F1000Research

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

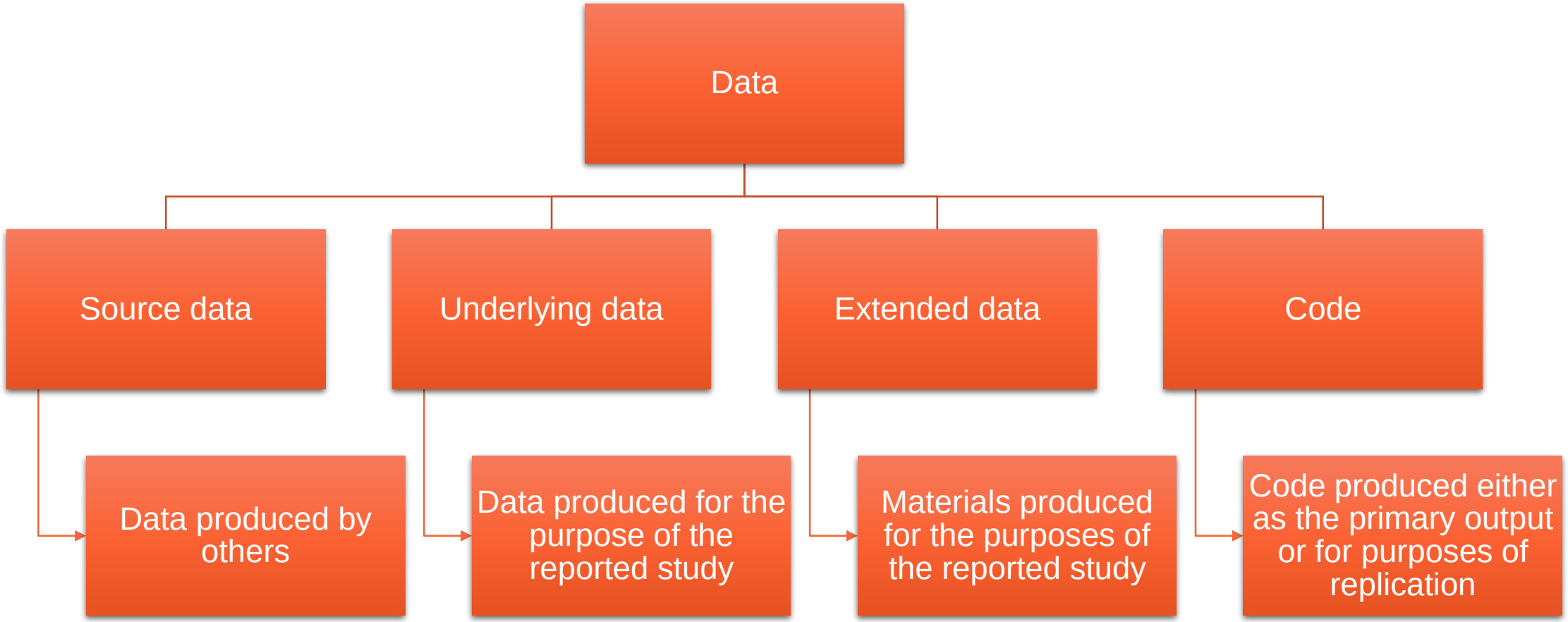
FAIR Data – an aspirational term outlining best practice for preserving research data. It should be Findable, Accessible, Interoperable and Reusable.

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)**

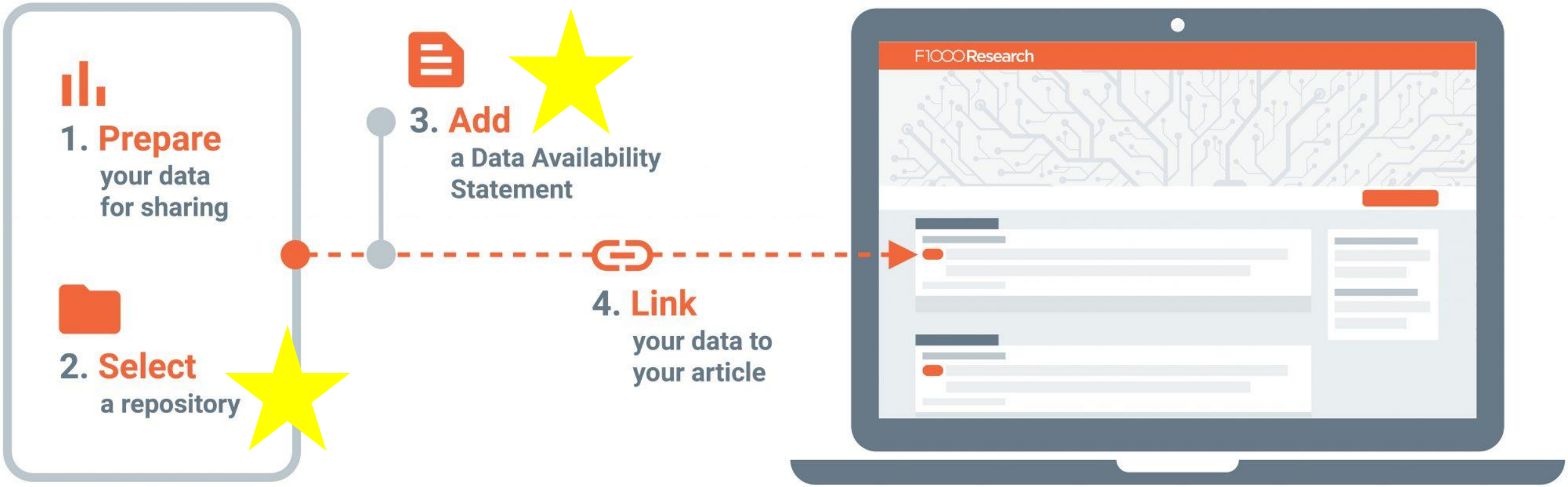
Repository – An online database for storing data and other research objects

What is data?


Data
Source data
Underlying data
Extended data
Code
Data produced by others
Data produced for the purpose of the reported study
Materials produced for the purposes of the reported study
Code produced either as the primary output or for purposes of replication

Open data in practice

Supporting your online publication



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

Data misconceptions #1

"I don’t have any data"

"Sharing isn't worth the extra effort"

"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"

F1000Research

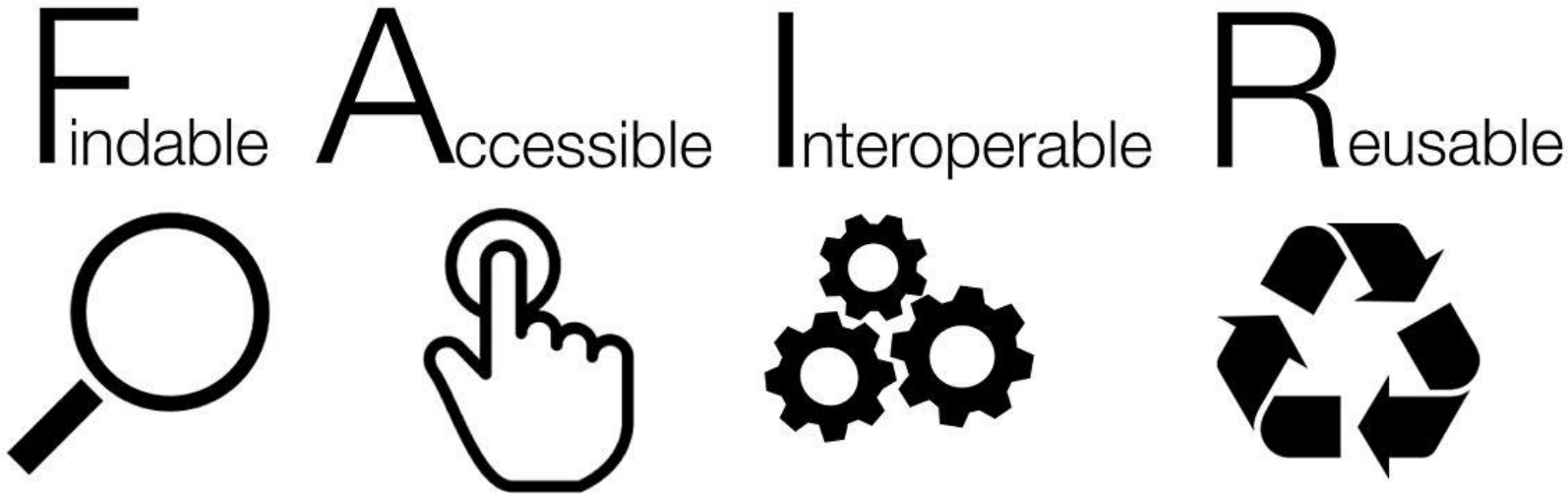
Open Data at F1000Research

F1000Research

Publish fast.
Openly.
Without restrictions.

F1000Research open data policies

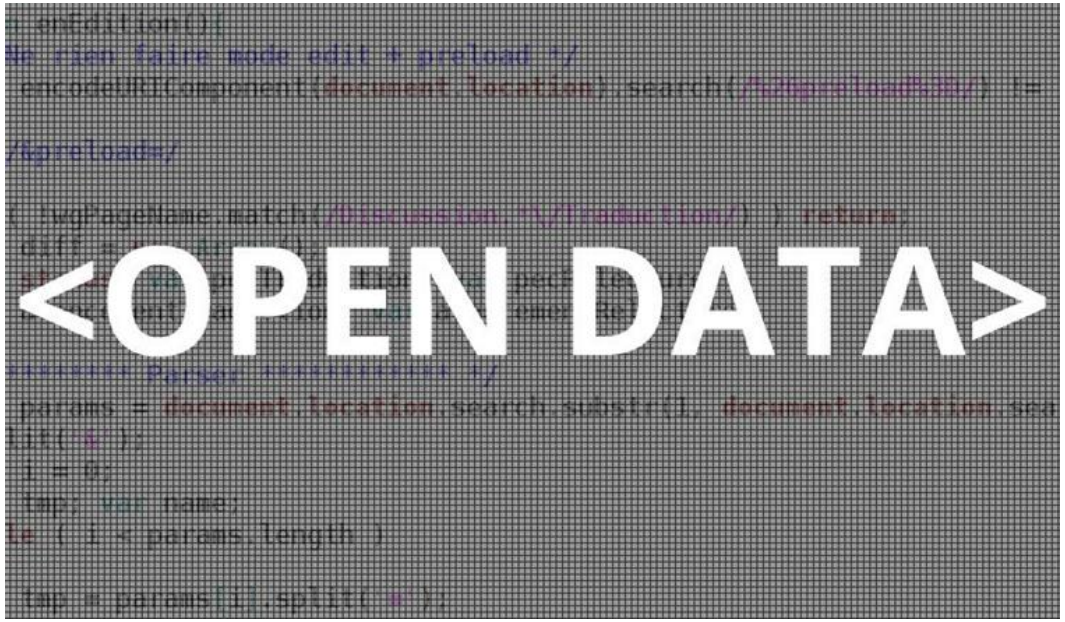
All articles on F1000Research that report original results should include the source data underlying the results, together with details of any software used to process the results.



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

Data Availability Statement

Data availability

Source data

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

Rwanda DHS 2015: https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf

Rwanda DHS 2000: https://dhsprogram.com/pubs/pdf/FR125/FR125.pdf

The Institute for Health Metrics and Evaluation (IHME) Low- and Middle-Income Country Neonatal, Infant, and Under-5 Mortality Geospatial Estimates 2000–2017 Local Burden of Disease data used in the current study are available for download from IHME: http://ghdx.healthdata.org/record/ihme-data/lmic-under5-mortality-rate-geospatial-estimates-2000-2017.

Underlying data

Data access is restricted to users with appropriate ethics approval from the committees listed in the Ethical Considerations section. A reader 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.kh189324x[22]

This project contains the following extended data:

- Interview guide, accessible here: https://ughe.org/wp-content/uploads/2021/02/1.-Interview-Guide_Exemplars-U5M.pdf.
- Tables and figures, accessible here: https://ughe.org/wp-content/uploads/2021/02/2.-Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf.

Reporting guidelines

DRYAD: StaRI 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.kh189324x[22]. Accessible here: https://ughe.org/wp-content/uploads/2021/02/3.-StaRI-checklist-IS-Exemplar-framework-methods.pdf.

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

Data availability

Underlying data

Zenodo: IRM raw data (video format) and dataset (csv) supporting platelet attachment to collagen IV or fibrinogen in percentage over time (related to Figure 1), https://doi.org/10.5281/zenodo.3774819[47].

Zenodo: Raw data, temporal profiling for platelet spreading dynamics (related to Figure 3). https://doi.org/10.5281/zenodo.3774823[48].

Zenodo: Raw data for microtubule extension IRM images (videos) and raw data set (csv) (related to Figure 4), https://doi.org/10.5281/zenodo.3774827[49].

Zenodo: Raw data (IRM videos) of Nocodazole experiments (videos) and raw dataset for statistical purposes (csv) (related to Figure 4), https://doi.org/10.5281/zenodo.3774835[50].

Zenodo: Nocodazole experiment low mag images, IRM, raw data. Platelets fixed, imaged by IRM in low magnification for counting purposes. Platelets are either control or treated with nocodazole, https://doi.org/10.5281/zenodo.3774843[51].

Zenodo: Raw data to support percentage of platelets in each morphological state, 1 hour post-platelet seeding (related to Figure 8), https://doi.org/10.5281/zenodo.3774845[52].

Zenodo: Dynamics of platelet spreading over time with/without treatments with manganese and thrombin (related to Figure 8). Raw images of platelets treated with and without Manganese and thrombin (tif, jpegs) and raw data set (csv), https://doi.org/10.5281/zenodo.3774849[53].

Zenodo: Un-cropped and unedited images/movies for all (DIC, movies, cryo-ET, SEM images). https://doi.org/10.5281/zenodo.3773437[54].

Extended data

Figshare: Differential dynamics of early stages of platelet adhesion and spreading on collagen IV- and fibrinogen-coated surfaces, https://doi.org/10.6084/m9.figshare.c.4944738[24].

This project contains the following extended data:

- **Figure S1. Platelet integrated activity.** Integrated activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase.
- **Figure S2. Interactions with the surface for collagen IV and fibrinogen.** The number of pixels interacting with the surface over time for the surfaces collagen IV and fibrinogen. Time in seconds.
- **Figure S3. Quantification and image analysis of platelet spreading, based on IRM live imaging for fibrinogen.** (A) Platelet spreading viewed by IRM, and the corresponding focal activity map, $\Delta IRM_t = IRM_t - IRM_{t+1}$. Positive values (yellow) imply local attachment; negative values (blue) imply local detachment (bottom right). One filopodia initially attaching and detaching (black arrow). Scale bar 2 μm (B) Integrated tapping activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase. (C) Total number of pixels 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 event. Right images, correspond IRM images. Scale bar 2 um.
- **Movie S1**. Shows the accumulated number of transitions from interaction to not interacting with the surface at every pixel over time.
- **Movie S2**. Shows an overlay of the highly active regions on top of the IRM images over time on collagen IV.
- **Movie S3**. Shows an overlay of the highly active regions on top of the IRM images over time on fibrinogen.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability

IRM spreading dynamics source code available from: https://github.com/assafZaritskyLab/IRM-Spreading-Dynamics

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.3770506[21]

License: GNU 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.

Rwanda DHS 2015: https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf

Rwanda DHS 2000: https://dhsprogram.com/pubs/pdf/FR125/FR125.pdf

The Institute for Health Metrics and Evaluation (IHME) Low- and Middle-Income Country Neonatal, Infant, and Under-5 Mortality Geospatial Estimates 2000–2017 Local Burden of Disease data used in the current study are available for download from IHME: http://ghdx.healthdata.org/record/ihme-data/lmic-under5-mortality-rate-geospatial-estimates-2000-2017.

Underlying data

Data access is restricted to users with appropriate ethics approval from the committees listed in the Ethical Considerations section. A reader 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.kh189324x[22]

This project contains the following extended data:

- Interview guide, accessible here: https://ughe.org/wp-content/uploads/2021/02/1.-Interview-Guide_Exemplars-U5M.pdf.
- Tables and figures, accessible here: https://ughe.org/wp-content/uploads/2021/02/2.-Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf.

Reporting guidelines

DRYAD: StaRI checklist for 'Development and application of a hybrid implementation research framework to understanc success in reducing under-5 mortality in Rwanda' https://doi.org/10.5061/dryad.kh189324x[22]. Accessible here: https://ughe.org/wp-content/uploads/2021/02/3.-StaRI-checklist-IS-Exemplar-framework-methods.pdf.

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

Data availability

Underlying data

Zenodo: IRM raw data (video format) and dataset (csv) supporting platelet attachment to collagen IV or fibrinogen in percentage over time (related to Figure 1), https://doi.org/10.5281/zenodo.3774819[47].

Zenodo: Raw data, temporal profiling for platelet spreading dynamics (related to Figure 3). https://doi.org/10.5281/zenodo.3774823[48].

Zenodo: Raw data for microtubule extension IRM images (videos) and raw data set (csv) (related to Figure 4), https://doi.org/10.5281/zenodo.3774827[49].

Zenodo: Raw data (IRM videos) of Nocodazole experiments (videos) and raw dataset for statistical purposes (csv) (related to Figure 4), https://doi.org/10.5281/zenodo.3774835[50].

Zenodo: Nocodazole experiment low mag images, IRM, raw data. Platelets fixed, imaged by IRM in low magnification for counting purposes. Platelets are either control or treated with nocodazole, https://doi.org/10.5281/zenodo.3774843[51].

Zenodo: Raw data to support percentage of platelets in each morphological state, 1 hour post-platelet seeding (related to Figure 8), https://doi.org/10.5281/zenodo.3774845[52].

Zenodo: Dynamics of platelet spreading over time with/without treatments with manganese and thrombin (related to Figure 8). Raw images of platelets treated with and without Manganese and thrombin (tif, jpegs) and raw data set (csv), https://doi.org/10.5281/zenodo.3774849[53].

Zenodo: Un-cropped and unedited images/movies for all (DIC, movies, cryo-ET, SEM images). https://doi.org/10.5281/zenodo.3773437[54].

Extended data

Figshare: Differential dynamics of early stages of platelet adhesion and spreading on collagen IV- and fibrinogen-coated surfaces, https://doi.org/10.6084/m9.figshare.c.4944738[24].

This project contains the following extended data:

- **Figure S1. Platelet integrated activity.** Integrated activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase.
- **Figure S2. Interactions with the surface for collagen IV and fibrinogen.** The number of pixels interacting with the surface over time for the surfaces collagen IV and fibrinogen. Time in seconds.
- **Figure S3. Quantification and image analysis of platelet spreading, based on IRM live imaging for fibrinogen.** (A) Platelet spreading viewed by IRM, and the corresponding focal activity map, $\Delta IRM_t = IRM_t - IRM_{t+1}$. Positive values (yellow) imply local attachment; negative values (blue) imply local detachment (bottom right). One filopodia initially attaching and detaching (black arrow). Scale bar 2 μm (B) Integrated tapping activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase. (C) Total number of pixels 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 event. Right images, correspond IRM images. Scale bar 2 um.
- **Movie S1.** Shows the accumulated number of transitions from interaction to not interacting with the surface at every pixel over time.
- **Movie S2.** Shows an overlay of the highly active regions on top of the IRM images over time on collagen IV.
- **Movie S3.** Shows an overlay of the highly active regions on top of the IRM images over time on fibrinogen.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability

IRM spreading dynamics source code available from: https://github.com/assafZaritskyLab/IRM-Spreading-Dynamics

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.3770506[21]

License: GNU General Public License v3.0

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.

Rwanda DHS 2015: https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf

Rwanda DHS 2000: https://dhsprogram.com/pubs/pdf/FR125/FR125.pdf

The Institute for Health Metrics and Evaluation (IHME) Low- and Middle-Income Country Neonatal, Infant, and Under-5 Mortality Geospatial Estimates 2000–2017 Local Burden of Disease data used in the current study are available for download from IHME: http://ghdx.healthdata.org/record/ihme-data/lmic-under5-mortality-rate-geospatial-estimates-2000-2017.

Underlying data

Data access is restricted to users with appropriate ethics approval from the committees listed in the Ethical Considerations section. A reader 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.kh189324x[22]

This project contains the following extended data:

- Interview guide, accessible here: https://ughe.org/wp-content/uploads/2021/02/1.-Interview-Guide_Exemplars-U5M.pdf.
- Tables and figures, accessible here: https://ughe.org/wp-content/uploads/2021/02/2.-Extended-Data-Tables-1a-and-1b-Table-2-Figure-1.pdf.

Reporting guidelines

DRYAD: StaRI 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.kh189324x[22]. Accessible here: https://ughe.org/wp-content/uploads/2021/02/3.-StaRI-checklist-IS-Exemplar-framework-methods.pdf.

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

Findable and Accessible

Interoperable

Reusable

Data availability

Underlying data

Zenodo: IRM raw data (video format) and dataset (csv) supporting platelet attachment to collagen IV or fibrinogen in percentage over time (related to Figure 1). https://doi.org/10.5281/zenodo.3774819[47].

Zenodo: Raw data, temporal profiling for platelet spreading dynamics (related to Figure 3). https://doi.org/10.5281/zenodo.3774823[48]

Zenodo: Raw data for microtubule extension IRM images (videos) and raw data set (csv) (related to Figure 4), https://doi.org/10.5281/zenodo.3774827[49].

Zenodo: Raw data (IRM videos) of Nocodazole experiments (videos) and raw dataset for statistical purposes (csv) (related to Figure 4). https://doi.org/10.5281/zenodo.3774835[50].

Zenodo: Nocodazole experiment low mag images, IRM, raw data. Platelets fixed, imaged by IRM in low magnification for counting purposes. Platelets are either control or treated with nocodazole. https://doi.org/10.5281/zenodo.3774843[51].

Zenodo: Raw data to support percentage of platelets in each morphological state, 1 hour post-platelet seeding (related to Figure 8). https://doi.org/10.5281/zenodo.3774845[52].

Zenodo: Dynamics of platelet spreading over time with/without treatments with manganese and thrombin (related to Figure 8). Raw images of platelets treated with and without Manganese and thrombin (tif, jpegs) and raw data set (csv), https://doi.org/10.5281/zenodo.3774849[53].

Zenodo: Un-cropped and unedited images/movies for all (DIC, movies, cryo-ET, SEM images). https://doi.org/10.5281/zenodo.3773437[54].

Extended data

Figshare: Differential dynamics of early stages of platelet adhesion and spreading on collagen IV- and fibrinogen-coated surfaces, https://doi.org/10.6084/m9.figshare.c.4944738[24].

This project contains the following extended data:

- **Figure S1. Platelet integrated activity.** Integrated activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase.
- **Figure S2. Interactions with the surface for collagen IV and fibrinogen.** The number of pixels interacting with the surface over time for the surfaces collagen IV and fibrinogen. Time in seconds.
- **Figure S3. Quantification and image analysis of platelet spreading, based on IRM live imaging for fibrinogen.** (A) Platelet spreading viewed by IRM, and the corresponding focal activity map, $\Delta IRM_t = IRM_t - IRM_{t+1}$. Positive values (yellow) imply local attachment; negative values (blue) imply local detachment (bottom right). One filopodia initially attaching and detaching (black arrow). Scale bar 2 μm (B) Integrated tapping activity of platelets: the mean absolute value |ΔIRM| at every time point. X-axis: Time in seconds. Y-axis: Platelet mean activity. Red dotted lines separate the phases: background, prior to platelet attachment, filopodial spreading phase, lamellipodial spreading phase, and the fully spread phase. (C) Total number of pixels 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 event. Right images, correspond IRM images. Scale bar 2 um.
- **Movie S1.** Shows the accumulated number of transitions from interaction to not interacting with the surface at every pixel over time.
- **Movie S2.** Shows an overlay of the highly active regions on top of the IRM images over time on collagen IV.
- **Movie S3.** Shows an overlay of the highly active regions on top of the IRM images over time on fibrinogen.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

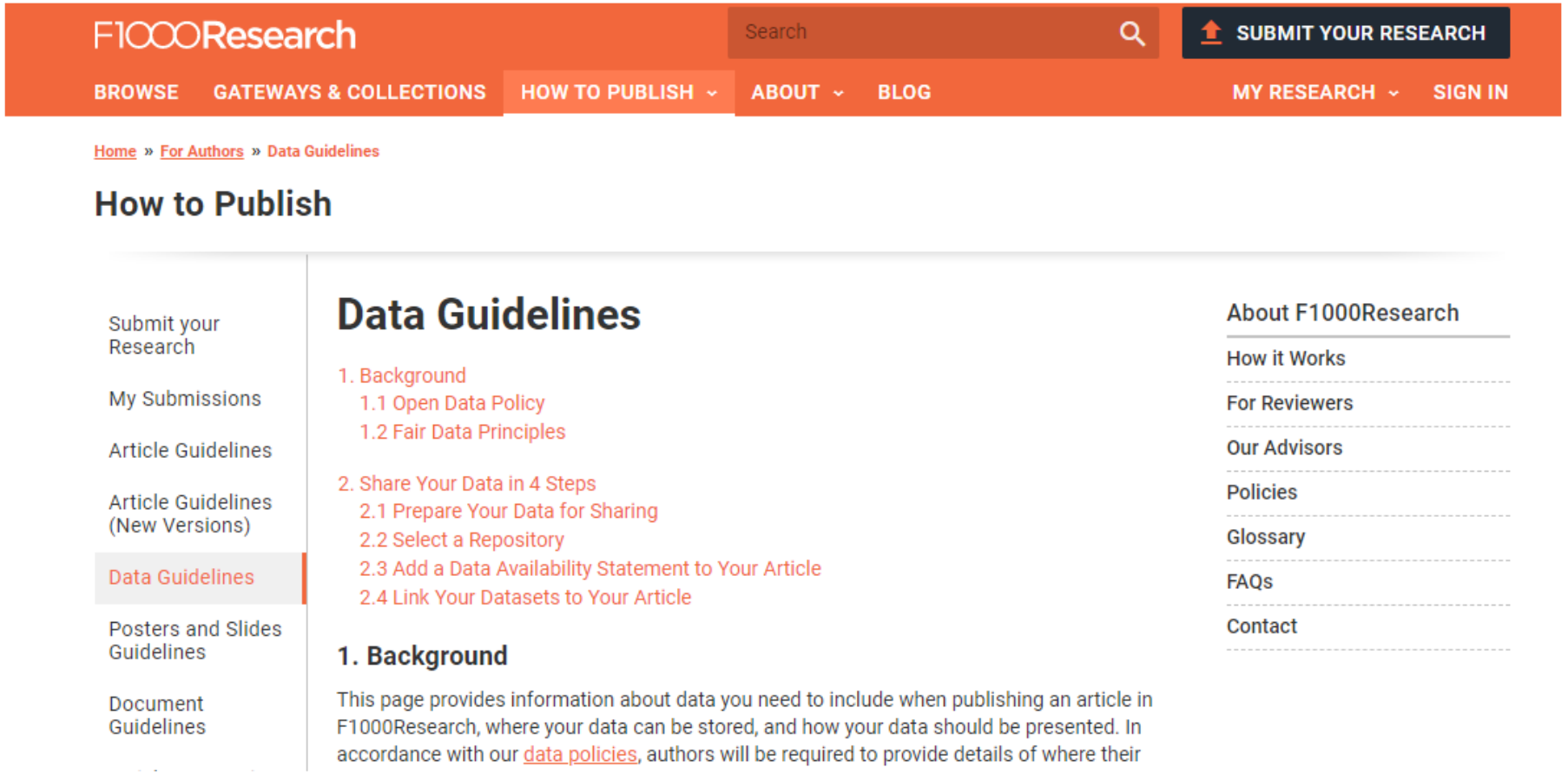
Software availability

IRM spreading dynamics source code available from: https://github.com/assafZaritskyLab/IRM-Spreading-Dynamics

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.3770506[21]

License: GNU General Public License v3.0

Open data support and resources



f1000research.com/for-authors/data-guidelines

research@f1000.com

Open data support and resources

Understanding Open Data

Everything you need to know about making your research data open and FAIR

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

F1000Research