ENCOURAGING DATA SHARING OUTSIDE OF MANDATES

ASHLEY FARLEY

PROGRAM OFFICER OF KNOWLEDGE & RESEARCH SERVICES

BILL & MELINDA GATES FOUNDATION

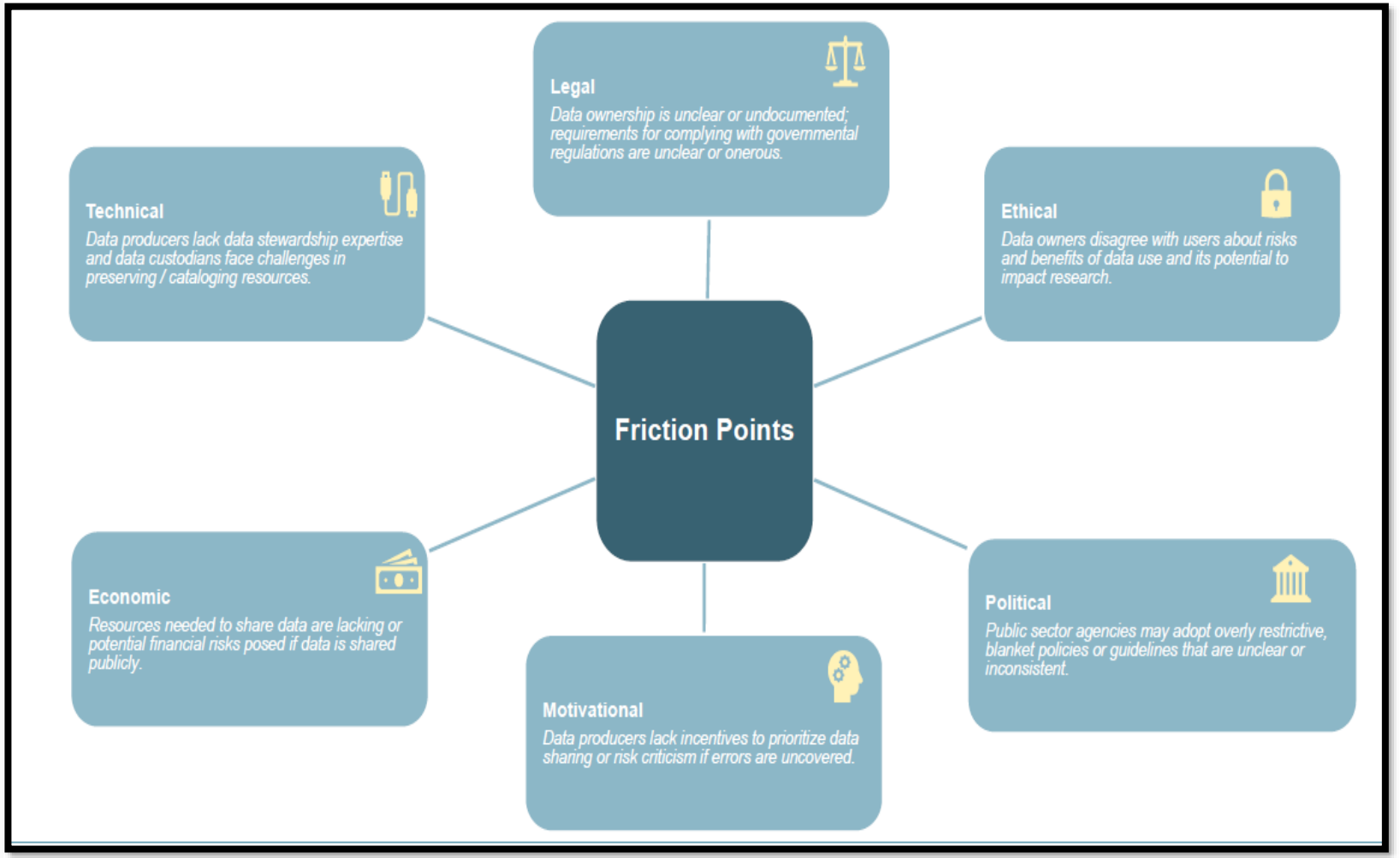
WHAT WE HAVE LEARNED

- Capacity/Resourcing Limitations
- Depends on where they publish as to quality of the data sharing
- Downstream versus Update

DATA SHARING FRICTION POINTS

Different disciplines are experiencing the same obstacles to data sharing

While friction points may be similar solutions may not be

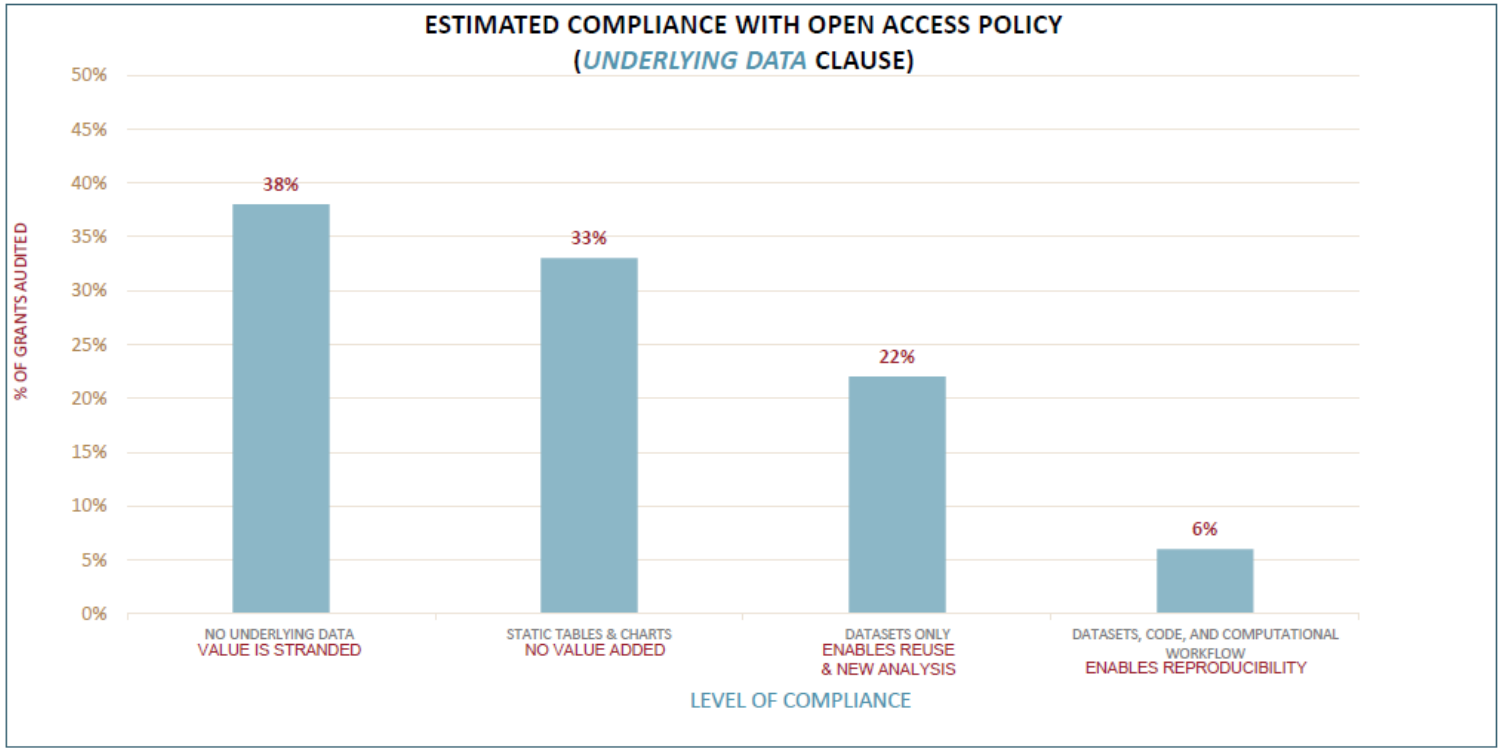


We estimate that **71% of grantees did not make their underlying data accessible** (stranded value) OR **simply attached static tables and figures** already embedded within the article (no value added). **22% of grantees made datasets used to perform analysis accessible** (enabling reuse and new analysis). **Only 6% of grantees made datasets, analysis code, and a description of the computational workflow accessible** (enabling reuse, new analysis and increasing the likelihood of reproducing the study).

In June 2019, the OOP completed an audit of underlying data for 140 grants yielding peer-reviewed articles in 2018.

The audit included manual checks of journal repositories to scan for datasets, code and computational workflow instructions accompanying articles or linked to separate data and code repositories.

Results from the audit are preliminary and represent ~ 7% of grants estimated to yield articles and subject to the Open Access policy in 2018. Without a clear view of which specific outputs are expected to be generated for each grant, compliance with the "underlying data" clause is difficult to verify.



'Data on request'

Sadly there is little change in the sharing of the data underlying publications. Only 9% of Covid-19-related research articles in Europe PMC have any data availability statement – which tells the reader where and under what conditions the data can be accessed – compared to 22% of all research articles published in 2020.

Even when authors do include a data availability statement, the criteria for accessing the data is often ambiguous. For example, a recent study of Covid-19 related preprints submitted to MedRxiv (the preprint server for Health Sciences) reported 8% of data availability statements mention 'reasonableness' as one of the criteria for granting access.

This behaviour is mirrored in data related to Covid-19 clinical trials. A recent study found that data is only being shared in a minority of cases (15.7%), with nearly half (47.8%) of the trial registry entries explicitly saying they are 'not willing to share data'.

These examples highlight that there is still significant work to be done to shift to make all research more open.

DATA AVAILABILITY STATEMENTS

- You can't know what you don't know
- Encourages thinking through the data sharing component
- Creates necessary linkages

GATES OPEN RESEARCH EXAMPLE

- Links to repository where data can be downloaded
- Explicitly states why some data can't be openly shared
- Includes what is necessary to gain access to safeguarded data

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

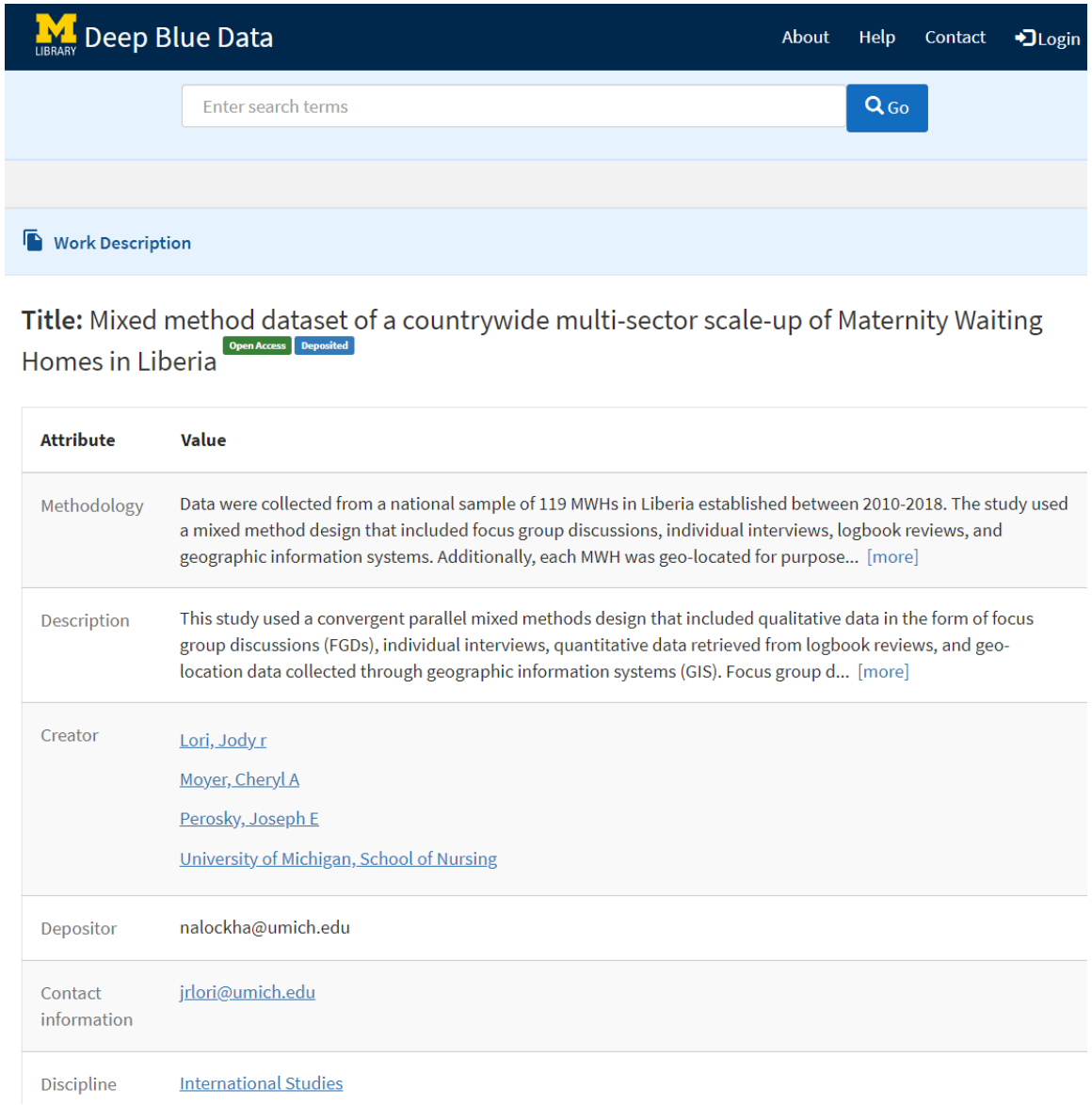
DRYAD: Development and application of a hybrid implementation research framework to understand success in reducing under-5 mortality in Rwanda. https://doi.org/10.5061/dryad.kh189324x[22]

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.

DEEP BLUE

- Depositing Assistance
- Robust Metadata
- External linkages



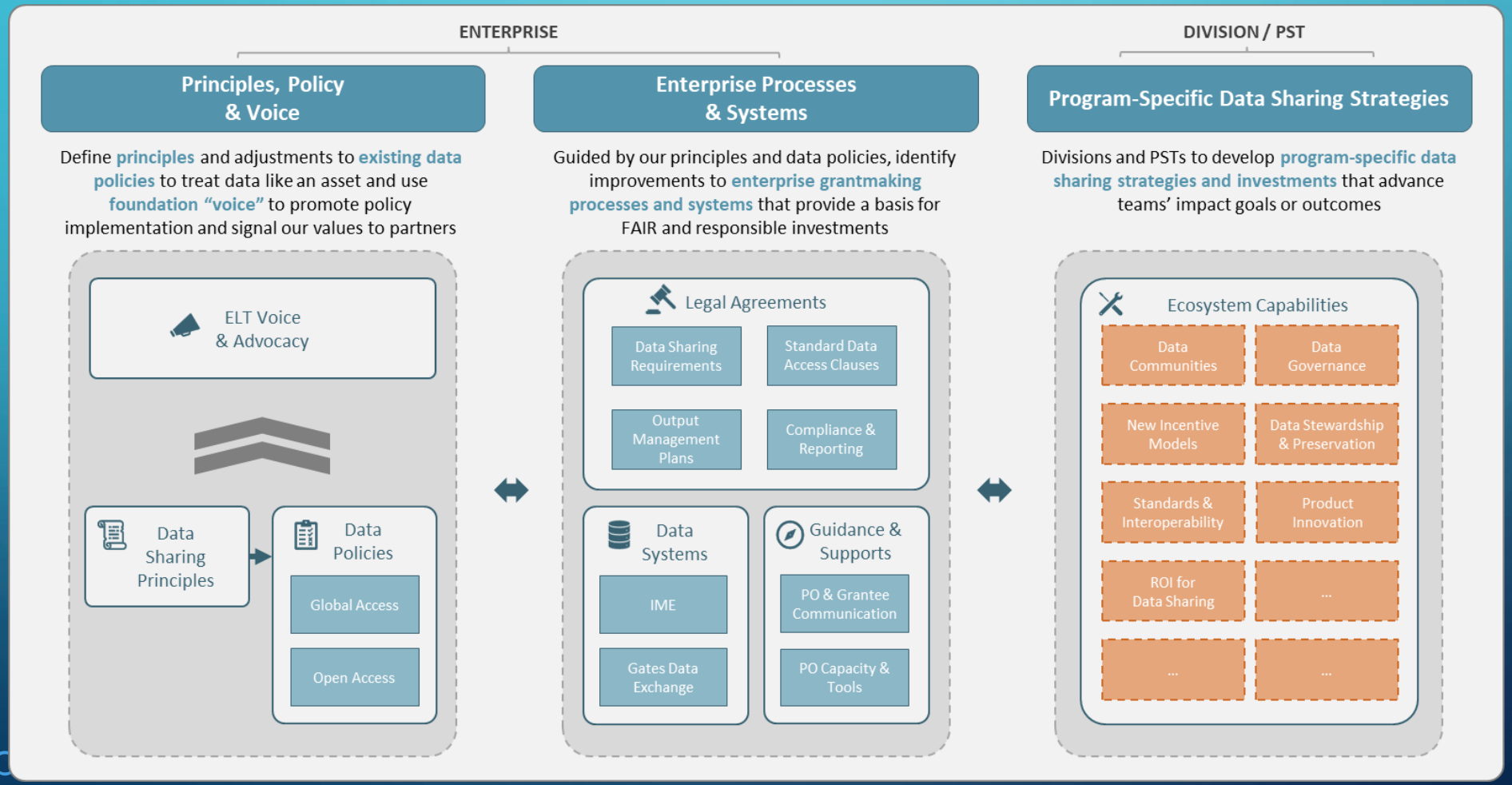
DEEP BLUE

- Articles need to point to the data
- Improve data searchability

Other Funding agency	The Bill & Melinda Gates Foundation
ORSP grant number	N023584
Keyword	Maternity Waiting Homes Liberia
Date coverage	2017-12 to 2018-06
Citations to related material	James, K.H., Perosky, J.E., McLean, K. et al. Protocol for geolocating rural villages of women in Liberia utilizing a maternity waiting home. BMC Res Notes 12, 196 (2019). https://doi.org/10.1186/s13104-019-4224-1 Coley, KM, Perosky, JE, Nyanplu, A, et al. Acceptability and feasibility of insect consumption among pregnant women in Liberia. Matern Child Nutr. 2020; 16:e12990. https://doi.org/10.1111/mcn.12990
Resource type	Dataset
Last modified	06/16/2020
Published	06/16/2020
Language	English
DOI	https://doi.org/10.7302/vs47-p951
License	http://creativecommons.org/licenses/by/4.0/

TREATING DATA AS AN ASSET

FOUNDATION DATA SHARING MODEL



HOW WE CAN DO BETTER

- Data Availability Statement requirements
- Partnering with publishers & partners who do this well
  - Deep Blue Example
- Recognize & promote those already doing this
- Showing the value of data management & sharing will encourage these practices without a forcing mechanism
- Treat Data as an Asset