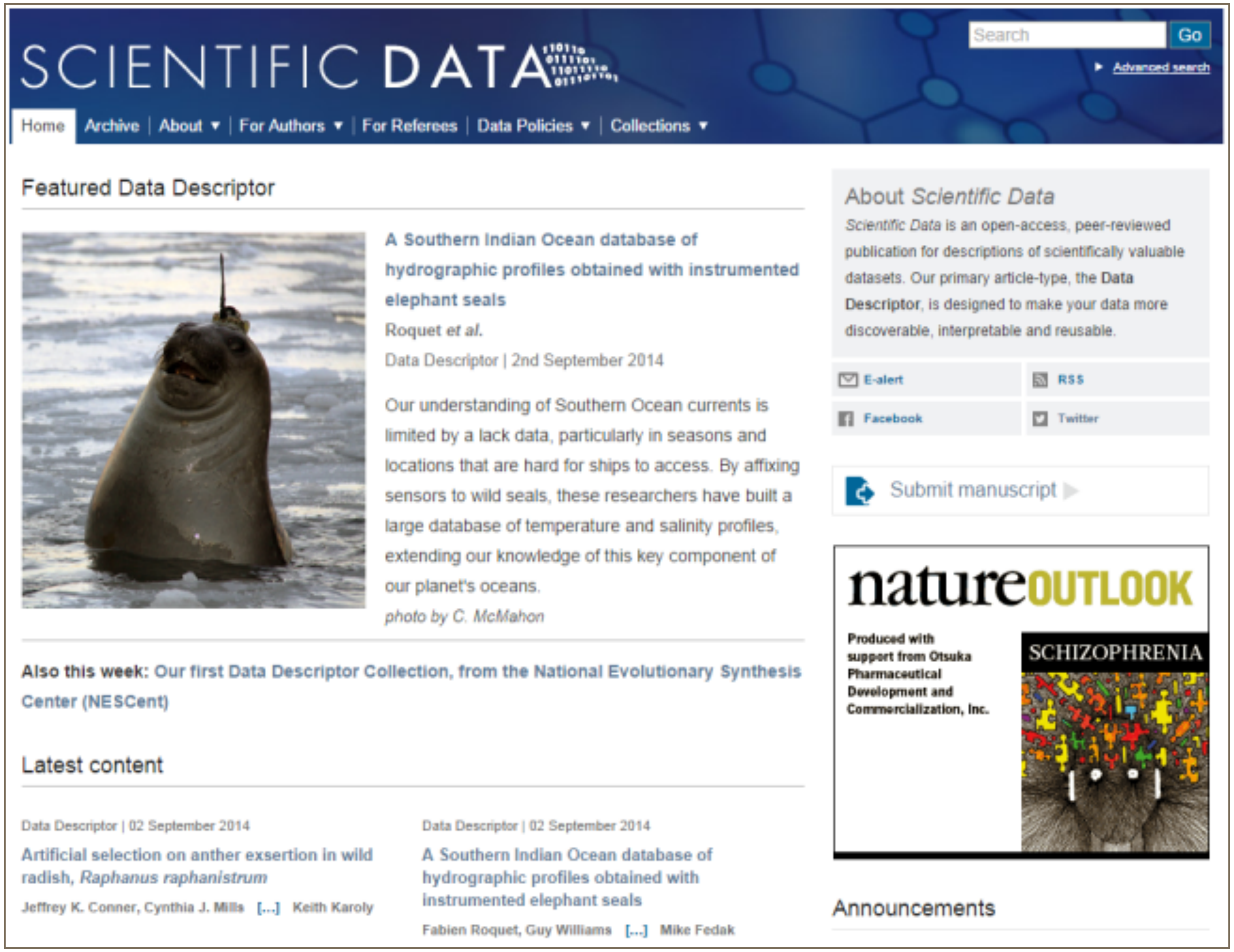
DATA PAPERS AND THEIR APPLICATIONS: DATA DESCRIPTORS IN *SCIENTIFIC DATA*

LCPD Workshop 12th September 2014

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

macmillan
Science and Education

Transforming Learning and Discovery



Featured Data Descriptor

A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals

Roquet *et al.*

Data Descriptor | 2nd September 2014

Our understanding of Southern Ocean currents is limited by a lack data, particularly in seasons and locations that are hard for ships to access. By affixing sensors to wild seals, these researchers have built a large database of temperature and salinity profiles, extending our knowledge of this key component of our planet's oceans.

photo by C. McMahon

Also this week: Our first Data Descriptor Collection, from the National Evolutionary Synthesis Center (NESCent)

Latest content

Data Descriptor | 02 September 2014

Artificial selection on anther exertion in wild radish, *Raphanus raphanistrum*

Jeffrey K. Conner, Cynthia J. Mills [...] Keith Karoly

Data Descriptor | 02 September 2014

A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals

Fabien Roquet, Guy Williams [...] Mike Fedak

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

natureOUTLOOK

Produced with support from Otsuka Pharmaceutical Development and Commercialization, Inc.

SCHIZOPHRENIA

Announcements

http://www.nature.com/sdata/

Scientific Data

Scope

An open access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the Data Descriptor, is designed to make your data more discoverable, interpretable and reusable.

Editorial team

Managing Editor (Andrew Hufton)
Editorial Curator (Victoria Newman)
Honorary Academic Editor (Susanna Sansone, Oxford)
Advisory Panel and Editorial Board

Open access article processing charge

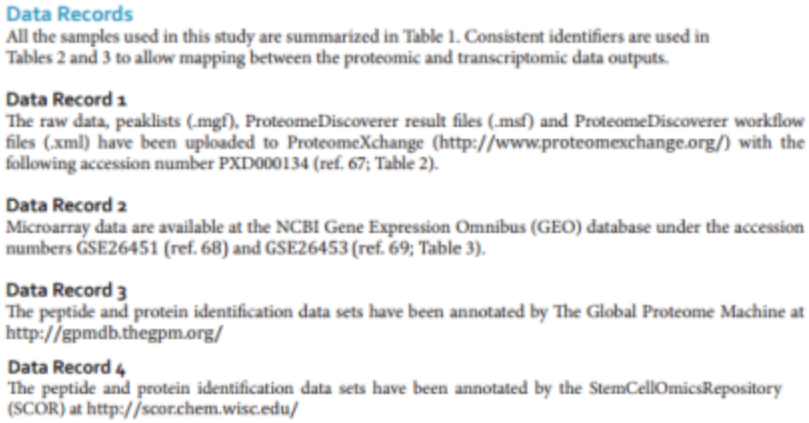
$1,000 USD / £650 GBP / €750 for each accepted article.

The 'Data Descriptor' article

Detailed descriptions of the methods and technical analyses supporting the quality of the measurements. Does not contain tests of new scientific hypotheses

Sections:

- Title
- Abstract
- Background & Summary
- Methods
- Technical Validation
- Data Records
- Usage Notes
- Figures & Tables
- References
- Data Citations



Data Records

All the samples used in this study are summarized in Table 1. Consistent identifiers are used in Tables 2 and 3 to allow mapping between the proteomic and transcriptomic data outputs.

Data Record 1

The raw data, peaklists (.mgf), ProteomeDiscoverer result files (.msf) and ProteomeDiscoverer workflow files (.xml) have been uploaded to ProteomeXchange (http://www.proteomexchange.org/) with the following accession number PXD000134 (ref. 67; Table 2).

Data Record 2

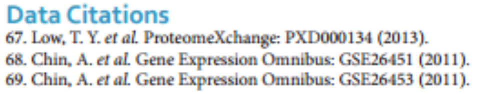
Microarray data are available at the NCBI Gene Expression Omnibus (GEO) database under the accession numbers GSE26451 (ref. 68) and GSE26453 (ref. 69; Table 3).

Data Record 3

The peptide and protein identification data sets have been annotated by The Global Proteome Machine at http://gpmdb.thegpm.org/

Data Record 4

The peptide and protein identification data sets have been annotated by the StemCellOmicsRepository (SCOR) at http://scor.chem.wisc.edu/



Data Citations

67. Low, T. Y. *et al.* ProteomeXchange: PXD000134 (2013).
68. Chin, A. *et al.* Gene Expression Omnibus: GSE26451 (2011).
69. Chin, A. *et al.* Gene Expression Omnibus: GSE26453 (2011).

Peer review at *Scientific Data*

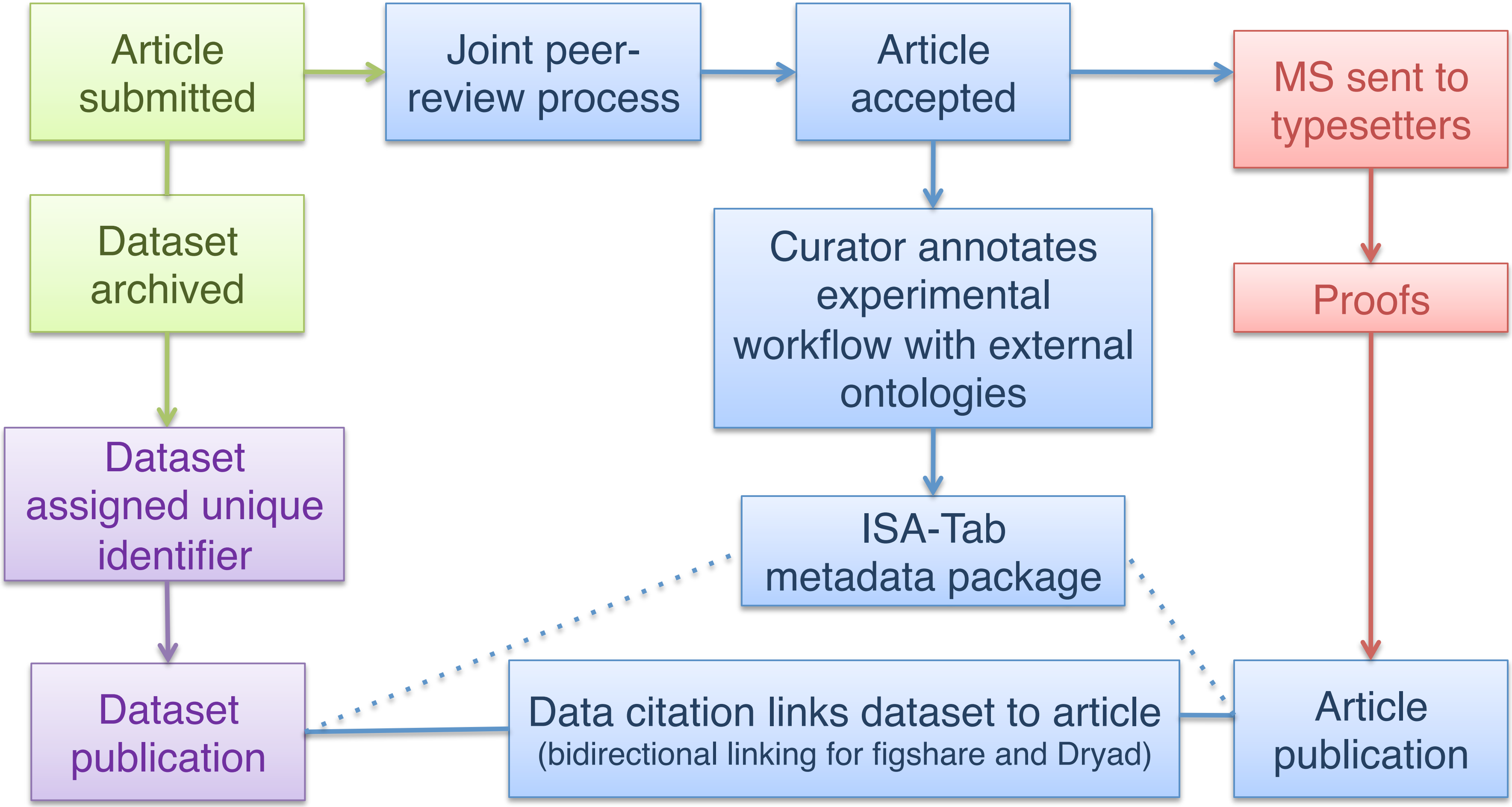
Focuses on:

- Completeness (can others reproduce?)
- Consistency (were community standards followed?)
- Integrity (are data in the best repository?)
- Experimental rigour technical quality (were the methods sound?)

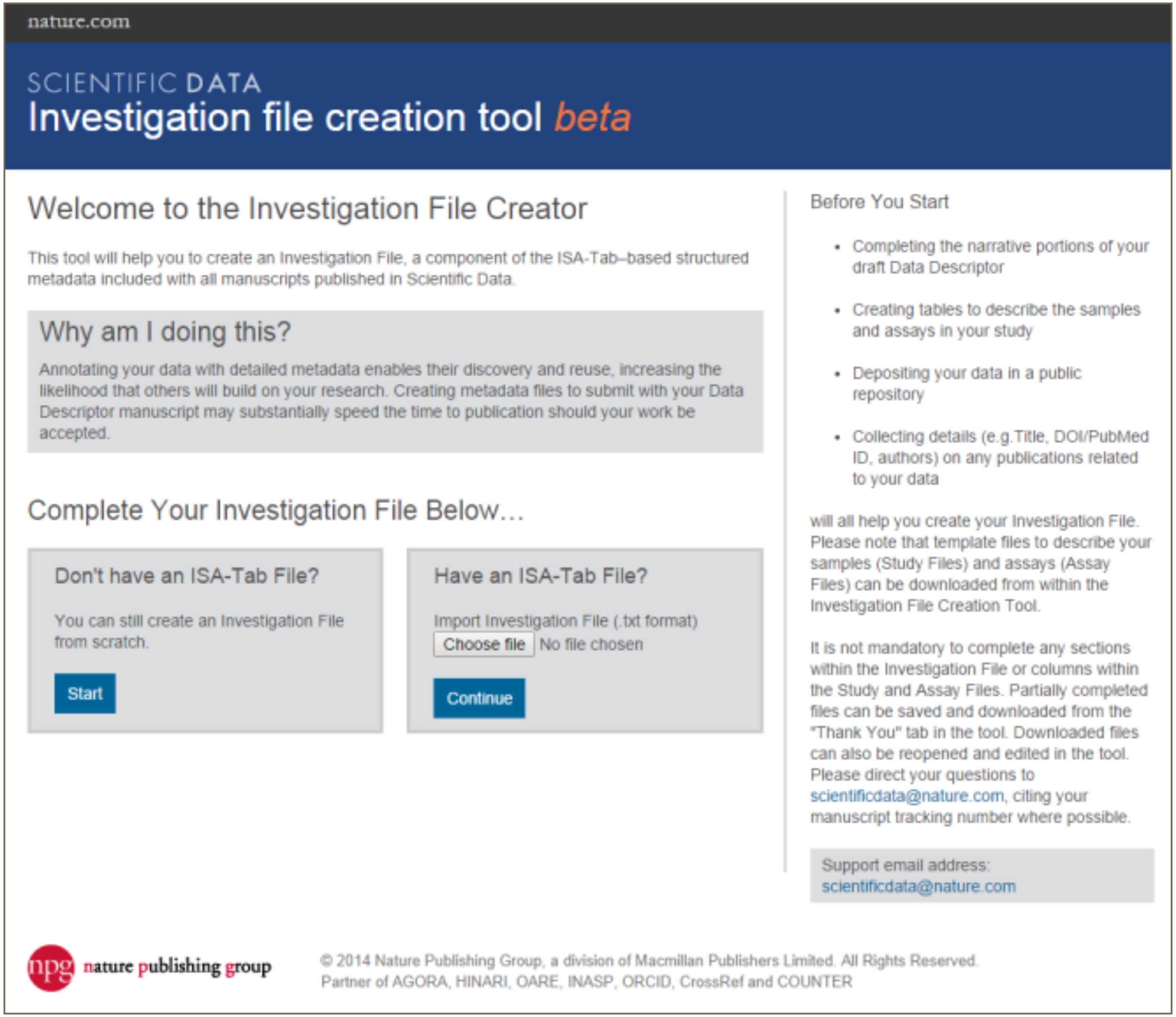
Does not focus on:

- Perceived impact/importance
- Size/complexity of data

Scientific Data workflow overview


Article submitted
Joint peer-review process
Article accepted
MS sent to typesetters
Dataset archived
Curator annotates experimental workflow with external ontologies
Proofs
Dataset assigned unique identifier
ISA-Tab metadata package
Dataset publication
Data citation links dataset to article
(bidirectional linking for figshare and Dryad)
Article publication

Green: author; Purple: repository; Blue: SciData; Red: production



SCIENTIFIC DATA

Investigation file creation tool *beta*

Welcome to the Investigation File Creator

This tool will help you to create an Investigation File, a component of the ISA-Tab–based structured metadata included with all manuscripts published in Scientific Data.

Why am I doing this?

Annotating your data with detailed metadata enables their discovery and reuse, increasing the likelihood that others will build on your research. Creating metadata files to submit with your Data Descriptor manuscript may substantially speed the time to publication should your work be accepted.

Complete Your Investigation File Below...

Don't have an ISA-Tab File?

You can still create an Investigation File from scratch.

Have an ISA-Tab File?

Import Investigation File (.txt format)

Before You Start

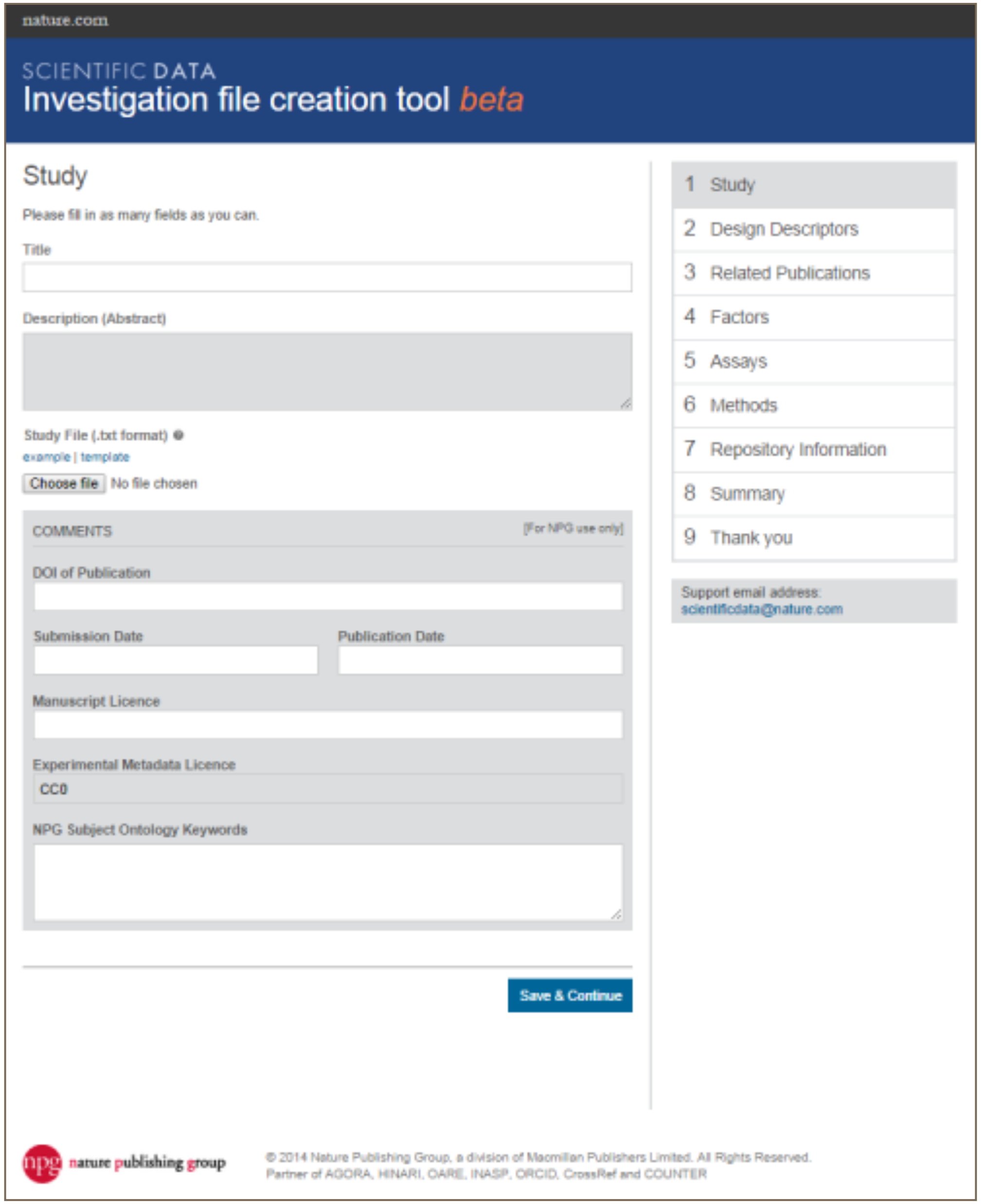
- Completing the narrative portions of your draft Data Descriptor
- Creating tables to describe the samples and assays in your study
- Depositing your data in a public repository
- Collecting details (e.g.Title, DOI/PubMed ID, authors) on any publications related to your data

will all help you create your Investigation File. Please note that template files to describe your samples (Study Files) and assays (Assay Files) can be downloaded from within the Investigation File Creation Tool.

It is not mandatory to complete any sections within the Investigation File or columns within the Study and Assay Files. Partially completed files can be saved and downloaded from the "Thank You" tab in the tool. Downloaded files can also be reopened and edited in the tool. Please direct your questions to scientificdata@nature.com, citing your manuscript tracking number where possible.

Support email address:
scientificdata@nature.com

© 2014 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Partner of AGORA, HINARI, OARE, INASP, ORCID, CrossRef and COUNTER



nature.com

SCIENTIFIC DATA

Investigation file creation tool *beta*

Study

Please fill in as many fields as you can.

Title

Description (Abstract)

Study File (.txt format)

example | template

Choose file No file chosen

COMMENTS [For NPG use only]

DOI of Publication

Submission Date

Publication Date

Manuscript Licence

Experimental Metadata Licence

CC0

NPG Subject Ontology Keywords

Save & Continue

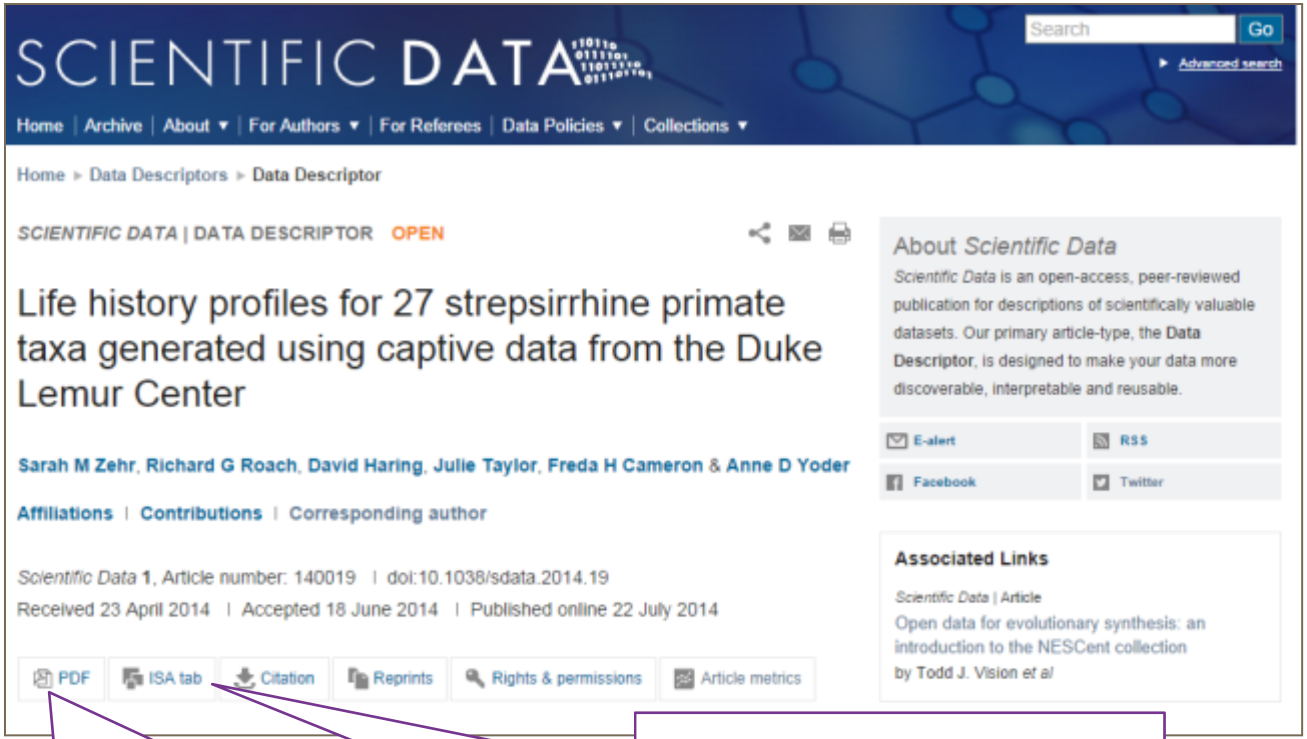
1 Study
2 Design Descriptors
3 Related Publications
4 Factors
5 Assays
6 Methods
7 Repository Information
8 Summary
9 Thank you

Support email address:
scientificdata@nature.com

nature publishing group

© 2014 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Partner of AGORA, HINARI, OARE, INASP, ORCID, CrossRef and COUNTER

The ‘Data Descriptor’ article



SCIENTIFIC DATA

Search Go

▸ Advanced search

Home | Archive | About ▾ | For Authors ▾ | For Referees | Data Policies ▾ | Collections ▾

Home ▸ Data Descriptors ▸ Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN

Life history profiles for 27 strepsirrhine primate taxa generated using captive data from the Duke Lemur Center

Sarah M Zehr, Richard G Roach, David Haring, Julie Taylor, Freda H Cameron & Anne D Yoder

Affiliations | Contributions | Corresponding author

Scientific Data **1**, Article number: 140019 | doi:10.1038/sdata.2014.19

Received 23 April 2014 | Accepted 18 June 2014 | Published online 22 July 2014

PDF | ISA tab | Citation | Reprints | Rights & permissions | Article metrics

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

E-alert | RSS

Facebook | Twitter

Associated Links

Scientific Data | Article

Open data for evolutionary synthesis: an introduction to the NESCent collection

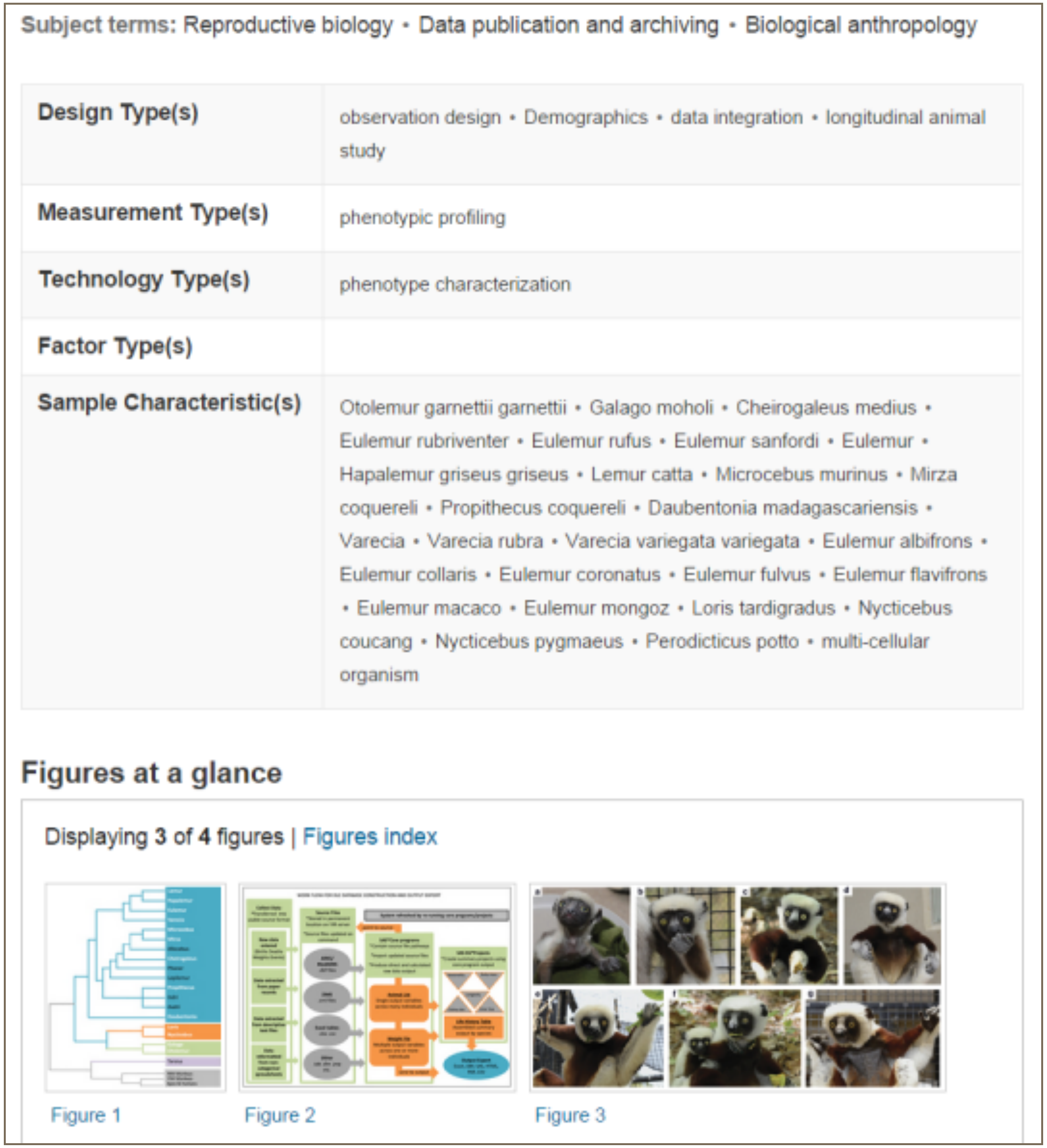
by Todd J. Vision *et al*

Article or *narrative* component

(PDF and HTML)

Experimental metadata or *structured* component

(in-house curated, machine-readable formats)



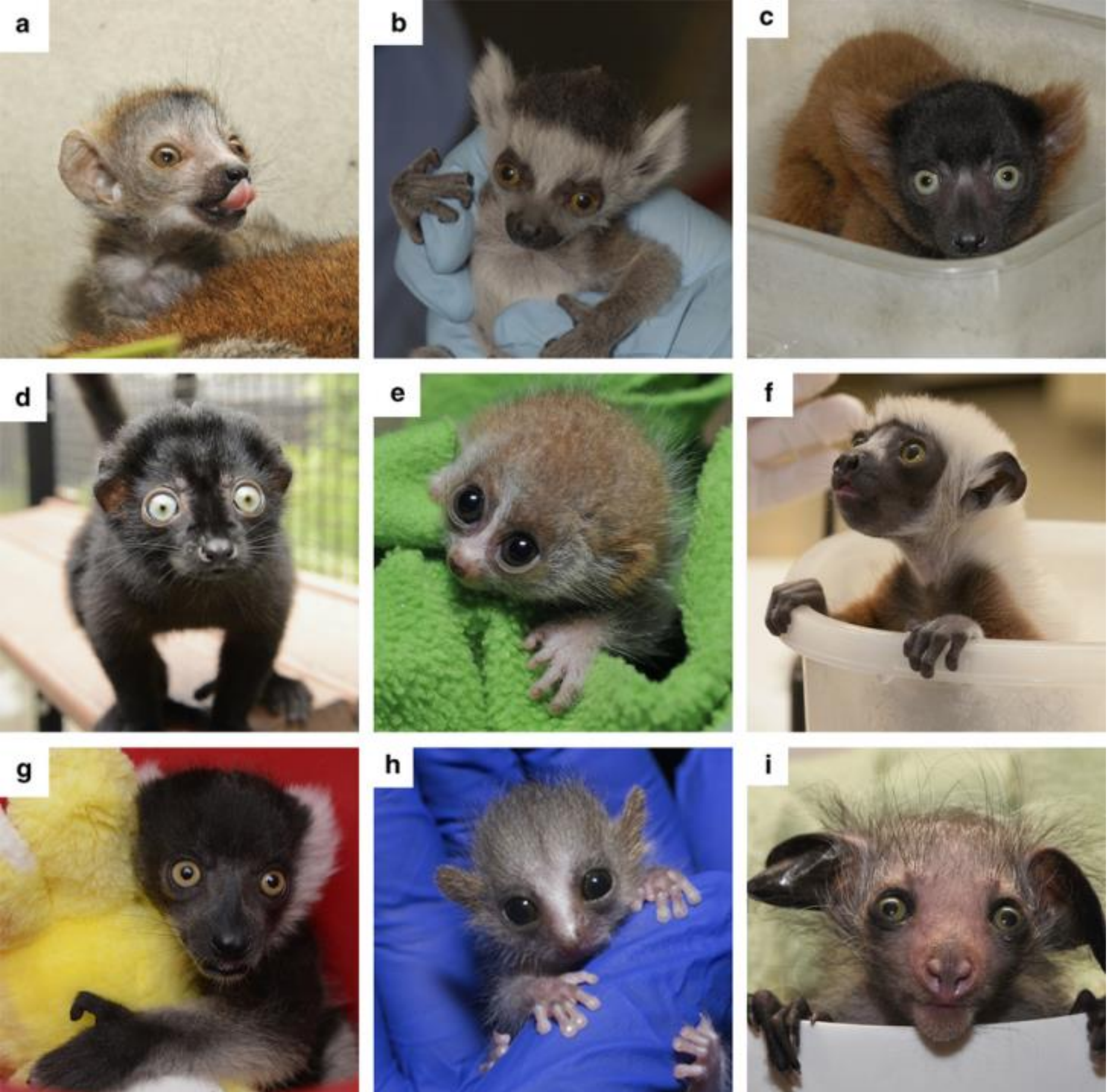
Subject terms: Reproductive biology • Data publication and archiving • Biological anthropology

Design Type(s)	observation design • Demographics • data integration • longitudinal animal study
Measurement Type(s)	phenotypic profiling
Technology Type(s)	phenotype characterization
Factor Type(s)	
Sample Characteristic(s)	Otolemur garnettii garnettii • Galago moholi • Cheirogaleus medius • Eulemur rubriventer • Eulemur rufus • Eulemur sanfordi • Eulemur • Hapalemur griseus griseus • Lemur catta • Microcebus murinus • Mirza coquereli • Propithecus coquereli • Daubentonia madagascariensis • Varecia • Varecia rubra • Varecia variegata variegata • Eulemur albifrons • Eulemur collaris • Eulemur coronatus • Eulemur fulvus • Eulemur flavifrons • Eulemur macaco • Eulemur mongoz • Loris tardigradus • Nycticebus coucang • Nycticebus pygmaeus • Perodicticus potto • multi-cellular organism

Figures at a glance

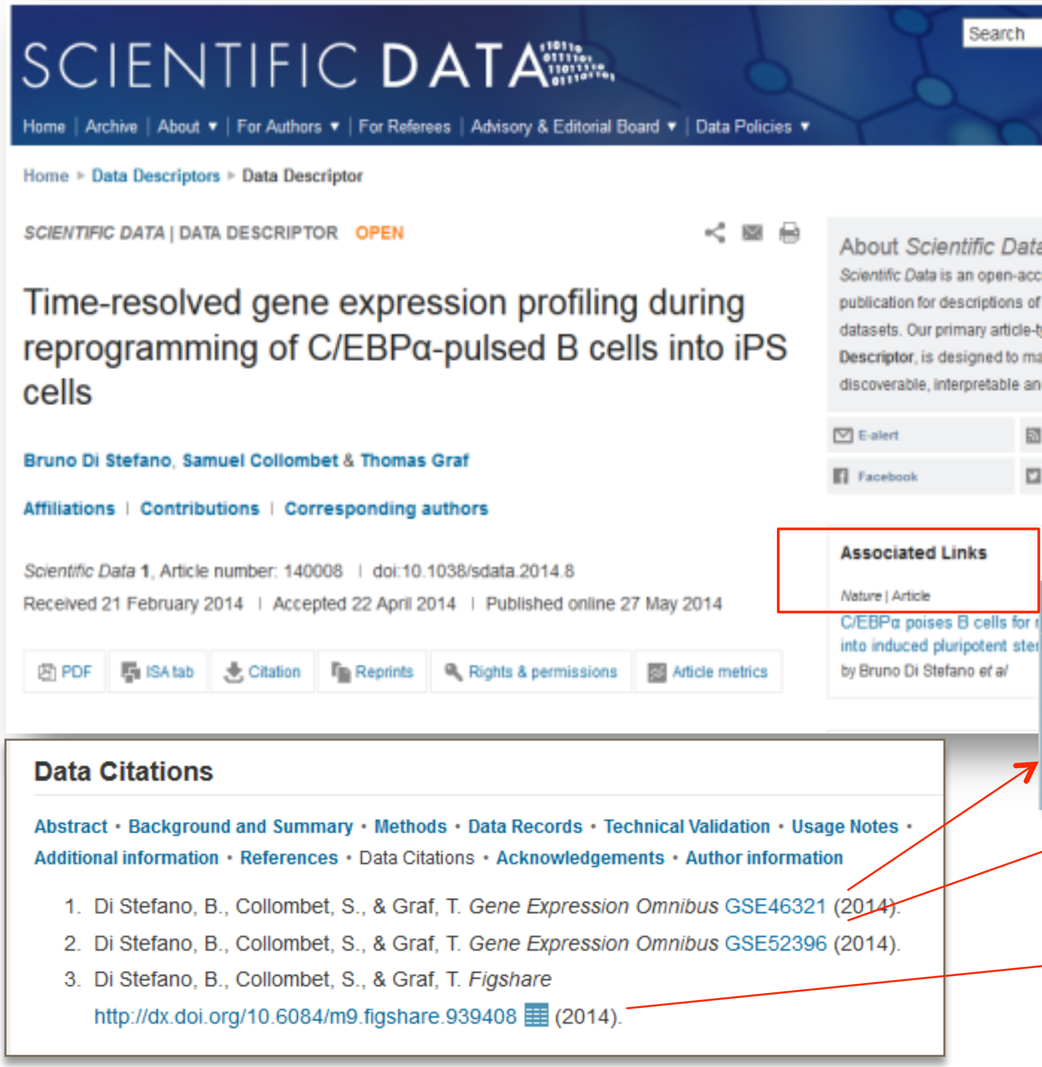
Displaying 3 of 4 figures | Figures index

Zehr et al. *Scientific Data* **1**, Article number: 140019 doi: 10.1038/sdata.2014.19


a
b
c
d
e
f
g
h
i

Stem Cells

SCIENTIFIC DATA



SCIENTIFIC DATA
Search
Go
Advanced search
Home | Archive | About ▾ | For Authors ▾ | For Referees | Advisory & Editorial Board ▾ | Data Policies ▾
Home ▸ Data Descriptors ▸ Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN

Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells

Bruno Di Stefano, Samuel Collombet & Thomas Graf

Affiliations | Contributions | Corresponding authors

Scientific Data **1**, Article number: 140008 | doi:10.1038/sdata.2014.8
Received 21 February 2014 | Accepted 22 April 2014 | Published online 27 May 2014

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

Associated Links

Nature | Article

C/EBPα poises B cells for r... into induced pluripotent ste...

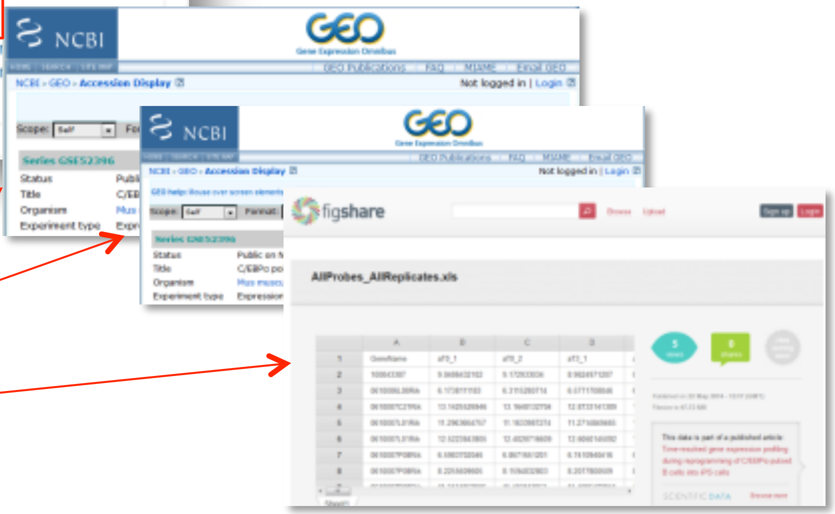
by Bruno Di Stefano *et al*

Data Citations

Abstract • Background and Summary • Methods • Data Records • Technical Validation • Usage Notes • Additional information • References • Data Citations • Acknowledgements • Author information

1. Di Stefano, B., Collombet, S., & Graf, T. *Gene Expression Omnibus* GSE46321 (2014).
2. Di Stefano, B., Collombet, S., & Graf, T. *Gene Expression Omnibus* GSE52396 (2014).
3. Di Stefano, B., Collombet, S., & Graf, T. *Figshare* http://dx.doi.org/10.6084/m9.figshare.939408 (2014).

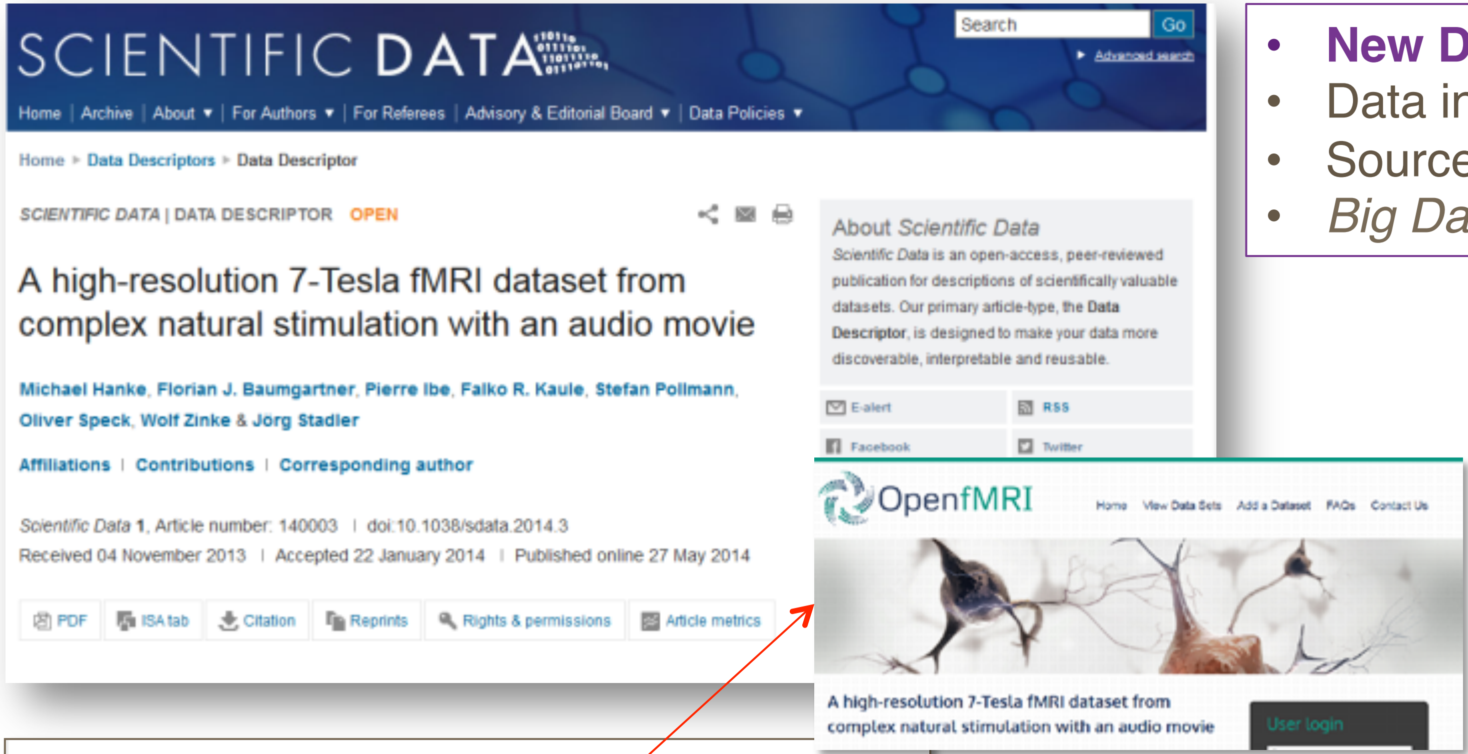
- ***Associated Nature Article***
- Data at **fig**share & NCBI GEO
- Integrated **fig**share data viewer



nature publishing group npg

Neuroscience

- **New Dataset**
- Data in OpenfMRI
- Source code in GitHub
- *Big Data*



SCIENTIFIC DATA
Search
Go
Advanced search
Home | Archive | About | For Authors | For Referees | Advisory & Editorial Board | Data Policies
Home ▸ Data Descriptors ▸ Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN

A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie

Michael Hanke, Florian J. Baumgartner, Pierre Ibe, Falko R. Kaule, Stefan Pollmann, Oliver Speck, Wolf Zinke & Jörg Stadler

Affiliations | Contributions | Corresponding author

Scientific Data **1**, Article number: 140003 | doi:10.1038/sdata.2014.3
Received 04 November 2013 | Accepted 22 January 2014 | Published online 27 May 2014

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie

Data Citations

Abstract • Background & Summary • Methods • Data Records • Technical Validation • Usage Notes • Additional information • References • Data Citations • Acknowledgements • Author information • Supplementary information

1. Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., Zinke, W., & Stadler, J. *OpenfMRI* ds000113 (2014).

Additional resources:

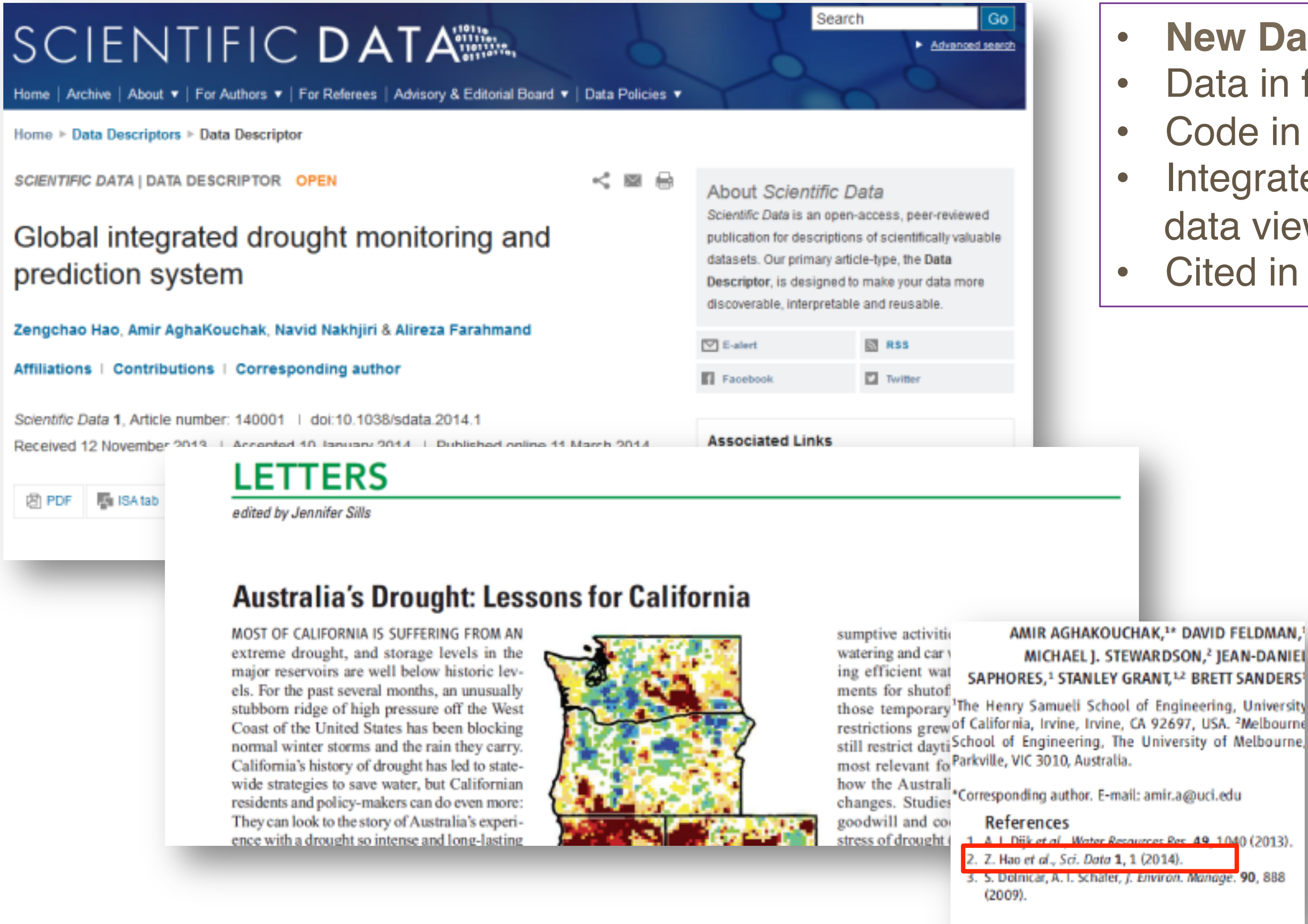
- More information and updates are made available at: http://www.studyforrest.org
- Source code repository: http://github.com/hanke/gumpdata
- Documentation for the source code: http://gumpdata.readthedocs.org

Code in GitHub

nature publishing group npg

Environmental

SCIENTIFIC DATA



SCIENTIFIC DATA

Search Go

Advanced search

Home | Archive | About ▾ | For Authors ▾ | For Referees | Advisory & Editorial Board ▾ | Data Policies ▾

Home ▸ Data Descriptors ▸ Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN

Global integrated drought monitoring and prediction system

Zengchao Hao, Amir AghaKouchak, Navid Nakhjiri & Alireza Farahmand

Affiliations | Contributions | Corresponding author

Scientific Data **1**, Article number: 140001 | doi:10.1038/sdata.2014.1

Received 12 November 2013 | Accepted 10 January 2014 | Published online 11 March 2014

PDF ISA tab

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

E-alert RSS

Facebook Twitter

Associated Links

LETTERS

edited by Jennifer Sills

Australia's Drought: Lessons for California

MOST OF CALIFORNIA IS SUFFERING FROM AN extreme drought, and storage levels in the major reservoirs are well below historic levels. For the past several months, an unusually stubborn ridge of high pressure off the West Coast of the United States has been blocking normal winter storms and the rain they carry. California's history of drought has led to statewide strategies to save water, but Californian residents and policy-makers can do even more: They can look to the story of Australia's experience with a drought so intense and long-lasting

sumptive activiti
watering and car
ing efficient wat
ments for shutof
those temporary
restrictions grew
still restrict dayti
most relevant fo
how the Australi
changes. Studies
goodwill and co
stress of drought

AMIR AGHAKOUCHAK,[1*] DAVID FELDMAN,[1]
MICHAEL J. STEWARDSON,[2] JEAN-DANIEL
SAPHORES,[1] STANLEY GRANT,[1,2] BRETT SANDERS[1]

[1]The Henry Samueli School of Engineering, University of California, Irvine, Irvine, CA 92697, USA. [2]Melbourne School of Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.

*Corresponding author. E-mail: amir.a@uci.edu

References

1. A. I. Dijk *et al.*, *Water Resources Res.* **49**, 1040 (2013).
2. Z. Hao *et al.*, *Sci. Data* **1**, 1 (2014).
3. S. Dolnicar, A. I. Schafer, *J. Environ. Manage.* **90**, 888 (2009).

- **New Dataset**
- Data in **fig**share
- Code in **fig**share
- Integrated **fig**share data viewer
- Cited in *Science*

The right licence

Data Descriptor article: Licensed under one of two Creative Commons licenses, by author choice:

CC BY   CC BY NC

Metadata: released under the **CC0 waiver** to maximize reuse and aid data miners

Data: depends on public repositories. Partner repositories figshare and Dryad both use the **CC0 waiver**.

Repository criteria

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

1. Broad support and recognition within their scientific community
2. Ensure long-term persistence and preservation of datasets
3. Provide expert curation
4. Implement relevant, community-endorsed reporting requirements
5. Provide for confidential review of submitted datasets
6. Provide stable identifiers for submitted datasets
7. Allow public access to data without unnecessary restrictions

Thank you

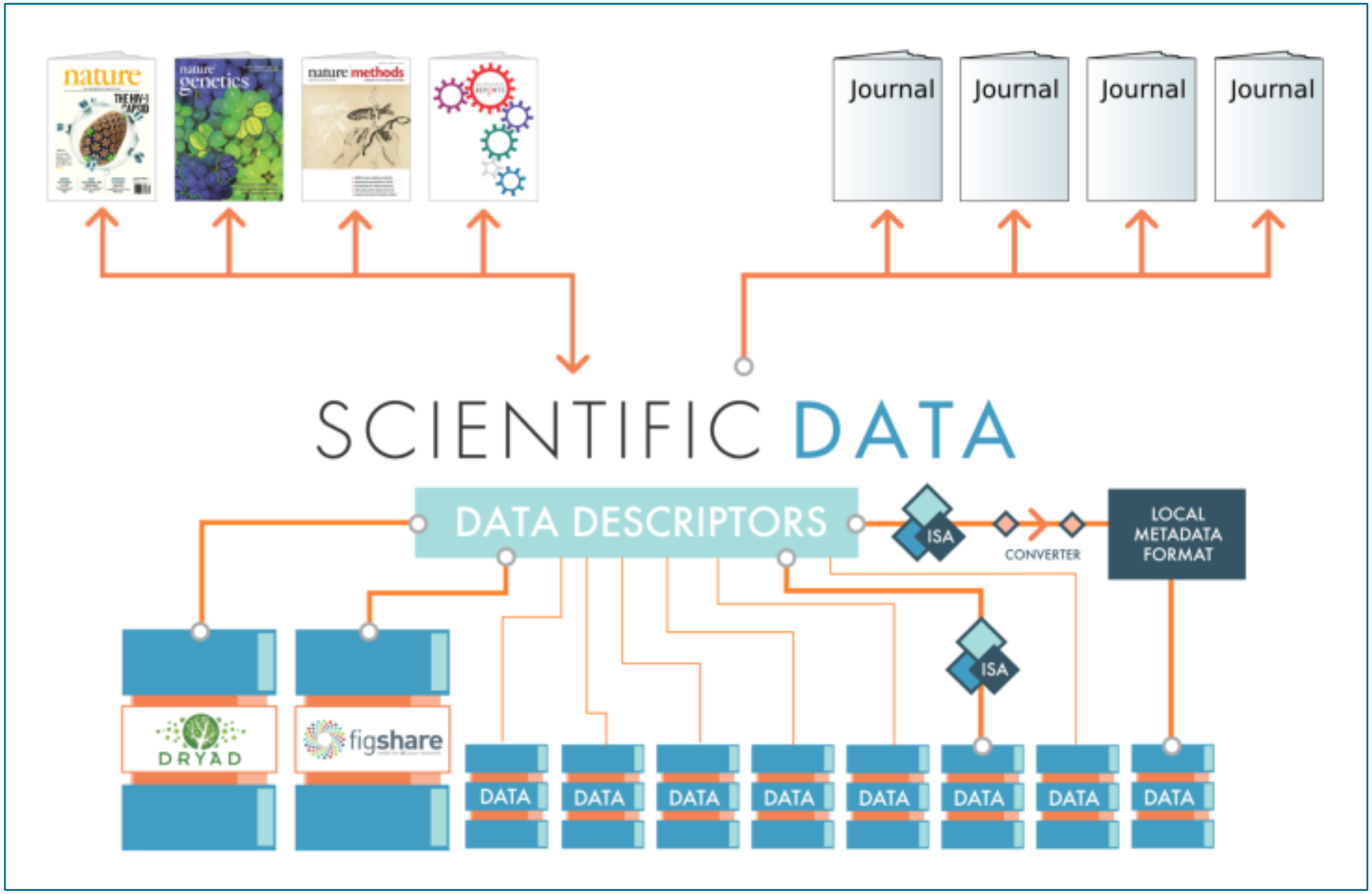
For more information
please contact

IAIN HRYNASZKIEWICZ
Head of Data and HSS Publishing,
Open Research

M: +44 (0)7814 290576
T: +44 (0)207 0146753
E: iain.hrynaszkiewicz@nature.com

macmillan
Science and Education

Transforming Learning and Discovery



nature

nature genetics

nature methods

Journal

Journal

Journal

Journal

SCIENTIFIC DATA

DATA DESCRIPTORS

ISA

CONVERTER

LOCAL METADATA FORMAT

ISA

DRYAD

figshare

DATA

DATA

DATA

DATA

DATA

DATA

DATA

DATA

Get Credit for Sharing Your Data

Publications will be indexed and citeable.

Open-access

Creative Commons licenses (CC-BY/CC-BY-NC) for the main Data Descriptor. Each publication supported by CCO metadata.

Focused on Data Reuse

All the information others need to reuse the data; no interpretative analysis, or hypothesis testing

Peer-reviewed

Rigorous peer-review focused on technical data quality and reuse value

Promoting Community Data Repositories

Not a new data repository; data stored in community data repositories