



# Enabling Data Citation and Metrics

Working with the community, evolving the Elsevier infrastructure

Mike Taylor

# Part 1: Towards Metrics for Research Data

## A Short History of Research Data Citation

- 2014 – FORCE11 Data Citation Principles
- 2015 – Starr et al: “Achieving human and machine accessibility of cited data in scholarly publications”
- 2016 – Data Citation Pilot <https://www.force11.org/group/data-citation-implementation-pilot-dcip>
- Publishers Early Adopters - <https://www.force11.org/group/data-citation-implementation-pilot-dcip/eg3-dcip-publisher-early-adopters>

## A Short (But Fast) History of Research Data Metrics

- 2013-4 – Is there anybody out there?
- 2014-5 – Explosive growth in interest – RDA, FORCE, CASRAI, BioCADDIE, “Making Data Count”, NISO
- 2016 – NISO recommendations published

## Who are NISO, and what are their data metrics recommendations?

- “National Information Standards Organization”
- Focus on interoperability between information systems in the academic / publisher space
- Sloan grant to investigate potential standards for altmetrics
- [http://www.niso.org/topics/tl/altmetrics\\_initiative/](http://www.niso.org/topics/tl/altmetrics_initiative/)

## What are their data metrics recommendations and how did we reach them?

- Second phase initiated working group to look at “metrics for non-traditional outputs”
- Co-chair: Dr Kristi Holmes of NWU, Mike Taylor.
- Members include publishers, librarians, infrastructure specialists, domain experts, JISC
- Tightly co-ordinated efforts with CASRAI, RDA, etc
- Principle outcome is to recommend report dataset download usage by using COUNTER compliant formulations, and that funders support repositories to collect and report download metrics
- No link yet ☹️ will tweet in next week or two @herrison @tac\_NISO

## **Part 2: Evolving the Elsevier Infrastructure**

## Our Commitment to Research Data

*“Research Data is of the utmost importance for researchers in their advancement of knowledge. Therefore also the proper discovery and thus citation of such data is something that Elsevier fully supports – and will offer to researchers wherever possible”.*

**Philippe Terheggen**

Managing Director

Elsevier Journals

January 2016

## The two challenges: develop new product, integrate existing platforms

Existing platforms:

- Researcher submissions
- Journal data policies
- Publishing platforms (principally ScienceDirect) carries Data Journals and Articles
- Existing articles have ancillary data
- Data articles have citations, data articles are cited
- XML workflow, supplier management
- Data linking programme

New products:

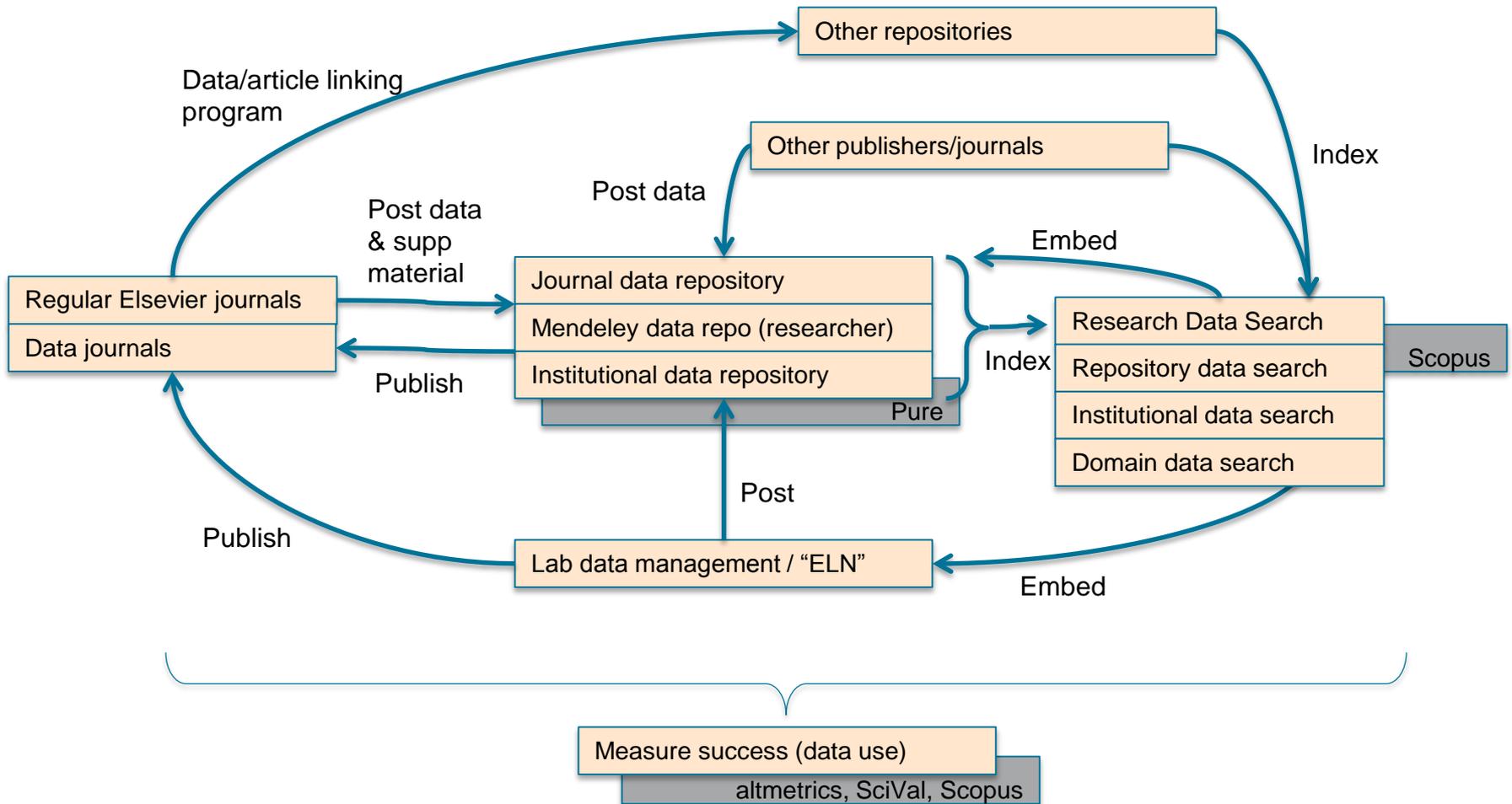
- Mendeley data repository
- Repository data search

## Building the research data infrastructure

Our ambitions for 2016:

- (Most) Elsevier journals to promote data publishing with data policies
- Our submission system to support data citations
- Our submission system to support data submissions
- Data objects correctly recognized throughout Elsevier infrastructure
- Data citations correctly identified throughout Elsevier infrastructure
- Data citations displayed correctly
- Plan to support DataCite DOIs throughout Elsevier platforms
- Plan to generate data metrics for Mendeley Data and Data Search
- Communication, community and socialization plans

# Our ambition for integration



## Data metrics and citation

Metrics create the value loop.

- Co-chair NISO Working Group on ‘non-traditional research outputs’
- Draft recommendation on data downloads will be released for public comment this month
- Final recommendations will be published this Summer
  
- Data metrics based on citation can’t happen until the infrastructure has been made
- Research Metrics team will address data citation metrics as soon as we are able!
  
- Draft recommendations on downloads are based around COUNTER recommendations
- Identifying was a ‘data object’ is a requirement (in a corpus of 60M documents in Scopus)

## A basket of metrics – there is no *one* metric

Our core values include commitments to working with the community, to using open formulations, and providing a selection of metrics for different research outputs.

- **Article level metrics** – citation counts, field-weighted relative citation index, scholarly activity, social activity, etc
- **Journal level metrics** – citation-based ranking, ‘PageRank’ type metric, new (simple) citation-based metric coming later this year, scholarly activity, social activity

## In Conclusions: we are not alone!

- We are here because we have an interest in research data sets
- FORCE11 are leading the conversation around data citation - <https://www.force11.org/group/data-citation-implementation-pilot-dcip> - groups for publishers, repositories, identifiers...
- NISO have recommendations in the download space (link to be supplied...)
- Working with the community to develop standards and best practice will accelerate acceptance of research data metrics
- A useful article (that I wrote) <https://www.elsevier.com/connect/data-citation-is-becoming-real-with-force11-and-elsevier>