australia: 41% SHARING, 59% NOT SHARING

Germany: 55% SHARING, 45% NOT SHARING

WILEY

So, how much is this all worth?

那麼，這些值多少錢？

"potential value of research data repositories for Australia might be at least $1.8 billion and possibly up to $5.5 billion per annum"

Table 2   The annual value accruing within Australia from curating and openly sharing public research data (summary of estimates)

Estimate	Labour Costs (Lower Bound)	Total Expenditure (Upper Bound)
Current value of data in public research	**$1.9 billion to $2.9 billion**	**$4.0 billion to $6.2 billion**
- Use value (cost of data activity time)	$2.0 billion to $2.9 billion	$4.1 billion to $6.2 billion
- Estimated return on investment (at 40%)	$1.9 billion to $2.9 billion	$4.0 billion to $6.0 billion
Potential value of data repositories	**$1.8 billion to $2.6 billion**	**$3.7 billion to $5.5 billion**
- Efficiency impacts	$1.4 billion to $2.1 billion	$3.0 billion to $4.5 billion
- Additional (re)use return on investment	$370 million to $495 million	$690 million to $1.0 billion
Unrealised upside of curation & sharing	**$1.4 billion to $2.4 billion**	**$2.9 billion to $4.9 billion**

Note: Lower bound estimates are based on the labour-cost share and upper bounds estimates on total research funding and expenditure.
Source: Authors' analysis.

'Open Research Data', Report to the Australian National Data Service (ANDS) November 2014
John Houghton Victoria Institute of Strategic Economic Studies & **Nicholas Gruen** Lateral Economics

WILEY

Data Publication
數據出版

WILEY

How data is published
如何出版數據

WILEY

The Data Publication Pyramid
數據出版金字塔

Current State
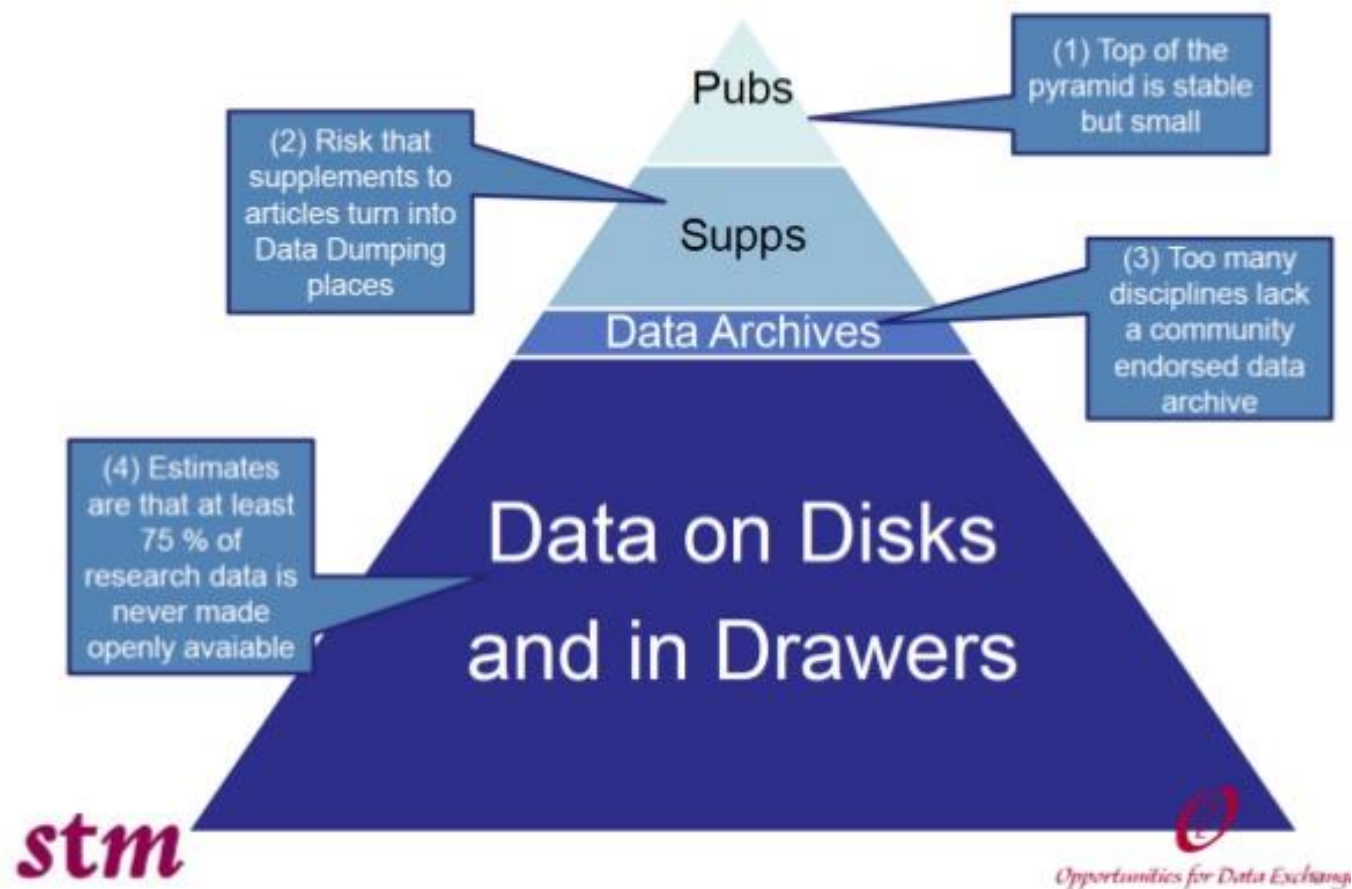
WILEY

The Data Publication Pyramid

數據出版金字塔

The Ideal Pyramid
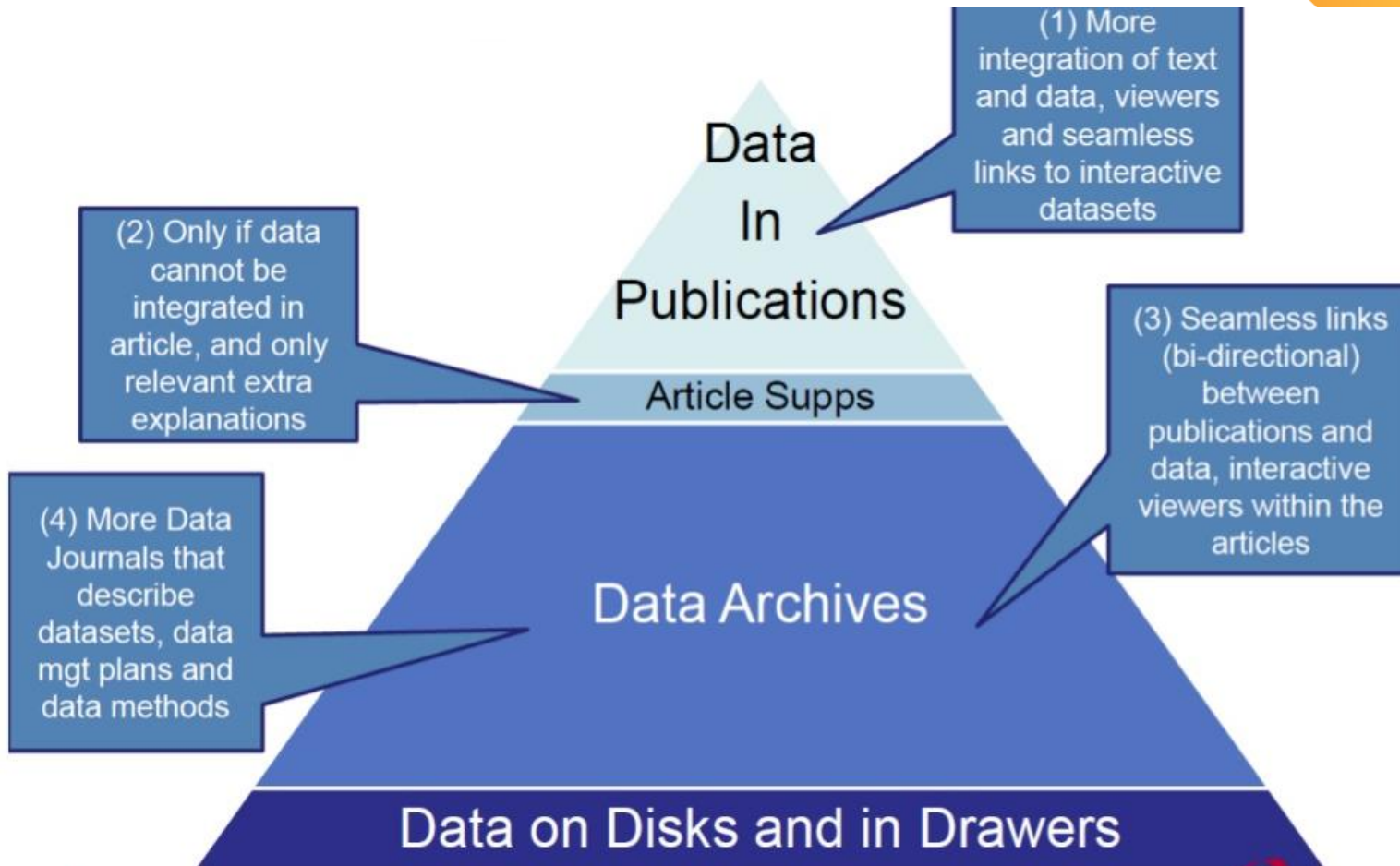
WILEY

Opportunities for Data Exchange

Keys to publishing data
出版數據關鍵

1. Availability

2. Discoverability

3. Interpretability

4. Reusability

5. Citability

6. Curation

7. Preservation

WILEY

Keys to publishing data
出版數據關鍵

1. Availability	
2. Discoverability	Able to find and use data
3. Interpretability	
4. Reusability	
5. Citability	Able to reference data
6. Curation	
7. Preservation	Data are looked after

WILEY

WILEY

Wiley Data Citation Policy
Wiley數據引用政策

"In recognition of the significance of data as an output of research effort, **this journal requires data to be cited in the same way as articles**. This is appropriate for data held within institutional or more general data repositories. It is not intended to take the place of community standards such as in-line citation of Genbank accession codes.

If citing or making claims based on data, please refer to the data in the text and provide a formal citation in the reference list. We recommend the format proposed by the Joint Declaration of Data Citation Principles: Authors; Year; Dataset title; Data repository or archive; Version (if any); Persistent identifier (e.g. DOI)."

Wiley Data Sharing Policy
Wiley數據分享政策

"…**expects that data supporting the results in the paper will be archived in an appropriate public repository.** Whenever possible the scripts and other artefacts used to generate the analyses presented in the paper should also be publicly archived. Exceptions may be granted at the discretion of the editor for sensitive information such as human subject data or the location of endangered species. Authors will be able to complete a data accessibility statement to be published with their paper. Further guidance is available at http://olabout.wiley.com/WileyCDA/Section/id-828082.html"

WILEY

How do I share my data?

1 **Review the Policy**

Review the data sharing policy on your selected journal's homepage.

2 **Submit your Data**

- Login or register for a ScholarOne account to submit your manuscript and data.

- During the file upload step in the submission process, choose "Data Files" from the drop-down menu to upload your data.

- Your data will be available for peer review, if you submit data file(s) by the time review commences.

- If you plan to take advantage of this service Data files must be submitted by article acceptance.
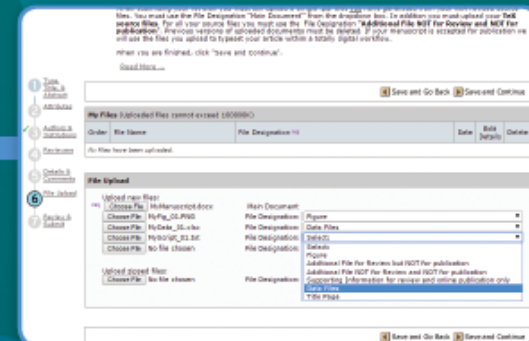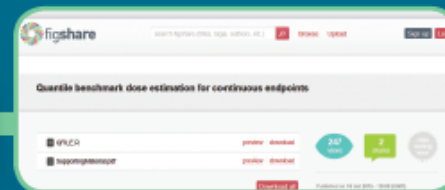
3 **Publication**

- If your article is accepted for publication, your data file(s) is automatically deposited into the journal's figshare data repository without charge or further work.

- On publication your data is made publicly available under a CC0 license on figshare.

4 **View your Data**

- Your article appears on figshare, with links back to your article on Wiley Online Library.

- Your published article on Wiley Online Library has a link to the data on figshare along with an accessibility statement.

WILEY

Open Science

Facilitating faster and more effective research discovery

Driven through state-of-the-art publishing technology

Accessibility of publicly funded research	Reproducibility and verification of researcher data, methodology and reporting standards	Collaborative, interoperable and global dissemination standards	More inclusive and networked research practices	Integration of researcher identification and evaluation tools
Open Access	Open Data	Open Standards	Open Collaboration	Open Recognition & Reward
ADVANCED SCIENCE — GOLD / GREEN — CHORUS	figshare — DRYAD	STANDARD ANALYTICS		ORCiD — KUDOS — Altmetric — publons

WILEY

Thank you

WILEY