

Outputs of the NISO Alternative Assessment Metrics Project

*A Recommended Practice of the
National Information Standards Organization*

NISO RP-25-2016 Alternative Assessment Metrics Project

About NISO Recommended Practices

A NISO Recommended Practice is a recommended "best practice" or "guideline" for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Published by

National Information Standards Organization (NISO)
3600 Clipper Mill Road
Suite 302
Baltimore, MD 21211
www.niso.org

Copyright © 2016 by the National Information Standards Organization

All rights reserved under International and Pan-American Copyright Conventions. For noncommercial purposes only, this publication may be reproduced or transmitted in any form or by any means without prior permission in writing from the publisher, provided it is reproduced accurately, the source of the material is identified, and the NISO copyright status is acknowledged. All inquiries regarding translations into other languages or commercial reproduction or distribution should be addressed to: NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211.

ISBN: 978-1-937522-71-1

Contents

Section 1: Altmetrics Definitions and Use Cases	1
1.1 Purpose and Scope.....	1
1.2 A Definition of Altmetrics	1
1.2.1 What is Altmetrics?	1
1.2.2 Scholarly Impact and the Role of Altmetrics in Research Evaluation	1
1.3 Main Use Cases.....	2
1.3.1 Persona #1: Librarians.....	3
1.3.2 Persona #2: Research Administrators	3
1.3.3 Persona #3: Member of a Hiring Committee	4
1.3.4 Persona #4: Member of a Funding Agency	4
1.3.5 Persona #5: Academics / Researchers	5
1.3.6 Persona #6: Publishers / Editors	5
1.3.7 Persona #7: Media Officers / Public Information Officers / Journalists	6
1.3.8 Persona #8: Content Platform Provider	6
Section 2: Alternative Outputs in Scholarly Communications	7
2.1 Background and Context	7
2.2 Alternative Scholarly Outputs.....	7
2.3 Implications for Future Research	7
Section 3: Study and Recommendations on Data Metrics	9
3.1 Summary.....	9
3.2 Key recommendations	9
3.3 Background and Context	10
3.4 Data Metrics Definitions	11
3.4.1 What is a Published Dataset?.....	11
3.4.2 What is Data Citation?	11
3.4.3 What is Data Usage?	11
3.4.4 What are altmetrics for data?.....	12
3.4.5 What are persistent identifiers for data?	12
3.5 Organizations Involved in Research Metrics	12
3.5.1 Method	12
3.5.2 Consulted bodies	12
3.6 Findings	13
3.7 Recommendations	14
3.7.1 Access to Research Data Metrics.....	14
3.7.2 Data Citation	15
3.7.3 Machine-actionable Persistent Identifiers	15
3.7.4 Required Metadata	15
3.7.5 Landing Pages	15
3.7.6 Reference Lists	15
3.7.7 Research Data Usage Statistics	15
3.7.8 Altmetrics for Datasets.....	16
Section 4: Persistent Identifiers in Scholarly Communications	17
4.1 Background and Context	17
4.2 Persistent Identifiers in Scholarly Communications Document	17
4.3 Implications for Future Research	18
Section 5: Altmetrics Data Quality Code of Conduct.....	19

NISO RP-25-2016 Alternative Assessment Metrics Project

5.1	Purpose and Scope.....	19
5.2	Data Quality Code of Conduct Terminology.....	19
5.3	Recommendations	19
5.3.1	Transparency	19
5.3.2	Replicability	20
5.3.3	Accuracy	20
5.4	Annual Report	20
Appendix A: NISO Altmetrics Working Group C "Data Quality" Code of Conduct Self-Reporting Table.....		21
Appendix B: NISO Altmetrics Working Group C "Data Quality" Code of Conduct Self-Reporting Table: Samples		24
Appendix C: Glossary		70
Appendix D: Combined Bibliography.....		72



Documents

Document Sharing Hub

Foreword

About this Recommended Practice

Altmetrics are increasingly being used and discussed as an expansion of the tools available for measuring the scholarly impact of research in the knowledge environment. The NISO Alternative Assessment Metrics Project was begun in July 2013 with funding from the Alfred P. Sloan Foundation to address several areas of limitations and gaps that hinder the broader adoption of altmetrics. This document, the output from the project, was created by three working groups.

- “Working Group A” extensively studied the altmetrics literature and other communications and discussed in depth various stakeholders' perspectives and requirements for these new evaluation measures.
- “Working Group B” created documents that are intended to help users better understand the landscape of data metrics and thus offer recommendations toward improvements, and to help organizations that wish to use altmetrics effectively communicate about them with each other and with those outside the community.
- “Working Group C” studied and discussed issues of data quality in the altmetrics realm, an essential aspect of evaluation before metrics can be used for research and practical purposes.

NISO Business Information Topic Committee

This recommended practice is part of the portfolio of the Business Information Topic Committee. At the time the Topic Committee approved this recommended practice for publication, the following individuals were committee members:

Yanick Beaudoin

International Development Research Centre

Greg Raschke

North Carolina State University Libraries

Anne Campbell (co-chair)

EBSCO Information Services

Angela Riggio

University of California, Los Angeles

Todd Carpenter

National Information Standards Organization
(NISO)

Siôn Romaine

University of Washington

Suzanne Hopkins

Thomson Reuters

Christine Stamison (co-chair)

NERL Program, Center for Research
Libraries

Nettie Lagace

National Information Standards Organization
(NISO)

Gavin Swanson

Cambridge University Press

Stuart Maxwell

Scholarly iQ

Elizabeth Winter

Georgia Tech University Libraries

NISO RP-25-2016 Alternative Assessment Metrics Project

NISO Altmetrics Initiative Working Group Members

The following individuals served on NISO Altmetrics Initiative Working Group A, responsible for the content of Section 1 of this recommended practice:

Rachel Borchardt
American University Library

Brigitte Jörg
Thomson Reuters

Robin Chin Roemer (co-chair)
University of Washington

Martha Kyrillidou
Principal, Martha Kyrillidou & Associates

Dianne Cmor
Nanyang Technological University

Jean Liu
Altmetric

Rodrigo Costas
Centre for Science and Technology Studies,
University of Leiden

Joshua Lupkin
Tulane University

Tracey DePellegrin
Genetics Society of America

Beth Martin
University of North Carolina Charlotte,
J. Murrey Atkins Library

Sharon Dyas-Correia
University of Toronto

Kim Mitchell
SAGE Publications

Martin Fenner
DataCite

Sharon Parkinson
Emerald Publishing Group

Karen Gutzman
Galter Health Sciences Library at
Northwestern University

Isabella Peters
Leibniz Information Centre for Economics

Michael Habib (co-chair)
Independent, scholarly communications,
publishing, library markets

Ginny Pittman
Wolters Kluwer

Kazuhiro Hayashi
NISTEP (National Institute of Science &
Technology Policy, Japan)

Sheila Yeh
University of Denver University Libraries

NISO RP-25-2016 Alternative Assessment Metrics Project

The following individuals served on NISO Altmetrics Initiative Working Group B, responsible for the content of Sections 2, 3 and 4 of this recommended practice:

David Baker
CASRAI

John Bell
Dartmouth College Library

Jane Burpee
McGill University

Frank Cervone
University of Illinois at Chicago

Jody DeRidder
University of Alabama Libraries

Martin Fenner
DataCite

Susan Heckethorn
Los Alamos National Laboratory

Kristi Holmes (co-chair)
Galter Health Sciences Library at
Northwestern University

David Kernohan
Jisc

Kim Mears
Augusta University

Ehsan Mohammadi
Galter Health Sciences Library at
Northwestern University

Fiona Murphy
Murphy Mitchell Consulting Ltd.

Joy Painter
California Institute of Technology Library

Gordon Rogers
Thomson Reuters

Jochen Schirrwagen
Bielefeld University

Mike Taylor (co-chair)
Elsevier

**Liaisons from CASRAI Research Dataset-
Level Metrics:**

Sunje Dallmeier-Tiessen
CERN

Rebecca Lawrence
Faculty of 1000

NISO RP-25-2016 Alternative Assessment Metrics Project

The following individuals served on NISO Altmetrics Initiative Working Group C, responsible for the content of Section 4 of this recommended practice:

Euan Adie
Altmetric

Scott Chamberlain
rOpenSci

Tilla Edmunds
Thomson Reuters

Martin Fenner
DataCite

Gregg Gordon
Social Science Research Network
(SSRN)

Kornelia Junge
John Wiley & Sons, Ltd.

Stefanie Haustein (co-chair)
Université de Montréal

Stuart Maxwell
Scholarly iQ

Angelia Ormiston
Johns Hopkins University Press

Maria Stanton
American Theological Library
Association (ATLA)

Greg Tananbaum (co-chair)
Scholarly Publishing and Academic
Resources Coalition (SPARC)

Joe Wass
Crossref

Zhiwu Xie
Virginia Tech University Libraries

Zohreh Zahedi
Centre for Science and Technology
Studies, University of Leiden

Acknowledgements

The Altmetrics Initiative Working Groups wish to acknowledge those outside the formal working group membership who contributed to this effort:

Working Group A: Mike Showalter, Plum Analytics

Working Group B: Karen Gutzman, Galter Health Sciences Library at Northwestern University

Trademarks, Services Marks

Wherever used in this standard, all terms that are trademarks or service marks are and remain the property of their respective owners.

Section 1: Altmetrics Definitions and Use Cases

1.1 Purpose and Scope

This section represents the output of the working group tasked with the following action items:

1. To come up with specific definitions for the terms commonly used in alternative assessment metrics, enabling different stakeholders to talk about the same thing.
2. To identify the main use cases for altmetrics and the stakeholder groups to which they are most relevant, and to develop a statement about the role of alternative assessment metrics in research evaluation.

1.2 A Definition of Altmetrics

Over the past several years, as altmetrics has developed as a practice, some terms and conventions have emerged among stakeholders. Below is a discussion of how various terms are understood by these practitioners.

1.2.1 What is Altmetrics?

Altmetrics is a broad term that encapsulates the collection of multiple digital indicators related to scholarly work. These indicators are derived from activity and engagement among diverse stakeholders and scholarly outputs in the research ecosystem, including the public sphere.

The inclusion in the definition of altmetrics of many different outputs and forms of engagement helps distinguish it from more established citation based metrics. At the same time, it leaves open the possibility of the complementary use of these conventional metrics, including for purposes of gauging scholarly impact. However, the development of altmetrics in the context of alternative assessment sets its measurements apart from conventional instances of citation-based scholarly assessment.

1.2.2 Scholarly Impact and the Role of Altmetrics in Research Evaluation

Scholarly impact is a concept based largely upon the values of different research stakeholders, and continues to evolve over time. Clarifying the concept of impact within the context of a given community is therefore important in avoiding misinterpretations of altmetrics. As such, to avoid being overly limiting, we focus on the current and potential uses for altmetrics, including its use in research evaluation.

The diversity of stakeholders, and the diverse ways of being impactful, makes a narrow definition of impact impractical. For stakeholders invested in conventional methods of scholarly communication, impact may be synonymous with citation based metrics, while for stakeholders with strong interests in societal change, such metrics may be inadequate indicators of impact. For stakeholders interested in the broad influence of scholarly outputs, altmetrics may offer insight into impact by calculating an output's reach, social relevance, and attention from a given community, which may include members of the public sphere.

Citations, usage, and altmetrics are all important and imperfect indicators of the values reflected by the term scholarly impact. Just as with conventional citation based assessments, it is inadvisable to use altmetrics as an uncritical proxy for scholarly impact because the attention paid to a research

output or the rate of its dissemination may be unclear until combined with qualitative information.

Additionally, it is important to recognize that data quality and indicator construction are key factors in the evaluation of specific altmetrics. Indicators that do not transparently conform to recommended standards are difficult to assess, and thus may be seen as less reliable for purposes of measuring influence or evaluation. Likewise, the manner in which indicators are presented within relevant tools has an important effect on the ability of users to evaluate altmetrics and on evaluating the usefulness of altmetrics. For example, if it is straightforward to benchmark the altmetrics of an article with those of other similar articles to help judge whether performance is strong, this contributes to the positive evaluation of altmetrics.

1.3 Main Use Cases

Use cases for altmetrics are driven by the different stakeholders in the research ecosystem, many of whom interact directly with one another, and some of whom overlap on an individual basis—as in the case of a researcher who is also a member of a hiring committee. The deployment of personas helps to highlight the different ways that these stakeholders collect, develop, and consume altmetrics, as well as the potential commonalities between altmetrics’ stakeholders’ needs, goals, and usages.

The following table presents some of the major use cases for altmetrics by describing eight primary stakeholder personas. There may be interrelationships between personas, as one person can serve many roles, and those roles can interact with each other. To further explain and contextualize the relationships between the stakeholders, each use case has been tagged according to three overarching themes.

1. **Showcase achievements:** Indicates stakeholder interest in highlighting the positive achievements garnered by one or more scholarly outputs.
2. **Research evaluation:** Indicates stakeholder interest in assessing the impact or reach of research.
3. **Discovery:** Indicates stakeholder interest in discovering or increasing the discoverability of scholarly outputs and/or researchers.

NISO RP-25-2016 Alternative Assessment Metrics Project

1.3.1 Persona #1: Librarians

Persona	Use case	Theme(s)
As a librarian, I want to...	Add value to my existing institutional repositories by encouraging researchers to deposit their works. For example, by creating a report to showcase frequency of views and downloads.	Showcase achievements
	Showcase the performance of my institution's scholarly outputs (or the outputs of a particular author).	Showcase achievements
	Increase authors' awareness of the scholarly and societal impacts of their scholarly outputs, as well as increase institutional awareness of such outputs in general.	Showcase achievements
	Monitor usage and decide to which journals and content my institution should subscribe.	Discovery Research evaluation
	Support both faculty and the university administration in their promotion and tenure exercises by offering a range of recognized impact-report services.	Showcase achievements Research evaluation
	Advise faculty/researchers on possible ways to improve upon the attention paid toward, and reach of, their work.	Showcase achievements Discovery

1.3.2 Persona #2: Research Administrators

Persona	Use case	Theme(s)
As a research administrator, I want to...	Showcase the achievements of my organization to other stakeholders. For example, I want to demonstrate the achievements of my institution's researchers to potential hires, students, collaborators, and other researchers.	Showcase achievements
	Support researchers in applying for competitive funding by effectively showcasing positive achievements of their research outputs.	Showcase achievements
	Gauge the performance and achievements of my institution's scholarly outputs.	Research evaluation

NISO RP-25-2016 Alternative Assessment Metrics Project

	Predict and determine the return on investment of my institution's research.	Research evaluation
	Compare/benchmark the performance and achievements of departments and/or groups within my institution.	Research evaluation
	Identify potential collaborators at other institutions with whom to partner with on grant applications and other projects.	Discovery

1.3.3 Persona #3: Member of a Hiring Committee

Persona	Use case	Theme(s)
As a member of a hiring committee, I want to...	Showcase my institution or organization in the best light to potential recruits.	Showcase achievements
	Evaluate potential employees and assess their achievements on the broadest range of aspects possible.	Research evaluation
	Identify new talent whom I may want to recruit.	Discovery

1.3.4 Persona #4: Member of a Funding Agency

Persona	Use case	Theme
As a member of a funding agency, I want to...	Evaluate the previous achievements of academics/researchers who are applying for funding.	Research evaluation
	Evaluate the broader impacts (attention drawn, engagement caused, or influence) of research that I funded. For example, citations in media reports are influential to policy makers and general public.	Research evaluation
	Identify trends in public interest or need, or new and emerging topics, so that I can decide in which research areas to invest.	Discovery
	Showcase the returns of investment of my organization to other stakeholders by, for example, - Demonstrating to the general public that their donations have been used appropriately and effectively. - Showing politicians and government bodies that their funding has been used appropriately and effectively.	Showcase achievements

NISO RP-25-2016 Alternative Assessment Metrics Project

1.3.5 Persona #5: Academics / Researchers

Persona	Use case	Theme(s)
As an academic or researcher, I want to...	Assess the reach, engagement, and influence of my own research outputs, by, for example, incorporating altmetrics in my portfolio to complement my other accomplishments.	Showcase achievements
	Assess the reach, engagement with, and influence of the research outputs of my peers, by, for example, writing a letter for the tenure of a researcher at another university.	Research evaluation
	Comply with reporting requests or mandates from funders, department heads, research administrators, etc.	Research evaluation
	Choose to publish in a journal that will provide the maximum exposure of my work to relevant audiences.	Discovery
	Choose to contribute to a publication whose metrics or qualitative data can be tracked to help me assess the reach, engagement with, and influence of my work.	Research evaluation Showcase achievements
	Discover influential research that is important/interesting in my field.	Discovery
	Identify potential collaborators and connections between research.	Discovery
	Discover where research is being discussed and potentially join into the conversation.	Discovery

1.3.6 Persona #6: Publishers / Editors

Persona	Use case	Theme(s)
As a publisher or editor, I want to...	Demonstrate the reach, engagement with, and influence of research published in my journal.	Showcase achievements
	Use insights from attention assessment and other metrics to help make editorial decisions about themes or topics upon which to focus.	Research evaluation
	Encourage authors to publish in my journal by providing them with information on attention, metrics, and other qualitative information about their research. For example, I	Showcase achievements

NISO RP-25-2016 Alternative Assessment Metrics Project

	want to encourage authors to publish in my journal by demonstrating the promotional efforts that can be made by my publication on behalf of the authors.	Research evaluation
	Identify general trends that the public is interested in so that I can decide what research areas to target in future publications.	Discovery

1.3.7 Persona #7: Media Officers / Public Information Officers / Journalists

Persona	Use case	Theme(s)
As a media officer, public information officer, or journalist, I want to...	Promote research that my institution or organization has produced, in order to maximize reach and engagement. For example, I want to encourage people to interact with a blog post that someone has written about a major research study.	Showcase achievements
	Determine whether my press campaigns about my institution's or publication's research output have been successful.	Showcase achievements
	Discover ways to enhance the exposure of my institution's or publication's research outputs.	Discovery Showcase achievements
	Identify popular and newsworthy papers or topics to cover.	Discovery

1.3.8 Persona #8: Content Platform Provider

Persona	Use case	Theme(s)
As a content platform provider, I want to...	Help readers to find content that is interesting, useful, and/or relevant to them by showing them the conversations about that content. For example, I want to offer sorting, filtering, limiting, etc. according to the attention given to that subject by various audiences, or according to the discussion generated by it on certain media platforms.	Discovery
	Help authors to see an aggregated view and analysis of all the attention, metrics, and qualitative information about their research.	Showcase achievements

Section 2: Alternative Outputs in Scholarly Communications

2.1 Background and Context

The NISO Persistent Identifiers and Alternative Outputs Working Group investigated alternative research outputs and application of persistent identifiers in the scholarly environment in order to clearly identify and track research outputs and their relationships. The composition of the Working Group was developed to represent a balance of parties that are interested in unconventional research outputs and identifiers, with just under half of the members coming from academic institutions and libraries. Publishers and other organizations in the scholarly ecosystem make up the majority of the rest of the group's membership. The organizations represented are drawn from bodies that have an interest in research output metrics.

2.2 Alternative Scholarly Outputs

The NISO Scholarly Outputs table (see Google document at <https://sites.google.com/a/niso.org/scholarlyoutputs/>) is a first attempt at a comprehensive list of research outputs, including the traditional academic publication and extending to more alternative outputs. These outputs may fall within the scope of assessment when developing metrics to evaluate the impact of scholarly activity, with the acknowledgement that meaningful impact can go far beyond conventional publishing workflows and often involves the rich array of scholarly products that are created during the research process. These output types are grouped by class and alphabetized with a brief description and documentation of known current efforts (and by whom they are being undertaken). Relevant links are listed where available, and most entries have been assigned a focus area to group them by similar contextual uses. The focus areas are: Basic Sciences; Capacity; Code and Software; Communications; Data; Education and Training Materials; Events; Grey Literature; Images, Diagrams, and Video; Industry; Instruments, Devices, and Inventions; Methodologies, Publications; Regulatory, Compliance, and Legislation; Standards; and Other.

This work is not complete, as the very nature of scholarly activity continually evolves. However, the rich array of outputs represented in this table help to better establish the breadth and depth of scholarly work that may be produced by an investigator or by a research team. Through this effort, the working group hopes to generate discussion about how we may begin to leverage integrated data, persistent identifiers, and automated workflows to better capture and track the full complement of research activity, as is possible for publication data.

2.3 Implications for Future Research

Future work in this area requires a more comprehensive inventory of output types and the work presented here can serve as a springboard for these efforts to gain their perspectives about research outputs of interest (e.g., in the UK and Australia). This future comprehensive inventory should include additional stakeholders, including funders from countries where research assessment exercises are underway. Other key areas of work include the integration of various research assessment frameworks¹; a formal assessment of the extent of nontraditional research output types not

¹ Graham, K.E.R., H.L. Chorzempa, P.A. Valentine, & J. Magnan. "Evaluating Health Research Impact: Development and Implementation of the Alberta Innovates – Health Solutions Impact Framework." *Research Evaluation* 2012; 21(5):354-67. doi: 10.1093/reseval/rvs027.

yet managed with persistent identifiers, but necessary for long-term management of access and relationships; and creation of a priority list for incorporating these output types into existing information systems across the research spectrum and workflow. These activities could help guide development of comprehensive alternative metrics measures and methodologies reaching far beyond the traditional academic publication.

-
- Manville C., S. Guthrie, M.-L. Henham, B. Garrod, S. Sousa, A. Kirtley, S. Castle-Clarke, & T. Ling. "Assessing impact submissions for REF2014: An evaluation." Product Page. Santa Monica, CA: RAND Corporation, 2015 April 17. Report No.: RR-1032-HEFCE.
- Morgan Jones, M. & J. Grant. "Making the Grade: Methodologies for Assessing and Evidencing Research Impact." In: Dean A, Wykes M, Stevens H, editors. *7 Essays on Impact*. Exeter, UK: University of Exeter; 2013.
- Panel on Return on Investment in Health Research, 2009. *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. Canadian Academy of Health Sciences, Ottawa, ON, Canada. http://www.caahs-acss.ca/wp-content/uploads/2011/09/ROI_FullReport.pdf
- Sarli, C.C., E.K. Dubinsky, & K.L. Holmes. "Beyond Citation Analysis: A Model for Assessment of Research Impact." *Journal of the American Medical Library Association* 98, no. 1 (2010):17-23. doi:10.3163/1536-5050.98.1.008. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801963/>

Section 3: Study and Recommendations on Data Metrics

3.1 Summary

Working Group B, formed to address work areas related to unconventional research outputs and identifiers, first focused efforts upon a better understanding of data metrics. This focus is especially timely given the emphasis placed upon open science approaches, as well as mandates by funders to make research products openly accessible. This section of the Recommended Practice describes the current state of data citations, as determined through a series of conversations with key stakeholder organizations such as CASRAI, DataCite, JISC, and others; representatives from other ongoing efforts; and within the working group itself. Section 3 also presents a set of recommendations for data metrics, directed toward the spectrum of groups working with research data, including institutions and repository managers, international research organizations, and funders.

3.2 Key recommendations

- Metrics on research data should be made available as widely as possible.
- Data citations should be implemented following the Force11 *Joint Declaration of Data Citation Principles* (<https://www.force11.org/group/joint-declaration-data-citation-principles-final>), in particular:
 - Use machine-actionable persistent identifiers
 - Provide metadata required for a citation
 - Provide a landing page
 - Data citations should go into the reference list or similar metadata.
- Standards for research-data-use statistics need to be developed. They should be based on the COUNTER Code of Practice (<https://www.projectcounter.org/code-of-practice-sections/general-information/>) but should also take into consideration some special aspects of research data usage. There should be two formulations for data download metrics, to examine both “human” downloads and research-focused non-human agents.
- Research funders should provide mechanisms to support data repositories in implementing standards for interoperability and obtaining metrics.
- Data discovery and sharing platforms should support and monitor “streaming” access to data via API queries.

There have been recent attempts to define and identify large-scale data queries, but there is, as yet, no consensus in this area. This form of data is therefore explicitly excluded from these recommendations. For more on this group’s recommendations, see “Findings” (3.6).

3.3 Background and Context

Research data is now understood to be a primary goal of academic endeavor. This section of the Recommended Practice reports on the state of metrics relating to published research data, and makes recommendations to those stakeholders operating in this area of work.

Publishing datasets is an integral part of scholarly communication, and libraries need to consider digital datasets alongside journal articles and other resources. However, it has been reported that data are often used in publications without being properly cited. One of the reasons for lack of data citation in the literature is the lack of standards for data publication. In order to increase the visibility of datasets, the international consortium known as DataCite (<https://www.datacite.org>) was created. It allocates unique digital object identifiers (DOIs) and metadata for digital and physical objects with a focus on research data. The unique identifiers and associated metadata provided by DataCite should promote a culture of reusing data as there is a significant correlation between data documentation quality and data reuse satisfaction. However, the absence of widely accepted rules for citing data as separate academic artifacts has been identified as a significant reason behind the lack of research data citations. In response to the challenge, in 2014 FORCE11: The Future of Research Communication and e-Scholarship and others published the *Joint Declaration of Data Citation Principles* with the aim of increasing data citation adoption.

Researchers have been attempting to measure the research impact of datasets for the last two decades. In recent years, scientometricians, researchers in various domains, and funding organizations have explored ways to capture the metrics around datasets to use as evidence for identifying the research impact of data. The main motivation for capturing data citations is to credit scientists and data publishers for their contribution in creating, managing, and curating research data, and to provide evidence of reuse to funders.

Linking datasets to other forms of publications and growing transparency in scholarly communication ecosystem are additional major reasons for measuring metrics around datasets. Although measuring impacts of research data is complicated, some efforts have started to identify data citations. From a conventional citation based perspective, Thomson Reuters launched the Data Citation Index (http://wokinfo.com/products_tools/multidisciplinary/dci) in 2012, which indexes datasets and their citations from main repositories across different disciplines. Using statistics for searches, viewing, and downloads, Ingwersen and Chavan² suggested a Data Usage Index, which could reveal the impact of datasets from novel points of view. In the same way, Fear³ came to a conclusion that the impact of scholarly datasets cannot be measured through a single indicator, and as a result, suggested multiple metrics for measuring the value of datasets. Strasser, Kratz, and Lin⁴ reported that data citation was still underused, and that the second most valued metric after data citation would be derived from repository download data.

²Ingwersen, P., and V. Chavan. "Indicators for the Data Usage Index (DUI): An Incentive for Publishing Primary Biodiversity Data Through Global Information Infrastructure. BMC Bioinformatics 12 (Suppl 15), S3. (2011). <http://doi.org/10.1186/1471-2105-12-S15-S3>

³Fear, K. M. "Measuring and anticipating the impact of data reuse." University of Michigan. (2013). <http://deepblue.lib.umich.edu/handle/2027.42/102481>

⁴Strasser, C., John Ernest Kratz, and Jennifer Lin "Make Data Count - Unit 1 Final Report." (2015). Figshare. <http://dx.doi.org/10.6084/m9.figshare.1328291>

3.4 Data Metrics Definitions

3.4.1 What is a Published Dataset?

In the context of dataset metrics, a published dataset is considered a managed collection of data that has been made available for access and/or download (<http://www.w3.org/TR/vocab-dcat/#class-dataset>). It has ideally been assigned a persistent identifier and is described by metadata: "...a dataset may be generated as part of some scientific investigation, whether tabulated from observations, generated by an instrument, obtained via analysis, created through a mashup, or enhanced or changed in some manner" (<http://www.w3.org/TR/hcls-dataset>). Datasets are not necessarily 'collections of numbers': different disciplines have different investigatory methodologies. In this case, a dataset might be a transcript of an interview, a set of digitized images or some other artefact of the research process. See Section 2, "Alternative Outputs in Scholarly Communications."

3.4.2 What is Data Citation?

Data citation is the practice of providing a formal, structured reference in a scholarly work to a dataset, realized as citation to the full bibliographic reference information of the dataset. By way of contrast is the issue of granular citation, e.g., the citation of some subset of a dataset, whether a data subcomponent or a particular value (<http://www.codata.org/task-groups/data-citation-standards-and-practices>).

In 2014, the Joint Declaration of Data Citation Principles was released by the Data Citation Synthesis Group within FORCE11 (<https://www.force11.org/group/joint-declaration-data-citation-principles-final>). These guiding principles emphasize the role of research data in scientific communication by enabling citation to facilitate credit and attribution for contributions to data generation, and support data reuse. While more and more organizations and individuals endorse this declaration, it lacks practical guidance for implementation.

In order to create guidance for those who wish to implement the Joint Declaration of Data Citation Principles, the Force11 Data Citation Implementation Group (<https://www.force11.org/group/data-citation-implementation-group>) was formed and is working on topics such as:

- machine-readable data citations in the JATS XML schema for journal articles
- a list of identifier schemes conformant to the Joint Declaration
- use cases of publishing workflows for data deposition, identifier validation, and citation

3.4.3 What is Data Usage?

Data usage refers to the act of accessing and downloading a published research dataset (https://rd-alliance.org/sites/default/files/case_statement/RDA_WDS_WG_Publishing_Bibliometrics.pdf). COUNTER is the standard for defining usage of networked electronic resources, including journals, databases, books, and reference works. A corresponding standard for datasets does not yet exist. Many aspects of the COUNTER Code of Practice could also be applied to research data, as was done in the Making Data Count project for about 150,000 research datasets (<http://mdc.lagotto.io>).

The COUNTER Code of Practice explicitly ignores data derived from non-human sources, whether pieces of software that harvest content automatically, or search-engine robots. As non-human agents become more prevalent in the reuse of scientific data, this aspect has to be considered for data usage statistics, an important step beyond the existing COUNTER Code of Practice. For example, a researcher could develop an application that automatically downloads a fresh copy of a dataset. It is important to note the value of this form of access and not to automatically discount it. Furthermore, as APIs become more commonplace, access to research datasets may become more fragmented and more frequent over time, and metrics should be developed to recognize these trends. The standard

counters the issue of repeated, failed attempts at downloads by treating repeated download activities within a certain time frame as single events.

Another key difference in properties between research datasets and text documents such as journal articles is that data more frequently change over time. It is therefore important to specify the version of a dataset that led to a particular result. Datasets are also often made available in multiple parts or files, making it harder to aggregate usage statistics.

Strasser, Kratz, & Lin explained that many repositories were unable to report any data relating to download activity, and that the development and adoption of robust reporting tools and standards, including the ANSI/NISO standard SUSHI and NISO technical document SUSHI-lite (http://www.niso.org/workrooms/sushi/sushi_lite/), are needed for adoption.

3.4.4 What are altmetrics for data?

On the surface, altmetrics for data are conceptually and technically not really different from altmetrics for journal articles. Research data can accumulate similar metrics to other types of scholarly outputs (i.e., mentions of data in a blog post or tweets that share an article can be retrieved using the same mechanisms for both types of outputs).

There are a number of data-specific altmetrics that exist, as well. Output-specific altmetrics that can get at data reuse and experimentation (such as GitHub forks) and collaboration patterns (GitHub collaborators) already exist, and it's possible that newer data altmetrics may soon be realized (for example, a data dependency page rank akin to Impactstory's Depsy software dependency page rank, which could identify the extent to which research data is reused/remixed in other influential datasets and studies, or text-mining for direct links to datasets that appear in journal articles, as Depsy does for software).

3.4.5 What are persistent identifiers for data?

Identifiers play an important part in the development of metrics, whether they relate to an article (e.g., a Crossref DOI), a dataset (i.e., a DataCite DOI), or an individual (e.g., an ORCID).

We recommend that attention be paid to the question of identifying a specific subset or components of a dataset. This scenario is common, and is a particular challenge for data that are continuously growing (e.g., time series data) or otherwise dynamic. For example, both DataCite and Crossref have established methods for identifying components within the documents identified by a single DOI, for example data components within a single published data set. These should become interoperable (or ideally a single agreed standard) as the field matures. See Section 4, “Persistent Identifiers in Scholarly Communications.”

3.5 Organizations Involved in Research Metrics

3.5.1 Method

The approach of Working Group B was to investigate existing best practices and standards that might exist in the area of unconventional research metrics. The working group therefore researched related organizations, inviting some to meetings of the working group or seeking other representation.

3.5.2 Consulted bodies

The organizations involved in the consultation process were:

- **COUNTER** (<http://www.projectcounter.org/>). Formed in 2002 to enable standards in reporting journals and database downloads, COUNTER is a UK-based international organization. It is supported by vendors, intermediaries, and librarians.

- **RDA** (<http://www.rd-alliance.org>). The Research Data Alliance exists to build the social and technical bridges needed to facilitate the open sharing of data.
- **CASRAI** (<http://casrai.org>). The Consortia Advancing Standards in Research Administration Information is an international organization dedicated to improving the flow of information among research stakeholders.
- **Jisc** (<http://www.jisc.ac.uk>). Jisc is a UK-based organization that provides infrastructure and other support to UK higher-education establishments. Jisc continues to be active in the area of data repositories and data sharing more generally.
- **DataCite** (<http://www.datacite.org>). DataCite manages DOIs and metadata, with a focus on research data.
- **BioCADDIE** (<http://biocaddie.org>). BioCADDIE is a project that aims to build a data-discovery index of research data to enhance the discovery and use of such data.
- **California Digital Libraries** (<http://www.cdlib.org>), **PLOS** (<https://www.plos.org>), and **DataONE** (<https://www.dataone.org>), partners in Making Data Count (<http://mdc.lagotto.io/>)

3.6 Findings

The issues surrounding data citation, and the recommendations for how to deal with it going forward, have been well described in the *Joint Declaration of Data Citation Principles*, and in the implementation recommendations by Starr et al.⁵. Our position is to recommend these developing standards and to encourage the promotion of data citation and the citability of research datasets.

The COUNTER Code of Practice provides a set of well-recognized recommendations to count the download of research articles. COUNTER recommendations include such parameters as excluding non-human downloads (recognized by the header information), and eliminating fast repetition of downloads from the same machine. Putting forward such parameters is done in order to reduce the number of downloads counted that originate from automated web crawlers, or from attempts to artificially increase the number of downloads. However, we recognize that in the case of data downloads, there will be a number of legitimate use-cases in which machines will be downloading datasets; for example, via the CURL command-line utility, or via file-handling wrappers. Furthermore, it may be useful to make a distinction between useful web crawlers, such as products that index and search research datasets, and the ordinary web-link-following crawlers.

The RDA Publishing Data Bibliometrics Working Group is part of the RDA/WDS Publishing Data Interest Group. It liaises with other RDA working groups, such as the Publishing Data Services Working Group, which founded the DLISERVICE data-literature interlinking service (<http://dliservice.research-infrastructures.eu/>). Although this group has not published any recommendations or standards, Callaghan, Carpenter, & Kratz⁶ published the results of a survey conducted in 2014 that indicated that lack of standards in counting downloads was a key obstacle in the use of the data.

⁵ Starr J. et al. Achieving Human and Machine Accessibility of Cited Data in Scholarly Publications. *PeerJ Computer Science* 1 (1). (2015). <http://doi.org/10.7717/peerj-cs.1>

⁶ Callaghan, S, T. Carpenter, and J. E. Kratz (2015). "Walk Softly and Carry a Large Carrot: How To Give Credit for Academic Work." In: FORCE2015, 12-13 January 2015, Oxford, UK. http://cedadocs.badc.rl.ac.uk/1125/1/FORCE2015_poster_final.pdf

CASRAI has a related Data Level Metrics working group (http://ref.casrai.org/Research_Dataset-Level_Metrics). After discussions with the NISO working group, it was decided that the two groups would maintain a number of bilateral relationships, and that CASRAI would focus on data quality indicators, rather than data usage and sharing.

Jisc, as a component of its “Research at Risk” activity to support good research data management, is piloting an implementation of the existing IRUS-UK service (<http://www.irus.mimas.ac.uk/>) that will report on, at a file level, downloads of material in institutional and subject area research data repositories. IRUS (and the experimental IRUSdataUK, <http://rdmetrics.jiscinvolve.org/wp/2016/01/11/a-note-about-irusdatauk/>) is an implementation of the COUNTER standard, which can be plugged into nearly all common repository software.

BioCADDIE has an interest in research data metrics. Rather than start a new discussion, co-chairs of the relevant BioCADDIE workgroup have agreed to participate with the existing projects, namely NISO, CASRAI, THOR, and Making Data Count.

Strasser, Kratz, & Lin (2015) surveyed 247 interested parties. They concluded that although citations are the key measure of data impact, few data repositories have implemented relevant standards and working practices; however, most do count data downloads. The authors reported that while data citation is regarded by researchers as the prime way of recognizing data use, consistency and adoption have been weak spots. It is these areas that the FORCE11 declarations hope to address. Despite interest in alternative (social sharing) metrics for publications, there is little interest in this area for data at present. However, the authors report that download counts are considered valuable, and that these are used by dataset creators who collect them.

Having had discussions with people involved in hosting very large streaming data, members of the NISO working group wish to recognize that data metrics of the sort described in this document refer to discrete objects of published research datasets, rather than unprocessed streaming data of the type created, managed, and served by large data centers.

3.7 Recommendations

Researchers, funders, institutions, data repositories, and publishers increasingly expect research data to be made publicly available in dedicated data repositories, and to have access to tools and platforms that support them in understanding the ways in which their data is being used. With the significant and global interest in research data and research data reuse has come the need to more formally assess this data reuse, and to develop best practices for research data metrics.

Metrics for research data must be seen in the bigger context of bibliometrics for other research outputs. Many of same principles that apply to metrics for research articles, books, or book chapters can also be applied to research data.

These recommendations, therefore, are primarily addressed to institutions and repository managers, international bodies working with research data, and funders.

3.7.1 Access to Research Data Metrics

Just as it is now generally required that research data be made available by default (<https://www.force11.org/group/joint-declaration-data-citation-principles-final>), data about the use of this data should be created and maintained in such a way that it can be made available as widely as possible.

3.7.2 Data Citation

The *Joint Declaration of Data Citation Principles* (JDDCP, <https://www.force11.org/group/joint-declaration-data-citation-principles-final>) provides important principles that should be followed for data citation. Work on implementing these data citation principles is ongoing, e.g., in the FORCE11 Data Citation Implementation Pilot (DCIP, <https://www.force11.org/group/data-citation-implementation-pilot-dcip>), the RDA/WDS Publishing Data Interest Group, and the COPDESS project (<http://www.copdess.org/>). The following general recommendations are common to these activities.

3.7.3 Machine-actionable Persistent Identifiers

Data citation requires “a persistent method for identification that is machine actionable, globally unique, and widely used by a community”⁷. In some communities, such as the life sciences, it is common to use identifiers local to a particular system that are not machine actionable across systems. Work is ongoing—for example, in the DCIP project—to provide globally unique, machine-actionable identifiers to these communities, e.g., via the resolving services identifiers.org (<http://identifiers.org>) or n2t.net (<http://n2t.net>).

3.7.4 Required Metadata

The minimal metadata required for a data citation (e.g., persistent identifier, authors, title, publisher, publication date) should be made available by data repositories. This should be done via the landing page for the dataset in human and machine-readable format, e.g., using Dublin Core and PRISM HTML meta tags similar to those used when observing best practices for journal articles (<https://scholar.google.de/intl/en/scholar/inclusion.html#indexing>). The required metadata should also be registered with the persistent identifier when the persistent identifier supports metadata, e.g., when using a DOI.

3.7.5 Landing Pages

Following the recommendations by Starr et al, the identifier included in a citation should point to a landing page or set of pages rather than to the data itself. The landing page should persist even if the data is no longer accessible.

3.7.6 Reference Lists

Data citations should be included in the metadata of the citing resource. For many textual resources, such as journal articles and book chapters, a natural place for these data citations would be the reference list. The current practice of citing data in the text without additional metadata limits adoption of data citation because a) these data citations might not be available where access to the full text is restricted, and b) they can be difficult to find in the text when using automated text mining.

3.7.7 Research Data Usage Statistics

Research data sharing platforms, including data repositories, should work on standards and best practices for collecting and reporting standard usage statistics for research data. This work should be based on the COUNTER code of practice and on the lessons learned with usage statistics for journal

⁷ <https://www.force11.org/group/joint-declaration-data-citation-principles-final>, principle 4 (unique identification)

articles, but should take into account the special considerations that apply to research data. The Research Data Alliance (RDA), COUNTER, and NISO are important organizations that should also be involved in this standardization work.

Formulations should recognize the rise of non-human download agents used in research, and report on activity appropriately. We suggest that there be two formulations: one examining only “human” downloads, and another including downloads identified as being from genuinely research-focused non-human agents. As currently happens, downloads from known nonconstructive and vexatious agents should be excluded from statistical reports.

By default, repositories should use standard persistent identifiers, and record download counts, header information, and deposit and creation times. These parameters are required to support COUNTER-compliance, and represent what should become standard practice in research data management. Consistent with the recommendations for data citation, machine-actionable persistent identifiers and landing pages should be used for research data usage statistics.

As research funders begin to standardize approaches to research data management support, grants and other assistance mechanisms should be made available to support data repositories in implementing standards for research data usage statistics.

3.7.8 Altmetrics for Datasets

There currently seems to be a lack of interest in altmetrics for data in the community. This matches trends we see for altmetrics for journal articles, as wellⁱ. However, there may be those in the community that do find altmetrics data important to their work (as we have seen for altmetrics for other types of scholarly outputs).

Studies of altmetrics for data have been limited. We encourage more investigation into existing altmetrics for data and for as-yet-undetermined metrics, as well. One altmetrics provider, Altmetric LLP, has offered its data for analysis—approximately 0.3 percent (~15,500) of the > 6 million outputs Altmetric tracks are datasets from the figshare, Dryad, Zenodo, and Pangea data repositories. Analysis of data from other providers that track data-specific metrics like those mentioned above (Section 3.4) could yield further insights.

Section 4: Persistent Identifiers in Scholarly Communications

4.1 Background and Context

The use of Persistent Identifiers has increased as scholarly communications have become increasingly digital. In particular, use of the Digital Object Identifier (DOI), which began in 2000, has become pervasive when identifying scholarly articles, and latterly research data and alternative research products. In common with other persistent identifiers, the DOI Foundation obliges certain metadata and linking technologies to be adopted by its users. The consequence of using semantic and other standard approaches is that open infrastructures can be built to support interoperability and to support open science. For example, the ORCID organization, DataCite, and Crossref are able to map documents, data, and researchers using DOIs and ORCIDs.

Recently there has been an increase in the number of kinds of persistent identifiers used in scholarly communications. Persistent identifiers enable linking of persons, organizations, and scholarly outputs to each other – an essential component of altmetrics. By listing and detailing them in this document, we hope to encourage efforts to support open science and interoperability, and to measure, evaluate, and report on the effectiveness of research infrastructure and communication. Our ultimate goal is to promote and facilitate the broad use of persistent identifiers whenever possible.

The focus for the NISO Persistent Identifiers and Alternative Outputs Working Group included investigation of alternative research outputs and application of persistent identifiers in the scholarly environment to clearly identify research outputs and their relationships. The composition of the Persistent Identifiers and Alternative Outputs Working Group was developed to represent a balance of parties who are interested in unconventional research outputs and identifiers, with just under half of the 33 members coming from academic institutions and libraries. Publishers and other organizations in the scholarly ecosystem make up the majority of the rest of the group's membership. The organizations represented are drawn from bodies that have an interest in research output metrics.

4.2 Persistent Identifiers in Scholarly Communications Document

The persistent identifiers in Scholarly Communications Document (see <https://sites.google.com/site/nisopersistentids/>) is an environmental scan of common persistent identifiers that are used across a variety of scholarly domains to identify research outputs of any known type. Persistent identifiers may be applied to content at multiple levels of granularity, from links to a subset of a dataset to links to aggregated content. The purpose of this document is to raise awareness of the scope and complexity of persistent identifier use across systems, in the hopes of promoting and facilitating the use of persistent identifiers.

The scholarly environment is growing and evolving at a breakneck pace. Work that is accomplished with this attempt to identify and characterize ongoing efforts regarding persistent identifiers will become obsolete almost as soon as it is published. For these reasons, and to promote ongoing engagement with the broader scholarly community, this document is available online at <https://sites.google.com/site/nisopersistentids/>. The document provides a sorted set of known persistent identifiers, with a source, links to schema information or project site, domain relevance, and additional notes about the identifier, where available. The domains vary to include computing, general, government, life sciences, locations, mathematics, media, objects, persons, physical sciences, publishing, and standards. Selection criteria for this initial work included commonly used persistent