RESEARCH DATA ALLIANCE

RDA Data Usage Metrics Working Group

Overview of Recommendations

April 2022

research data sharing without barriers

rd-alliance.org

RDA Data Usage Metrics Working Group

Proposition (Berlin, 2018):

- Adoption of Data Usage metrics is necessary for the recognition of research data as a first-class research output.
- Researchers, funders, publishers and institutions will all benefit from metrics that can reliably reflect impact

RDA Data Usage Metrics WG Journey

Data Usage Metrics at Repositories: A Survey

Jouneau, Thomas; Bruno, Ian; Lowenberg, Daniella https://doi.org/10.5281/zenodo.3471489

Results of a survey undertaken by the RDA Data Usage Metrics Working Group during February and March 2019 and presented at the 13th RDA Plenary Meeting in Philadelphia on 3 April 2019.

Data Usage Metrics: Potential Pitfalls

Potential pitfalls of Data Usage Metrics based on patterns observed at a major scientific data repository presented at the Botswana RDA Plenary, 2018. https://bit.ly/RDA-DUM-P12

Data Usage Metrics: Addressing Shortcomings

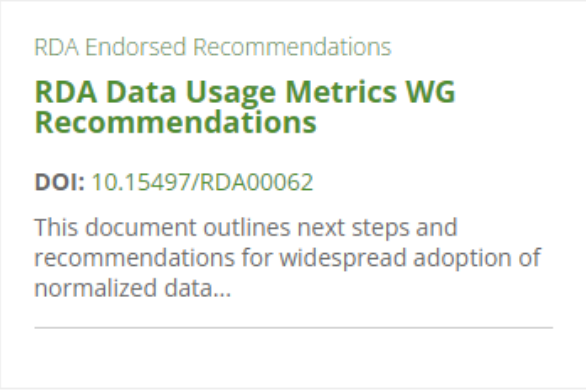
Exploration of Mitigations and Research Questions prompted by potential shortcomings of Data Usage Metrics presented at Helsinki RDA Plenary, 2019 by **Tobias Weber**.

https://vimeo.com/367997861 https://gitlab.lrz.de/RDA/data-usage-metrics-wg/-/blob/master/shortcomings.md

RDA Data Usage Metrics Working Group

Recommendations (2021):

- Lowenberg, D., Jouneau, T., & Bruno, I. (2021). RDA Data Usage Metrics WG Recommendations. Research Data Alliance. https://doi.org/10.15497/RDA00062.



Now officially endorsed!

Data Usage Metrics Recommendations (Condensed)

- **The community is interested in this topic**
  - It is complicated and there are activation barriers.
- **There are tools that can help that should be adopted where appropriate**
  - COUNTER Code of Practice for reporting views and downloads
  - Make Data Count recommendations
  - DataCite Usage Reports.
- **Data usage is nuanced** (defaulting to a “data impact factor” would be a really bad idea)
  - There is still work to be done to understand and overcome shortcomings
- **There should continue to be a Data Usage Metrics community within RDA**
  - It should work closely with other initiatives such as Scholix.

21 April 2022

MDC in the context of RDA Data Usage Metrics WG

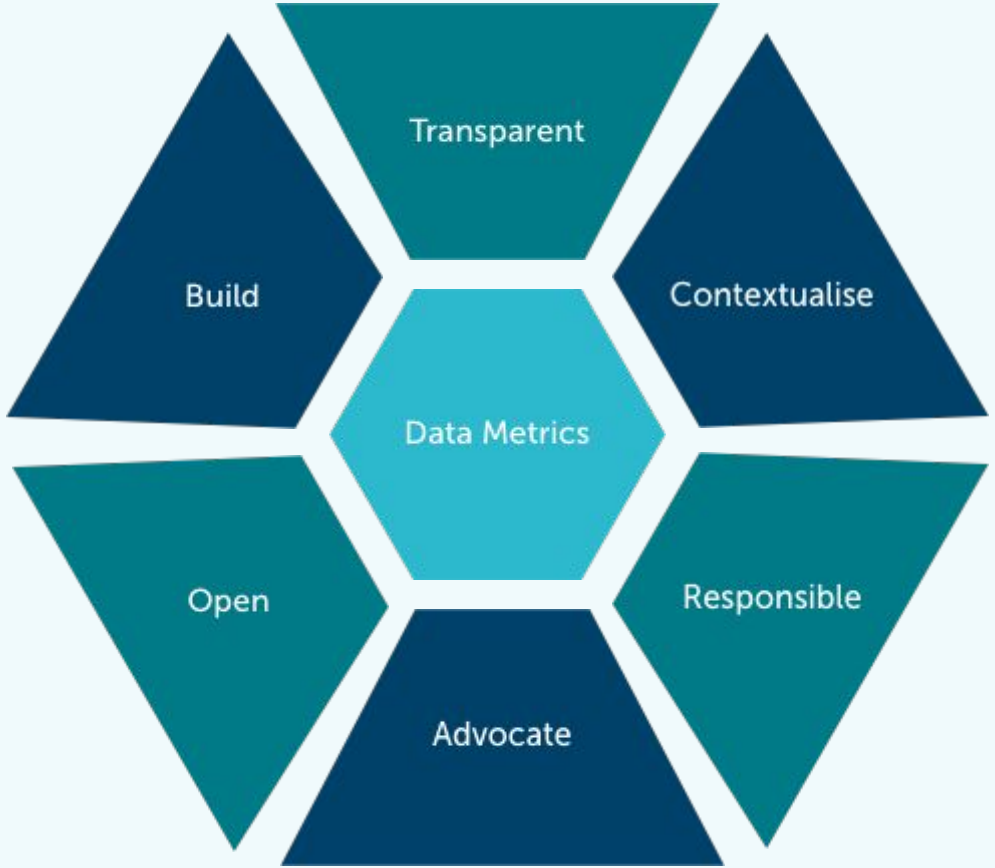
@makedatacount

What?
Make Data Count is a scholarly change movement committed to ensuring that the way data is used and cited is open, transparent, and responsible.

Who?
We are supported by a collective of organizations and individuals dedicated to the development of open data metrics.

How?
We build open infrastructure and community-based standards. We advocate through collaborative networks and adoption campaigns. We contextualize with evidence-based bibliometric studies.

Why?
We exist to support and inspire stakeholders across scholarly communications community to make assessment of research data a priority in policy and practice.



Make Data Count is an initiative

We are a collective of organizations and individuals invested in various components of data metrics

Scholix & COUNTER RD are standards

Scholix is a framework for submitting data citations to Crossref and COUNTER CoP RD is a standard for processing usage

MDC uses standards to develop data metrics

We advocate for and house best practices (above) and seek to be a driver of responsibly created open data metrics

This Much We Know is True

As a Community, We Have a Responsibility to Make Data Count.

1
Open data metrics are powerful measures for incentivizing the sharing of data that advances science for public good.

2
Having an open corpus of views, downloads, and citation counts is essential for infometric, scientometric, and bibliometric researchers to contextualize how datasets are used and reused across disciplines, career status, and other variables.

3
A comprehensive and responsible credit system for researchers must include open, fully auditable data metrics.

4
Normalization of data usage and data citation counts are the required first steps towards open data metrics.

5
Meaningful data metrics need to be based on bibliometric evidence and open community standards to avoid misguided, incomplete or biased metrics.

6
A true understanding of the investment, reach, and impact made in publicly accessible research data is only possible with open, transparent, and responsible data metrics.

Data Metrics is a Journey.

Where are we now?

Data Usage

Log processing service

DataCite building prototype, update at RDA Scholix/Usage joint session

COUNTER CoP v2 WG

Get in touch if you would like to participate or contribute

Complex metadata

Campaign for repositories to provide more consistent and thorough metadata

Perfect is the Enemy of Good

1

Basics are ready - start now

2

Work through complexities - evolve with space

3

Adoption now is essential to get to data metrics

Contextualizing Usage & Citation

Bibliometrics & infometrics

What needs to happen next?

Next steps to data citation metrics

→ **Make data citations the norm.**
Engage all stakeholders so that data is formally cited in references lists.

→ **Improve metadata.**
Get repositories to collect more and better metadata.

→ **Conduct mixed-methods research.**
Create evidence on data sharing and citation patterns across disciplines.

Let’s Make Data Count

Data Citation: Let’s Choose Adoption Over Perfection
https://doi.org/10.5281/zenodo.4701079

Open Data Metrics: Lighting the Fire
https://opendatametrics.org

Make Data Count
https://makedatacount.org

daniella.lowenberg@ucop.edu

@makedatacount

DataCite Usage Tracker

Connecting research, identifying knowledge

Matt Buys and Kristian Garza

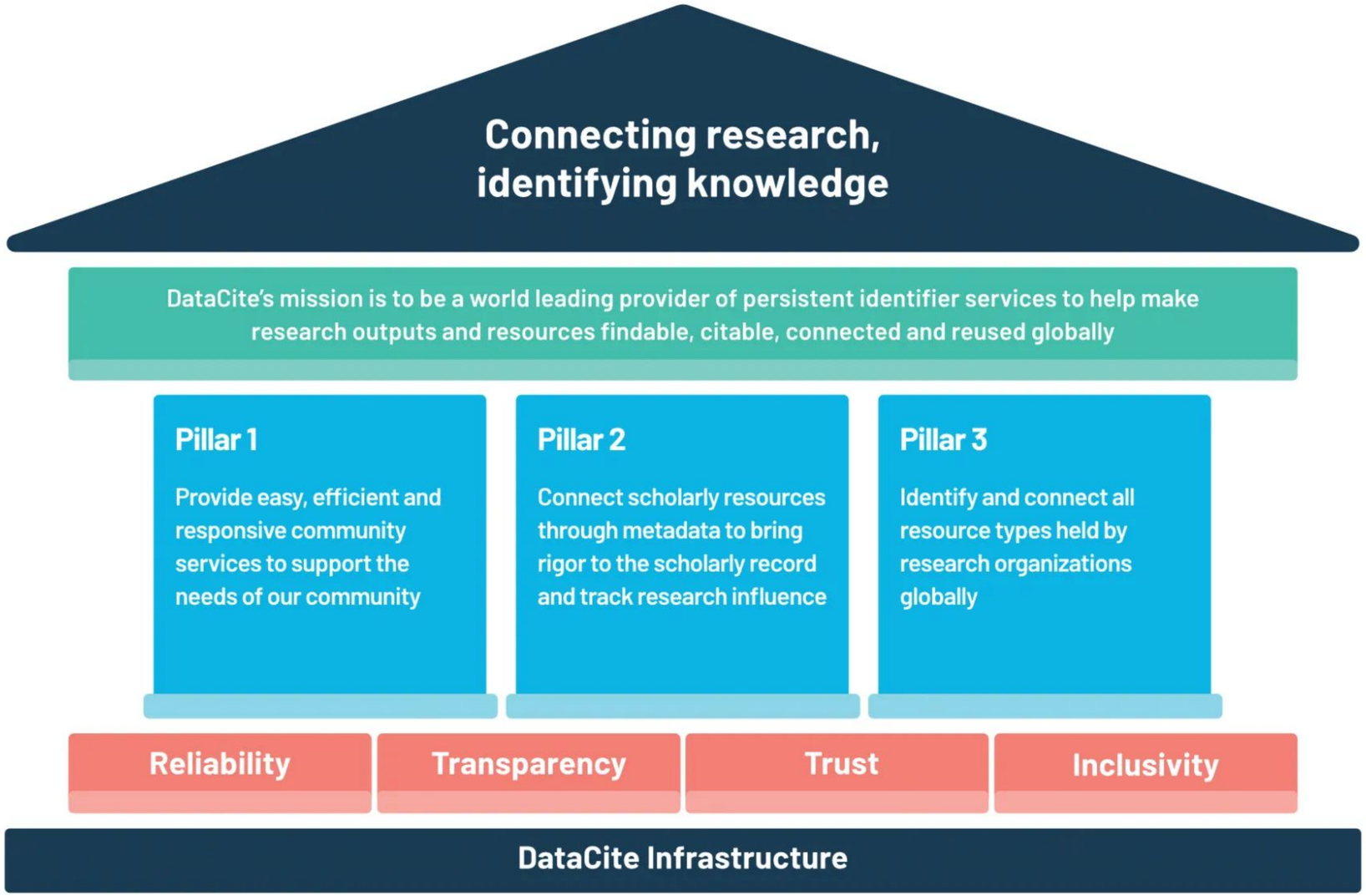
RDA Usage

April 21, 2022

DataCite
FIND, ACCESS, AND REUSE DATA

@datacite

Vision to Reality

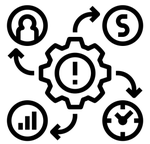


Activities supporting usage

Connect scholarly resources through metadata to bring rigor to the scholarly record and track research influence

Improving metadata quality, member insights, next schema version.

Enhancing PID Graph relational metadata, dashboards and visualizations.



Providing better statistics, data metrics and usage log processing.

Supporting FAIR implementation and workflows.

Enhancing discovery solutions.

Today: Processing data

Usage According to the CoP

4.5K
LoC

January

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

February

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28					

March

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

April

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

May

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

June

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

July

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

August

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

September

Su	Mo	Tu	We	Th	Fr	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

October

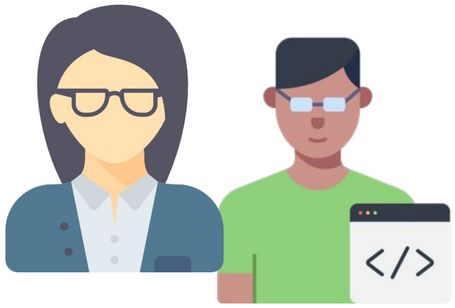
Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

November

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

December

Su	Mo	Tu	We	Th	Fr	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

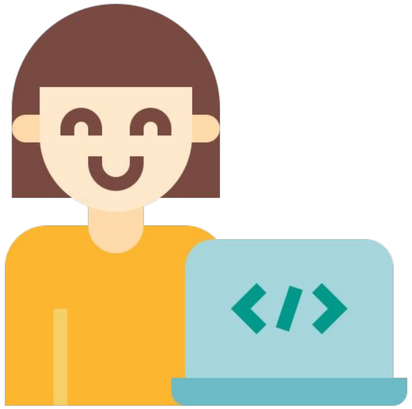


Usage Reporting Adoption

Adoption is the biggest challenge going forward for data usage metrics, and is the focus of the work in the MDC project.

What if?

>10
LoC



Directions for MDC Initiative

- Provide a hosted service
- Base on existing open source solution
- Use Javascript tracker, not log file processing
- Collecting repository usage stats using a Javascript tracker
- Never store any personally identifying information

What do you need to do?

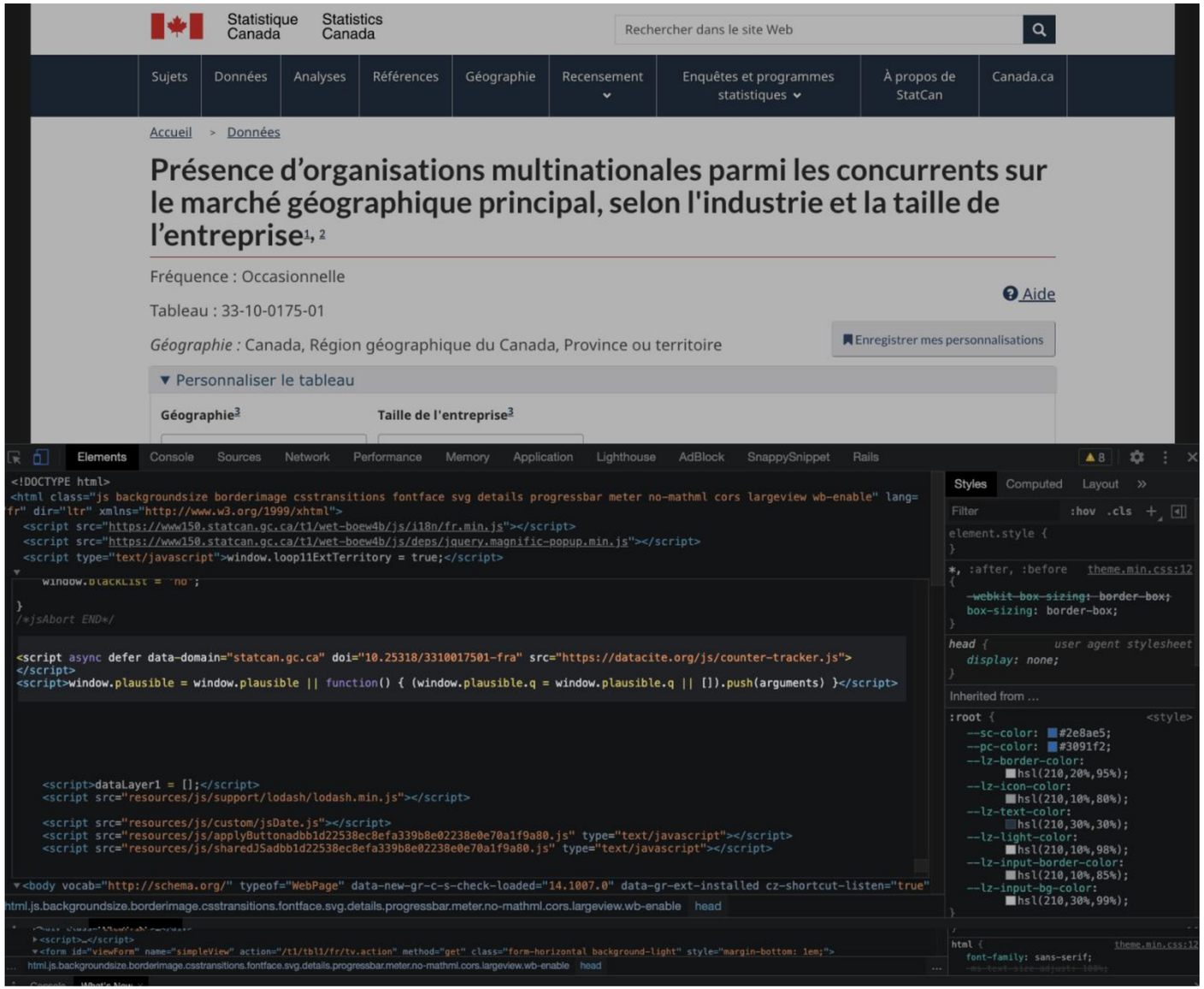
- Request/Enable usage capturing with Datacite
- Get your tracking ID.
- Add the following snippet (with your tracking ID and DOI) in the <head/> of landing pages.

```
<!-- Track View -->
<script defer data-doi="10.5072/1234"
        data-repoid="98d72415735ca2782a1"
        src="https://cdn.jsdelivr.net/npm/@datacite/datacite-tracker"></script>

<!-- Track Download -->
<script defer data-doi="10.5072/1234"
        data-repoid="98d72415735ca2782a1"
        data-metric="download"
        src="https://cdn.jsdelivr.net/npm/@datacite/datacite-tracker"></script>
```

Example

One just needs to paste the tracking code into DOI landing pages and put the DOI name in the tracker field.



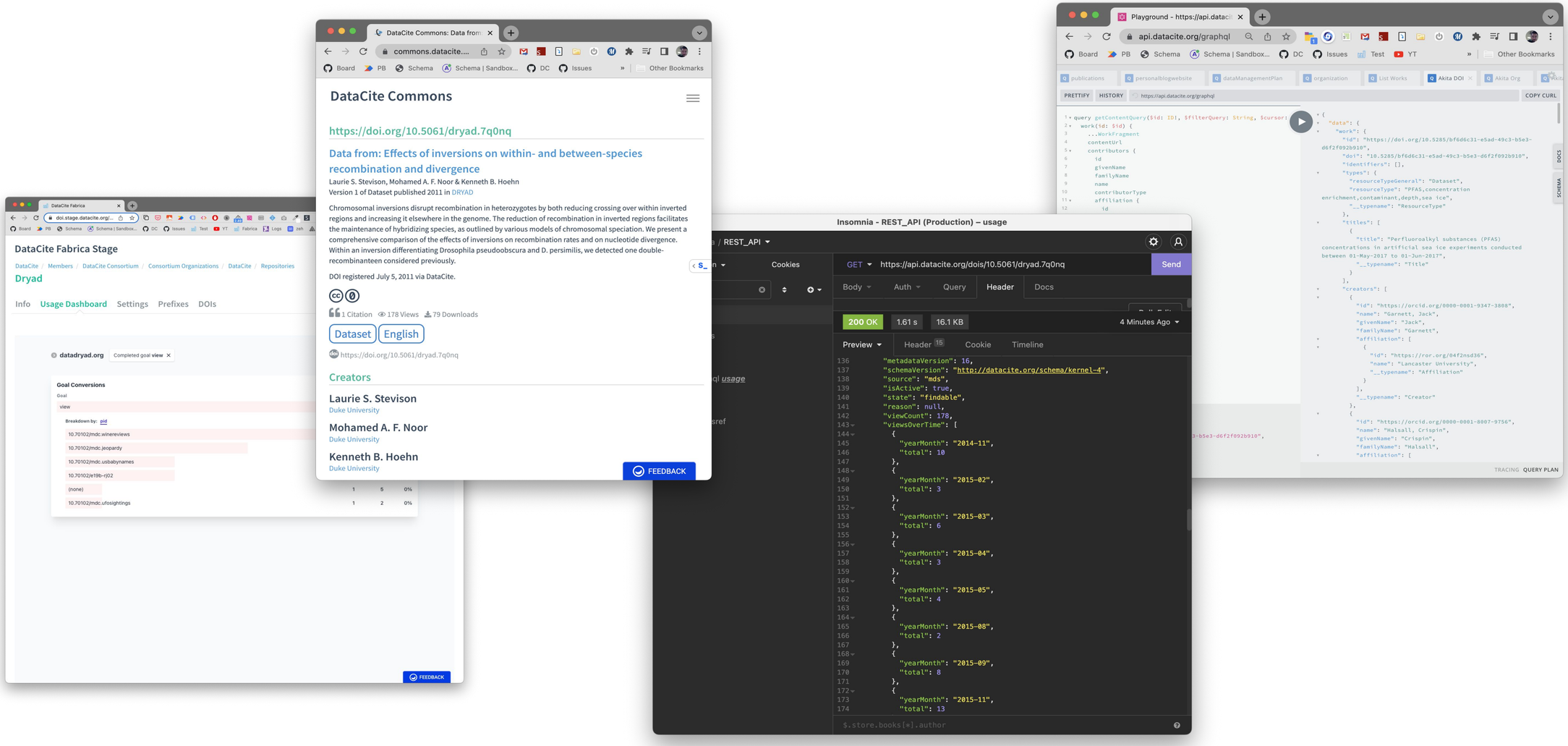
DublinCore or schema.org in your landing pages?

No need to add the DOI name. Our snippet will grab the DOI name automatically.

```
<!-- Track View -->
<script defer
        data-repoid="98d72415735ca2782a1"
        src="https://cdn.jsdelivr.net/npm/@datacite/datacite-tracker"></script>

<!-- Track Download -->
<script defer
        data-repoid="98d72415735ca2782a1"
        data-metric="download"
        src="https://cdn.jsdelivr.net/npm/@datacite/datacite-tracker"></script>
```

Accessing usage stats will be possible via many interfaces

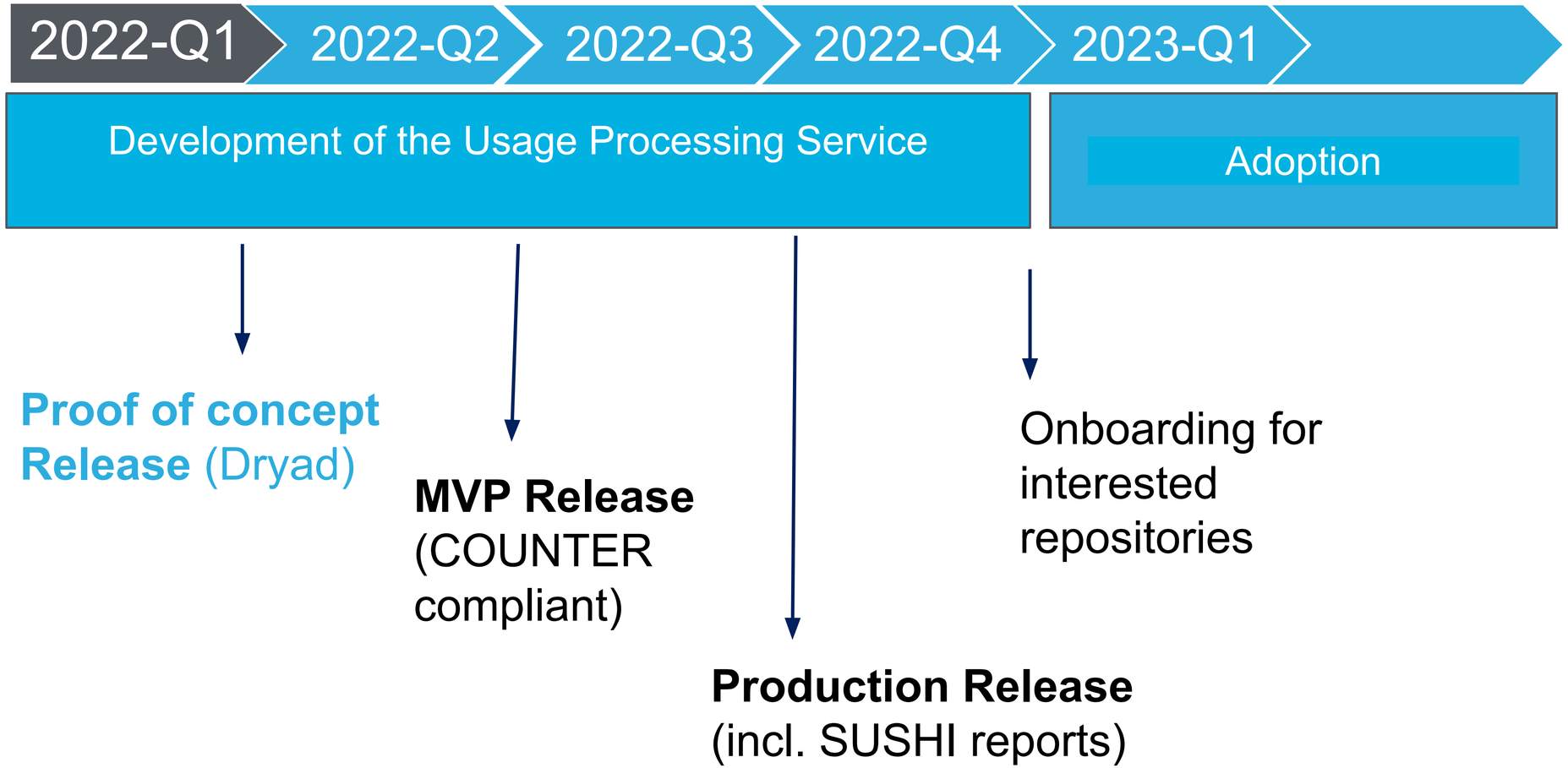


Usage Tracker: Proof of Concept

- It contains a “slice” of the full functionality.
- Currently planned for testing with Make Data Count Partners (Dryad)
- Usage counts will be only visible via Fabrica.

Usage Tracker: Next Steps

https://bit.ly/3uZiyVx



DataCite

FIND, ACCESS, AND REUSE DATA

Get in touch!

Email us:
info@datacite.org

Follow us:
@datacite

Talk to us:
pidforum.org

Read about us:
datacite.org

Get support:
support.datacite.org
support@datacite.org