

CNI Spring 2017 Membership Meeting
Albuquerque, May 3-4, 2017

Making Data Count

Promoting Open Data Through
Usage and Impact Tracking

#mdc

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

Буров 1986 — В. А. Буров. Заметки о новгородских берестяных грамотах // Древности славян и Руси. С. 218–227.
Буров 1988 — В. А. Буров. О родословии бояр Мишенинчей-Скопых (по материалам раскопа 1951–1962 гг.) // Древности славян и Руси. С. 232–236.
Бьёрнфлатен 1983 — J. I. Bjørnflaten. On the History of the Alternation /K → C/ in the Nominal Flexion of the Three E-tions of Old Norse. In: The 13th International Congress of Slavists: Kiev 1983. Tromsø: University of Tromsø, 1983. P. 315–338.
Бьёрнфлатен 1988 — J. I. Bjørnflaten. Dialectal Sound Change in Old Norse. In: The Case of the Second Regressive Palatalisation of Velar Consonants. Oslo, 1988. P. 315–338.
Бьёрнфлатен 1990 — J. I. Bjørnflaten. The birch bark letters. In: Acta Universitatis Upsalienensis. P. 315–338.
Бэклунд 1959 — A. Bäcklund. Personal Names in Medieval Scandinavia. Stockholm, 1959.
Бэклунд 1956 — A. Bäcklund. Could Old Russian Feminine Personal Names be of Norse Origin? European Review, XXXV, 1956. P. 255–258.
Вайян — А. Вайян. Руководство по старославянскому языку. М.: Высшая школа, 1988.
Валк 1956 — С. Н. Валк. Новые грамоты о новгородских берестяных грамотах XV в. // Исторический архив, 1956, № 1. С. 232–236.
Варл. — Вкладная грамота Варлаама Хутынского (см. Бэклунд 1959).
Васильев 1905 — Л. Л. Васильев. К истории звука ѣ в старославянском языке // ИОРЯС. 1905. Т. 10. Кн. 2.
Васильев 1908 — Л. Л. Васильев. О влиянии нейотированных гласных на развитие слога // ИОРЯС. 1908. Т. 13. Кн. 3.
Васильев 1909 — Л. Л. Васильев. Одно соображение в записках о старославянском языке // ИОРЯС. 1909. Т. 14. Кн. 3. Август.
Веселовский — С. Б. Веселовский. Ономастикон (древнерусские имена). М.: Наука, 1974.
Вермеер 1986 — W. Vermeer. The rise of the North Russian dialects. In: Acta Universitatis Upsalienensis. P. 503–515.
Вермеер 1991a — W. Vermeer. The mysterious North Russian dialect. In: Acta Universitatis Upsalienensis. P. 271–295.
Вермеер 1991b — W. Vermeer. The reflex of Proto-Indo-European *s in the North Russian dialects. In: Acta Universitatis Upsalienensis. P. 271–295.
Вермеер 1996 — W. Vermeer. Aspects of the oldest Finnish birchbark documents 292, 56 and 403 // Studies in West-Finnic Languages 16, 1991. P. 315–359.
www.flickr.com/photos/quinnanya/5124103275

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

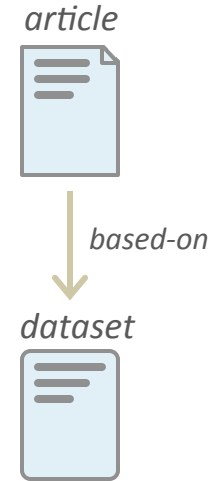
- Two year project supported by Alfred P. Sloan Foundation
- Build upon earlier NSF-funded work by CDL, DataONE, and PLOS (EAGER #1448821)
mdc.lagotto.io
- COUNTER code of practice for data usage reporting
- DataCite-hosted DLM service
- Integration of new DLM sources and clients
- Outreach and advocacy



Why isn't this just simple?

■ Datasets are different from articles

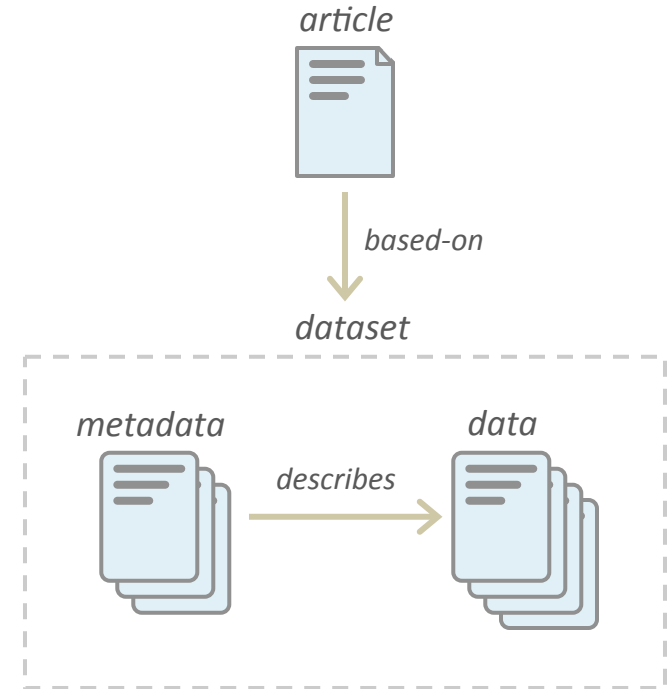
- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



Why isn't this just simple?

■ Datasets are different from articles

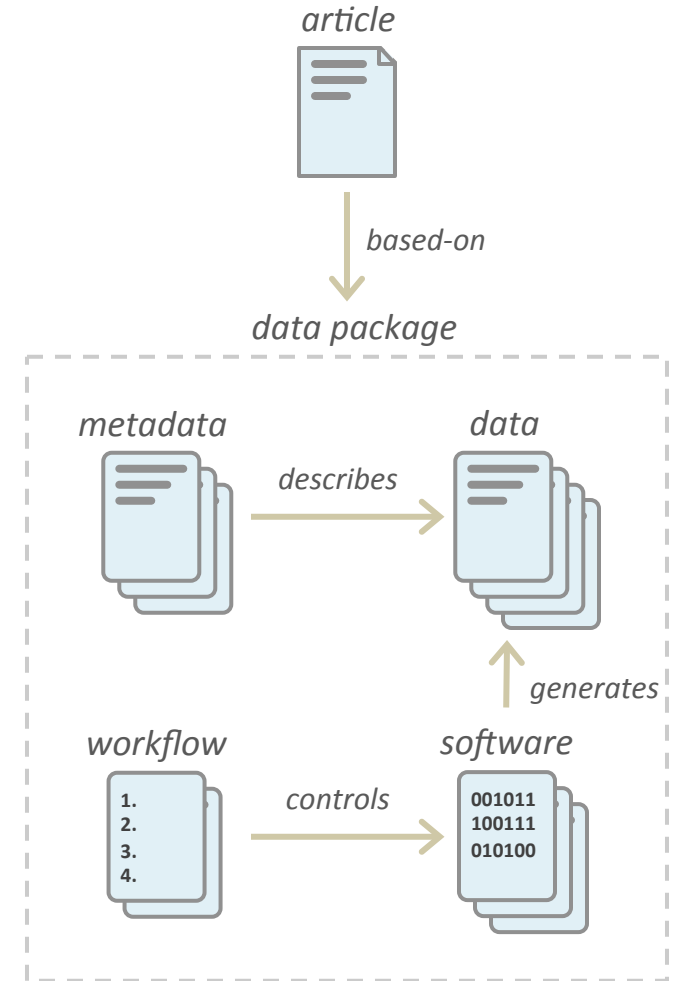
- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



Why isn't this just simple?

■ Datasets are different from articles

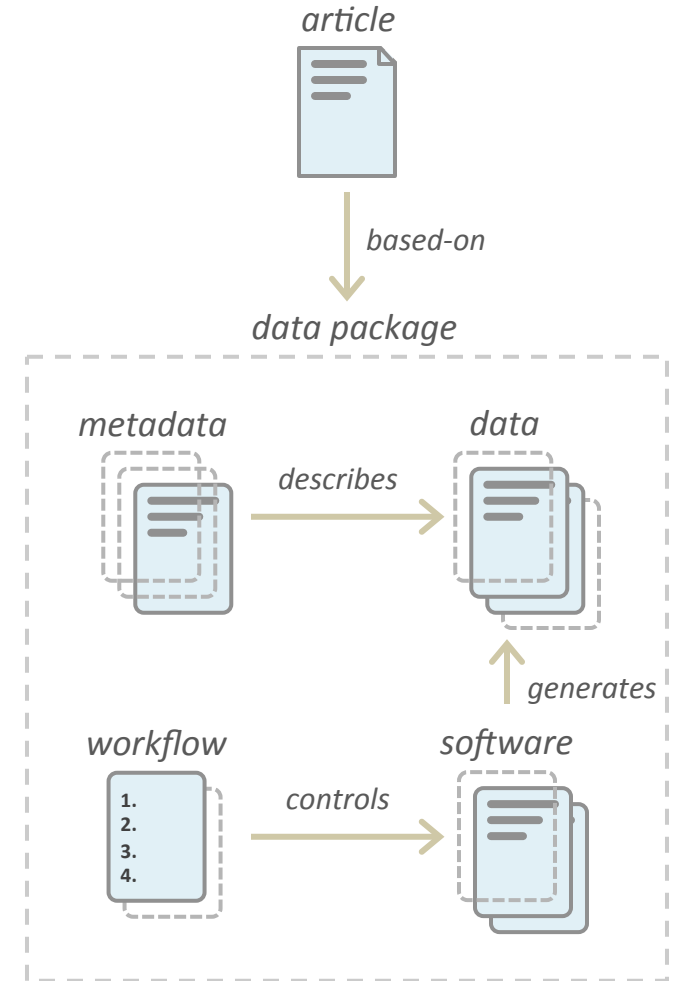
- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



Why isn't this just simple?

■ Datasets are different from articles

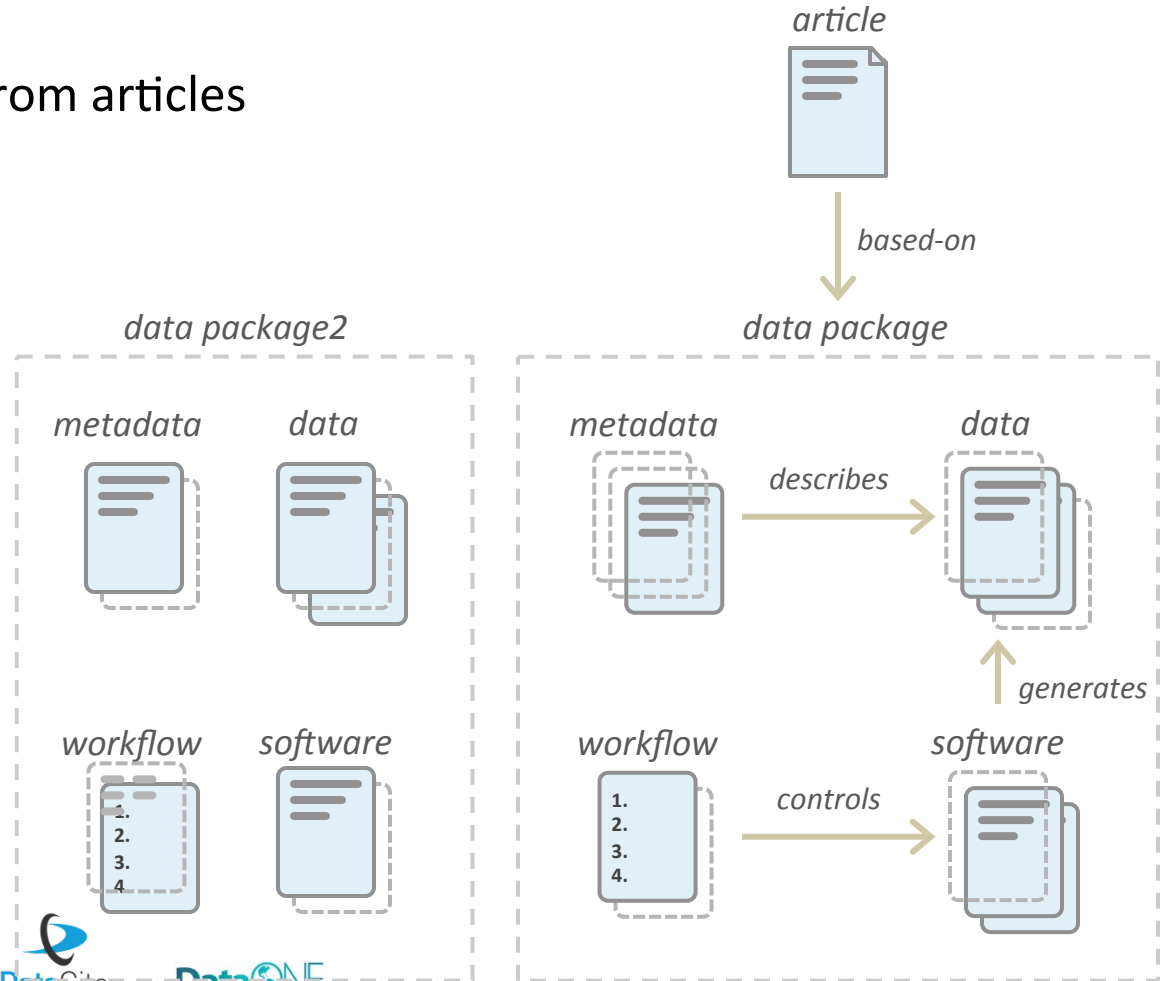
- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



Why isn't this just simple?

■ Datasets are different from articles

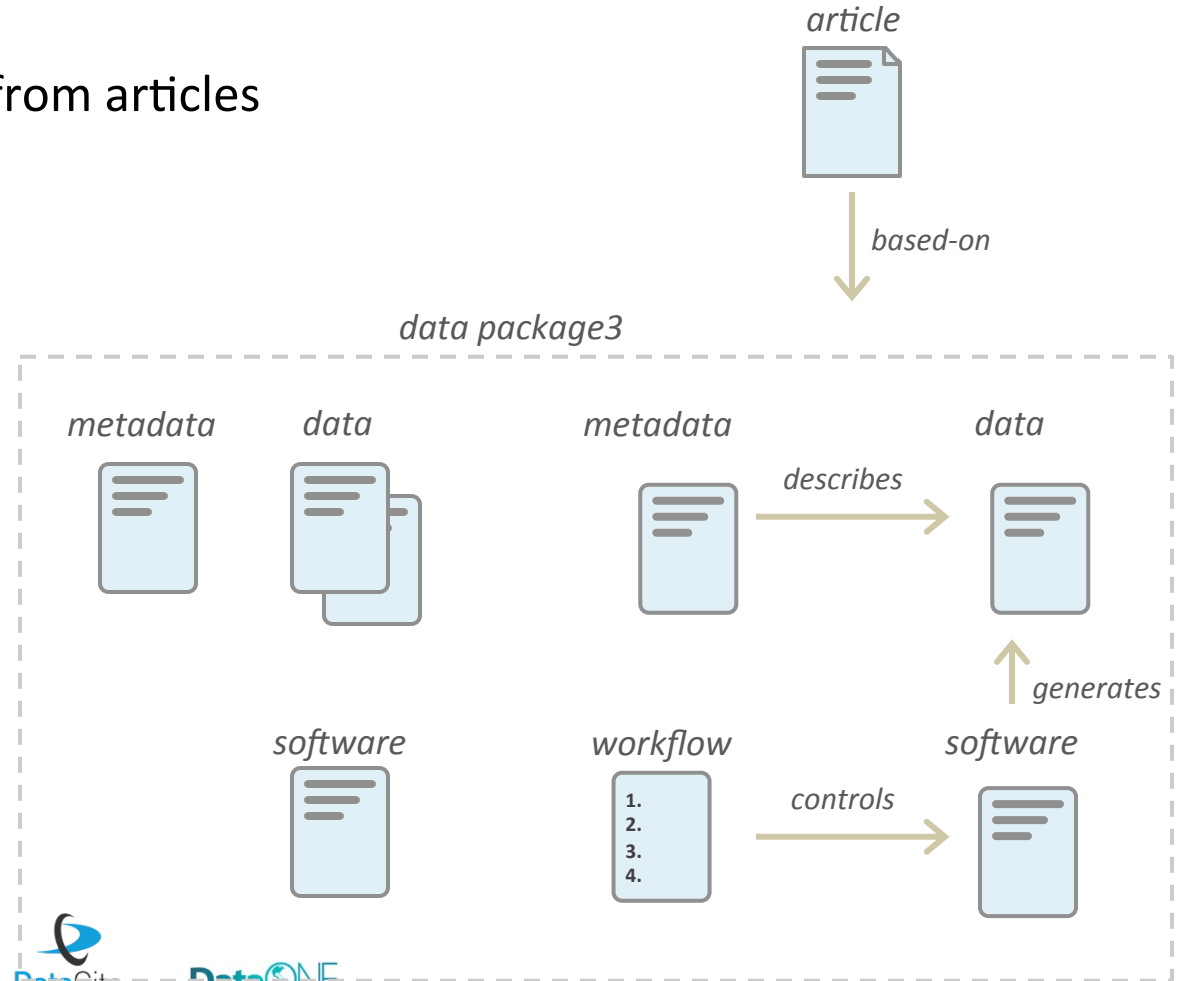
- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



Why isn't this just simple?

■ Datasets are different from articles

- ▶ *Atomicity*
- ▶ *Versioning*
- ▶ *Non-congruency*
- ▶ *Composability*



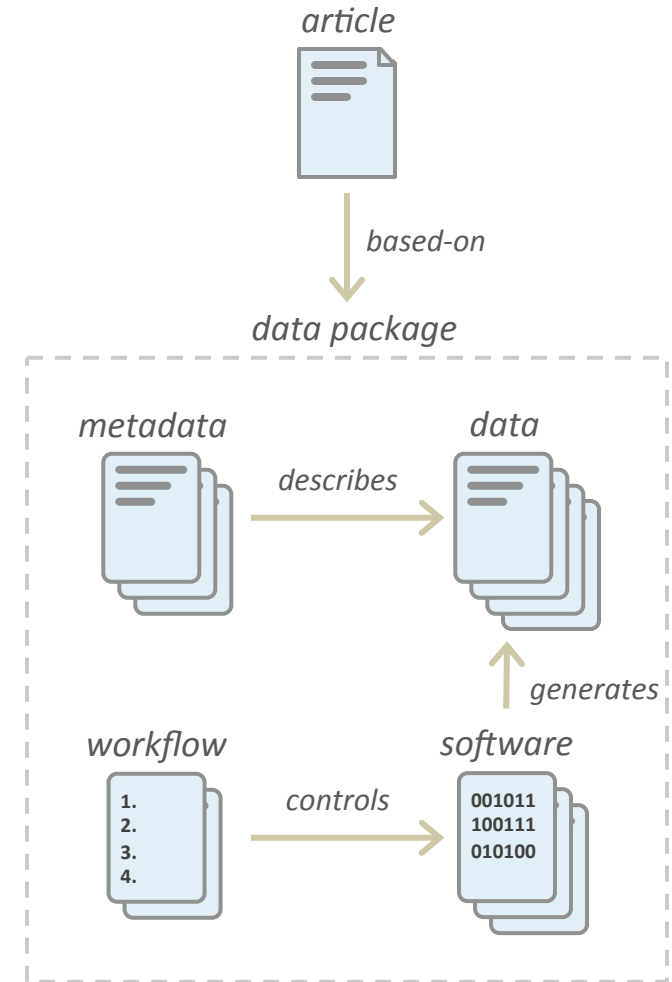
Why isn't this just simple?

■ 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](https://doi.org/10.1038/sdata.2015.39)

■ Filtering robotic activity

■ Granularity and version-aware *accounting* and *reporting*



The COUNTER

COUNTER's library, publisher and Code of Practice.

The Code of Practice enables publishers and libraries to compare data received and the value of the electronic resources.

Release 4 is the current Code of Practice.

Information about Release 5 Draft

Code of Practice

Sections

General Information

- 1.1 Purpose
- 1.2 Scope
- 1.3 Application
- 1.4 Strategy
- 1.5 Governance
- 1.6 Definitions
- 1.7 Versions
- 1.8 Auditing and COUNTER compliance
- 1.9 Relationship to other standards, protocols and codes
- 1.10 Making comments on the Code of Practice

Definitions of terms used

- SUSHI
- Usage Reports
- Data Processing
- Auditing
- Compliance

General Information

COUNTER serves librarians, vendors and publishers with usage statistics. The COUNTER Code of Practice defines these data elements, output reports, and the statistics and reports designated by the Code of Practice.

Information about the Release 5 Draft

1.1 Purpose

The purpose of the COUNTER Code of Practice is to ensure that data by establishing open, international standards that are consistent, credible and comparable.

1.2 Scope

This COUNTER Code of Practice defines the major categories of e-resources (e-books, e-journals, etc.) at an international level. In doing so, it defines the content and format of usage reports to avoid duplicate counting when international data are combined.

1.3 Application

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

Lagotto: Data

DLM Status Works Sources Publishers

Status 12 Sep 07:05 UTC

Works

151,767

API responses

last hour

46



DLM sources and clients

- Exemplar CDL Dash and DataONE Search clients
dash.cdlib.org
search.dataone.org
- Portable, pluggable client widgets

DataONE

About

News

Participate

Resources

DATAONE SEARCH:

Search

Summary

Timothy Assal

Username: assal

Contributor for 7 months
since Nov. 17, 2015

14  contributions

166  downloads

304  views

6  citations

9  saves

10  mentions

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

www.niso.org/apps/group_public/document.php?document_id=16553

- 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*

Alfred P. Sloan Foundation

www.sloan.org

- *Josh Greenberg*

COUNTER

www.projectcounter.org

- *Lorraine Estelle*

DataONE

www.data.org

- *Amber Budden*
- *Matthew Jones*
- *Dave Vieglais*



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

