Making Data Count
Promoting Open Data Through Usage and Impact Tracking

Stephen Abrams
UC Curation Center
California Digital Library

@slabrams
Citation lies at the heart of the academic enterprise

- Underlies scholarly communication
- Facilitates discovery and reuse
- Provides attribution
- Assigns credit
Well-established social/technical support for bibliographic citation

- CrossRef DOIs
  www.crossref.org
- Abstracting and indexing
- COUNTER usage reporting
  www.projectcounter.org
- Altmetrics
- Normative scholarly practice
What about data?

- Should be accepted as 1st-class scholarly outputs, but ...
- No widely-accepted standards for citation, logging, or reporting
- Little infrastructure support
- Insufficient awareness
- Need for data-level metrics (DLM)
Making Data Count

- Two year project supported by Alfred P. Sloan Foundation
- Build upon earlier NSF-funded work by CDL, DataONE, and PLOS (EAGER #1448821)
  mdc.lagotto.io
- COUNTER code of practice for data usage reporting
- DataCite-hosted DLM service
- Integration of new DLM sources and clients
- Outreach and advocacy
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability

Diagram:
- Article based-on data package
- Data package describes metadata
- Metadata generates data
- Data controls workflow
- Workflow generates software
Why isn’t this just simple?

- Datasets are different from articles
  - Atomicity
  - Versioning
  - Non-congruency
  - Composability

“Data citations should facilitate identification of ... specific data that support a claim.... [and] should include ... specific timeslice, version, and/or granular portion”

www.force11.org/group/joint-declaration-data-citation-principles-final
COUNTER code of practice

- Relevant metrics
  - Citations
  - Downloads
  - Page views
  - Social mentions

Kratz and Strasser (2015), Scientific Data 2 (150039) doi: 10.1038/sdata.2015.39

- Filtering robotic activity
- Granularity and version-aware accounting and reporting
**DLM service**

- Event Data hosted by DataCite
  
  eventdata.datacite.org

- Based on Lagotto ALM platform
  
  github.com/lagotto

- Aggregating usage data and references from repository, A&I, full-text corpora, and social media sources

- APIs for building future added-value services

- Cooperation with CrossRef Event Data
  
  eventdata.crossref.org
DLM sources and clients

- Exemplar CDL Dash and DataONE Search clients
  
dash.cdlib.org
  search.dataone.org

- Portable, pluggable client widgets
Outreach and advocacy

- BoF at RDA 10th Plenary, Montreal, September 2017
- Workshop at RDA 12th Plenary, Fall 2018
- FORCE11
  www.force11.org
- IRIS UK
  iris.mimas.ac.uk
- NISO data metrics
- Scholix
  www.scholix.org
- Work with Re3data to add DLM-compliance to repository registry
  www.re3data.org
Teamwork

California Digital Library
www.cdlib.org
- Stephen Abrams
- John Chodacki
- Daniella Lowenberg

DataCite
www.datacite.org
- Patricia Cruse
- Martin Fenner
- Laura Garcia Rueda

COUNTER
www.projectcounter.org
- Lorraine Estelle

DataONE
www.data.org
- Amber Budden
- Matthew Jones
- Dave Vieglais

Alfred P. Sloan Foundation
www.sloan.org
- Josh Greenberg
Making Data Count

mdc.lagotto.io

- COUNTER code of practice for tracking and reporting data usage, reference, and impact
- DataCite Event Data as a central hub for an expanding scholarly DLM ecosystem
- Tools and exemplars for integrating new DLM sources and clients
- Build a community of common concern and practice

☑ Social/technical infrastructure to measure and report on the reach and impact of research data
Thank you!

Making Data Count
Promoting Open Data Through Usage and Impact Tracking

mdc.lagotto.io
#mdc