

asiocedit

How to widely share your research data

*AsiaEdit Author Success Webinar
10 March 2022*



*Dr Trevor Lane, DPhil, PGCELT
Education Consultant*

asiocedit

1 Data ethics & accessibility

2 Journal data sharing policies


3 Tips for data publishing

1 Data ethics & accessibility asiocedit

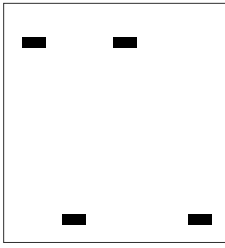
Data sources

Which data sources should be shared?

A Original files




↓



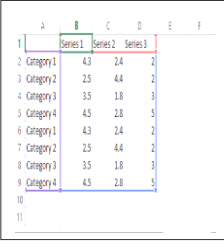
✓

Transparency

B Underlying data




↓



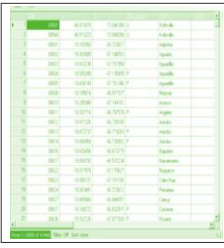
✓

Verification

C IPD: individual patient/
participant data (anonymised)



↓



✓


Reuse → Open Research, Open Data

How to widely share your research data 3

1 Data ethics & accessibility asiocedit

Open Research / Science


<https://www.ox.ac.uk/research/support-researchers/open-research>



UNIVERSITY OF
OXFORD

Open Research / Science / Scholarship


is the practice of research that enables others to **collaborate, validate and contribute**, where research outputs, data, protocols, lab notes and other research processes are *freely available*, under terms that enable **reuse, redistribution and reproduction** of the research and its underlying data and methods.



2022
MARCH
05
International
Open Data Day

- Open access
- Open data
- Reproducibility
- Research integrity
- Responsible metrics

- **Open access** is free, online access to research publications
- **Open data** is sharing data for scrutiny and reuse, to drive transparency, improve cooperation and strengthen the quality of research
- **Research integrity** requires adherence to consistently high standards, together with openness and transparency, thereby securing public trust in, and support for, research, while also ensuring that research findings can be used as effectively as possible



DORA San Francisco Declaration on Research Assessment
<https://sfdora.org/>


How to widely share your research data 4

Open Science

1 Data ethics & accessibility

asiocedit

https://unesdoc.unesco.org/ark:/48223/pf0000379949_locale=en
United Nations Educational, Scientific and Cultural Organization



UNESCO Recommendation on Open Science

Open Science


...combines various movements and practices aiming to make multilingual scientific knowledge **openly available, accessible and reusable** for everyone, to **increase scientific collaborations and sharing** of information for the benefits of science and society, and to **open the processes** of scientific knowledge creation, evaluation and communication to societal actors *beyond the traditional scientific community*.

Open scientific knowledge

Open science infrastructures & communication

Open engagement of societal actors

Open dialogue with other knowledge systems



- [COVID-19 Clinical Research Coalition](#)
- [Data Sharing Working Group](#)

How to widely share your research data
5

Data management

1 Data ethics & accessibility

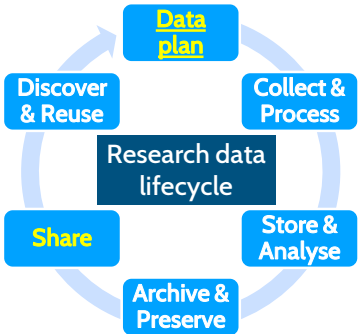
asiocedit

Ethical obligation to share data

Duke & Porter. *BioScience*, 2013;63(6):483–489. <https://doi.org/10.1525/bio.2013.63.6.10>


Replication	Original data need to be available for reanalysis
Data integrity	Multiple users to detect any data deficiencies
Preservation	Systematic archiving

Progress	Discover, access & analyse diverse datasets
Public trust	...in results of scientific research
Human rights	...to scientific progress & applications



Data management & sharing plan

Protocols	Study purpose, budget/team, data collection, format	Analyses	Data processing, models, tests/software, codebook
Handling/Storage	Records, privacy, management, quality control, data lock, secure archiving	Sharing/Accessibility	Metadata/data/code formats, journals/databases, DOIs, sharing conditions



Build your Data Management Plan
<https://dmptool.org/>

How to widely share your research data
6

1 Data ethics & accessibility asiocedit

Data ethics 1

Unethical data reporting/sharing

A Dishonest crop/splice

Data not shown

B Dishonest edit, outlier removal

Data not shown

C Publishing/sharing identifying information

D Publishing/sharing without consent/permission

E Duplicate publication

How to widely share your research data 7

1 Data ethics & accessibility asiocedit

Data ethics 2

Unethical data collection/analysis

Fabrication

Falsification

Plagiarism

Ethics/Consent

- No approval
- No consent
- Participation
- Publication
- Sharing

Underpowered

Sampling bias

Cherry picking

Hiding data, methods

Fishing trip

P-hacking

HARK = hypothesising after results are known

How to widely share your research data 8

Public data sharing

1 Data ethics & accessibility

asicedit

- Compulsory (by journal, funder)?
- Free / paid?
- Digital object identifier (DOI)?
- (Creative Commons, CC) licence?
- Updates/versioning? Embargo?
- "FAIR" principles?

How to widely share your research data

9

asicedit

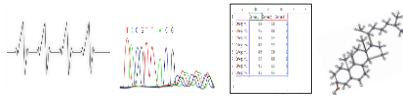
- 1 Data ethics & accessibility
- 2 **Journal data sharing policies**
- 3 Tips for data publishing

How to widely share your research data

10


2 Journal data sharing policies asiocedit

Data types



Experimental data

- Data points in tables, spreadsheets
- Biological sequences
- Diagrams, schematics, chemical structures
- Anonymised tracings/images, scans
- Photographs (microscopy, blot, gel)
- (Statistical) Code, algorithms, apps
- Software



Observational data

- Anonymised surveys
- Texts, transcriptions
- Field notes, interview notes
- Text coding/themes
- Maps, charts
- Video, audio
- Photographs
- Archival documents

United States Office of Management and Budget:
Research data is defined as the **recorded factual material** commonly accepted in the scientific community as **necessary to validate research findings**
...NOT: trade secrets, identifiable personal/medical data

How to widely share your research data 11


2 Journal data sharing policies asiocedit

FAIR principles


<https://www.go-fair.org/fair-principles/>

What is FAIR data & metadata?

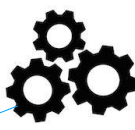
F
Findable




A
Accessible



I
Interoperable



R
Reusable



Persistent identifier

Eg, Digital object identifier (**DOI**), from DataCite, Crossref

<https://doi.org/10.1000/182>
doi:10.1000/182

Discoverable

Public open access (OA) repositories

<https://repositoryfinder.datacite.org/>

Single interface

- Common programming language
- Controlled vocabulary
- Non-proprietary

Complete & useful

- Creative Commons (CC) licence for reuse
- Complete, well-described data & metadata
- Context (protocol, codebook, code)

Eg, CSV, JSON, JSON lines, Parquet, R, Python, rtf / txt text, TIFF, MPEG4, SAS, SPSS, GZ, BZ2, LZ4, LZMA, ZSTD

How to widely share your research data 12

Reuse licences

Journal data sharing policies

asiocedit

https://en.wikipedia.org/wiki/Creative_Commons_license#media:File:Creative_commons_license_spectrum.svg
<https://creativecommons.org/about/cclicenses/>

Creative Commons (CC) user licences

CC0 (aka CC Zero) is a public dedication tool, which allows creators to give up their copyright and put their works into the worldwide public domain. CC0 allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions.

- Works funded by some governments
- Open Data

BY ⓘ Credit must be given to the creator

SA ⌚ Adaptations must be shared under the same terms

NC ₤ Only noncommercial uses of the work are permitted

ND Ⓣ No derivatives or adaptations of the work are permitted

Example: © 2022. This work is licensed under a [CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

© may belong to publisher or author
SHERPA/RoMEO – database of publisher copyright policies

How to widely share your research data
13

Open access journals

Journal data sharing policies

asiocedit

“Open access” (OA) journal types

Gold

Online articles immediately free to readers, eg, under CC licence; Author may have to **pay APC** (article processing charge)

Free to read

Diamond/Platinum

Online articles immediately free to readers, eg, under CC licence; **no APC**

Free to read

Free to publish

Bronze

Online articles immediately free to readers without CC licence

Black

Paywalled articles are illegally made free / pirated

Toll access

- **Subscription journal:** Free to publish some articles may be OA / delayed OA

- **Hybrid OA journal:** pay APC

DOAJ
Directory of OA Journals
<https://doaj.org>

1. Digital preservation
2. Persistent ID, eg, DOIs
3. Metadata
4. Users can make derivatives:
[CC BY](#)
[CC BY-SA](#)
[CC BY-NC](#)
[CC BY-NC-SA](#)
5. Licence info shown in articles
6. Author ©
7. Authors can self-archive, inc. **green open access**

How to widely share your research data
14

Journal transparency


2 Journal data sharing policies

asiocedit

<https://www.topfactor.org/>

What is the TOP Factor?

- Journal evaluation based on Transparency and Openness Promotion (TOP)
- Metric that reports the steps that a journal is taking to implement open science practices



JOURNAL	TOTAL TOP FACTOR
1 Meta-Psychology	27
2 Advances in Methods and Practices in Psychological Science	25
3 Peer Community In Registered Reports	25
4 Personality Science	24
5 Comprehensive Results in Social Psychology	23
6 Cortex	23
7 European Journal of Personality	23
8 Collabra	20
9 Journal of Personality	20
10 Journal of Personality and Social Psychology	20

Items per page: 10 1-10 of 10/2

TOP items

- Data citation
- Data transparency
- Analysis Code transparency
- Materials transparency
- Design & Analysis Reporting Guidelines
- Study Preregistration
- Analysis Plan Preregistration
- Replication studies

TOP levels

- Not implemented or mentioned
- Level I:** eg, states availability status
- Level II:** eg, must be uploaded to online repository
- Level III:** eg, must be uploaded to online repository & will be verified/reproduced

How to widely share your research data 15

Journal policies

2 Journal data sharing policies

asiocedit

Data transparency

Raw data

- Declare** image edits
- Submit **original files** with paper for inspection or for publication as supplementary files
- Submit **raw data** when asked by journal

Sharing data

- Submit underlying/minimal dataset as **supplementary file**
- Journal may prefer upload of dataset to **public repository**

Citing all data

- Cite & link** to datasets used, with DOI, in reference list (even if CCO/public domain)
- State in illustrations and/or Acknowledgement if © **permission** was obtained for use of third-party data

Data availability statement

- State if dataset is **available** publicly or not & where/link
- State how/when reader can **access data** or how to request access from author

Complete reporting

- Use Equator Network or journal **reporting checklist**
- Submit **protocol/data plan** as supplementary file
- Upload protocol** (repository or www.protocols.io)
- Publish protocol** paper or cohort profile paper
- Submit protocol** as Registered Report before study, then full report (see <https://www.cos.io/initiatives/registered-reports>)
- Register** clinical trial online, then summary of results (eg, www.clinicaltrials.gov)

How to widely share your research data 16

Data submission 2 Journal data sharing policies asiocedit

scientific **data**

Authors **must deposit their datasets** in an appropriate repository prior to peer review. ...**As part of the peer-review process**, editors and referees evaluate the appropriateness of the repository chosen to hold the dataset, the completeness of the deposited datasets, and their alignment with community standards. Authors will be required to **release their datasets publicly** when the Data Descriptor is published.

<https://www.nature.com/data/policies/editorial-and-publishing-policies#data-deposition>

Data policies

PLOS **ONE**

PLOS journals require authors to **make all data necessary to replicate their study's findings publicly available** without restriction at the time of publication. When specific **legal or ethical restrictions** prohibit public sharing of a data set, **authors must indicate how others may obtain access to the data.**

<https://journals.plos.org/plosone/s/data-availability>

PNAS

To allow others to replicate and build on work published in PNAS, authors **must make materials, data, and associated protocols, including code and scripts, available** to readers upon publication.... **Any restrictions** on access to the data, or any parts of it, must be disclosed in detail at submission. Authors are **encouraged to deposit laboratory protocols** and include their DOI or URL in the methods section of their paper.

<https://www.pnas.org/author-center/editorial-and-journal-policies#materials-and-data-availability>

How to widely share your research data 17

Data sharing 2 Journal data sharing policies asiocedit

<https://www.springernature.com/gp/authors/research-data-policy/research-data-policy-types>

Data sharing types

SPRINGER NATURE

- 1 • Data sharing and data citation is **encouraged**
- 2 • Data sharing and evidence of data sharing **encouraged**
- 3 • Data sharing **encouraged** and statements of data availability required
- 4 • Data sharing, evidence of data sharing and peer review of data **required**


How to widely share your research data 18

asiocedit

- 1 Data ethics & accessibility
- 2 Journal data sharing policies
- 3 Tips for data publishing**

How to widely share your research data 19

Data availability 3 Tips for data publishing asiocedit

 **Is this a good data availability statement?**

Data Availability
The data used in this study are openly available (CC0 licence) on our university's repository at Figshare (DOI: xxxxxxxxxx) [Reference 12]. ✓

Data Availability
The data used in this study are the private property (CC BY-NC-ND licence) of the authors and are for institutional use only (URL: www.university.repository/restricted). ✗

How to widely share your research data 20

Availability statements 1

3 Tips for data publishing

asiocedit

<https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>

Common data availability statements

The Greek Hippocratic texts used in this study are available to the public under a Creative Commons license at A Digital Corpus for Graeco-Arabic Studies: <https://www.graeco-arabic-studies.org/texts.html>.... The datasets for linguistic analysis are available upon request.
Karimullah, K.I. *Humanit Soc Sci Commun* 7, 27 (2020). <https://doi.org/10.1057/s41599-020-0511-7>

All data generated or analysed during this study are included in this published article (and its additional files). The ChIP-seq data have been deposited in the Gene Expression Omnibus database [GEO:GSE83330] (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=avylwvwgwtmrlmz&acc=GSE83330>).
Kim, K.W., Thakur, N., Piggott, C.A. et al. *BMC Biol* 14, 104 (2016). <https://doi.org/10.1186/s12915-016-0320-z>

The datasets analysed during the current study are available in the Dataverse repository: <http://dx.doi.org/10.7910/DVN/HEWGDD>. These datasets were derived from the following public domain resources: <http://www.census.gov/popest/data/index.html>....
Shutters, S., Muneeppeerakul, R. & Lobo, J. *Palgrave Commun* 1, 15010 (2015). <https://doi.org/10.1057/palcomms.2015.10>

The datasets generated during and/or analysed during the current study are **available in the [name] repository**, [DOI or persistent weblink to dataset].

The datasets generated during and/or analysed during the current study are **available from the corresponding author on reasonable request**.

All data generated or analysed during this study **are included in this published article** [and its supplementary information files].

[Special data type] have been deposited in [repository name] with the **accession code [code with link]**.

These datasets were derived from the following **public domain resources**: [list resources and their URLs].

How to widely share your research data 21

Availability statements 2

3 Tips for data publishing

asiocedit

<https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>

Special situations

The datasets analyzed during the current study are available from the Web of Science repository, owned by Thomson Reuters (<http://scientific.thomson.com/isi/>) but restrictions apply to the availability of these data, which were used under license from Thomson Reuters, and so are not publicly available. Data are however available from the authors upon reasonable request and permission of Thomson Reuters.
Palla, G., Tibély, G., Mones, E. et al. *Palgrave Commun* 1, 15016 (2015). <https://doi.org/10.1057/palcomms.2015.16>

All requests for raw and analyzed data and materials are promptly reviewed by The University of Texas MD Anderson Cancer Center to verify if the request is subject to any intellectual property or confidentiality obligations. Patient-related data not included in the paper might be subject to patient confidentiality. Any data and materials that can be shared will be released via a Material Transfer Agreement....
Deng, Q., Han, G., Puebla-Osorio, N. et al. *Nat Med* 26, 1878–1887 (2020). <https://doi.org/10.1038/s41591-020-1061-7>

...**restrictions apply** to the availability of these data, which were used under license, and so are not publicly available. Data are available from the authors on **reasonable request and with permission** of [third party name]

The data that support the findings of this study are **not openly available due to** [reasons of sensitivity e.g. human data] and are available from the corresponding author on **reasonable request** [include information on the data's location, e.g. in a **controlled** access repository where relevant / by special agreement/ethics approval]

Data sharing **not applicable** to this article as **no datasets were generated** or analysed during the current study

How to widely share your research data 22

Data publications 1

3 Tips for data publishing

asiocedit

Data papers to promote data reuse

F1000Research

“Data Notes” (and “Genome Notes”)


- Sharing datasets in a discoverable, useable and reproducible way
- Fostering new collaborations across disciplines
- Giving appropriate credit to data producers with a citable publication
- F1000Research helps with deposition in the most appropriate data repository and facilitates rigorous peer review
- Main Body: *Intro, Materials & Methods (can link to protocols.io), Dataset validation/Ethics/Clinical Trials*
- Data (and Software) Availability (+ link + licence); Reporting Guidelines (if applicable); Consent (if applicable)... NO ANALYSES/CONCLUSIONS

Eg, [A dataset for the perceived vulnerability to disease scale in Japan before the spread of COVID-19](#) [version 2; peer review: 2 approved]


<https://f1000research.com/for-authors/article-guidelines/data-notes>

“Database articles”

BMC Medicine




BMC
Bioinformatics



Botanical
Studies

“Data papers”



ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

The International Journal of
Volume 41 Number 2 February 2022

Robotics Research

How to widely share your research data

23

Data publications 2

3 Tips for data publishing

asiocedit

Data journals



(GIGA)ⁿ
SCIENCE

SCIENTIFIC
DATA



Journal of
open health data



(GIGA)byte

Journal of
open archaeology data


Journal of Open
Humanities Data











FmJ
Freshwater Metadata Journal

How to widely share your research data

24

Data description paper

3 Tips for data publishing

asicedit

https://www.nature.com/sdata/publish/for-authors

Data paper example

scientific data

“Data Descriptors”

- Background & Summary (700 words max)
- Methods (unlimited length)
- **Data Records** (unlimited length)
- **Technical Validation** (unlimited length)
- **Usage Notes** (unlimited length)
- **Code Availability**; References
- Figures (3/manuscript), Tables (10/manuscript)

scientific **data**

OPEN

DATA DESCRIPTOR

COVID Border Accountability Project, a hand-coded global database of border closures introduced during 2020

Mary A. Shiraeff¹, Cara Hoot¹, Mark A. Weiss¹, Sarah Nasser¹, Nicholas Lazar^{1,2}, Elizabeth Bering¹, Erin Strangle¹, Lukas Fedden¹, Noah Ruak Taylor¹, Cayleigh Jackson¹, William V. Anka-Bradstreet¹, Leah Mott¹, Matthew Amos¹, Maggie Shuman¹, Mary Louise Misdorf¹, Johanna Smeere¹, Suzanne E. Brantley¹, Luke L. Schenck¹, Colin Lewis-Roch¹ & COBAP Team¹

Usage notes: ...In sum, the **COBAP** database will assist both biomedical scientists with modeling viral transmission patterns for future pandemics and social scientists and policymakers with assessment on whether prior travel restrictions were effective. ...We plan to record end dates to all policies introduced in 2020; and as resources allow, we will collect and verify new policies introduced in 2021....

Code availability: Codes for the database, raw data outputs and data visualizations appearing on our website are available on our project GitHub (<https://github.com/COBAPteam/COBAP>)....

References: 24. Shiraeff, M. A. et al. The COVID Border Accountability Project (**COBAP**): Mapping Travel and Immigration Policy Responses to COVID-19. *Harvard Dataverse* <https://doi.org/10.7910/DVN/U6DJAC> (2020).

CC BY: Shiraeff et al. [COVID Border Accountability Project, a hand-coded global database of border closures introduced during 2020](https://doi.org/10.1038/s41598-021-01038-0). *Sci Data* 8, 253 (2021). <https://doi.org/10.1038/s41598-021-01038-0>

How to widely share your research data

25

Data attribution & credit

3 Tips for data publishing

asicedit

Citation elements

Dataset

Smith, T.W., Marsden, P.V., & Hout, M. (2011). *General social survey, 1972–2010 cumulative file* (ICPSR31521-v1) [data file and codebook]. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi: 10.3886/ICPSR31521.v1. Accessed 2 January 2022. Referenced in [doi:.....].

Data paper & supporting data

Cushman D, Bennett O, Berka R, et al. An overview of the National COVID-19 Chest Imaging Database: data quality and cohort analysis. *Gigascience* Volume 10, Issue 11, November 2021, giab076, <https://doi.org/10.1093/gigascience/giab076>

- **Availability of Source Code:** The codebase for the data warehouse is open source and available through the NHSX GitHub: Project: [covid-chest-imaging-database](https://github.com/nhsx/covid-chest-imaging-database)
- 53. Cushman D, Bennett O, Berka R, et al. Supporting data for “An overview of the National COVID-19 Chest Imaging Database: data quality and cohort analysis.”. *GigaScience Database*. 2021. <http://dx.doi.org/10.5524/100927>

Creator & Date of publication

Dataset title & version [resource types]

Publisher/Distributor/Platform name

ID or DOI / Handle / ARK or URL link

[Deposit date] [Access date]

[Link to data description paper or study]

How to widely share your research data

26


3 Tips for data publishing asiocedit

Citation styles

Dataset & software citation

PNAS

1. E. van Sebille, M. Doblin, [Data from](#) "Drift in ocean currents impacts intergenerational microbial exposure to temperature." [Figshare](#). Available at <https://dx.doi.org/10.6084/m9.figshare.3178534.v2>. Deposited 15 April 2016.
2. C. Reynaud et al., Tomography. [Zenodo](#). <https://doi.org/10.5281/zenodo.3712368>. Deposited 15 July 2020.

 National Library of Medicine

3. Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI. The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence [[dataset](#)]. 2015 Jun 30 [cited 2015 Oct 23]. [Dryad Digital Repository](#). Available from: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.47v3c> Referenced in doi: 10.7554/eLife.08347
4. RxNorm [[dataset](#)]. April 4, 2016 Full Monthly Release. 2016 Apr 4 [cited 2016 Apr 18]. Available from: <https://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html>
5. Hayes B, Tesar B, Zurow K. OTSoft: Optimality Theory Software. Version 2.3.2 [[software](#)]. 2013 Jan 14 [cited 2015 Feb 14]. Available from: <https://linguistics.ucla.edu/people/hayes/otsoft/>.

How to widely share your research data 27

3 Tips for data publishing asiocedit

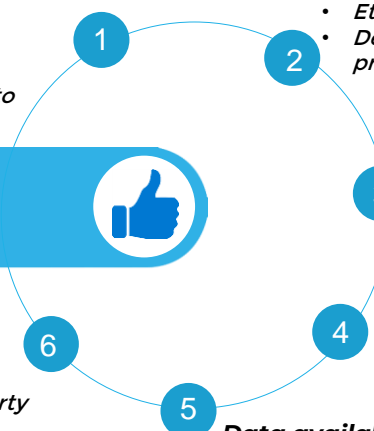
General tips

Check journal/funder guidelines:

- Requirements on data sharing, repository & user licence
- Funder requirements of open access publishing
- Give editor/ reviewers access to (embargoed) dataset

Avoid:

- Uploading datasets only as supplementary files
- Signing away ©; using third-party data/materials illegally
- Media promotion of unpublished papers; check journal embargoes



Ethics:

- Data management plan; data ownership, credit data "authors"
- Ethical approval, consent; trial registration
- Declare conferences, data papers, preprints

Strategy:

- Publish protocol paper (or submit Registered Report)
- Upload anonymised dataset & publish data description paper
- Publish full research paper

Complete context:

- Use /upload reporting guidelines; data plan/protocol transparency
- Sample details (N, inclusion/exclusion)
- Primary vs secondary outcomes

Data availability statement:

- Match journal/funder policy; permissions
- Location + Link + Licence?
- Cite data source

How to widely share your research data 28

1 Data ethics & accessibility

2 Journal data sharing policies

3 Tips for data publishing

asiaedit

Thank you!

