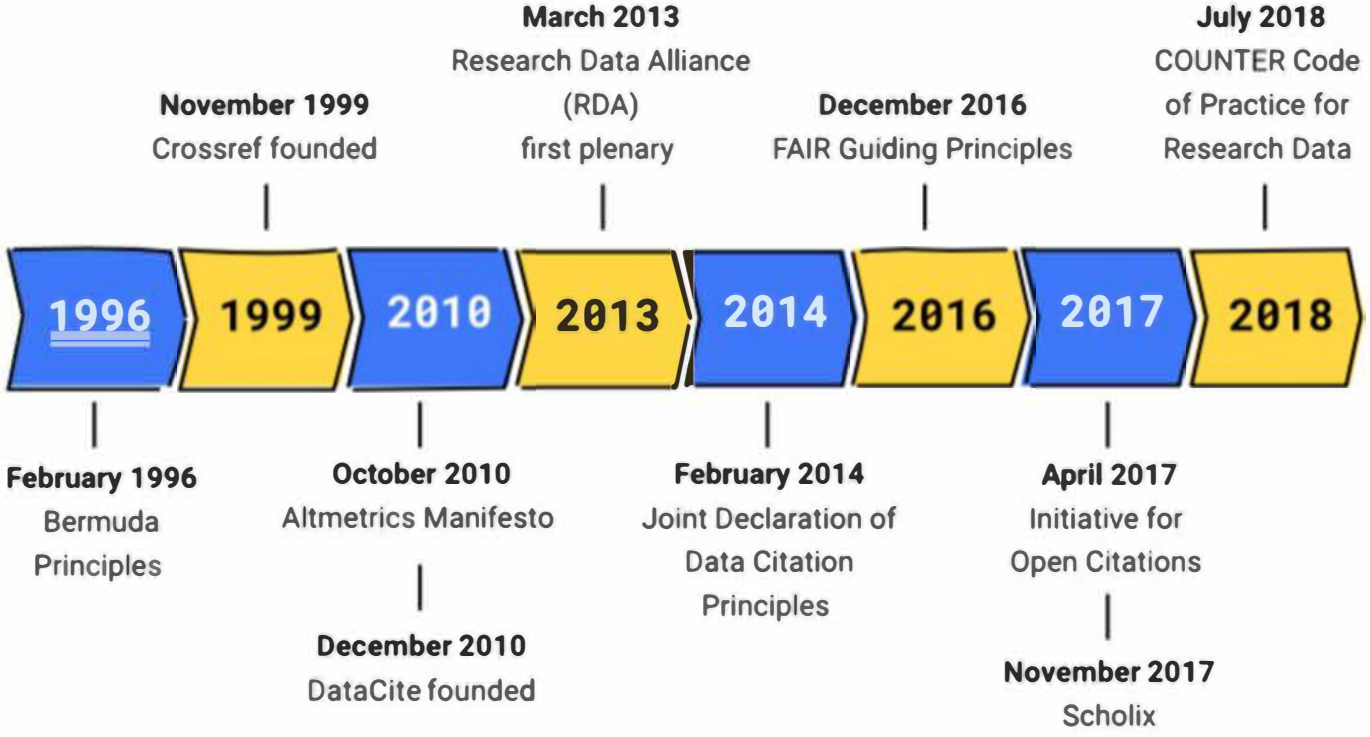
Current State of Research Data Metrics

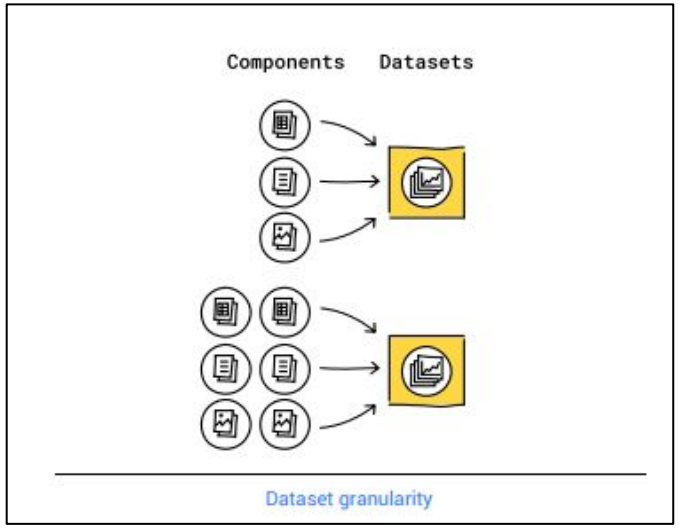
Daniella Lowenberg
@danilowenberg
Make Data Count Lead
Dryad Product Manager
California Digital Library

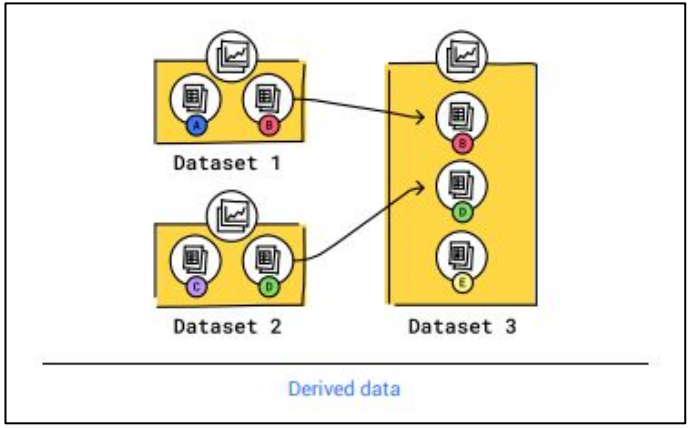
We *value* data, so how do we *evaluate* data?

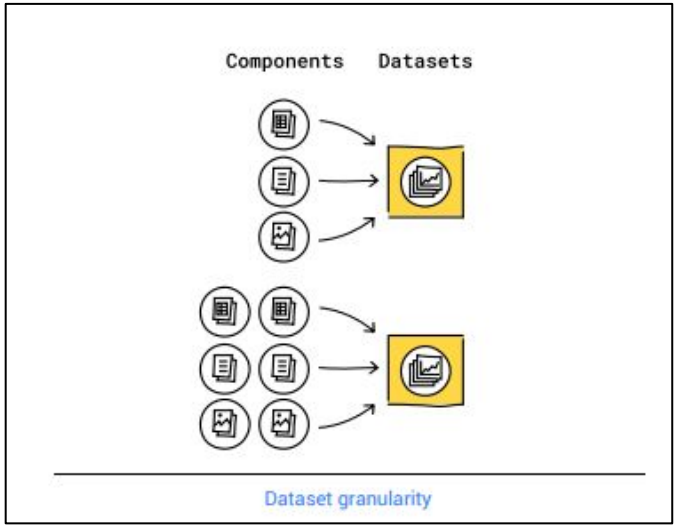


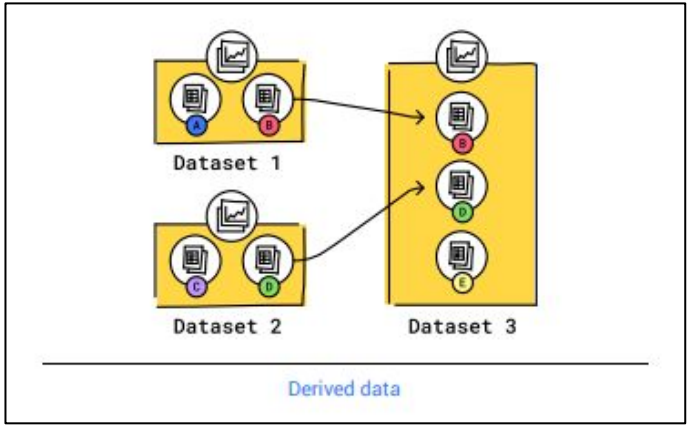
Open data metrics milestones

Data are complicated

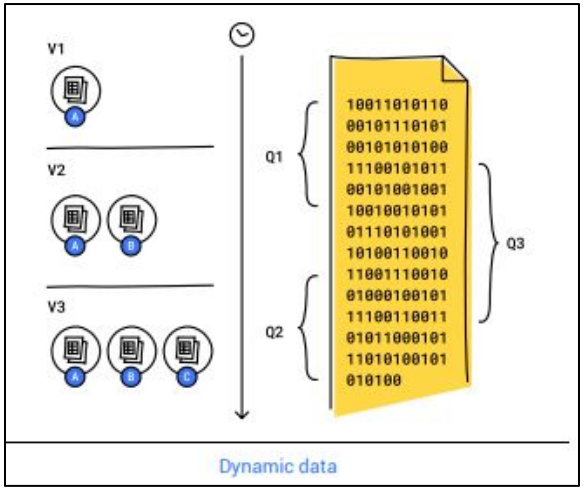

Components
Datasets
Dataset granularity


Dataset 1
Dataset 2
Dataset 3
Derived data

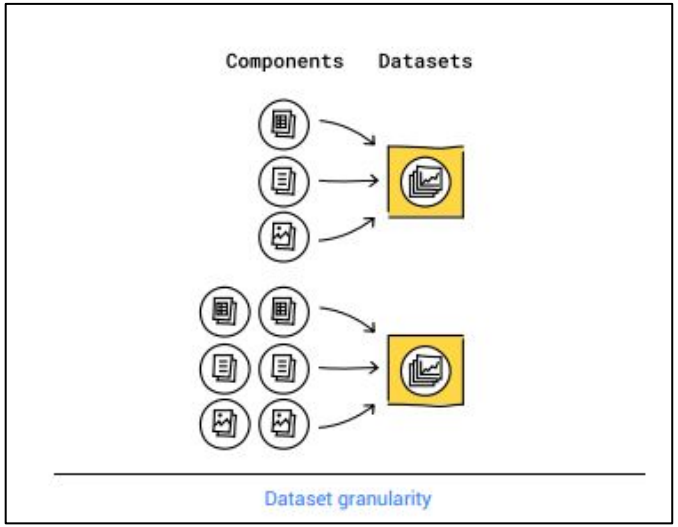

Components
Datasets
Dataset granularity



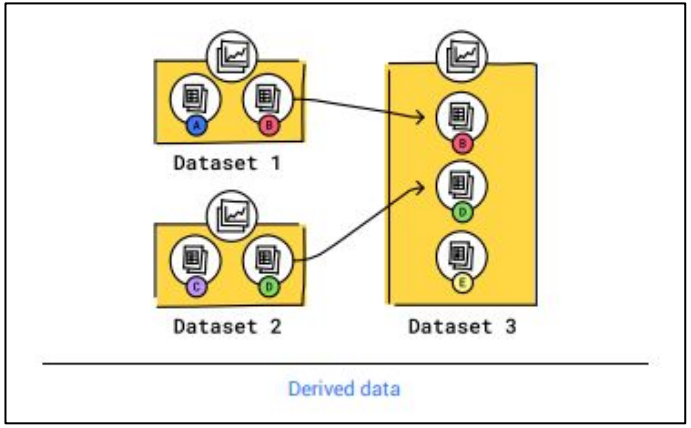
Derived data

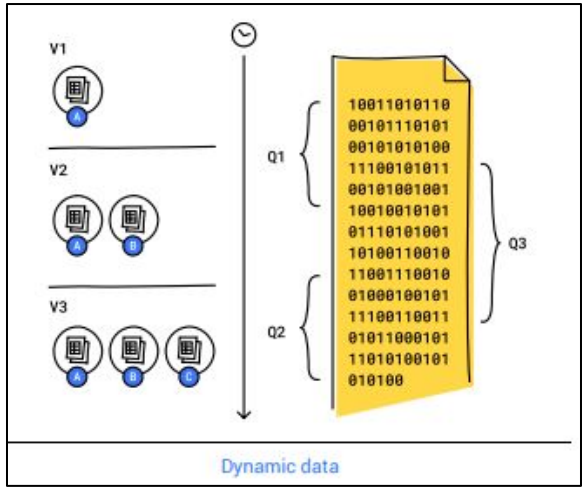


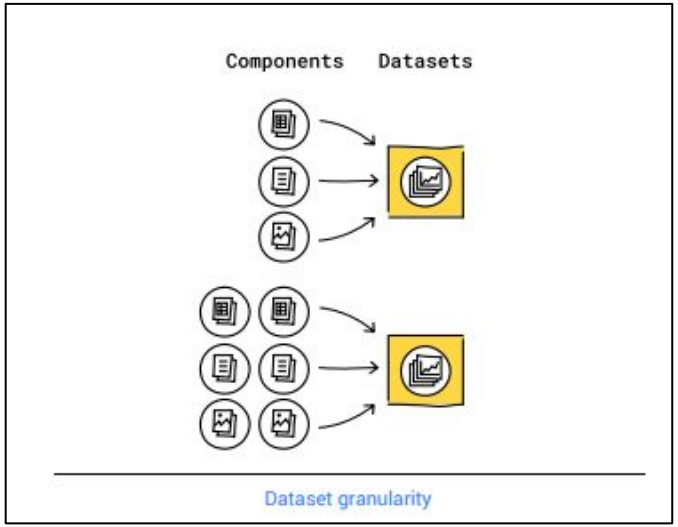
Dynamic data

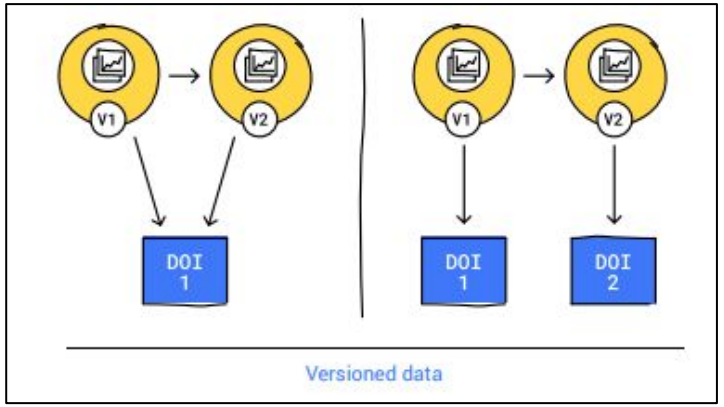


Dataset granularity


Dataset 1
Dataset 2
Dataset 3
Derived data


V1
V2
V3
Q1
Q2
Q3
10011010110
00101110101
00101010100
11100101011
00101001001
10010010101
01110101001
10100110010
11001110010
01000100101
11100110011
01011000101
11010100101
010100
Dynamic data


Components
Datasets
Dataset granularity


V1
V2
DOI 1
V1
V2
DOI 1
DOI 2
Versioned data

Data usage

Are they comparable?

Are they comparable?

2,496	1,311
views	downloads

33174 views

6967 downloads

30	4
views	downloads

COUNTER
CONSISTENT CREDIBLE COMPARABLE

The COUNTER Code of Practice for Research Data

The Code of Practice for Research Data Usage Metrics standardizes the generation and distribution of usage metrics for research data, enabling for the first time the consistent and credible reporting of research data usage.

COUNTER welcomes input and feedback from the community on this first iteration, so that it can be further developed and refined.

A downloadable PDF is now available in the download section below.

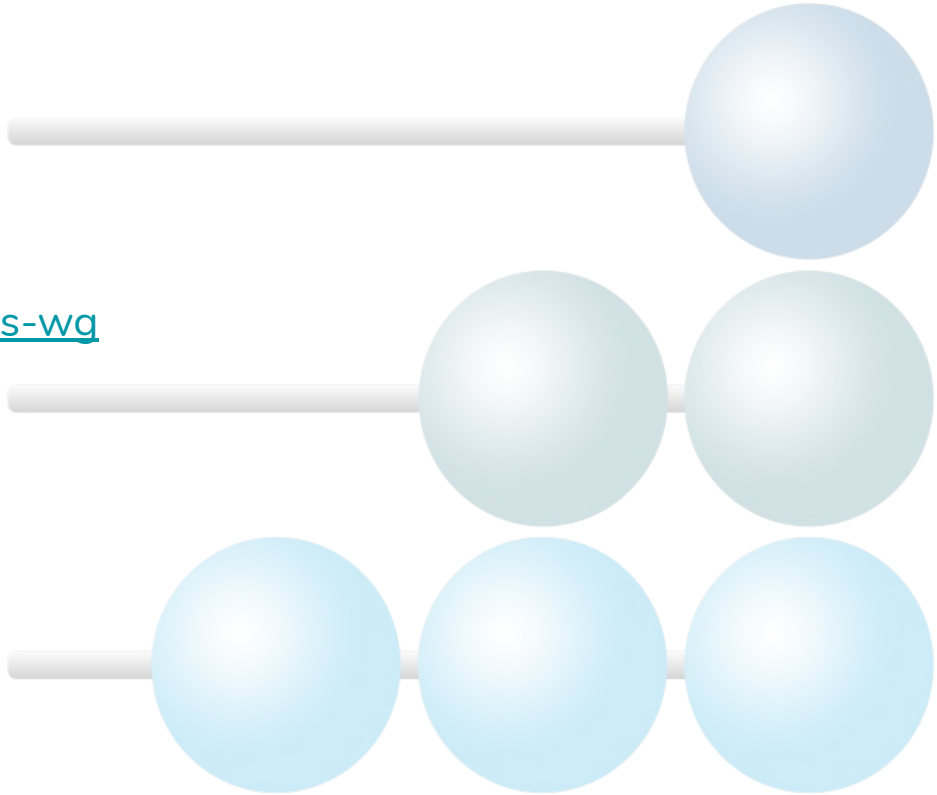
Data Usage Initiatives

RDA Data Usage Metrics WG

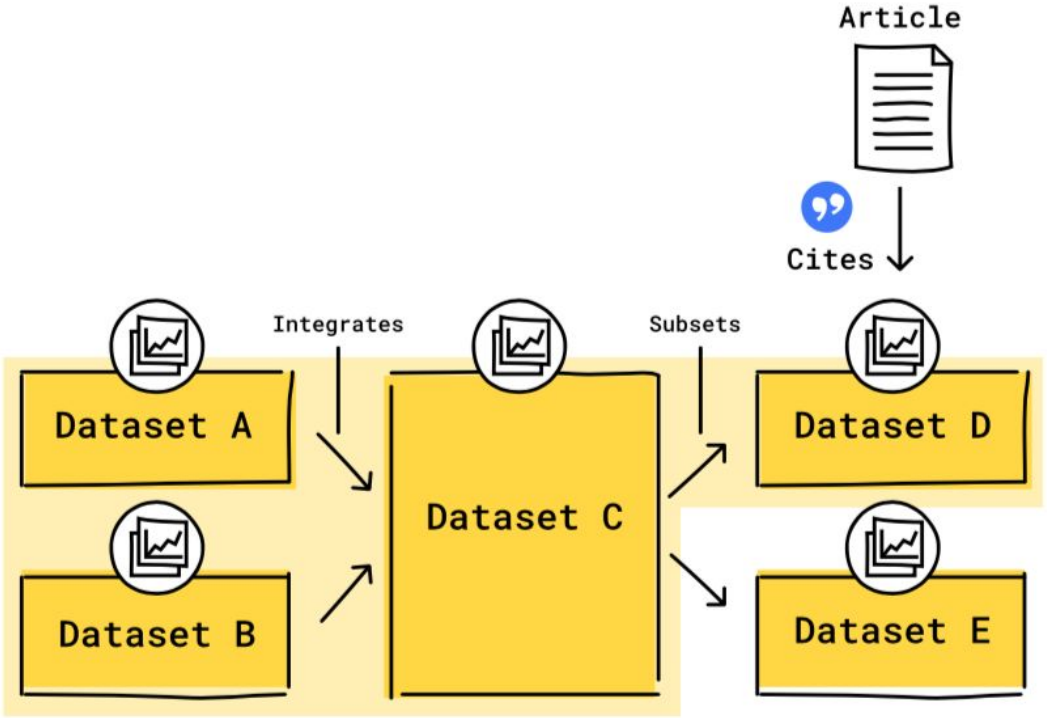
https://rd-alliance.org/groups/data-usage-metrics-wg

Make Data Count

makedatacount.org



Data citation

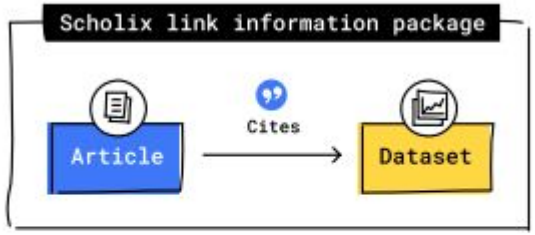


Data citation examples

a) An article cites a dataset, b) a dataset is derived from two other datasets, c) subsets of a dataset are generated.

Joint Declaration Of Data Citation Principles

Publisher Data Citation Frameworks



A Scholix Link Information Package

The package contains information about the two objects, and information about the nature of the link and the link package itself

Crossref

ESIP Introduces New Data and Software Citation Guidelines Aimed at Making Earth Science Research FAIR

GBIF #CiteTheDOI

Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.
Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G
Nat Genet, 25(1):25-29, 30 Apr 2000
Cited by: 14480 articles | PMID: 10802651 | PMCID: PMC3037419
+ Add to export list — Free to read

Fast and accurate short read alignment with Burrows-Wheeler transform.
Li H, Durbin R
Bioinformatics, 25(14):1754-1760, 17 May 2009
Cited by: 12845 articles | PMID: 19451168 | PMCID: PMC2705234
+ Add to export list — Free to read & use

PLINK: a tool set for whole-genome association and population-based linkage analyses.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC
Am J Hum Genet, 81(3):559-575, 24 Jul 2007
Cited by: 11927 articles | PMID: 17701901 | PMCID: PMC1950838
+ Add to export list — Free to read

Quantifying the impact of public omics data.

Perez-Riverol Y[1], Zorin A[1], Dass G[1], Vu MT[1], Xu P[2], Glont M[1], Vizcaíno JA[1], Jarnuczak AF[1], Petryszak R[1], Ping P[3], Hermjakob H[1]

CDL DataCite DataONE

These counts are not yet data metrics

How do we get there?

Support community initiatives around normalized data citation & data usage

Build on successes instead of reinventing the wheel

Infrastructure needs to be open

Numbers need to be transparent and standardized

Step 2: Transparent Infrastructure

We don’t know what is meaningful for data yet

We must not default to impact factor/ h-index

Step 3: Bibliometrics Principles

Curation ≠ peer review

Evaluation of data quality / scientific merit is not in numbers

Step 4: Qualitative Assessment

Metrics change behavior

We can’t be prescriptive to researchers

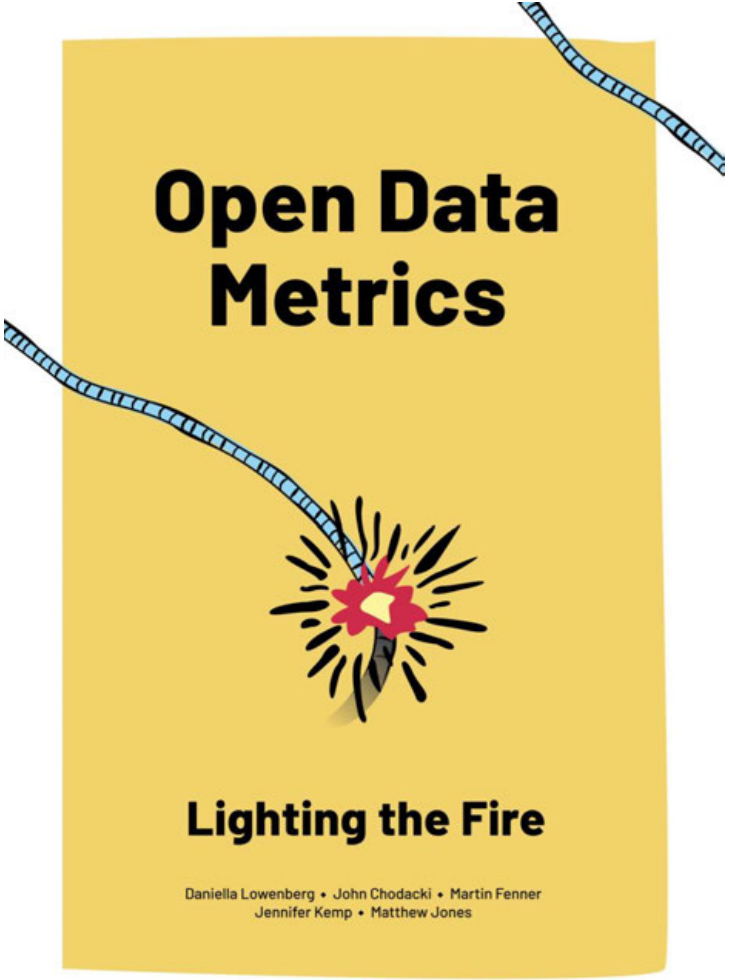
Get Involved

Benchmark counts & promote best practices for data usage and data citation

Join the conversation and present use cases

Collaborate with broader research stakeholders to build open data metrics

More Info


Open Data Metrics
Lighting the Fire
Daniella Lowenberg • John Chodacki • Martin Fenner
Jennifer Kemp • Matthew Jones

@makedatacount
@danilowenberg
makedatacount.org
opendatametrics.org

Alfred P. Sloan Foundation

University of California
CDL
California Digital Library

DataCite
FIND, ACCESS, AND REUSE DATA

DataONE