

データの再現性と再利用を
促進するデータジャーナル
—Scientific Dataを例に—

2016.2.8

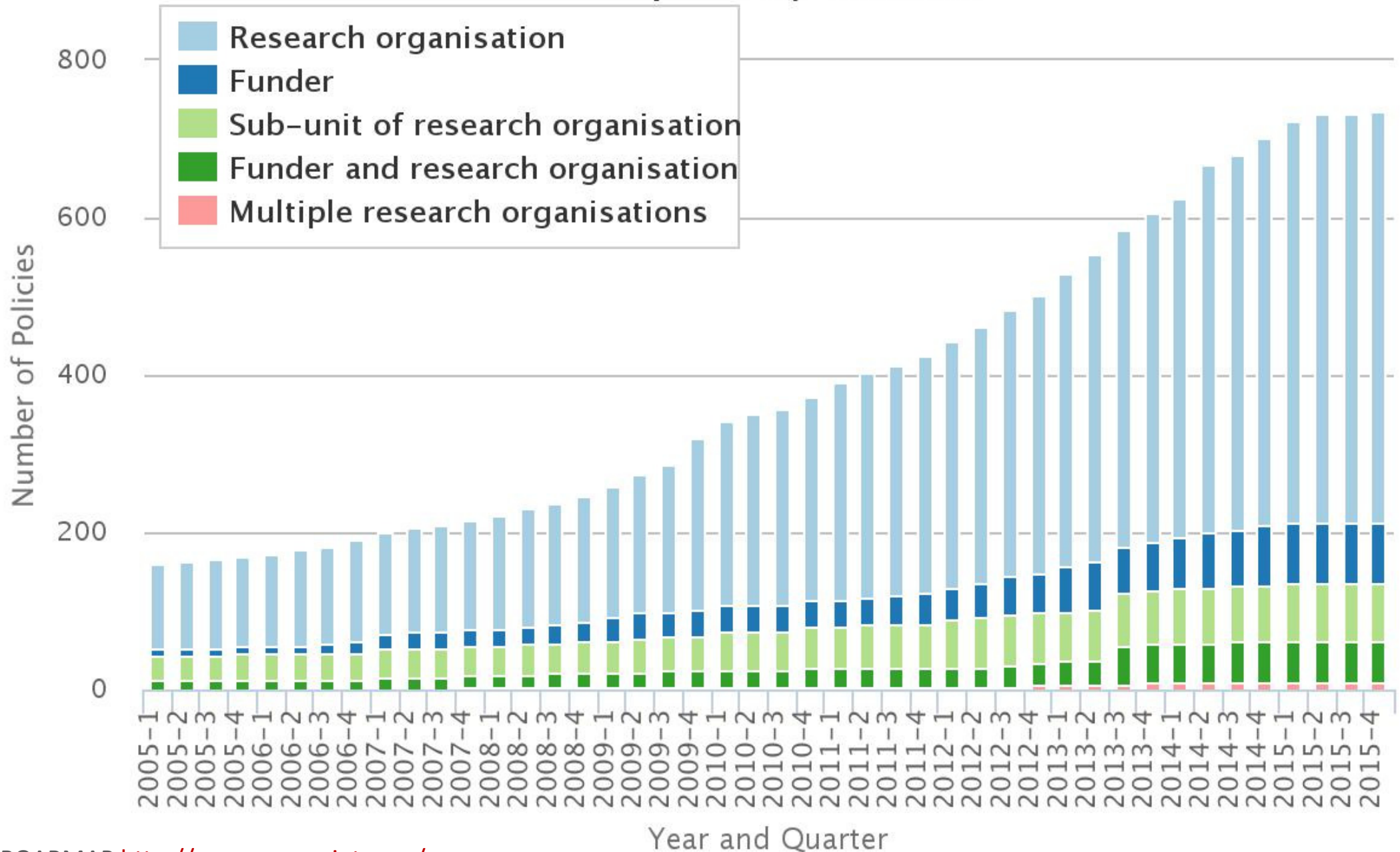
Open Research
Marketing Manager

新谷洋子

Nature Publishing Group

SPRINGER NATURE

Policies Adopted by Quarter



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>

なぜデータを公開するのか？

- 投資への費用対効果を増大（研究費と時間）
- 他者の目にさらすことで新しい仮説の検証を可能に
- 複数の研究結果を使ったメタ分析を効率的に行う
- 異なるデータセット間の統合を可能に
- 科学的発見の再現性や妥当性の検証
- 誤りや不正を未然に防ぐ

研究結果の透明性を高め、再現性を確証し、再利用や新しい発見を促すには論文をオープンアクセスにするだけでは不十分
→ 付随する他の成果物（データ、プロトコル、ソースコード）も自由に利用できる形で公開する必要がある

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データ公開に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
- 公開されているデータの使い勝手が悪い
- データ自体の査読は主要なジャーナルでは行わない

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

機会

価値のある研究データが死蔵されてしまう



データを出版し、データが世に出る機会を創出

インセンティブ

データを出版することで得られるインセンティブが不足



引用可能な出版物としてクレジット(業績)になる

有用性

公開されているデータが本当に役立つのか微妙



データ記述、データ自体も査読をして有用性を検証

データ論文とは？



- データについて記述したものを査読して出版することでデータの利活用を促進
 - データ記述を引用可能な文献（業績）にし、データを公開したことがクレジットになる
- データ論文を出版するデータジャーナルが誕生
GigaScience, Earth Systems Science Data, etc.

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

- 回答者 387 (329が現役の研究者)
- 物理学(24%), 地球・環境科学(21%), 生物学(20%), 化学(19%) その他(16%)

Scientific Data 市場調査 (2011年実施)

データ出版に関して研究者が望むことは？

- 96% データの認知度や見つけやすさの向上
- 95% データ自体のユーザビリティの向上
- 93% データを公開することによるクレジット
- 80% データセットやコンテンツの査読

Nature Publishing Group (2014): Data publication survey - raw data. [figshare](http://dx.doi.org/10.6084/m9.figshare.1234052).
<http://dx.doi.org/10.6084/m9.figshare.1234052>

**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の利用促進**

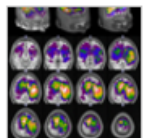
各研究コミュニティに特化した既存のデータリポジトリへのデータ公開をサポートします

Featured Data Descriptor



A collection of non-human primate computed tomography scans housed in MorphoSource, a repository for 3D data
Copes *et al.* | 2nd February 2016

Evolutionary and anthropological research often relies on exacting comparative measurements of bones and fossils. Here, the authors release a dataset comprised of 431 computed tomography scans of skulls from 59 non-human primate species, each supported by craniometric landmark data. These scans will help facilitate detailed morphological assessments, while also making these valuable museum specimens more widely accessible in digital form.



Data Descriptor | 02 February 2016

A structural and functional magnetic resonance imaging dataset of brain tumour patients

Cyril R. Pernet, Krzysztof J. Gorgolewski,

Dominic Job, David Rodriguez, Ian Whittle & Joanna Wardlaw



Data Descriptor | 02 February 2016

DNA methylation profiling of primary neuroblastoma tumors using methyl-CpG-binding domain sequencing

Anneleen Decock, Maté Ongenaert, Wim Van Crielinge, Frank Speleman & Jo Vandesompele

About *Scientific Data*

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

[Browse publications by date](#) ▶

[Browse publications by subject](#) ▶

✉ E-alert

📡 RSS

f Facebook

🐦 Twitter

[Submit manuscript](#) ▶

Announcements

Publishing Better Science through Better Data

Scientific Data's annual conference on sharing and publishing research data took place on 23 October.

[Video and slides for all talks now available!](#)

Scientific Data Updates

A blog by *Scientific Data*

Call for submissions: Zika virus related datasets
20 January 2015

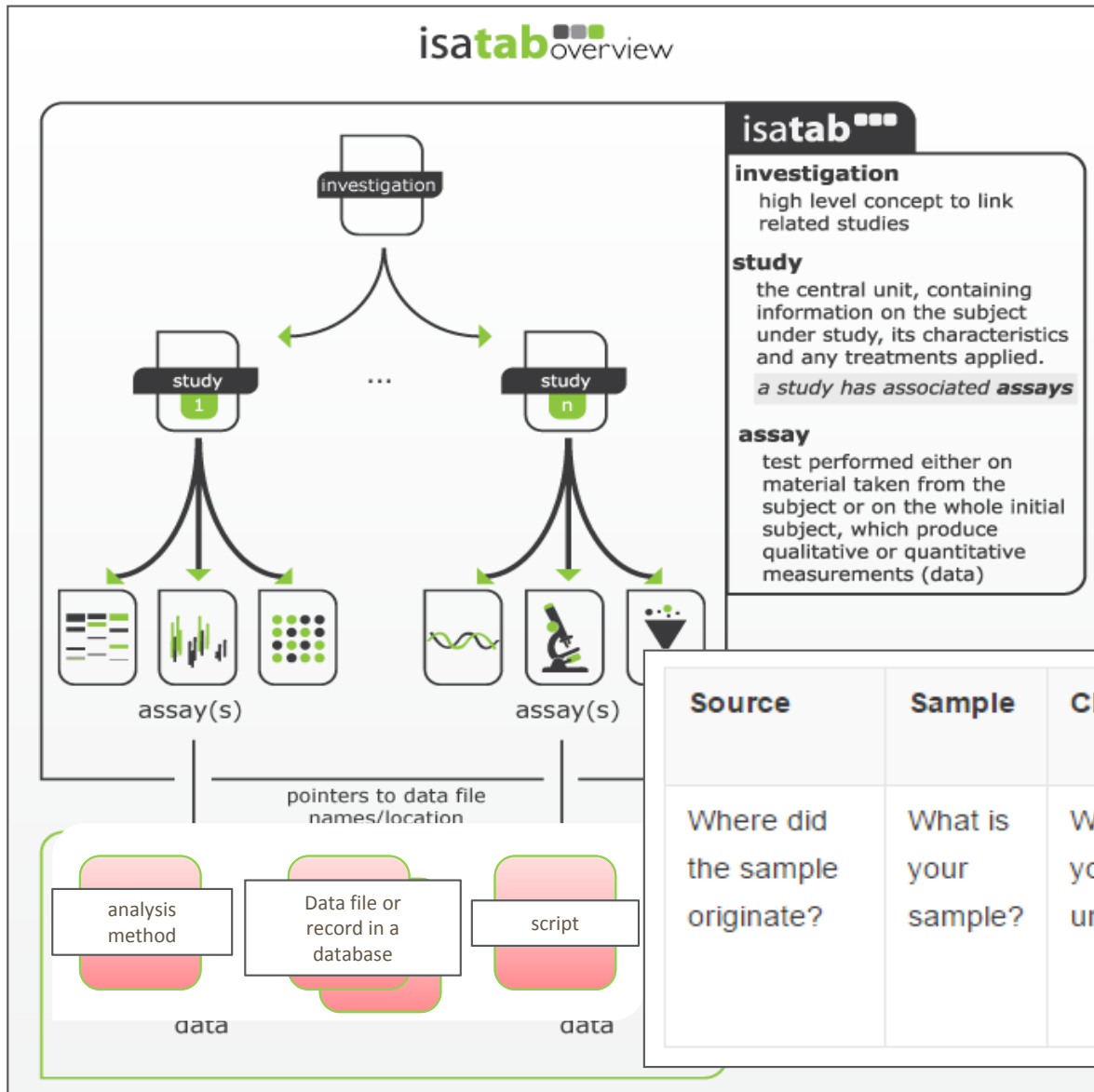
SCIENTIFIC DATA

- 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
- 新たな仮説の検証、知見、手法などは含まない
- 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

Data Descriptor 構造化されたメタデータ



社内キュレーション:

- 著者がデータを構造化されたコンテンツにする際のサポート
- 実験メタデータにセマンティックな解釈など付加価値をつける

Source	Sample	Characteristics	Protocol	Data
Where did the sample originate?	What is your sample?	What do users need to know about your source and samples to fully understand your work?	How did the samples become data?	Where is the data?

Data Descriptor

Human readable
representation of study
i.e. article (HTML & PDF)

Machine readable
representation of
study
i.e. metadata

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 Packer

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Scientific Data **2**, Article number: 150026 (2015)
Received 08 December 2014 | Accepted 25 April 2015 | Published 25 April 2015



PDF



ISA tab



Citation



Reprints



Rights & permissions



Article metrics

Abstract

[Abstract](#) • [Background & Summary](#) • [Methods](#) • [Data Records](#) • [Technical Validation](#) • [Usage Notes](#) • [Additional Information](#) • [References](#) • [Data Citations](#) • [Acknowledgements](#) • [Author information](#)

Camera traps can be used to address large-scale questions in community ecology by providing

Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR [OPEN](#)



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

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:

データの書誌情報

[Additional Information](#) • [References](#)

camera traps can be used to



SCIENTIFIC DATA

Menu

Home ▶ Data Descriptors ▶ Data Descriptor

SCIENTIFIC DATA | DATA DESCRIPTOR
OPEN

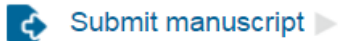
Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna

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



nature collections
Cancer immunotherapy
Presented by 

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

About Scientific Data

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

 E-alert

 RSS

 Facebook

 Twitter

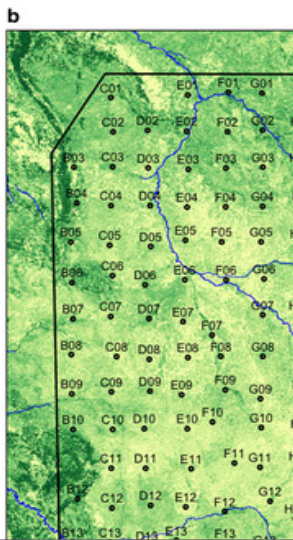
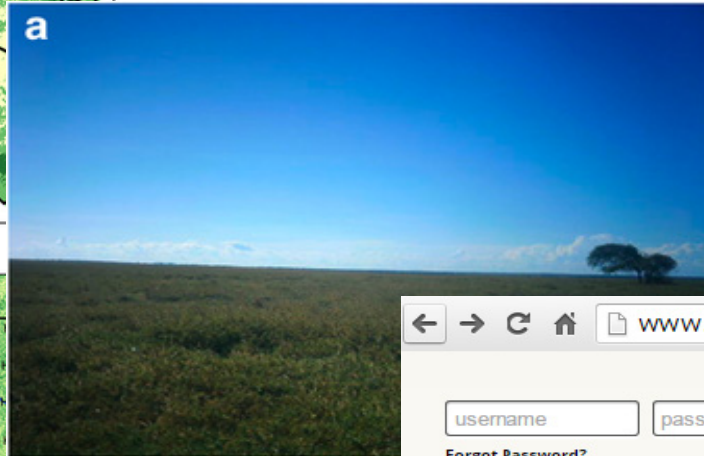
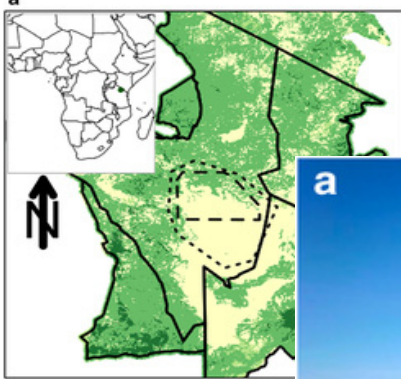
Data Citations

[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).

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

SCIENTIFIC DATA



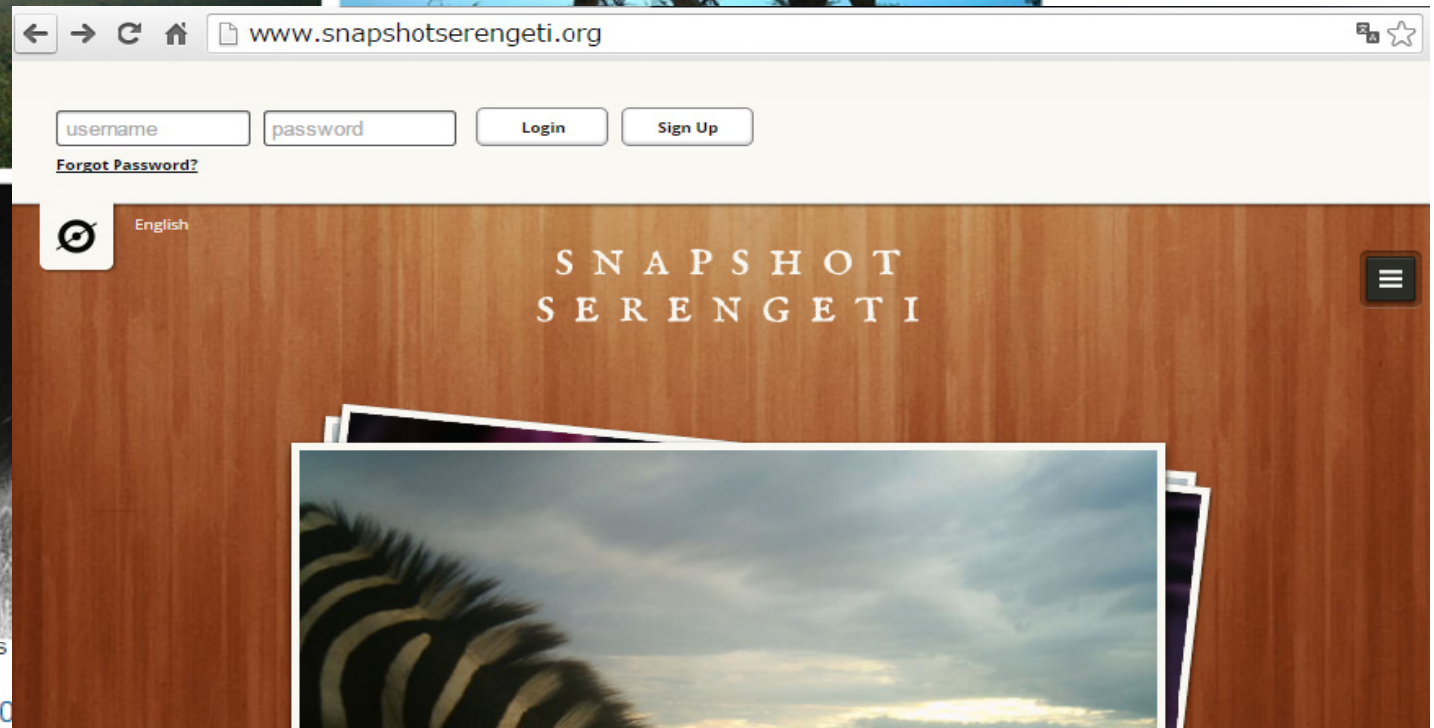
Data Citations

[Abstract](#) • [Background](#)

[Additional Information](#) • [References](#) • [Data Citations](#)

1. Swanson, A. *Dryad* <http://dx.doi.org/10.5061/dryad.5q1k1>

- New Dataset
- Data in Dryad
- Science
- Science



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

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).

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

About *Scientific Data*

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

E-alert

RSS

Facebook

Twitter



About ▾

For researchers ▾

For organizations ▾

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

SCIENTIFIC DATA

出版されたさまざまなData Descriptor

古いデータ
セット

Research | 26 May 2015 | OPEN

An archive of longitudinal recordings of the vocalizations of chimpanzees

Frans X. Plooij, Hetty van de Rijt-Plooij [...] Anne Pusey

キュレーションにより
集約されたデータ
リソース

計算に基づき作成され
たデータとソースコード
の公開

15 September 2015 | OPEN

of host-pathogen

on

h, Claire Risley

ata 2, 150049

Research | 25 March 2015 | OPEN

Sequence variants from whole genome sequencing of Icelanders

Daniel F Gudbjartsson, Patrick Sulem [...] Kari Stefansson

Scientific Data 2, 150011

大規模研究コン
ソーシアムによ
るデータセット

インパクトの高い論
文に付随するデータ

ch 2014

rated drous

Amir A

1, 1400

Research | 09 Decem

An open science resource for functional connectome

Xi-Nian Zeng

Scientific Data 2, 150047

著者が価値があると考え、他の人も使える
データを対象とします

単一実験による
データセット

Research |

The proteome

Jocelyn F. Kreitzer

Scientific Data 2, 150047

population

cibility in

elopment

一次データの登録

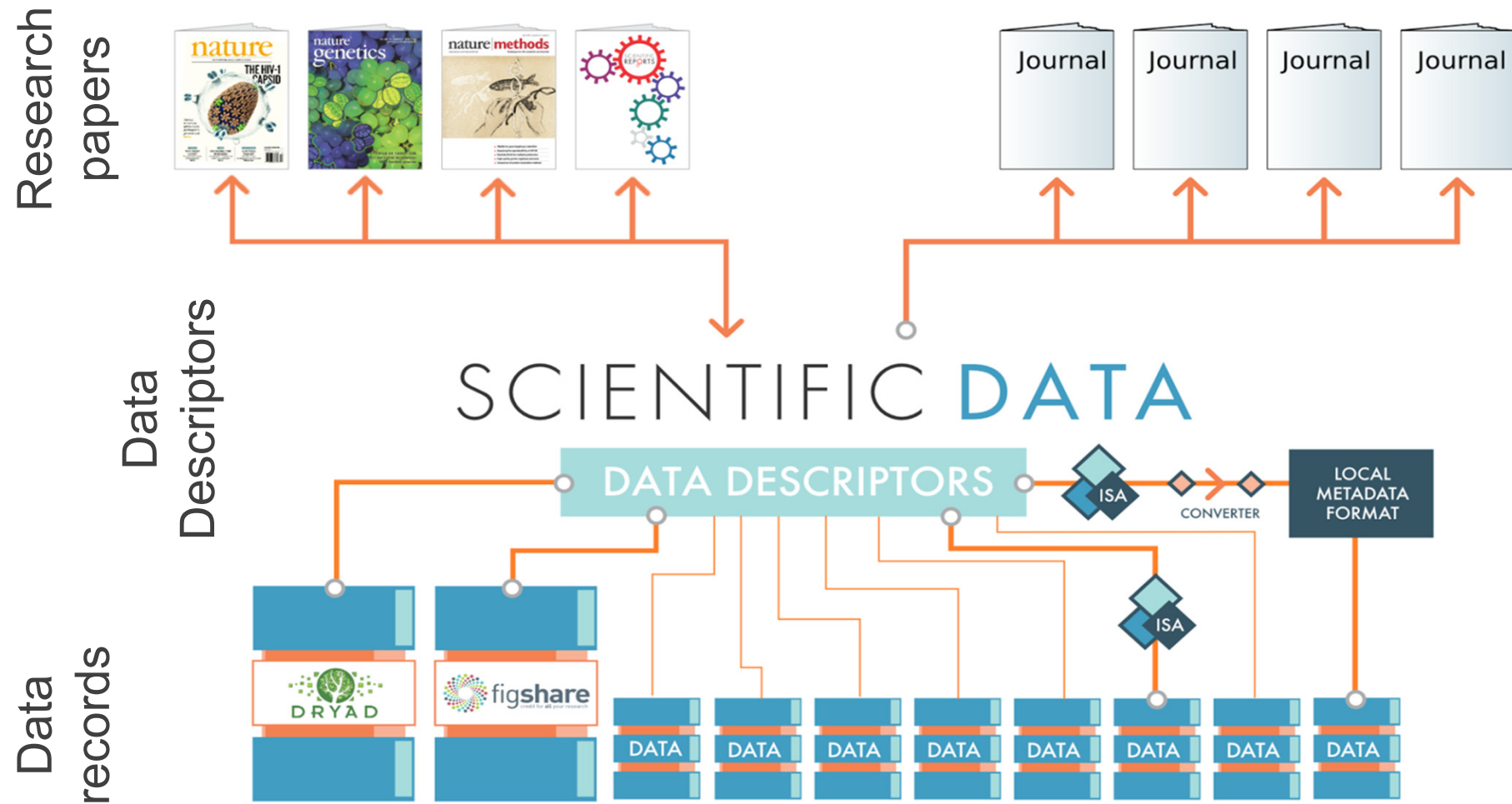
Scientific Dataはコミュニティに認知されたデータリポジトリにデータセットを登録することを推奨



- 現在 80近くのデータリポジトリと提携
- 生命科学、医学、化学、物理学、社会科学

*どのリポジトリに登録すべきか、著者にアドバイスもしています

データ出版のエコシステム



WHATS NEW

2015年最もアクセスされたData Descriptor:
Snapshot Serengeti, high-frequency annotated
camera trap images of 40 mammalian species in an
African savanna

Page views (25,533)



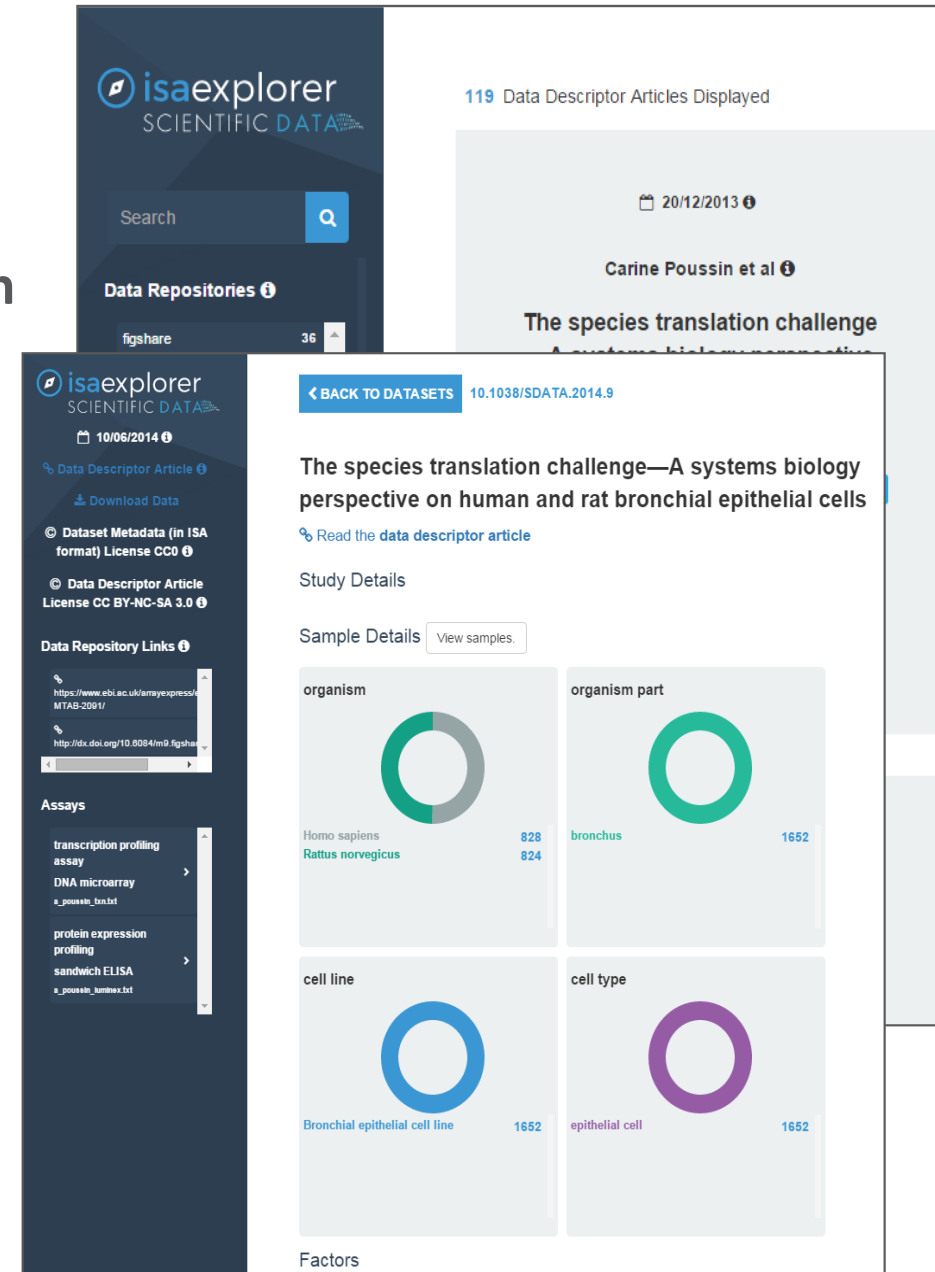
*上記数値は2016年2月8日現在のもの

2015年最も引用されたData Descriptor:
Global integrated drought monitoring and
prediction system

Publishing Better Science through Better Data (毎年開催)

ISA Explorer テスト運用開始 (2015.12--)

Data Curation Editor がRDAに合わせて来日



まとめ

- オープンサイエンス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
- 解決策としてデータ論文、データジャーナルが誕生
- 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を確保、再利用を促進
- 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

SCIENTIFIC DATA

Supported by



Honorary Academic Editor

Susanna-Assunta Sansone

Managing Editor

Andrew L. Hufton

Data Curation Editor

Varsha K. Khodiyar

Advisory Panel and Editorial Board

including senior researchers, funders,
librarians and curators

Visit nature.com/sdata

Email scientificdata@nature.com

Tweet @ScientificData