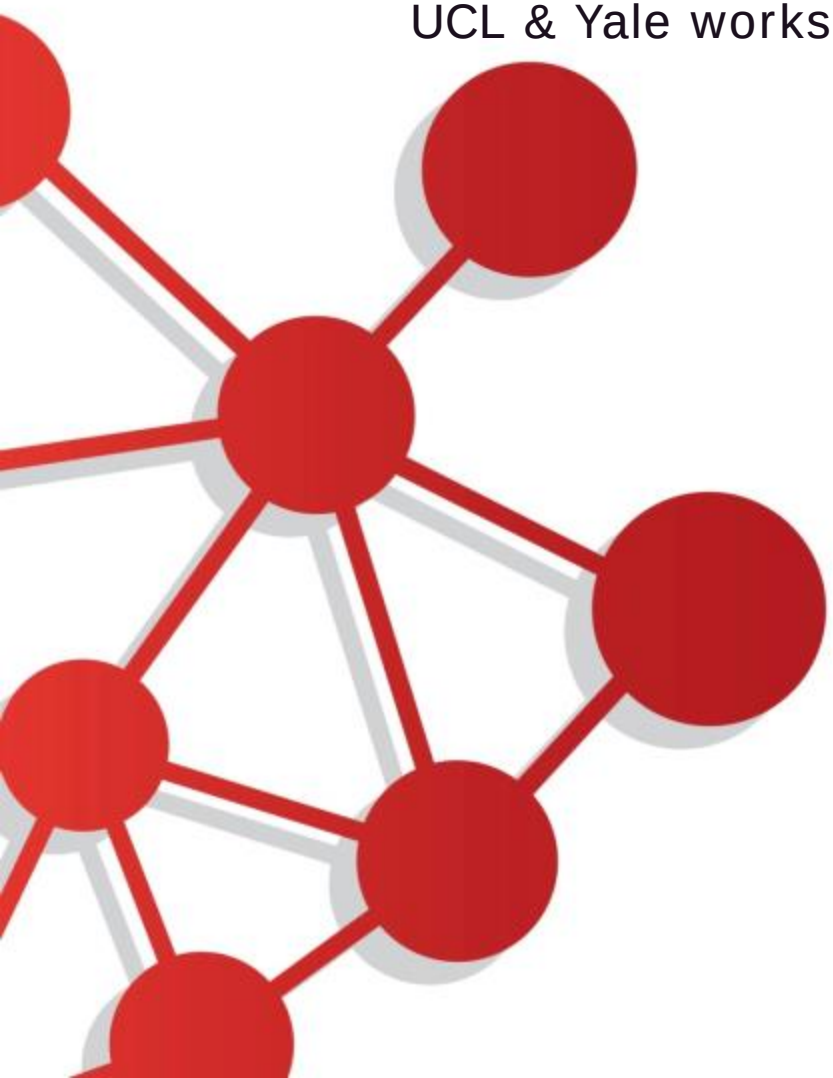
Publication and access to research datasets

UCL & Yale workshop, 9th October 2015



Iain Hrynaszkiewicz

Head of Data and HSS Publishing, Open Research
Nature Publishing Group & Palgrave Macmillan

iain.hrynaszkiewicz@nature.com

@iainh_z

nature publishing group npg

Publishers / journals and data access

- More reliable evidence – and papers
- Journal mission/goals
- Content innovation (and more use and reuse)
- Reliability (peer review)
- Discoverability and visibility (bibliographic databases)
- Permanence (content and links)
- Credit/incentives (article types and citations)
- Encouraging and implementing good practice and policies

Journal data policies

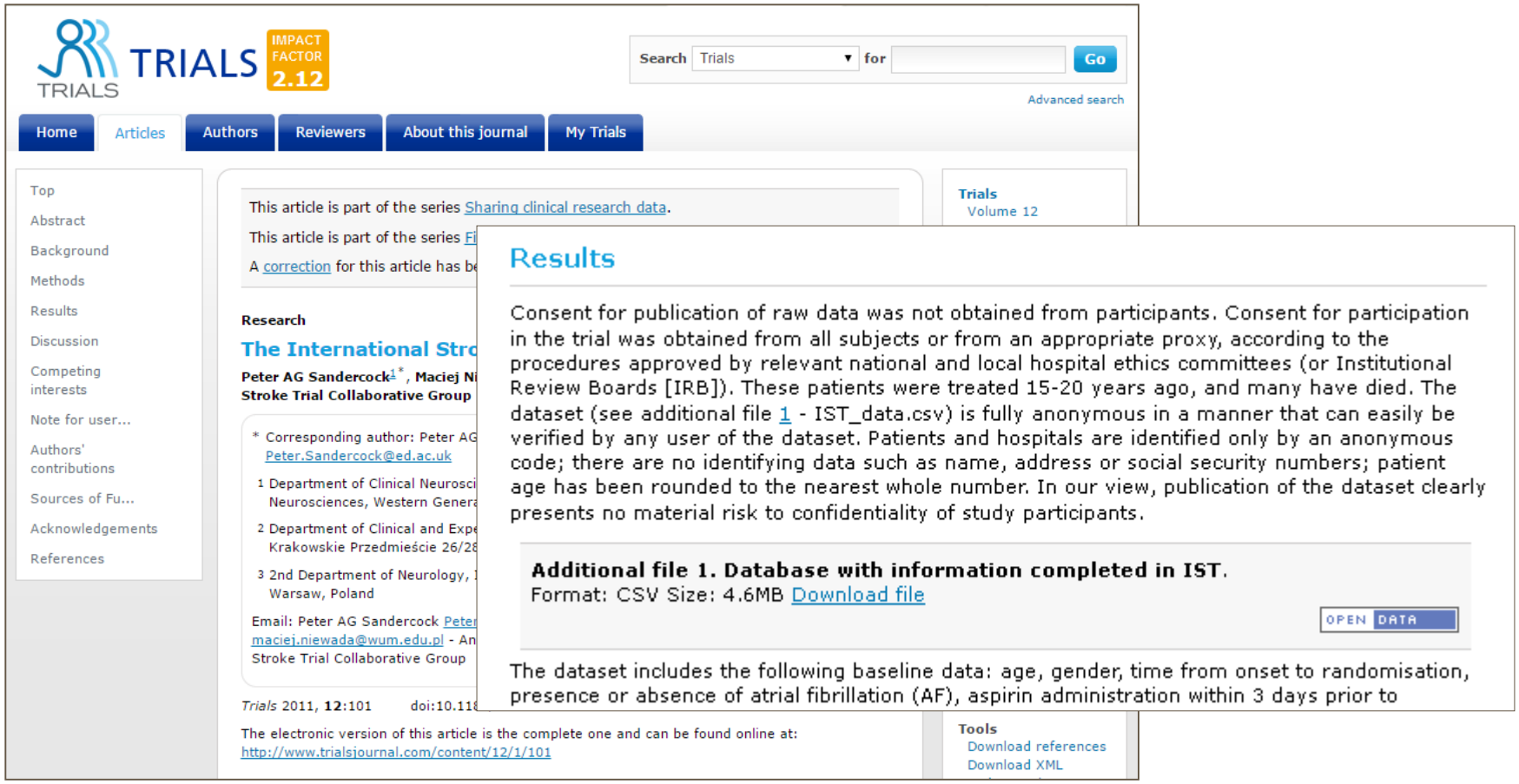
- Willingness to share stated (*Annals Internal Medicine*)
- Data sharing implied by submission (BioMed Central*)
- Data sharing implied as a condition of publication (Nature*)
- Mandated data sharing with statement in paper (PLOS, BMJ)
- Mandated data sharing with statement and link to data (non-medical journals e.g. ecology, animal genomics)
- Mandated open data as a condition of submission (*Scientific Data, GigaScience, F1000Research*)

**Minimum publisher requirement – some disciplines/journals may mandate open access to data*

STRONGER

1. Vines, T. H. *et al.* **Mandated data archiving greatly improves access to research data**. *FASEB J.* fj.12–218164– (2013). doi:10.1096/fj.12-218164

Data sharing via supplementary files



Sandercock *et al*: **The International Stroke Trial database.** *Trials* 2011, **12**:101
doi:10.1186/1745-6215-12-101

Data sharing via repository links

Research

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

BMJ 2015 ; 351 doi: http://dx.doi.org/10.1136/bmj.h4320 (Published 16 September 2015)
Cite this as: *BMJ* 2015;351:h4320

Article | Related content | Metrics | Responses | Peer review

Joanna Le Noury, research psychologist [1], John M Nardo, retired clinical assistant professor [2], David Healy, professor [1], Jon Jureidini, clinical professor [3], Melissa Raven, postdoctoral fellow [3], Catalin Tufanaru, research associate [4], Elia Abi-Jaoude, staff psychiatrist [5]

Author affiliations

Correspondence to: J Jureidini Jon.Jureidini@adelaide.edu.au

Accepted 3 August 2015

Data sharing via repository links

Research

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

BMJ 2015 ;
Cite this as: *B*

Article

Joanna Le N
professor[1], *J*
associate[4], *B*

Author affili

Corresponde

Accepted 3

planned (and, if relevant, registered) have been explained.

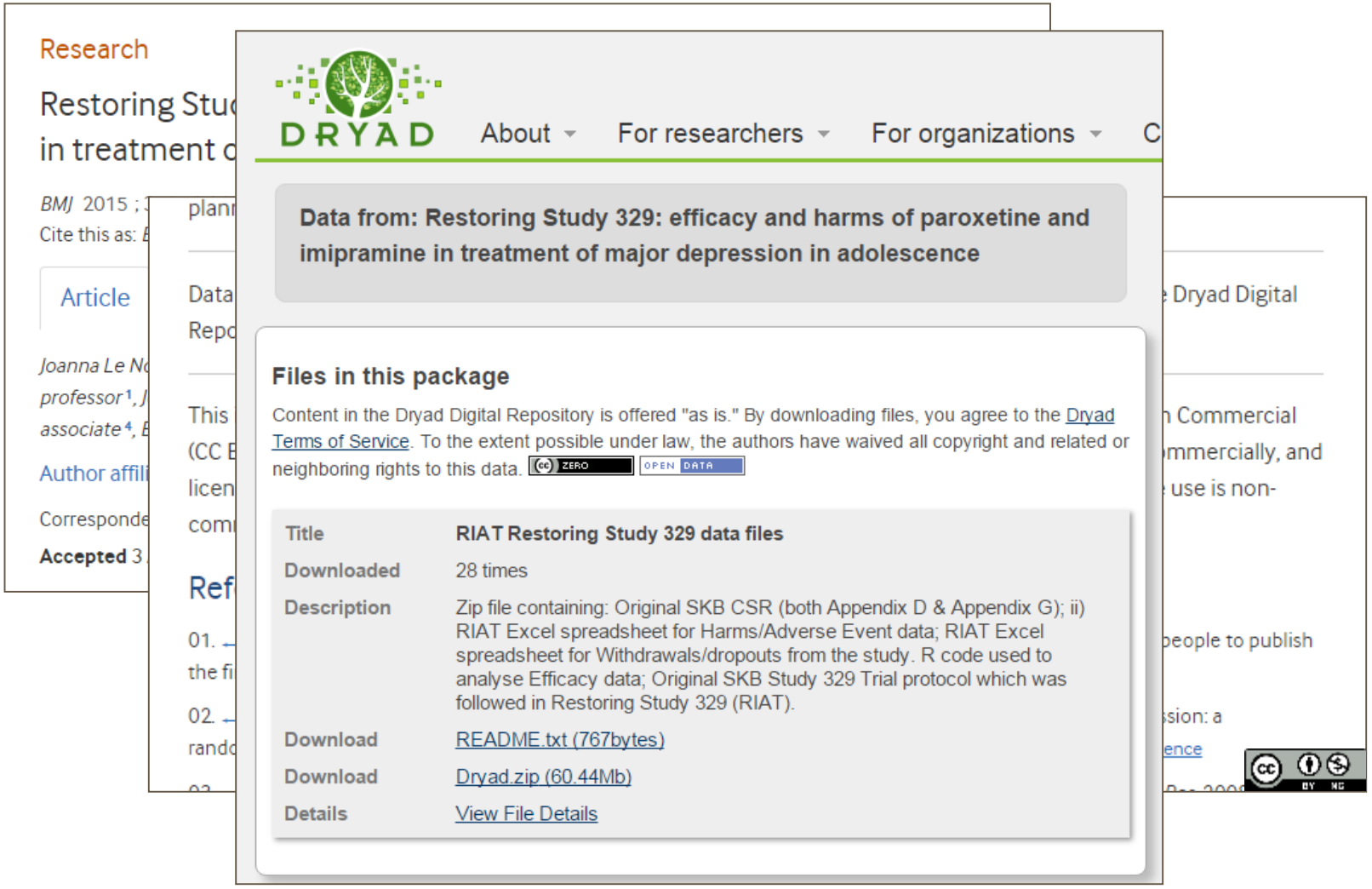
Data sharing: Clinical study reports, detailed data tables, and programming code are available on the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.bv8j6) and at www.Study329.org/.

References

01. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865. FREE Full Text

02. Keller MB, Ryan ND, Strober M, et al. Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry* 2001;40:762-72. CrossRef Medline Web of Science

Data sharing via repository links



List of potential patient identifiers in datasets

Identifier (information sources)	Comments
Direct	
Name [8 15]	
Initials [13]	
Address, including full or partial postal code [8-15]	
Telephone or fax numbers or contact information [8 10 12 15]	
Electronic mail addresses [8]	
Unique identifying numbers [8-15]	Generalised HIPAA items 7-10, 18
Vehicle identifiers [8]	
Medical device identifiers [8]	
Web or internet protocol addresses [8]	
Biometric data [8]	
Facial photograph or comparable image [8 10 11 13]	
Audiotapes [11]	
Names of relatives [10]	
Dates related to an individual (including date of birth) [8 9 11 15]	
Indirect—may present a risk if present in combination with others in the list	
Place of treatment or health professional responsible for care [10 15]	Could be inferred from investigator affiliations
Sex [9]	
Rare disease or treatment [10]	
Sensitive data, such as illicit drug use or "risky behaviour" [15]	
Place of birth [10 15]	
Socioeconomic data, such as occupation or place of work, income, or education [9 10 12 15]	MRC requirement is for "rare" occupations only
Household and family composition [15]	
Anthropometry measures [15]	
Multiple pregnancies [15]	
Ethnicity [9]	
Small denominators—population size of <100 [14]	
Very small numerators—event counts of <3 [14]	
Year of birth or age (this article)	Age is potentially identifying if the recruitment period is short and is fully described
Verbatim responses or transcripts [15]	

"...*datasets that contain* ***three or more indirect identifiers****, such as age or sex, should be reviewed by an independent researcher or ethics committee*"

Hrynaszkiewicz *et al.*, *BMJ* 2010;340:c181

Data on (reasonable) request - issues

- Meta-analysis fails to launch when <40% IPD available – unanswered requests and refusal to share

  Systematic Reviews 2014, **3**:97 doi:10.1186/2046-4053-3-97

- Poor availability of psychological research data (only 64/249 datasets available)

  American Psychologist, Vol 61(7), Oct 2006, 726-728. doi:10.1037/0003-066X.61.7.726

- Data received from 1/10 authors publishing in *PLOS Medicine* and *PLOS Clinical Trials*

  PLoS ONE 4(9): e7078. doi:10.1371/journal.pone.0007078

- **Sensitive data repositories (e.g. UKDA)**
  Permanence, curation, persistent identifiers, versioning
- **Data-on-request services (e.g. YODA)**
  Independent governance, scientific review and transparency of access requests, DUAs
- **Journals / publishers**
  Peer review, visibility, credit/citations, robust links

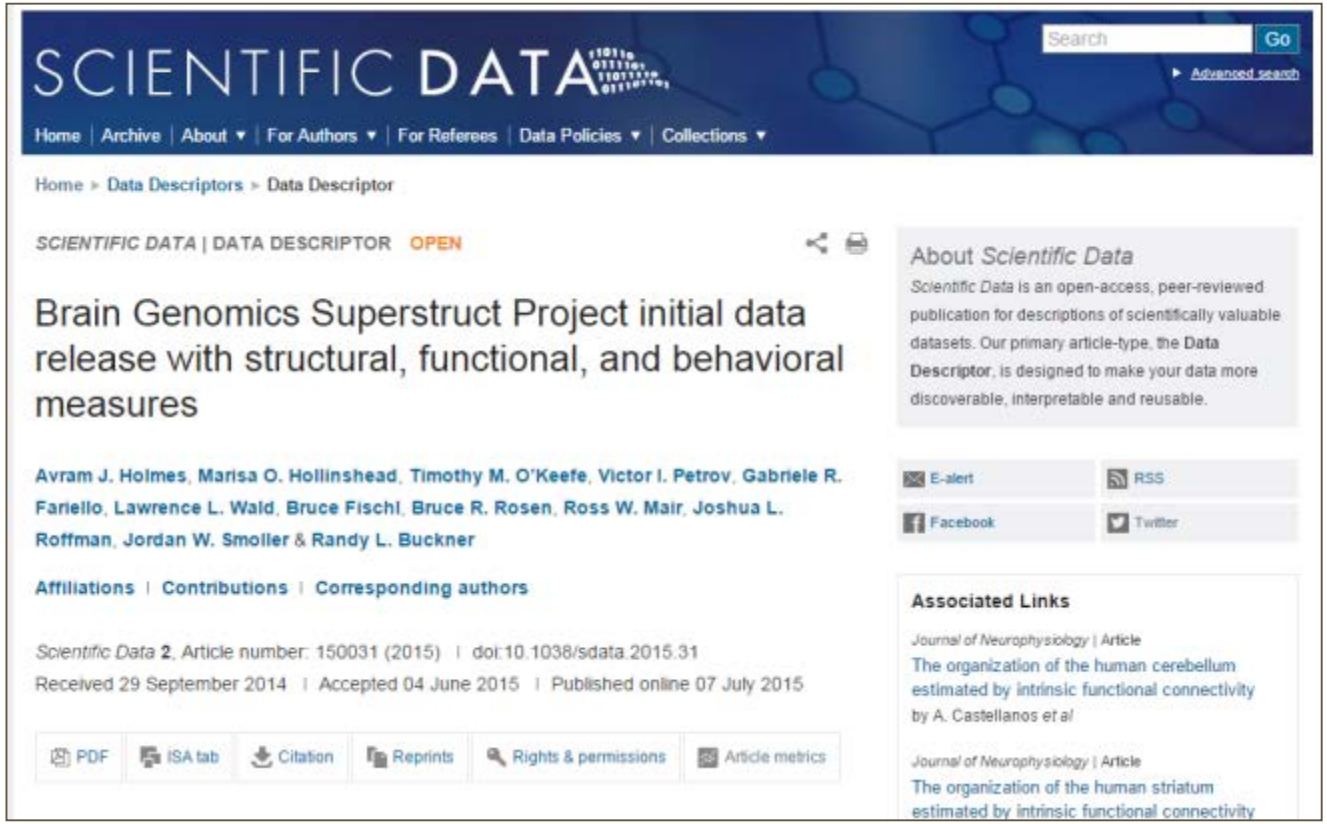
+

=

Better way to *publish* data on request

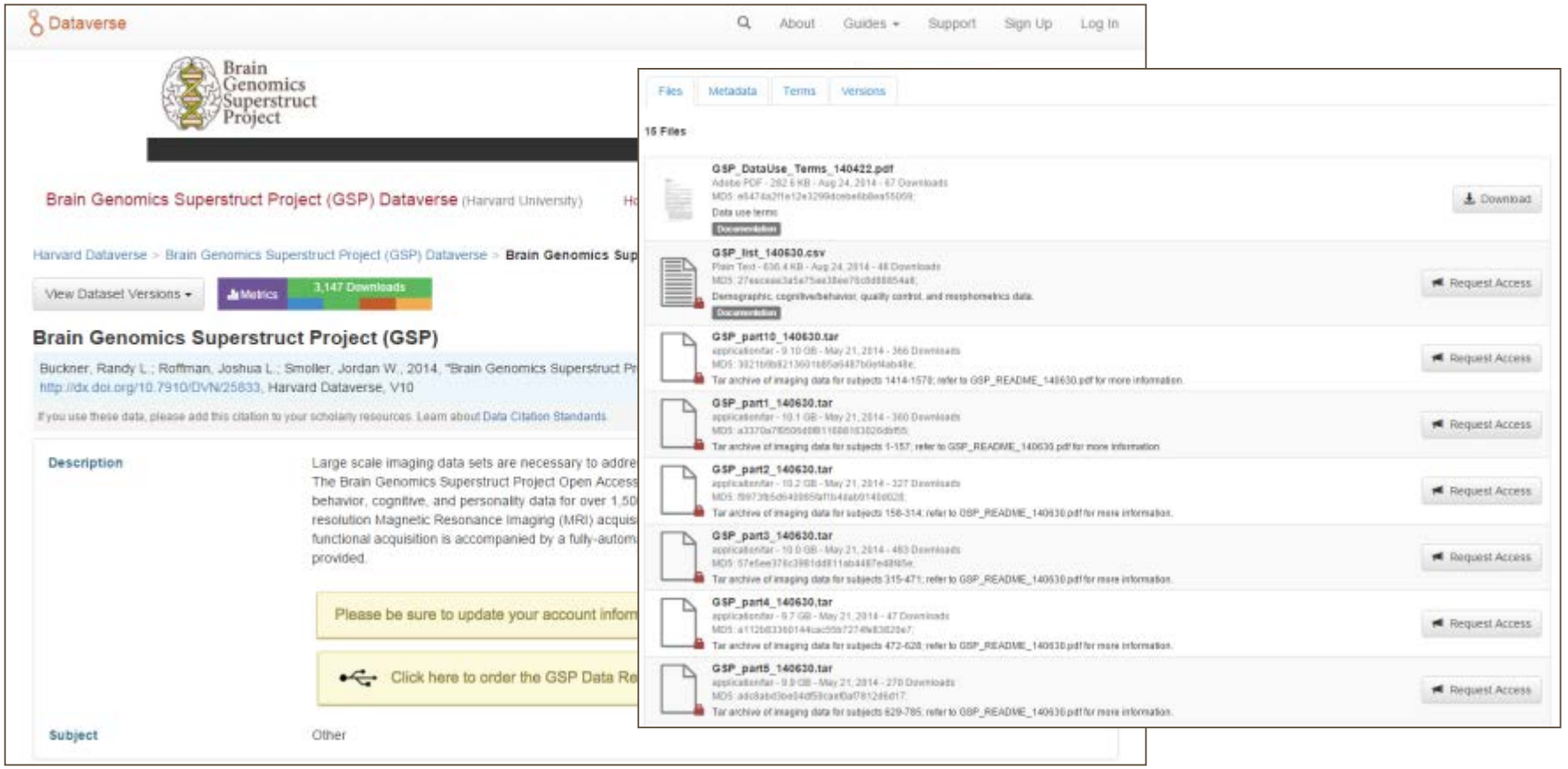
1. Hrynaszkiewicz, I., Khodiyar, V., Hufton, A. & Sansone, S. A. **Publishing descriptions of non-public clinical datasets: guidance for researchers, repositories, editors and funding organisations**. BioRxiv http://dx.doi.org/10.1101/021667 (2015).

Open access Data Descriptor



http://www.nature.com/articles/sdata201531

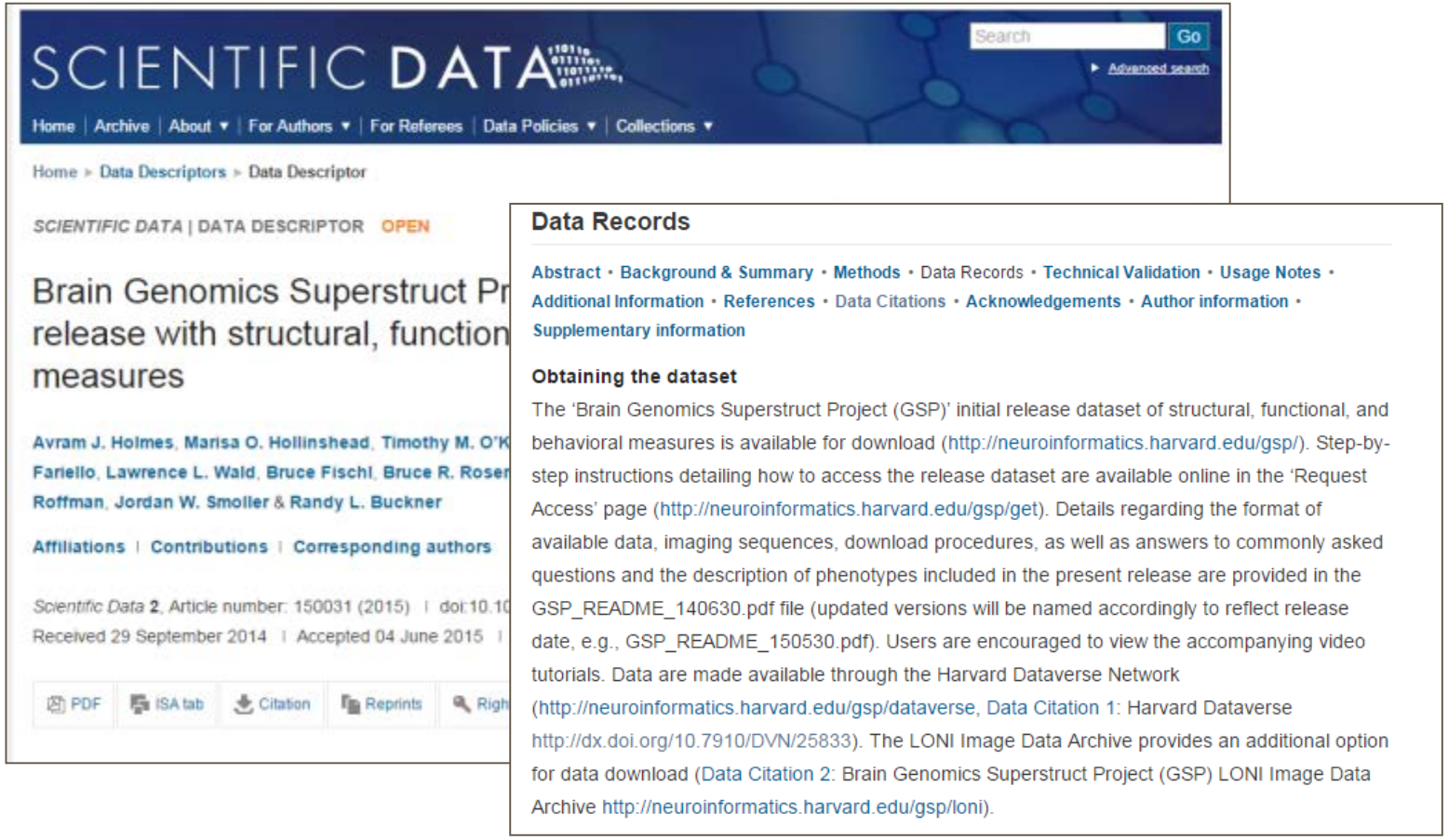
Linked to open access data



http://dx.doi.org/10.7910/DVN/25833

All approved repositories:
http://www.nature.com/sdata/data-policies/repositories

And restricted access data



http://www.nature.com/articles/sdata201531

Appendix: Repository approval criteria

- Supported and recognized by scientific community
- Ensure long-term persistence and preservation of datasets
- Provide data curation
- Implement community-endorsed reporting requirements
- Provide for confidential review of datasets
- Provide stable identifiers
- Allow public access to data without unnecessary restrictions

http://www.nature.com/sdata/about/faq#q21
http://www.nature.com/sdata/data-policies/repositories