

オープンサイエンスの動向と研究データ公開の推進における出版社の取り組み

データの再現性を促進する
Scientific Data

Open Research
Marketing Manager

新谷洋子

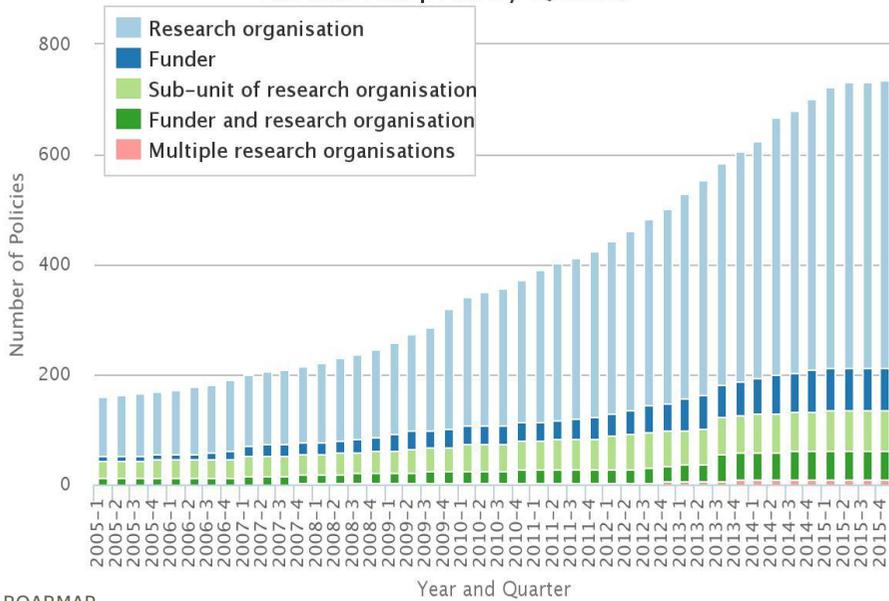
y.shintani@nature.com

12 November, 2015



Transforming Learning and Discovery

Policies Adopted by Quarter



ROARMAP
<http://roarmap.eprints.org/>

Highcharts.com

The Open Academic Tidal Wave



1. **Recommended open access to scholarly papers of publicly funded research**
2. **Recommended open access to all digital outputs of publicly funded research**
3. **Mandated open access to scholarly papers of publicly funded research**
4. **Mandated open access to all digital outputs of publicly funded research**
5. **Enforced, mandated open access to scholarly papers of publicly funded research**
6. **Enforced, mandated open access to all digital outputs of publicly funded research**

Hahnel, Mark (2015): Global funders who require data archiving as a condition of grants.
 figshare. <http://dx.doi.org/10.6084/m9.figshare.1281141>

科学データの公開にまつわるさまざまな課題

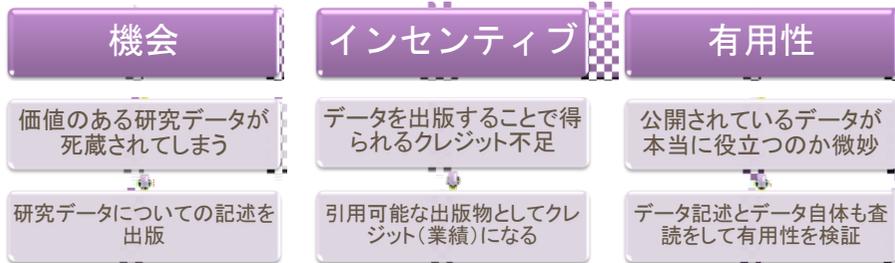
- すべての研究プロジェクトが研究成果に結びつくわけではない
- 研究データが死蔵されてしまう
- 価値のある情報が発見しにくい場所に置かれてしまう
- データリポジトリ不足またはデータリポジトリの認知不足
- データを公開するのは勿体無い
- 業績を出すことに負われて時間をかけられない
- データを整備して公開する作業に対する見返りが少ない
- 自身が公開したデータの不適正な再利用に関する懸念
- データ自体の査読は主要なジャーナルでは行わない
- 公開されているデータの使い勝手が悪い

機会

インセンティブ

有用性

データ共有に関する問題点と解決策



データ自体の有用性が検証済みの研究データを発見しやすく、検索しやすく、理解しやすく、つまり再現・再利用されやすい形で引用可能な業績(文献)として出版

5

macmillan
Science and Education

Scientific Data 市場調査 (2011年実施)

Key survey data

- 60%の研究者はすでに他の研究者とデータをシェアしている
- 50%はそういったシェアされたデータを月に一度は見ている
- 45%はどこにデータをポストしたらよいのかわからない
- 90%がデータジャーナルのコンセプトを肯定的に受け止めた
- 80%がデータジャーナルの存在がデータの共有を促すだろうと考えた

Survey respondent characteristics

- 387 respondents (329 active researchers)
- Physics (24%), Earth and environmental science (21%), Biology (20%) Chemistry (19%) Others (16%)

6

macmillan
Science and Education

SCIENTIFIC DATA



Credit : クレジット

引用可能な形で研究データを公開できるので、クレジットが得られます



Open : オープン

クリエイティブコモンズライセンスでデータの再利用を推進します



Reuse : 再利用

データ記述をキュレーションによって標準化し再利用に役立てます



Quality : 品質

各コミュニティの基準に基づいた厳格な査読と社内キュレーションで品質と再利用性を保証します



既存のデータリポジトリを促進

各研究コミュニティに特化した既存のデータリポジトリへのデータ公開をサポートします

7

macmillan
Science and Education

The screenshot shows the Scientific Data journal website. At the top, there is a search bar and navigation links. The main content area features a 'Featured Data Descriptor' with a photograph of a bird and a red flower. The text describes biological, ecological, conservation, and legal information for all species and subspecies of Australian birds. Below this, there is a 'Latest Content' section with two entries: 'A resource of ribosomal RNA-depleted RNA-Seq data from different normal adult and fetal human tissues' and 'Construction, complete sequence, and annotation of a BAC contig covering the silkworm chorion locus'. On the right side, there is an 'About Scientific Data' section, social media links, and a 'Submit manuscript' button. At the bottom, there are logos for Nature and Macmillan.

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

8

macmillan
Science and Education

SCIENTIFIC DATA

- 2014年5月創刊
- オンライン限定のオープンアクセス誌:誰でも利用可能
- 自然科学、物理学、社会科学のデータセットを掲載する査読誌
- 科学的に貴重なデータセットの記述を掲載
- 新しい形式のコンテンツ「Data Descriptor」を掲載

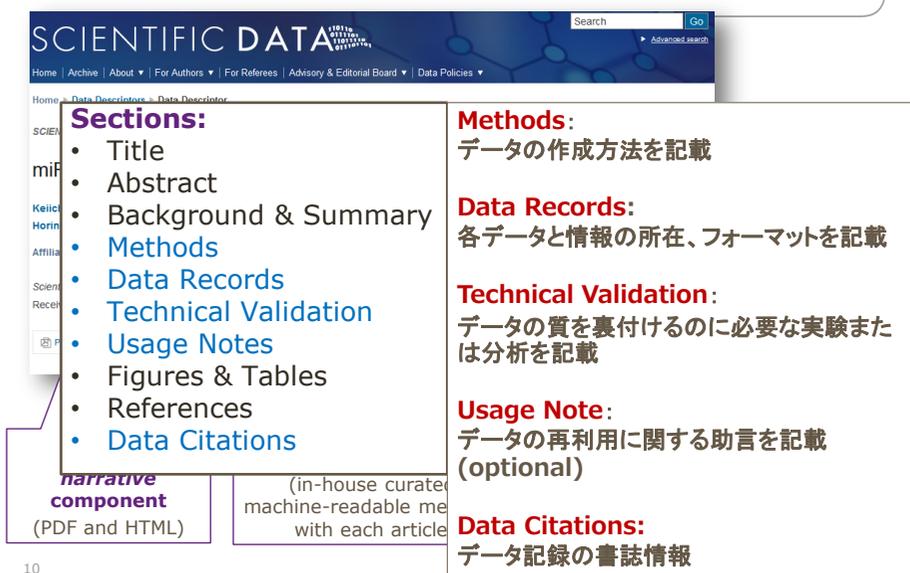
Data Descriptorとは？

- 実験・観察データセットの詳細な記述を提供するための新しい論文形式
- 論文ではなく“Data Descriptor”つまり、データについての説明を記述したもの
- データ収集の背景、データ所蔵場所、データ取得方法、データの品質を示す技術的分析、データ利用方法などの記述
- コンピューター処理可能な形式に、実験デザイン情報【表】をメタデータとして落とし込んだものが論文に付随

9

 macmillan
Science and Education

The 'Data Descriptor'



Sections:

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

Methods:
データの作成方法を記載

Data Records:
各データと情報の所在、フォーマットを記載

Technical Validation:
データの質を裏付けるのに必要な実験または分析を記載

Usage Note:
データの再利用に関する助言を記載
(optional)

Data Citations:
データ記録の書誌情報

narrative component
(PDF and HTML)

(in-house curated machine-readable metadata with each article)

10

SCIENTIFIC DATA

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/EBP α -pulsed B cells into iPS cells

Bruno Di Stefano, Samuel Collobet & 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

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

About Scientific Data
Scientific Data is an open-access, peer-reviewed publication for descriptions of scientific datasets. Our primary article-type, the Data Descriptor, is designed to make your dataset discoverable, interpretable and reusable.

Associated Links
Nature | Article
C/EBP α into inducible pluripotent stem cells
by Bruno Di Stefano, Samuel Collobet & Thomas Graf

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

- Di Stefano, B., Collobet, S., & Graf, T. *Gene Expression Omnibus* GSE46321 (2014)
- Di Stefano, B., Collobet, S., & Graf, T. *Gene Expression Omnibus* GSE52396 (2014)
- Di Stefano, B., Collobet, S., & Graf, T. *Figshare* <http://dx.doi.org/10.6084/m9.figshare.939408> (2014).

Associated Nature Article

- Data at **figshare** & NCBI GEO
- Integrated **figshare** data viewer

OpenfMRI | Home | New Data Sets | Add a Dataset | FAQs | Contact Us

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 & Jorg 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

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

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

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

Code in GitHub

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

11 <http://www.nature.com/articles/sdata20148>



SCIENTIFIC DATA

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 & Jorg 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

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

About Scientific Data
Scientific Data is an open-access, peer-reviewed publication for descriptions of scientific datasets. Our primary article-type, the Data Descriptor, is designed to make your dataset discoverable, interpretable and reusable.

Associated Links
Nature | Article
A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie
by Michael Hanke, Florian J. Baumgartner, Pierre Ibe, Falko R. Kaule, Stefan Pollmann, Oliver Speck, Wolf Zinke & Jorg Stadler

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

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

Code in GitHub

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

12 <http://www.nature.com/articles/sdata20143>



SCIENTIFIC DATA

Home > Data Descriptors > Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR

OPEN

Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna

Alexandra Swanson, Margaret Kosmala, Chris Lintott, Robert Simpson, Arfon Smith & Craig Packer

Affiliations | Contributions | Corresponding author

Scientific Data 2 Article number: 150026 (2015) | doi:10.1038/sdata.2015.26
Received 08 December 2014 | Accepted 29 April 2015 | Published online 09 June 2015

PDF ISA tab Citation Reprints Rights & permissions Article metrics

Data Citations

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

1. Swanson, A. *Dryad* <http://dx.doi.org/10.5061/dryad.5pt92> (2015).

DRYAD About - For researchers - For organizations -

Data from: Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna

- New Dataset
- Data in Dryad
- Citizen science
- Crowd sourcing

13

<http://www.nature.com/articles/sdata201526>



一次データの登録について

Scientific Dataはコミュニティに認知されたデータリポジトリにデータセットを登録することを推奨



約80のリポジトリと提携しています

SCIENTIFIC DATA

Home Archive About For Authors For Referees Data Policies Collections

Featured Data Descriptor

Biological, ecological, conservation and

About Scientific Data

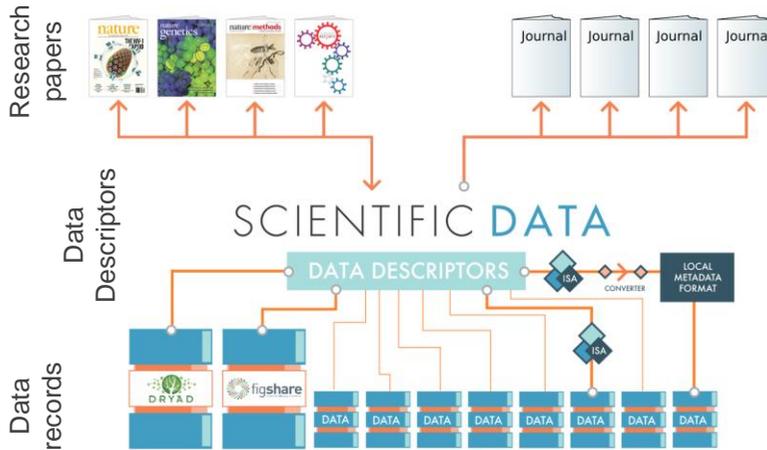
Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable

14

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



データ出版のエコシステム



15

macmillan
Science and Education

まとめ

- オープンサイエンス推進と研究データ公開義務化の波
- 研究データ公開における複数の課題
- 解決策としてのデータ論文、データジャーナル
- データを発見・検索・理解しやすいフォーマットにより再現性促進
- データを登録、保管するデータリポジトリとの協同によりデータ公開を促進
- 正しいデータ公開を促すことでオープンサイエンスを加速

16

macmillan
Science and Education