The State of Open Data

@MimiKeshani
Implementation Manager, figshare

What are data?

- “Facts gathered for reference or analysis” [Apple dictionary]
- Components of the research → inherently valuable products of research
- Data means different things to different people

What does it mean for data to be open?

- *"As open as possible, as closed as necessary"*

  EC Horizon2020 Guidelines

- FAIR – FORCE11 guiding principles to make data open
  - Findable
  - Accessible
  - Interoperable
  - Reusable

→ **Human and machine actionable**

Data FAIRport

Find, Access, Interoperate & Re-use Data

Find
Access
Interoperate
Re-use
Data

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtls.nl).
#A full list of authors and their affiliations appears at the end of the paper.

doi:10.1038/sdata.2016.18
http://www.datafairport.org/

What does it mean to be open?

- Data citation principles
  - Focus on credit, interoperability, flexibility and access to data.

→ Both FAIR and data citation principles are designed as guidance and do not give specifics of implementation

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [https://www.force11.org/group/joint-declaration-data-citation-principles-final]

What are the concerns of being open?

- Donation requires additional labour
- Scooping and potential for others' careers to progress ahead of your own
- Getting sued for making the wrong data available
- Analysis interrogation and errors spotted

So, do the benefits of sharing data overcome these worries?

Importance of being open

- Transparency breeds trust
  - Far beyond research – policy makers, industry, publishers, tax payers
- Free data access has greatly helped down stream capital flows [1][2]
- Build on work that has already been done – inherently improves reproducibility, replication and waste reduced

1 Sunder Rajan, Kaushik. 2006. Biocapital: The Constitution of Post-Genomic Life. Durham, NC: Duke University Press.
2 Kelty, Christopher M. 2008. Two Bits: The Cultural Significance of Free Software. Duke University Press.

Government & Funder Mandates

North America

- OSTP Memo
- NIH, NSF, NOAA, SI, and more…
- Tri-Agency Open Access Policy
- FASTR
- Other Agency/Funder Policies
  - **ie: Gates, Moore, Sloan, NEH, Howard Hughes**

UK / Europe

- Concordat on Open Research Data (HEFCE, RCUK, UUK, Wellcome Trust, EPSRC, and more)
- Horizon 2020
- OpenAIRE

Australia

- **ARC** – data management plan expectation from researchers
- **Australia National Data Service** – supporting the sector, Institutional Data Management Frameworks

Cost of being open

- Over the next 8 years, the amount of data created per year will increase 10X to over 160 zettabytes – (with 95% originating from the world of IoT – according to a 2016 IDC report).

- However, they estimate only 15% of that data can be stored

http://techblog.comsoc.org/2016/03/09/idc-directions-2016-iot-internet-of-things-outlook-vs-current-market-assessment/

Cost of being open

- NSF is rationalising investments
- EU have already denounced future costs of large biomed databases

→ calls for centralisation via programmes like ELIXIR.

→ Database curators and repositories must provide sustainable business models

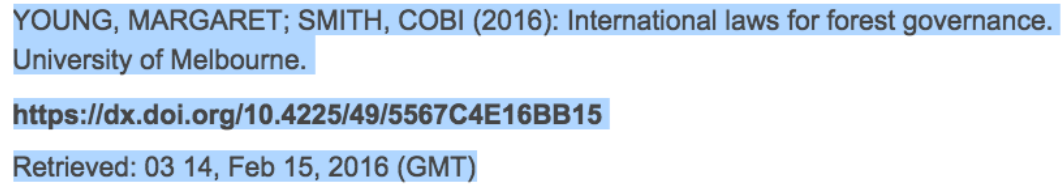
What is figshare?

- A data repository
- A place for academics to make their research openly available
- Cloud storage

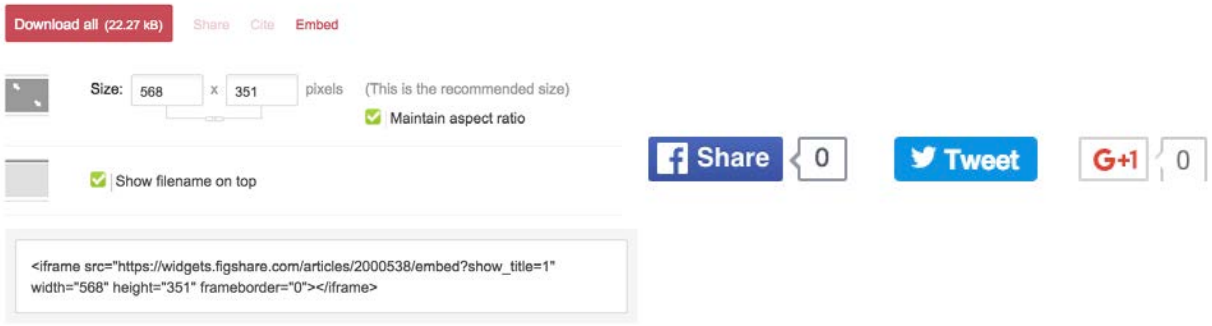
→ **Our ultimate focus to aid in the reproducibility, replication, and reuse of research data**

What is figshare?

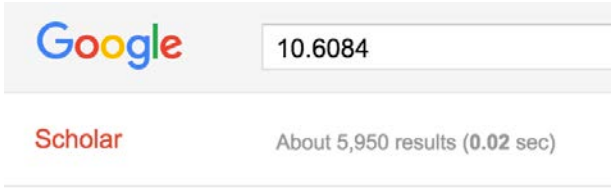
citable



shareable



discoverable



Citations

PIVlab - Time-Resolved Digital Particle Image Velocimetry Tool for MATLAB

Version 6 26.03.2015, 08:17 by William Thielicke, Eize J. Stamhuis

PIVlab is a time-resolved particle image velocimetry (PIV) software that does not only calculate the velocity distribution within particle image pairs, but can also be used to derive, display and export multiple parameters of the flow pattern. A user-friendly graphical user interface (GUI) makes PIV analyses and data post-processing fast and efficient.
Screen-capture video of the tool:
http://vimeo.com/10090907
Example analyses & videos can be found on the PIVlab website:

4278 views	616 downloads	14 citations

1

CATEGORIES

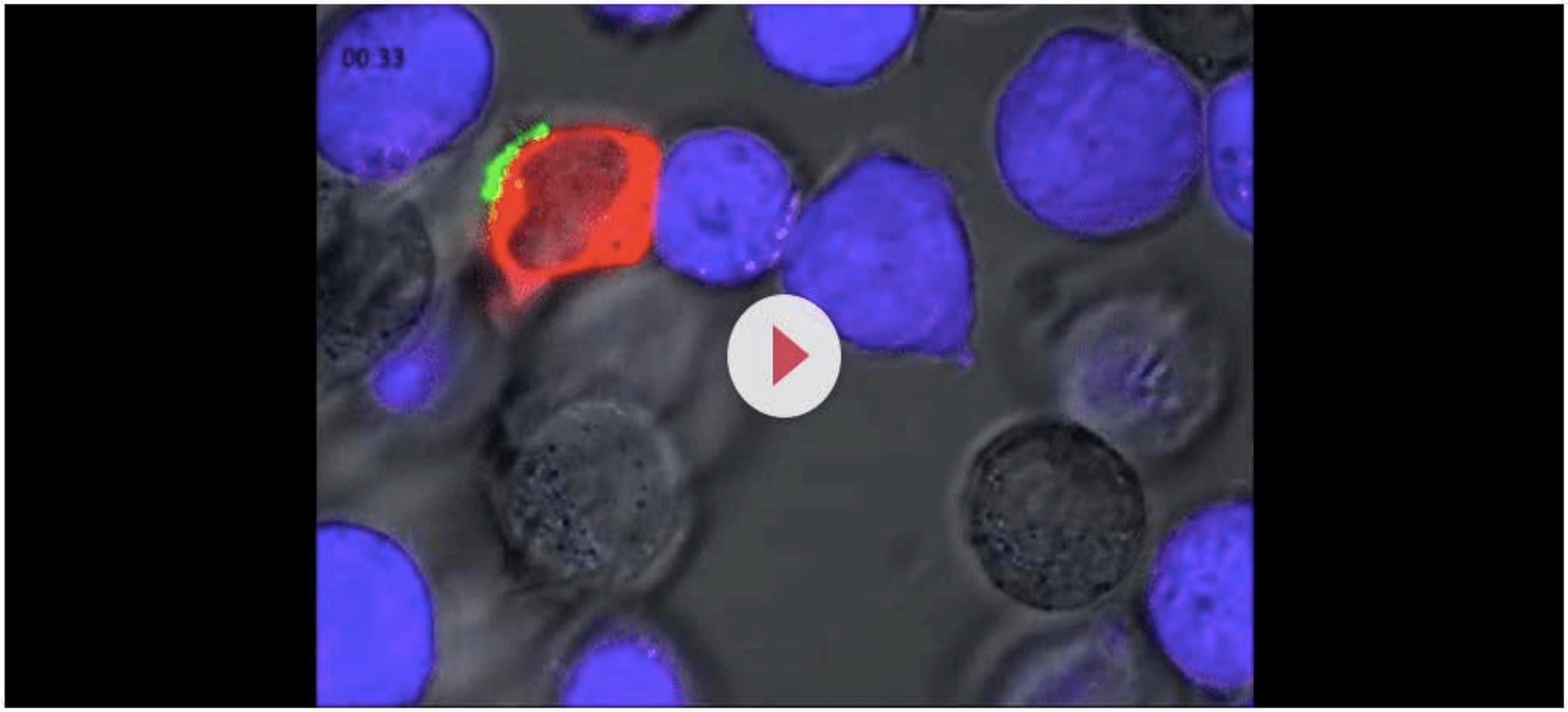
- Environmental Science
- Limnology
- Oceanography

William Thielicke

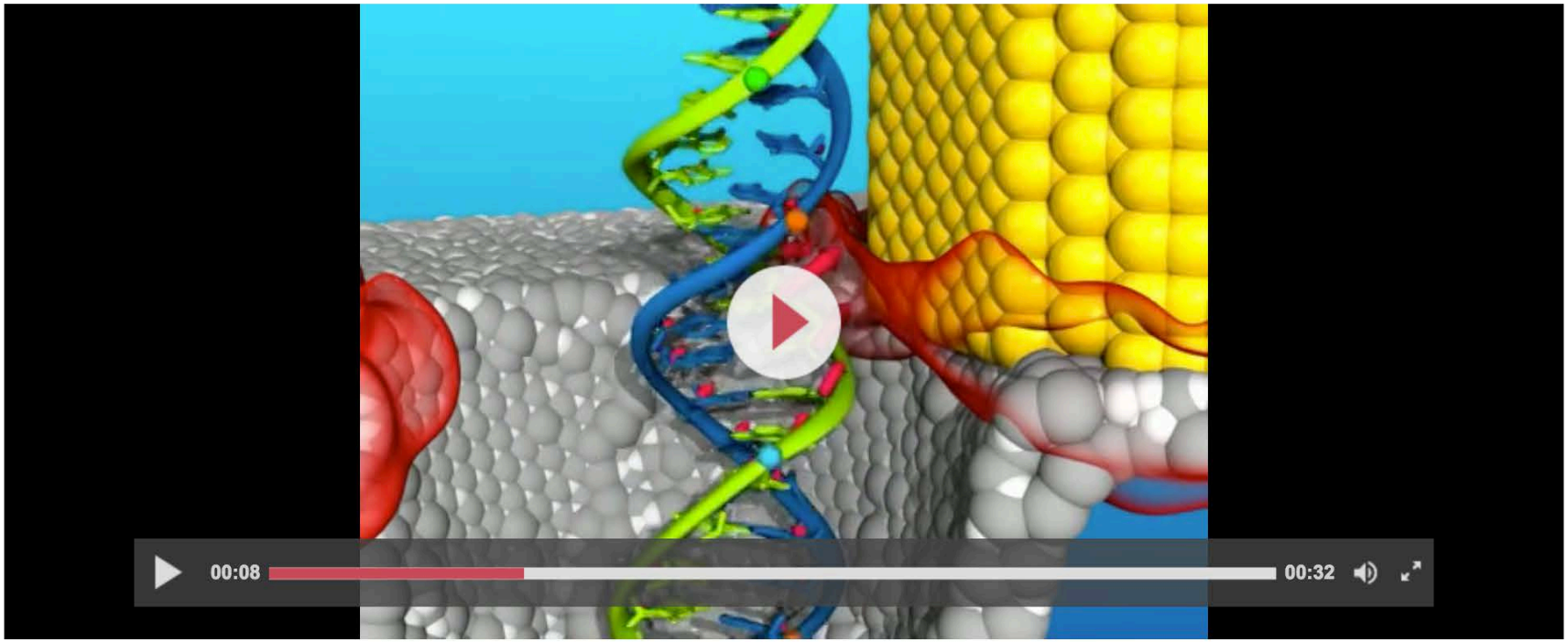
0000-0001-8866-9769

Research assistant / PhD student in Aerospace Engineering
Bremen, Germany

4278 item views	616 item downloads	14 citations



Video files…



Download (7.98 MB) Share Cite Embed

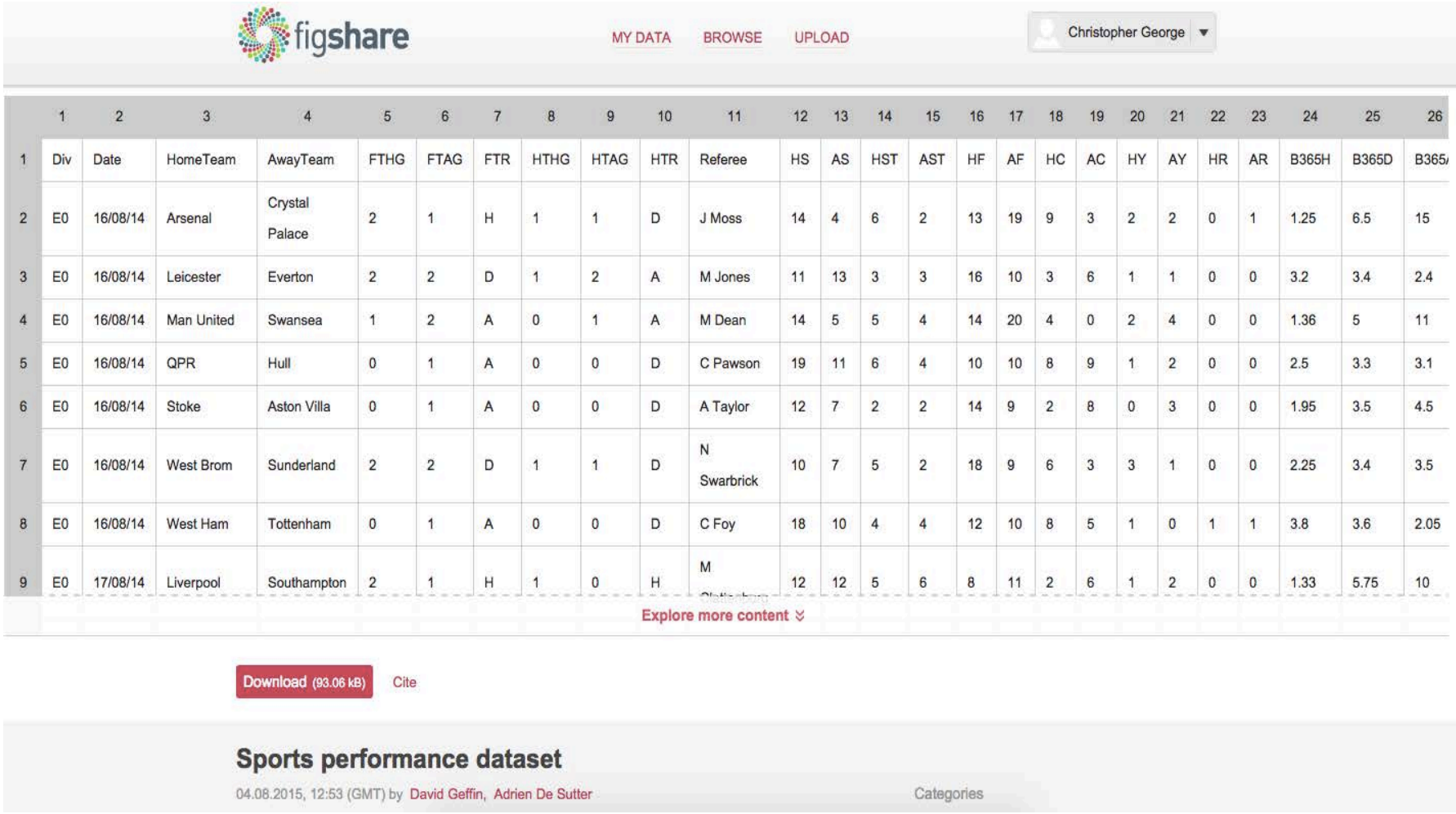
Belkin, Maxim; Chao, Shu-Han; P. Jonsson, Magnus; Dekker, Cees; Aksimentiev, Aleksei (2015): Plasmonic Nanopores for Trapping, Controlling Displacement, and Sequencing of DNA. ACS Publications.

https://dx.doi.org/10.1021/acsnano.5b04173.s002

Retrieved: 14 52, Nov 25, 2015 (GMT)

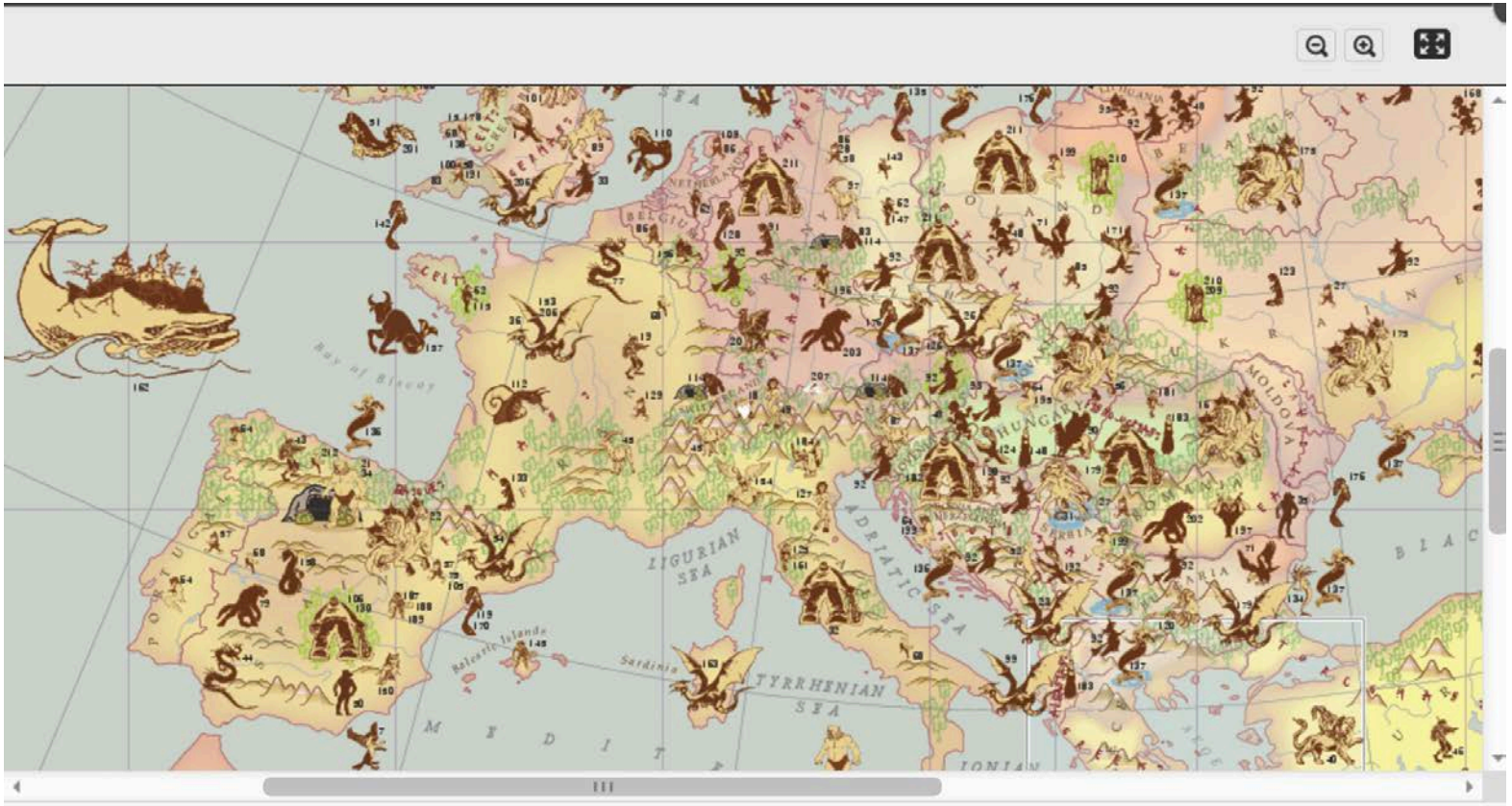
Place your mouse over the citation text to select it

Spreadsheet data..

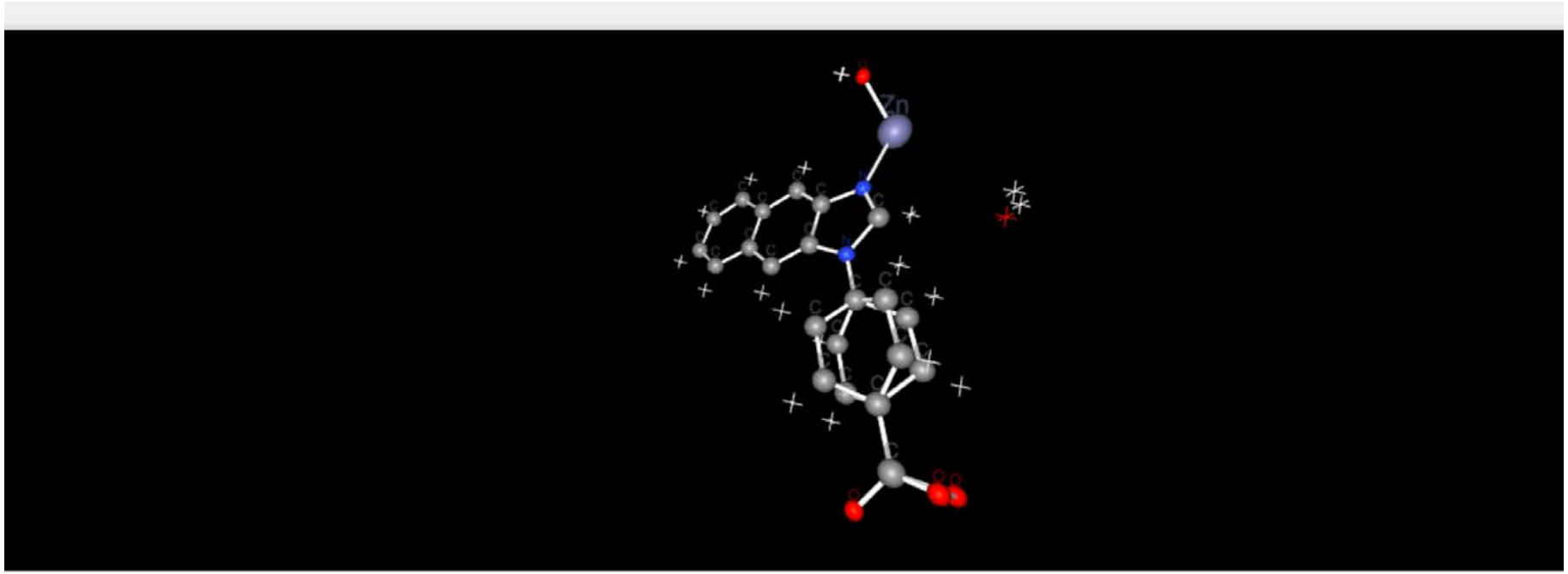


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR	B365H	B365D	B365
2	E0	16/08/14	Arsenal	Crystal Palace	2	1	H	1	1	D	J Moss	14	4	6	2	13	19	9	3	2	2	0	1	1.25	6.5	15
3	E0	16/08/14	Leicester	Everton	2	2	D	1	2	A	M Jones	11	13	3	3	16	10	3	6	1	1	0	0	3.2	3.4	2.4
4	E0	16/08/14	Man United	Swansea	1	2	A	0	1	A	M Dean	14	5	5	4	14	20	4	0	2	4	0	0	1.36	5	11
5	E0	16/08/14	QPR	Hull	0	1	A	0	0	D	C Pawson	19	11	6	4	10	10	8	9	1	2	0	0	2.5	3.3	3.1
6	E0	16/08/14	Stoke	Aston Villa	0	1	A	0	0	D	A Taylor	12	7	2	2	14	9	2	8	0	3	0	0	1.95	3.5	4.5
7	E0	16/08/14	West Brom	Sunderland	2	2	D	1	1	D	N Swarbrick	10	7	5	2	18	9	6	3	3	1	0	0	2.25	3.4	3.5
8	E0	16/08/14	West Ham	Tottenham	0	1	A	0	0	D	C Foy	18	10	4	4	12	10	8	5	1	0	1	1	3.8	3.6	2.05
9	E0	17/08/14	Liverpool	Southampton	2	1	H	1	0	H	M	12	12	5	6	8	11	2	6	1	2	0	0	1.33	5.75	10

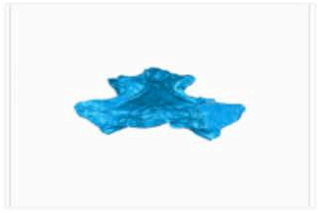
Maps…


LIGURIAN SEA
TYRRHENIAN SEA
ADRIATIC SEA
Balearic Islands
Sardinia
tjom_a_867544_sm4926.pdf – Download

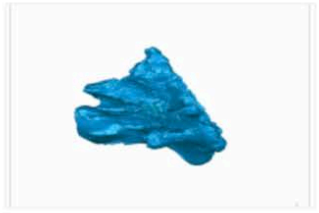
Discipline specific files…



3D printing files…



dorsal 4.pdf (44.81 MB)



cervical.pdf (31.06 MB)

dorsal 8.pdf (52.5 MB)

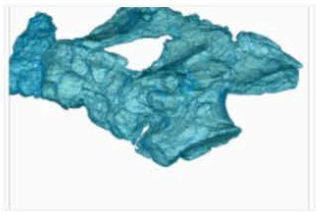
hindlimb.pdf (62.57 MB)

caudal series.pdf (128.18 MB)

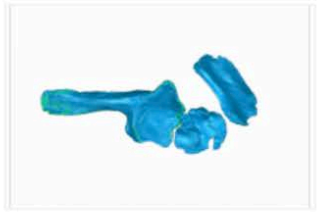
forelimb.pdf (84.08 MB)

view

hip.pdf (69.8 MB)



sacrum.pdf (62.06 MB)



shoulder.pdf (77 MB)

whole_skeleton_… .pdf (62.88 MB)

Download all (674.93 MB) Share Cite Embed

10 files

figshare for..

researchers

institutions

publishers

figshare as a commercial repository

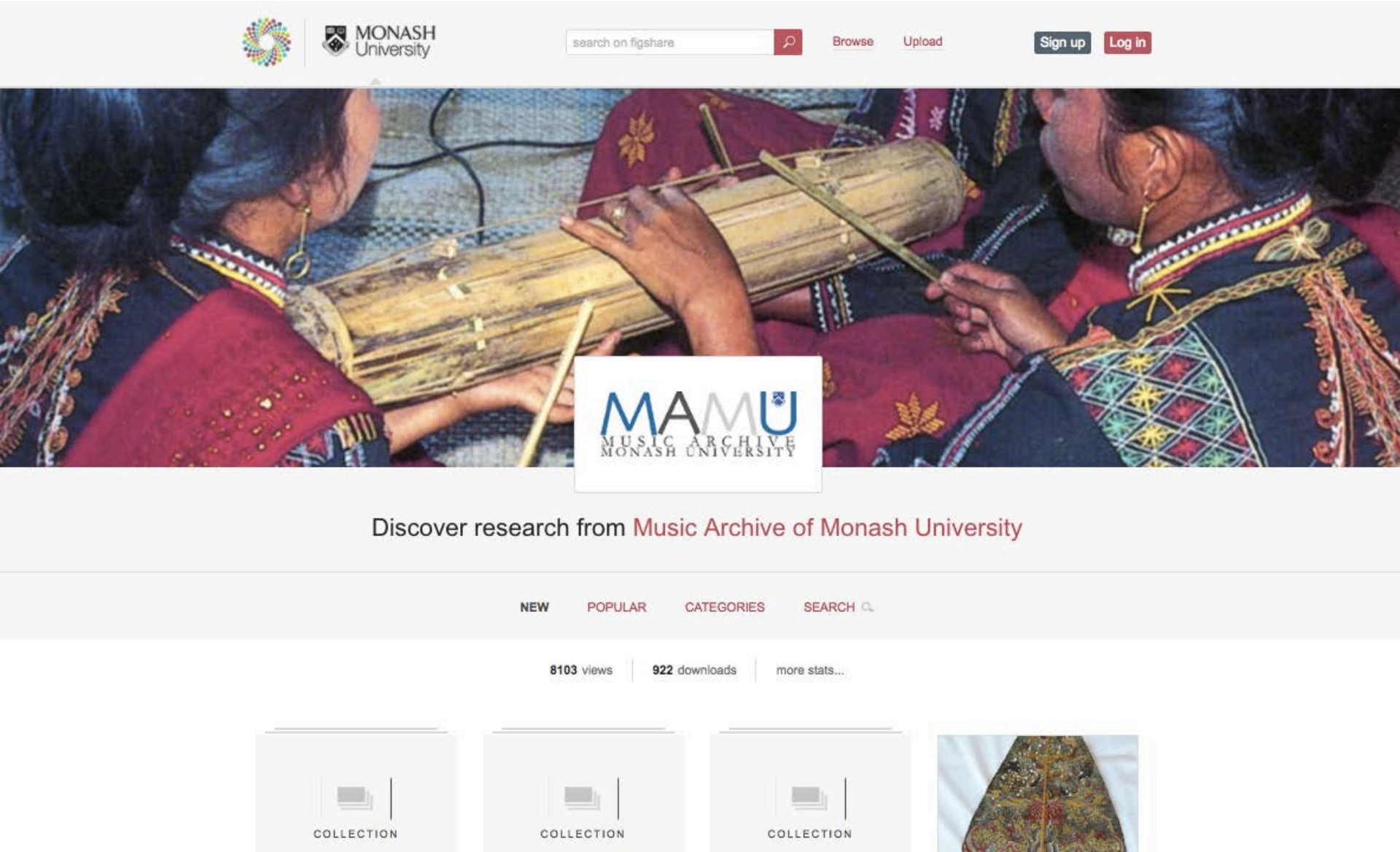
- Never paid access
- Everything is available through open APIs
- Figshare.com is FREE for any researcher publish their data
- A business model allows us to be sustainable
- If figshare goes bust → a ll code is held in escrow and will be made freely available
- Operate as an infrastructure / middleware layer → plug into local / institutional storage, do not own any of the data

→ focus is on being FAIR

figshare as a FAIR repository

- Contains FAIR Data Objects
- Licences clearly displayed
- Full and open description of all technologies, controlled vocabularies and formats used
- Gives identifiers and rich metadata according to the Data Citation Principles
- Has well maintained open API

figshare for institutions




Murdoch Childrens
Research Institute
Healthier Kids. Healthier Future.


Brunel
University
London


HSN
University College
of Southeast Norway


Middlesex
University
London


Stockholm
University


Loughborough
University


ST. EDWARD'S
UNIVERSITY


US
University of Sussex


Fontys
University of Applied Sciences


BATH
SPA
UNIVERSITY


UNIVERSITY OF AMSTERDAM


THE UNIVERSITY OF
MELBOURNE


Carnegie
Mellon
University


MONASH
University


UNIVERSITY OF
SURREY


THE UNIVERSITY
of ADELAIDE


The Open
University


ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON


LA TROBE
UNIVERSITY


St George's
University of London


University of
Salford
MANCHESTER


THE UNIVERSITY OF
AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND


The
University
Of
Sheffield.


Cranfield
UNIVERSITY


Universiteit Utrecht


Zuyd University
of Applied Sciences
ZU
YD

figshare for publishers

F1000Prime F1000Workspace

F1000Research Open for Science

SUBMIT YOUR RES

HOME SUBJECTS CHANNELS HOW TO PUBLISH ABOUT BLO

CrossMark click for updates

METHOD ARTICLE

Creating 3D visualizations of MRI data: A brief guide [version 1; referees: 3 approved]

Christopher R. Madan

Author affiliations

Grant information: The author(s) declared that no grants were involved in supporting this work.

incf This article is included in the Neuroinformatics channel.

METRICS

VIEWS 941

DOWNLOADS 197

Get PDF

Get XML

Cite

Expor

Track

Email

Share

Abstract

While magnetic resonance imaging (MRI) data is itself 3D, it is often difficult to adequately present the results papers and slides in 3D. As a result, findings of MRI studies are often presented in 2D instead. A solution is to create figures that include perspective and can convey 3D information; such figures can sometimes be produced by standard functional magnetic resonance imaging (fMRI) analysis packages and related specialty programs. However, many options cannot provide functionality such as visualizing activation clusters that are both cortical and subcortical (i.e., a 3D glass brain), the production of several statistical maps with an identical perspective in the 3D rendering, or animated renderings. Here I detail an approach for creating 3D visualizations of MRI data that satisfies all of these criteria. Though a 3D 'glass brain' rendering can sometimes be difficult to interpret, they are useful in showing a more overall representation of the results, whereas the traditional slices show a more local view. Combined, presenting both 2D and 3D representations of MR images can provide a more comprehensive view of the study's findings.

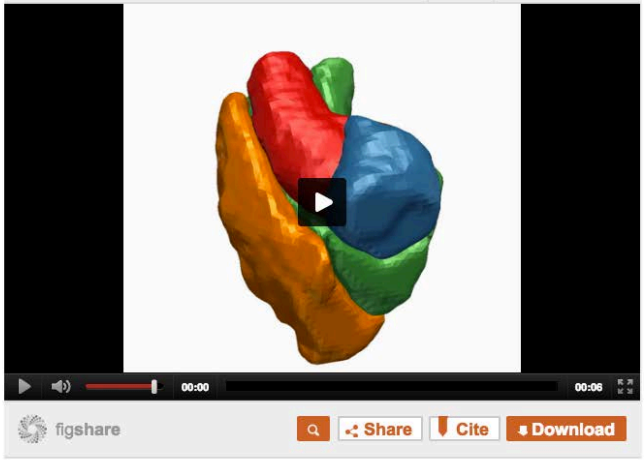
Render in 3D. Start ParaView and load the VTK file, as done previously. As shown in Figure 4B–E, the volumes can be rendered as points, wireframes, and surfaces. Furthermore, many settings can be customized to adjust the rendering properties, such as the lighting/reflectance properties shown in Figure 4D and 4E.

ParaView can also create cameras that move over time, allowing for the generation of animations of the structures rotating. This can be done using the "Animation View" panel in the bottom-center of ParaView: select "Camera", "Orbit", and then "+". The default settings for the camera positions are usually sufficient. If desired, the camera path can also be edited afterwards by inputting specific coordinates (the best way to preview the path is to simply press 'play' at the top and see how it looks). Even without rendering the animation itself, having a camera path allows for later reproduction of 3D renderings from the same camera positions.

Using a camera path, an animation can be rendered by going "File", "Save Animation". An example rendered video is shown in Movie 1. (Note, videos here were re-compressed with Handbrake [https://www.handbrake.fr; freely available for Windows and Mac] to reduce their file size).

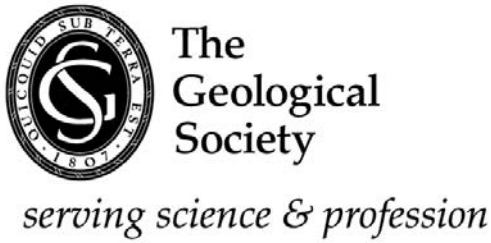
Movie 1. Rotating 3D animation of the anatomical ROI with the same render settings as used in Figure 4E.

58 views 0 shares




F1000
FACULTY of 1000


AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION


The
Geological
Society
serving science & profession


TUTA SUB AEGIDE PALLAS
1683


wellcome


SAGE
Publishing


PNAS


BRILL


WILEY


HARTWELL
JOURNAL OF MEDICINE


SciELO


Taylor & Francis Group
an informa business


THE ROYAL
SOCIETY


PLOS


KARGER

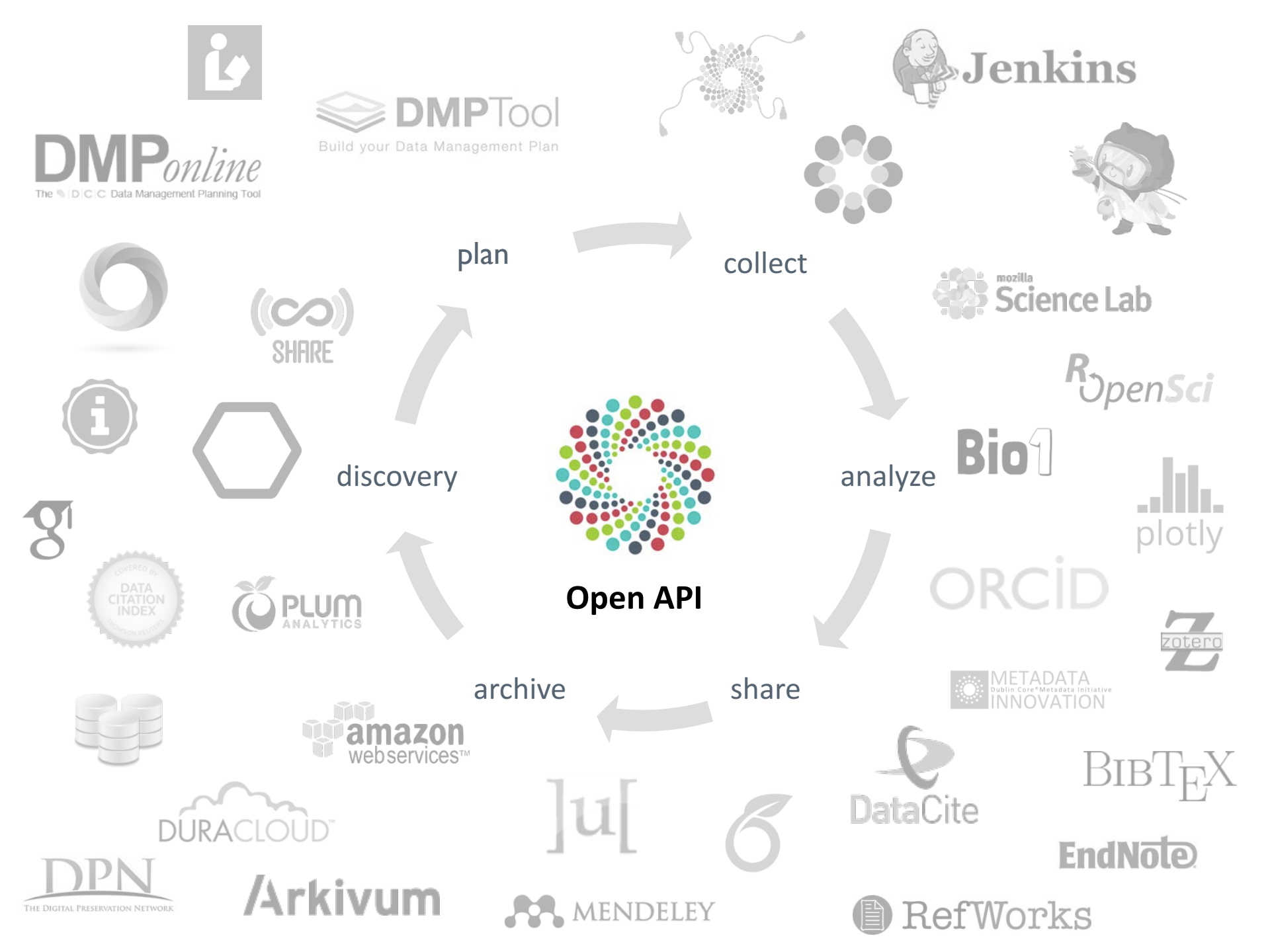

OSA
The Optical Society


ORCID


ACS
Chemistry for Life®


SPRINGER NATURE


ENDOCRINE
SOCIETY


DMPonline
The DCC Data Management Planning Tool
DMPTool
Build your Data Management Plan
Jenkins
SHARE
mozilla
Science Lab
ROpenSci
Bio1
plotly
ORCID
zotero
METADATA
Dublin Core Metadata Initiative
INNOVATION
BIBTEX
DataCite
EndNote
RefWorks
MENDELEY
Arkivum
DPN
The Digital Preservation Network
DURACLOUD
amazon
web services
PLUM
ANALYTICS
DATA
CITATION
INDEX
plan
collect
analyze
share
archive
discovery
Open API

The State of Open Data

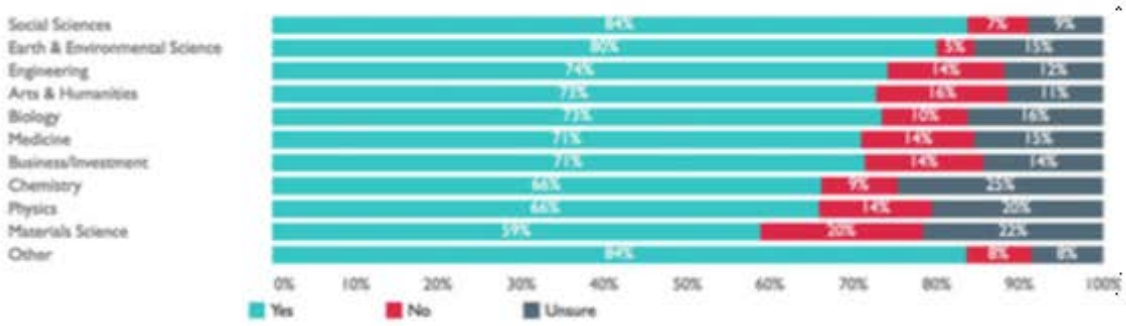
Treadway, Jon; Hahnel, Mark; Leonelli, Sabina; Penny, Dan; Groenewegen, David; Miyairi, Nobuko; Hayashi, Kazuhiro; O'Donnell, Daniel; Science, Digital; Hook, Daniel (2016): The State of Open Data Report. figshare. https://dx.doi.org/10.6084/m9.figshare.4036398.v1

The Open Data Survey and Results

Researchers are aware of open data

Figure 1 - Awareness of data that is free to access, reuse, repurpose and redistribute, n=1915

Figure 2 - Awareness of data that is free to access, reuse, repurpose and redistribute, by subject area, n=1436



The Open Data Survey and Results

Researchers are already reusing open data

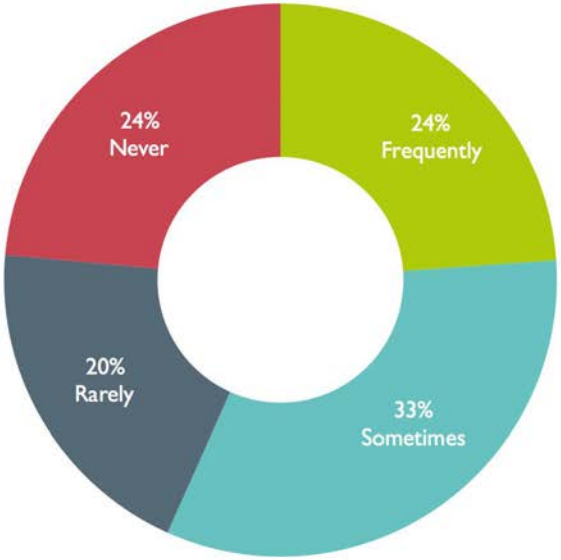


Figure 4 - Regularity with which respondents have made data free to access, reuse, repurpose and redistribute, n=1869

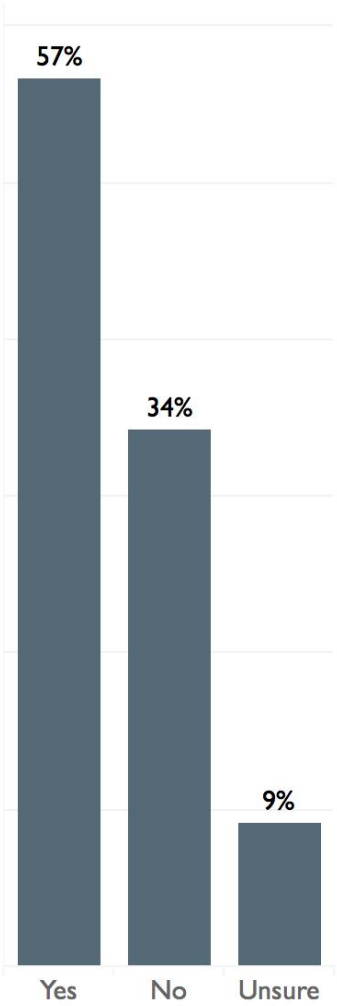


Figure 5 - Respondents who have reused data made free by others, n=1777

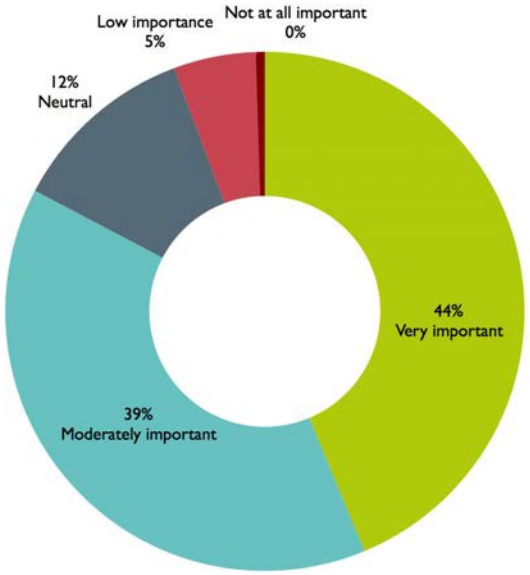
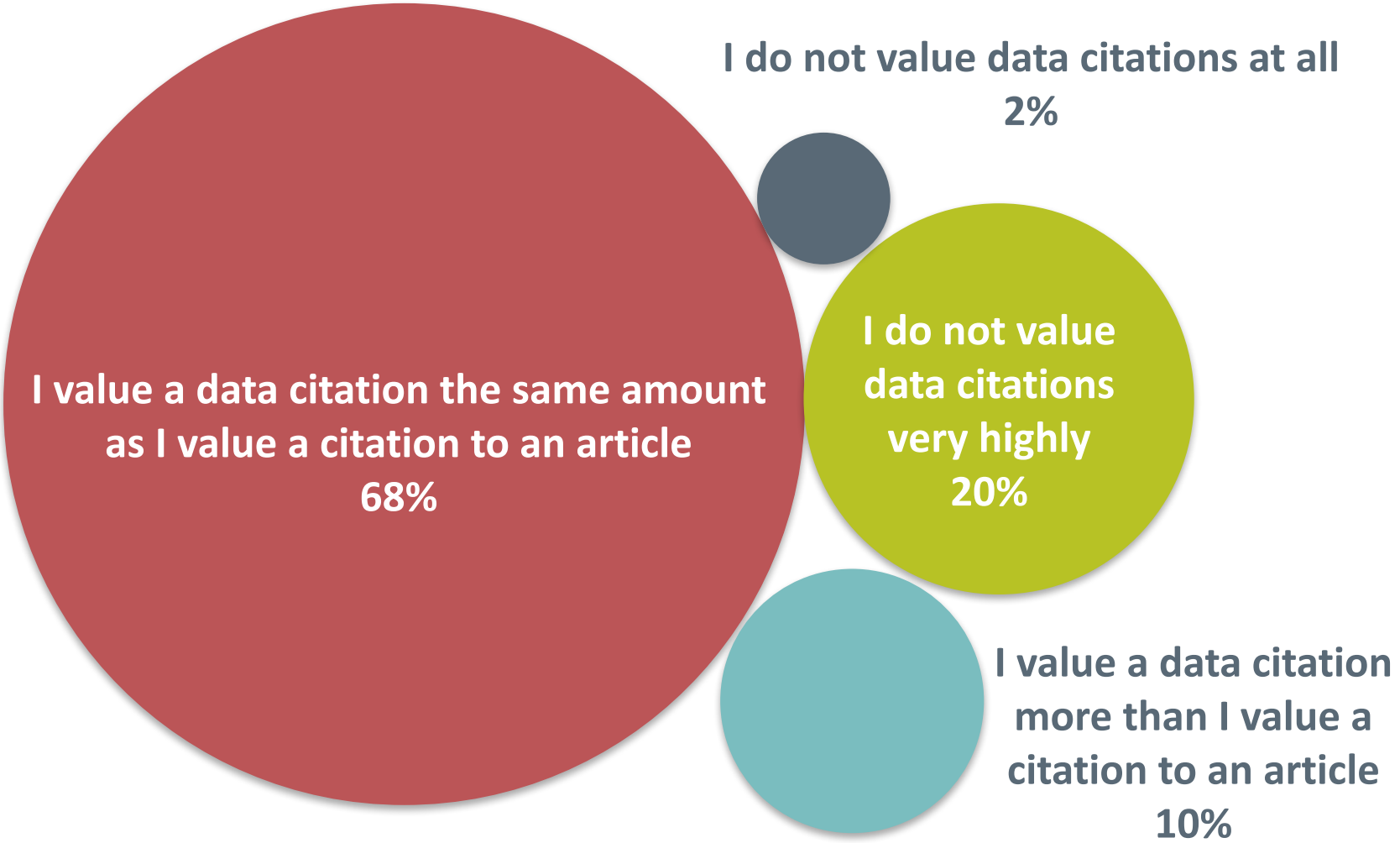


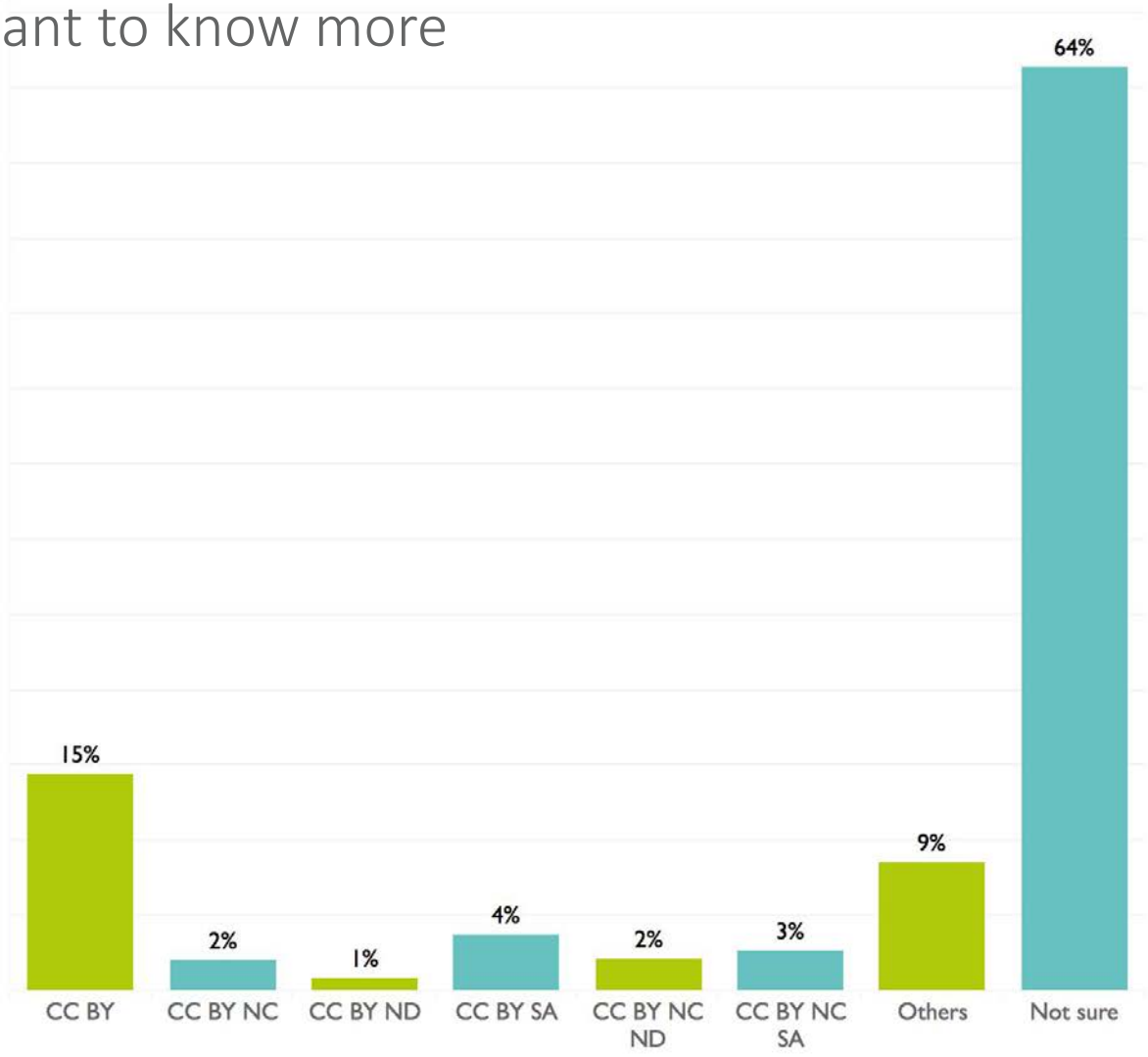
Figure 6 - Importance of freely available data to those who have reused it, n=1006

The State of Open Data survey results



The Open Data Survey and Results

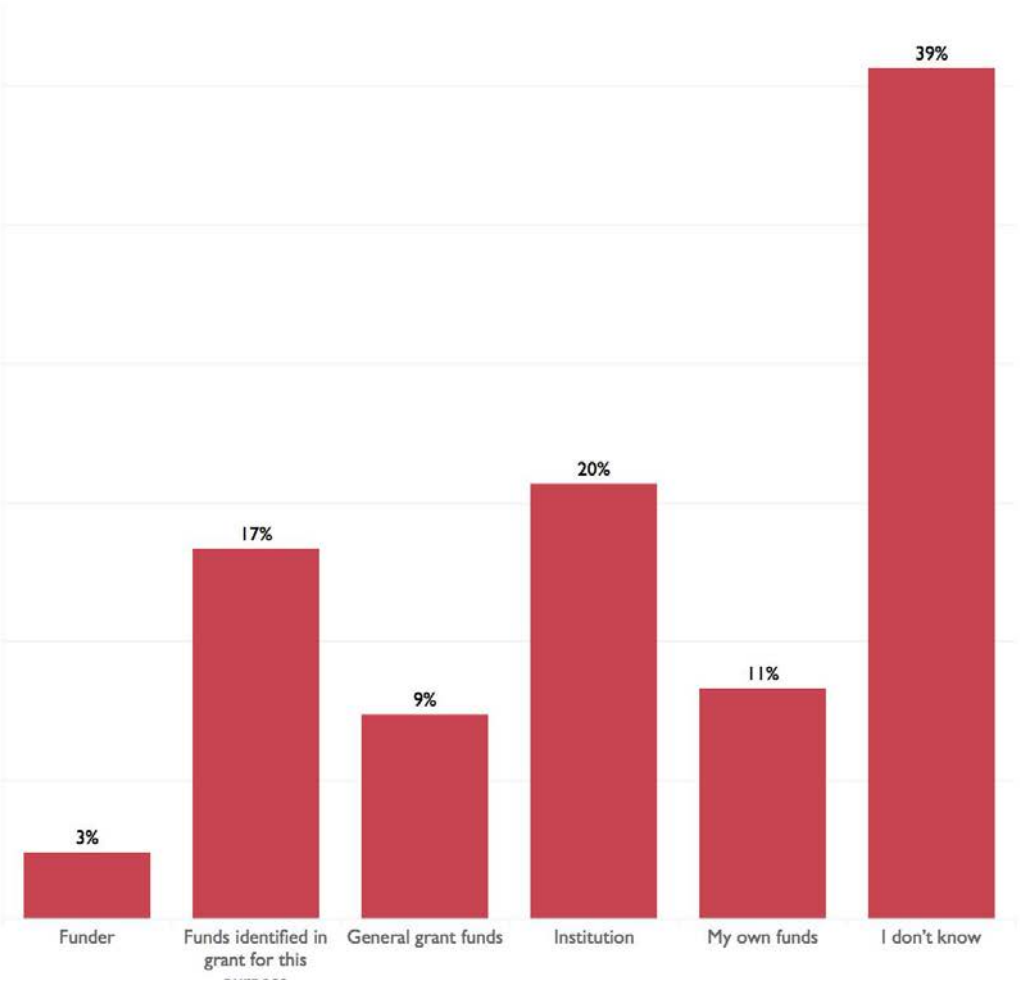
Respondents admit to uncertainty and gaps in their knowledge; they want to know more



The Open Data Survey and Results

Researchers are uncertain on who will meet the costs

Figure 9 - Respondent's knowledge of source of funds for making data openly available, n=1554



The Open Data Survey and Results

Respondents admit to uncertainty and gaps in their knowledge; they want to know more

Funder requirements		Institution requirements		Publisher requirements	
I don't know	20%	I don't know	25%	I don't know	31%
No policy	53%	No policy	40%	No policy	33%
Policy exists	27%	Policy exists	36%	Policy exists	35%

Figure 10 - Respondent's awareness of open data policies, by Funder n=1451, by Institution n=1338, by Publisher, n=1401

The Open Data Survey and Results

Respondents admit to uncertainty and gaps in their knowledge; they want to know more

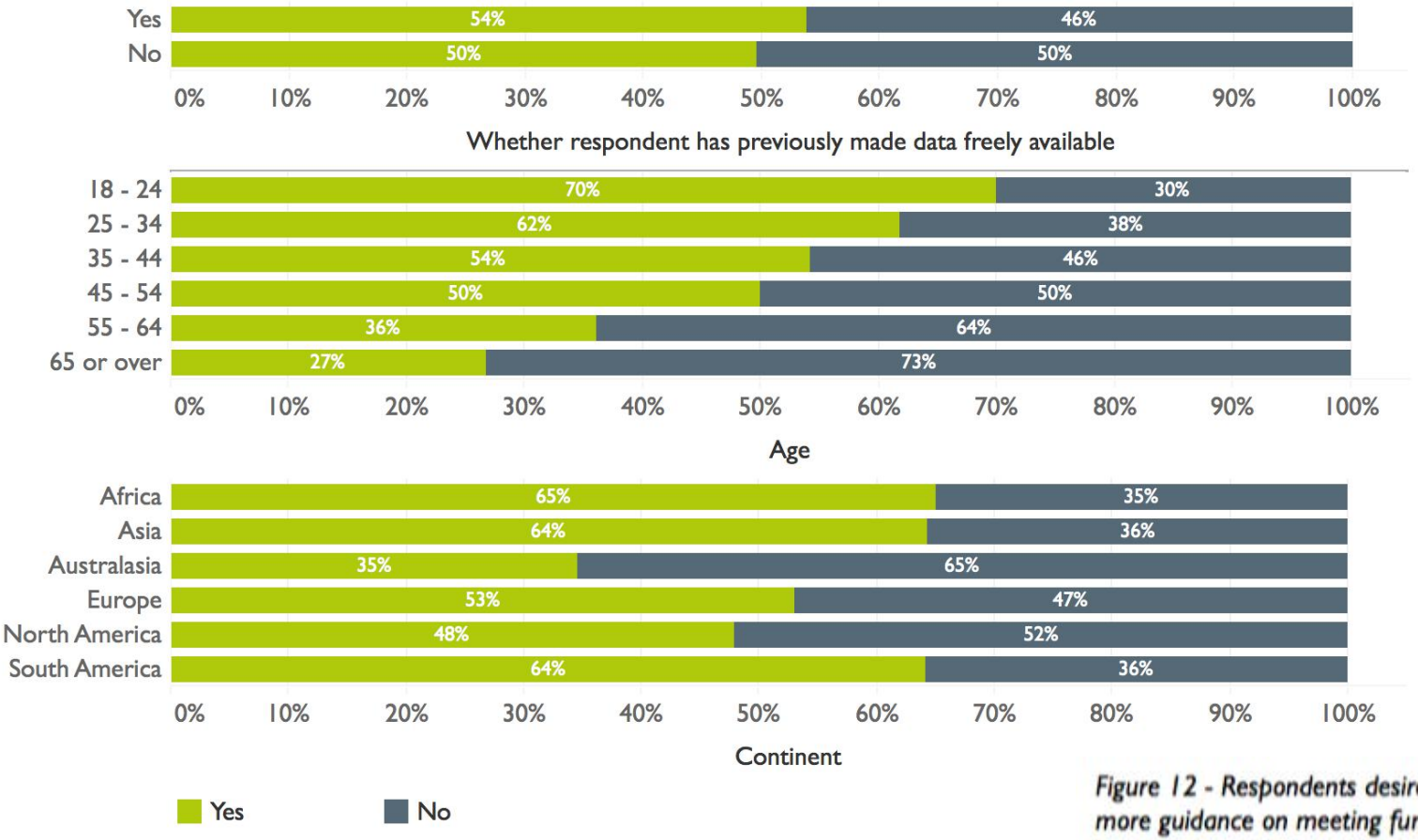


Figure 12 - Respondents desire for more guidance on meeting funder requirements, by data sharers n=1572, by age n=1428, by continent n=1430

The Open Data Survey and Results

What does the future hold?

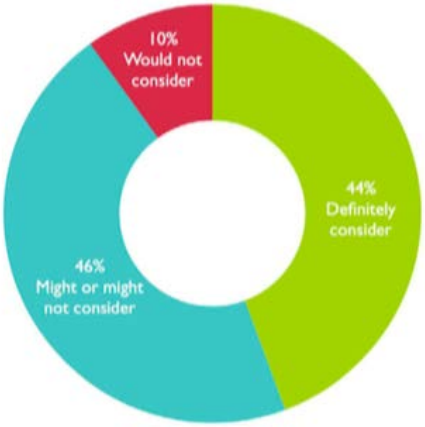
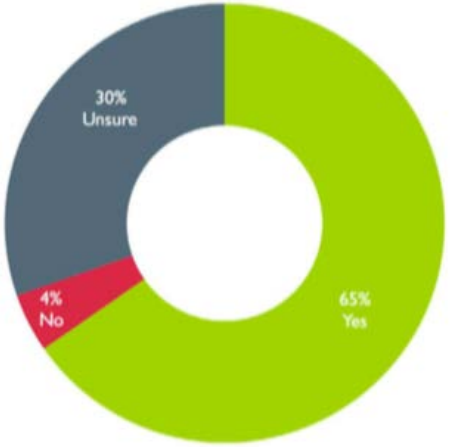


Figure 14 - For respondents who have never made data freely open, willingness to do so in future, n=233

Figure 16 - For respondents who have never made data freely open to date nor reused data others have made freely available, willingness to reuse data in future, n=228



Where to find the report

- The State of Open Data Report
- The State of Open Data Poster
- Raw data for the report

- We're looking for participants for 2017's report. If you'd like to help, please contact info@figshare.com

Summary

- Open data = FAIR data that can be cited according to the data citation principles
- Soon, funder mandates will mean data has to be open by default “as open as possible, as closed as necessary”
- Use FAIR repositories with sustainable business models
- Researchers are aware of open data → already making data available and reusing open datasets
- Researchers would welcome more guidance on open data policies and best practice

Thanks for your time

mimi@figshare.com **or** info@figshare.com

@figshare, @MimiKeshani

https://figshare.com

figshare