Open Data Essentials

26 May 2021, 9am BST / 1:30pm IST / 3pm WIB

Matt Cannon, Head of Open Research, Taylor & Francis
Guillaume Wright, Publisher, F1000Research
Agenda

- Introductions and welcome
- Definitions and terminology
- Data myths and misconceptions
- Open data at F1000Research
- Open data support and resources
Welcome and introductions

Matt Cannon – Head of Open Research, Taylor & Francis Group

Guillaume Wright – Publisher, F1000Research
Definitions and Terminology

Data – “The recorded information (regardless of the form or the media in which it may exist) necessary to support or validate a research project’s observations, findings or outputs, or which is required for legal, (funder), or regulatory compliance.”

Data Sharing – a broad term to cover many workflows and policies encouraging researchers to share their data

Open Data – describes a specific way of making data accessible to readers/researchers. Data will be stored in a repository under an open license (usually CC-BY or CC0). Equivalent to an open access article

FAIR Data – an aspirational term outlining best practice for preserving research data. It should be Findable, Accessible, Interoperable and Reusable.

Data Availability Statement (DAS) – a short section of your paper describing if/how data can be accessed by readers

Persistent Identifier (PID) - a unique reference number to ensure data can be located. Often a Digital Object Identifier (DOI)

Repository – An online database for storing data and other research objects
What is data?

Data

- Source data
  - Data produced by others

- Underlying data
  - Data produced for the purpose of the reported study

- Extended data
  - Materials produced for the purposes of the reported study

- Code
  - Code produced either as the primary output or for purposes of replication
Open data in practice

Supporting your online publication

1. Prepare your data for sharing
2. Select a repository
3. Add a Data Availability Statement
4. Link your data to your article
Selecting a data repository

There are many options available to preserve your data:

• Policy considerations – are there institutional, funder, subject guidelines on where to store?

• Technological considerations – what kind of persistent identifier (PID) do you need? do you need to embargo your data? A link to enable anonymous peer review?

• Are you interested in data curation? Some domain repositories will help with curating data

• Guidance available from libraries, publisher websites and third parties

• Encourage you to check www.fairsharing.org and www.re3data.org for repositories and data standards

• Look out for our "how to select a repository" guide
Data Availability Statements

A short section to describe how different data sources used in the study can be accessed, consider headings for different kinds of data (source, underlying etc)

• If the data is open:
  • must include a link to the data and its PID
  • ensure the data is cited in your references
• If you are not able to share the data – the statement should explain why
  • (legal, ethical, commercial reasons are common explanations)
  • "Open as possible, as closed as necessary"

Examples of suitable templates online

Open Data Toolkit
Links and resources for researchers on all aspects of open data
Data misconceptions #1

"Sharing isn't worth the extra effort"

"I don’t have any data"

"I don't have time to share my data"
Data misconceptions #2

"I don’t have permission to share"

"Data sharing is only for science research"
Open Data at F1000Research
Publish fast.
Openly.
Without restrictions.
All articles on F1000Research that report original results should include the source data underlying the results, together with details of any software used to process the results.
F1000Research open data policies

As open as possible, as closed as necessary

- Deposit data in repository
- Include data availability statement in article
- Software availability
- Reasonable exceptions
F1000Research open data policies

- **Exceptions:**
  - Ethical or security
  - Confidentiality or data protection
  - Large data or 3rd party data
  - Software availability
  - No “available upon request” option

- **In-article code and data:** visualisation and reanalyses widgets

- **Data and software:** can also be stand-alone peer reviewed articles
Open data compliance and guidance

1. Prepare your data for sharing

2. Select a repository

3. Add a Data Availability Statement

4. Link your data to your article
Data Availability Statement

Data availability

Source data

The Demographic and Health Surveys data used in the current study are available from the DHS website.


Underlying data

Data access is restricted to users with appropriate ethics approval from the committees listed in the Ethical Considerations section. A lead or reviewer may apply to the authors for access by providing a written description of background, reasons, and intended use. If the methodology does not violate the condition of informed consent under which the interview was conducted, and the proposal approved by UGHE and other relevant ethics boards, the user can obtain the data from the corresponding author and include one of the authors in the project and analysis.

Extended data

DRYAD Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda. https://doi.org/10.5061/dryad.kh1893242

This project contains the following extended data:

- Tables and figures, accessible here: https://ughe.org/wp-content/uploads/2021/04/2.-Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf

Reporting guidelines


This work is licensed under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication licence.

Data availability

Underlying data

- Data: raw data (vide format) and dataset (csv) supporting patient attachment to collagen IV or fibronectin in percentage over time (related to Figure 1). https://doi.org/10.5061/dryad.2774127
- Data: raw data (sputum inhibition details in IRM images (scaled) and raw data set (csv) related to Figure 4). https://doi.org/10.5061/dryad.2774148
- Data: raw data (IRM images) of nodosecell experiments (address and new dataset for statistical purposes (csv) (related to Figure 3). https://doi.org/10.5061/dryad.2774150
- Data: nodosecell experiment sputum images, IRM raw data. Planetes field, imaged by IRM in low magnification for counting purposes. Planetes are either control or treated with nodosecell. https://doi.org/10.5061/dryad.2774152
- Raw data: support percentage of patient in each morphological state. I have post-patellar scaling (related to Figure 3). https://doi.org/10.5061/dryad.2774153
- Data: Dynamics of plaque spreading over time with/without treatments with mungosfure and thymol (related to Figure 13). New images of plaque treated with or without mungosure and thymol (FITC, samples) and raw data set (csv). https://doi.org/10.5061/dryad.2774154
- Data: Uncompiled and renamed images/movie for all (DIF, movies, raw, sv). https://doi.org/10.5061/dryad.2774156

Extended data

Figure 5A: Differential dynamic of early stages of plaque adhesion and spreading on collagen IV and fibronectin coated surfaces. https://doi.org/10.5061/dryad.4YFQK/X.4ZD7Y8F

This project contains the following extended data:

- Figure 5A: Planar integrated activity: ingrowth activity of patient: the mean absolute value (MAV) at every time point. X-axis: Time in seconds. Y-axis: Planar mean activity. Red dotted lines separate the phases: background, prior to plaque attachment. Fibroplastic spreading phase, lamellapodial spreading phase, and the fully formed plaque.
- Figure 5B: Immunity of the surfaces for collagen IV and fibronectin. The number of patients interacting with the surface over time for the surfaces collagen IV and fibronectin. Time in seconds.
- Figure 5C: Quantification and image analysis of plaque spreading. Based on IRM live imaging for fibronectin: (A) Planar spreading traced by IRM, and the corresponding focal activity map. UYMA, ++U, +U, -U, ++U, and +++U. Positive values (yellow) imply local detachment: negative values (blue) imply local detachment (bottom right). One fibroblast initially attaching and detaching (black arrow). Scale bar 2 µm. (B) 2D integrated trapping activity of patients: mean absolute value (MAV) at every time point. X-axis: Time in seconds. Y-axis: Planars mean activity. Red dotted lines separate the phases: background, prior to plaque attachment. Fibroplastic spreading phase, lamellapodial spreading phase, and the fully spread phase. (C) Total number of patients interacting with the surface over time. Time in seconds. (D) Accumulated attachment and detachment over time shown by activity map, yellow means more attachment events, blue means fewer attachment events. Right images: control IRM images. Scale bar 2 µm.
- Movie S1: Shows the accumulated number of interactions from interaction to not interacting with the surface at every pixel over time.
- Movie S2: Shows an overview of the highly active region on top of the IRM images over time on collagen IV.
- Movie S3: Shows an overview of the highly active region on top of the IRM images over time on collagen IV.

Software availability

IRM spreading dynamics source code available from: https://github.com/massDNF/IRIRSpreading-Dynamics

Archived source code at time of publication https://doi.org/10.5061/dryad.2774159

License (Sh) General Public License v3.0
Data Availability Statement

Data availability

Source data

The Demographic and Health Surveys data used in the current study are available from the DHS website.

Underlying data

Data access is restricted to users with appropriate ethics approval from the committees listed in the Ethical Considerations section. A leader or reviewer may apply to the authors for access by providing a written description of background, reasons, and intended use. If the methodology does not violate the condition of informed consent under which the interview was conducted, and the proposal approved by UCSC and other relevant ethics boards, the user can obtain the data from the corresponding author and include one of the authors in the project and analysis.

Extended data

DIYAD: Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda. [DOI](https://doi.org/10.1093/heapro/23-2)

This project contains the following extended data:
- Tables and figures, accessible here: [https://uqhe.org/wp-content/uploads/2021/06/Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf](https://uqhe.org/wp-content/uploads/2021/06/Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf)

Reporting guidelines

DIYAD: STAR checklist for Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda. [DOI](https://doi.org/10.1093/heapro/23-2)

This work is licensed under a [CC0 1.0 Universal](http://creativecommons.org/publicdomain/zero/1.0/) (CC0 1.0) Public Domain Dedication license.
FAIR data in practice

Data availability

Source data

The Demographic and Health Surveys data used in the current study are available from the DHS website.


Underlying data

Data access is restricted to users with appropriate ethics approval from the committees of the Ethical Considerations section. A leader or reviewer may apply to the authors for access by providing a written description of background, reasons, and intended use. If the methodology does not violate the condition of informed consent under which the interview was conducted, and the proposal approved by UGHE and other relevant ethics boards, the user can obtain the data from the corresponding author and include one of the authors in the project and analysis.

Extended data

DRYAD: Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda.

https://doi.org/10.5061/dryad.13bph36

This project contains the following extended data:

- Tables and figures, accessible here: https://ughe.org/wp-content/uploads/2021/02-Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf

Reporting guidelines

DRYAD: Staff checklist for Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda.

https://doi.org/10.5061/dryad.3334776

This work is licensed under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license.

Data availability

Underlying data

- Zeneda: IMI raw data (video format) and dataset (csv) supporting plantar attachment to collagen IV or fibronectin in percentage over time (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Zeneda: Raw data for microtubule extension (IMI images) and raw data set (csv) (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Zeneda: Novicel (IMI images) and raw data set (csv) (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Zeneda: Raw data for microtubule extension (IMI images) and raw data set (csv) (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Zeneda: Raw data for microtubule extension (IMI images) and raw data set (csv) (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Zeneda: RAW data (IMI images) of Novicel experiments (open-source) and raw data set (csv) (related to Figure 1).
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)

Extended data

Figure 1 shows differential dynamics of early stages of plantar adhesion and spreading on collagen IV and fibronectin coated surfaces. https://doi.org/10.5061/dryad.Imo.baw.f4k

This project contains the following extended data:

- Figure S1. Planar integrated activity. Integrated activity of plantar, the mean absolute value (L2 norm) at every time point. A fast integral activity is used to separate the phases; background, prior to plantar attachment, fibroblast spreading phase, lamellipodium spreading phase, and the fully covered cell.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Figure S2. Immunostaining with the surface for collagen IV and fibronectin. The number of pixels interacting with the surface over time for the surfaces collagen IV and fibronectin. Time in seconds.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Figure S3. Quantification and image analysis of planar spreading, based on IMI live imaging for fibroblasts. (4) Planar spreading mixed with IMI and the corresponding focal adhesion map. GPA = IMI × Focal. Positive values (yellow) of the focal staining: negative values (blue) of the focal staining (dotted right). One fibroblast initially attaching and detaching focal adhesion. Scale bar 2 mm. (4) Integrated activity of plane means the measured value (L2 norm) at every time point. A fast integral activity is used to separate the phases; background, prior to plantar attachment, fibroblast spreading phase, lamellipodium spreading phase and the fully covered cell. Total number of pixels interacting with the surface over time. Time in seconds. (2) Accumulated attachment and detachment time over a specific activity map, yellow means more attachment events, blue means fewer attachment event. Right images correspond IMI images. Scale bar 2 mm.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Figure S4. Shows the accumulated number of transitions from interaction to no interaction with the surface at every time point.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Figure S5. Shows an overview of the highly active region on top of the IMI images over time on collagen IV.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)
- Figure S6. Shows an overview of the highly active region on top of the IMI images over time on collagen IV.
  https://doi.org/10.21236/s2021.51.8.29 (Zeneda 2021-01-26)

Software availability

RM spreading dynamics source code available from: https://github.com/lasseDahlbo/IAK/spreading-dynamics

Analyzed source codes at time of publication: https://doi.org/10.3210/s2021.51.8.29 (Zeneda 2021-01-26)

License: (GNU General Public License v3.0)
Open data support and resources

Data Guidelines

1. Background
   1.1 Open Data Policy
   1.2 Fair Data Principles

2. Share Your Data in 4 Steps
   2.1 Prepare Your Data for Sharing
   2.2 Select a Repository
   2.3 Add a Data Availability Statement to Your Article
   2.4 Link Your Datasets to Your Article

1. Background

This page provides information about data you need to include when publishing an article in F1000Research, where your data can be stored, and how your data should be presented. In accordance with our data policies, authors will be required to provide details of where their

f1000research.com/for-authors/data-guidelines
research@f1000.com
Here at F1000Research, we’re big advocates for open data. We believe that sharing research data can accelerate the pace of discovery, provide credit and recognition for authors, and even improve public trust in research (but more on that later).

We know the 21st century researcher has lots to think about – not least securing grant funding, conducting high quality research, and maximizing impact after publication. We’re asking you to add one more thing to this list: open data. Far from being another hoop to jump through, sharing your research data can bring a whole host of benefits to every stage of your research journey.

think.f1000research.com/open-data/
Any questions?

research@f1000.com

f1000research.com | think.f1000research.com/open-data