Postprint: New Standard for Research Data Usage Statistics and Its Application Case Study

Authors: Lin Weiming, Cymbidium

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] This study introduces the "Implementation Specification for Research Data Usage Statistics" jointly launched by Make Data Count and COUNTER, providing new metrics and perspectives for data-level measurement. [Method/Process] Through analysis of the standard text, this paper introduces the background, objectives, scope, relevant concepts, and core content of the specification, and through case studies, examines the application of the specification by Dash, DataONE, Zenodo, and seven other data repositories. [Results/Conclusion] Research data usage statistics possess unique characteristics; the launch of the specification can complement data citation and data altmetrics to describe a complete picture of research academic impact. Currently, few data repositories comply with this specification. To promote the application of data usage measurement, cooperation among different stakeholders—including standards organizations, researchers, institutional and data repositories, publishers, research institutions and funding agencies, and libraries—is required across the entire lifecycle of data generation, management, dissemination, and utilization.

Full Text

Preamble

Vol. 63 No. 16, August 2019, ChinaXiv Partner Journal

Research on New Standards for Research Data Usage Statistics and Application Cases

Lin Weiming, Ye Lan
Shenzhen University Library, Shenzhen 518060

Abstract

[Purpose/Significance] This paper introduces the *Code of Practice for Research Data Usage Metrics* jointly developed by MakeDataCount and

COUNTER, providing new indicators and perspectives for data-level metrics. [**Method/Process**] Through analysis of the standard text, this study presents the background, objectives, scope, relevant concepts, and core content of the specification. Case studies examine its implementation in Dash, DataONE, Zenodo, and seven other data repositories. [**Results/Conclusion**] Research data usage statistics possess unique characteristics. The introduction of this specification can complement data citation and altmetrics to provide a comprehensive description of research impact. Currently, few data repositories comply with this standard. To promote the application of data usage metrics, cooperation is required among diverse stakeholders—including standards organizations, researchers, institutional repositories and data repositories, publishers, research institutions and funding agencies, and libraries—across the entire lifecycle of data production, management, dissemination, and utilization.

Keywords: research data; usage metrics; usage statistics; usage data; data-level metrics
Classification Number: G251.4
DOI: 10.13266/j.issn.0252-3116.2019.16.004

With the development of data-intensive research paradigms, research data has become an increasingly important component of academic output. To encourage researchers to openly share and reuse research data, the academic community has begun calling for research data to be treated like research papers, integrating data into research evaluation frameworks and exploring data metrics and impact assessment. Usage statistics for research data are considered one of the most important indicators by researchers and other stakeholders, second only to data citation counts [1]. However, due to the lack of standardized specifications for collecting and obtaining research data usage statistics, these indicators have not yet fulfilled their potential. To address this gap, the MakeDataCount project collaborated with COUNTER (Counting Online Usage of Networked Electronic Resources) to develop and officially release the *Code of Practice for Research Data Usage Metrics* (Release 1) on June 5, 2018 [2]. This specification standardizes the generation and reporting of research data usage statistics, providing consistency, reliability, and interoperability for data repositories and platform providers, thereby promoting data reuse among repositories, libraries, funders, and other stakeholders. This paper introduces the background, core content, and application of this new standard for research data usage statistics and proposes recommendations for promoting data usage metrics based on stakeholder analysis.

1 Background of Data Usage Statistics

According to the definition by the National Information Standards Organization (NISO) [3], data usage refers to user actions of accessing and downloading publicly published datasets, encompassing statistics on data downloads, data access, dataset annotations, and similar activities. The proposal for data usage statistics emerged from stakeholder recognition of their importance, the need

for comprehensive data impact assessment, and the desire to fill existing gaps in data usage metrics.

1.1 The Importance of Data Usage Metrics

Data usage metrics can help researchers understand the attention their datasets receive before formal citations occur, serving as important data for research evaluation and incentivizing participation in data sharing and reuse. For institutions, these metrics enable monitoring of data usage trends to evaluate the effectiveness of data storage facilities, assess capacity requirements for storage and network systems, and create specialized collections for highly popular datasets. Data repositories can use them to understand usage patterns and the impact of specific datasets, while funding agencies can track how their funded research outputs (research data) contribute to scientific progress and society at large.

1.2 The Urgent Need for Comprehensive Data Impact Assessment

Although data citation metrics are currently the most prominent indicators in data metrics and impact assessment, they cannot fully capture the landscape of data reuse. The scientific literature community has already begun reflecting on the singularity and absolutism of citation as an academic evaluation metric, proposing instead comprehensive assessment of academic impact based on multiple approaches including citation, usage, and altmetrics. For research data, foreign studies have suggested establishing multiple dataset evaluation indicators. For example, K. M. Fear [4] argued that dataset evaluation should not rely on single metrics but should consider multiple factors such as data citation metrics, secondary impact (e.g., G-index), disciplinary breadth of data reuse, and data download counts. J. Bollen et al. [5] proposed usage-based impact measurement covering the entire research process, including citations, discovery, downloads, peer review emails, reading, and preservation. Therefore, the introduction of data usage metrics can complement data citation and altmetrics to provide a complete picture of data impact.

1.3 The Lack of Best Practices for Data Usage Metrics

Data citation, data usage (downloads and views), and altmetrics constitute the three major indicator systems for data metrics and impact assessment. Data citation metrics represent the earliest and most studied area, followed by altmetrics. Standards and specifications have already been published for both data citation and altmetrics. For instance, FORCE11 (The Future of Research Communication and e-Scholarship) released the *Joint Declaration of Data Citation Principles* in 2014 [6], and the Scholix project (A Framework for Scholarly Link Exchange) [7] established a framework to promote the exchange of link information between scholarly literature and data, helping researchers understand data in publications and literature citing data. Clarivate Analytics launched the Data Citation Index in 2012 to track and record citations of individual datasets.

NISO initiated an altmetrics project in 2013, with one working group focusing on altmetrics for non-traditional research outputs such as datasets and software [8]. However, best practices for data usage metrics are still being explored. To promote their application, the MakeDataCount project, funded by the NSF EAGER program since 2014 and building upon PLOS's Article-Level Metrics (ALM) project launched in 2009, initiated the Data-Level Metrics (DLM) pilot project using Lagotto [9]. Initial members included PLOS, the California Digital Library (CDL), and DataONE, though PLOS later withdrew and DataCite joined. This collaboration ultimately produced the *Code of Practice for Research Data Usage Metrics*, enabling different data repositories to provide usage data according to a unified specification—a milestone in understanding how research data are reused.

2 Current Research Status

Current research both domestically and internationally primarily approaches data-level metrics and data impact from holistic perspectives, with considerable attention focused on citation metrics for scientific data impact. Research specifically addressing usage statistics for data-level metrics remains limited.

2.1 Holistic Approaches to Data-Level Metrics and Data Impact

Foreign organizations and projects have conducted relevant research. For example, Knowledge Exchange's 2013 report *The Value of Research Data* analyzed data metrics concepts from cultural and technical perspectives, stakeholders in data sharing and metrics, and relevant repositories and tools [11]. The UK's Digital Curation Centre (DCC) published a 2015 report *How to Track the Impact of Research Data with Metrics*, discussing data metrics concepts, tools and services, and challenges [12]. The Consortia Advancing Standards in Research Administration Information (CASRAI) established a Dataset-Level Metrics Subject Group to bring together stakeholders to develop data-level metrics [13]. The Research Data Alliance (RDA) and World Data System (WDS) jointly formed the RDA/WDS Publishing Data Bibliometrics Working Group to study data metrics and services. These projects, mostly initiated between 2013-2015, have seen minimal updates recently. While they stimulated thinking about data-level metrics in the scientific community, no systematic and comprehensive evaluation methodology has emerged due to the complexity of data metrics. Domestic research has primarily focused on introducing foreign projects and developments, with Gu Liping [14] being among the first to introduce data-level metrics concepts, development, and applications; Wang Yiping [15] discussing the connotation, types, relationships, stakeholders, and evaluation methods of scientific data impact; and Meng Yang [16] analyzing similarities and differences between data metrics and bibliometrics.

2.2 Data Citation Approaches to Data-Level Metrics

Foreign research on data citation has been conducted extensively in both theory and practice, with organizations such as DataCite, the Research Data Alliance, and the DCC establishing data citation standards and principles, and tools like the Data Citation Index tracking citations. China has also released a national standard *Scientific Data Citation*, with studies by Ding Nan [17] and Xing Hongmei [18] analyzing social science data impact using the Data Citation Index.

2.3 Data Usage Approaches to Data-Level Metrics

The earliest formal proposal of data usage statistics as a distinct research object came from P. Ingwersen and V. Chavan [19] in 2009, who proposed the Data Usage Index (DUI) and constructed 14 indicators including search density, download density, usage impact, and interest impact based on data from GBIF (the biodiversity database). Unfortunately, these indicators only apply to the GBIF repository, and their scientific validity and universality require further investigation. Domestically, only Ding Pei [20] has addressed scientific data usage statistics, examining key processes in collection, standardization, cleaning, and reporting.

In summary, given that current research focuses more on citation metrics and less on usage indicators such as downloads and views, this paper introduces the *Code of Practice for Research Data Usage Metrics* developed by MakeDataCount and COUNTER to establish universal research data usage statistics, aiming to further leverage the role of usage metrics in data metrics.

3 Introduction to the Specification

The specification was developed through collaboration between three institutions with extensive experience in data management and the authoritative usage statistics organization COUNTER. The COUNTER project, launched in March 2002, was an international initiative to standardize the format, content, and terminology of usage data provided by database vendors to libraries, ensuring consistency, reliability, and interoperability for electronic journals, books, databases, and multimedia resources [21]. MakeDataCount, a Sloan Foundation-funded two-year project comprising the California Digital Library (CDL), DataCite, and DataONE, built upon this foundation. CDL, established by the University of California in 1997, includes the University of California Curation Center (UC3) as one of its four major programs, helping researchers and UC libraries manage, preserve, and access digital assets. DataCite, founded in late 2009, is an international non-profit organization providing permanent DOI identifiers for research data. DataONE (Data Observation Network for Earth), launched in August 2009 as an NSF-funded DataNet project, establishes a distributed framework and sustainable cyberinfrastructure for describing and discovering earth observation data. MakeDataCount and COUNTER members began discussing research data usage statistics in June 2017, ultimately releasing

the specification to enable standardized usage data across different repositories.

3.1 Dataset-Related Concepts

The specification tracks dataset usage. Four key concepts define datasets and their hierarchical relationships:

- **Dataset**: A collection of data published or held by an agent, provided with its metadata for access or download in one or more formats. In COUNTER terminology, this is a content item; synonymous terms include data package.
- **Component**: A part of a dataset that can be accessed or downloaded separately. Synonymous terms include data file and data granule.
- **Collection**: A collection of datasets. Related terms include catalog and repository.
- **Version**: A fundamental characteristic of datasets referring to multiple versions of a dataset. Changes in content or metadata, changes to one or more components, or changes to fixed attributes of components all generate different versions.

3.2 Referenced Standards

The specification draws upon multiple existing standards for electronic resource usage statistics, harvesting, citation, and altmetrics. It primarily references the *COUNTER Code of Practice Release 5* (July 2017) [23], which addresses usage statistics for journals, books, and other scholarly resources, making many definitions, processing rules, and reporting recommendations applicable to research data. It also references the SUSHI standard (ANSI/NISO Z39.93-2014: Standardized Usage Statistics Harvesting Initiative) [24] for automated collection of usage statistics, and the "Scholix metadata schema for the exchange of scholarly communication links" [25] to standardize dataset metadata, along with NISO's *Outputs of the NISO Alternative Assessment Metrics Project* (NISO RP-25-2016) [3] for recommendations on data metrics and persistent identifiers.

3.3 Specification Objectives, Scope, Relationship to COUNTER Release 5, and Governance

The specification aims to provide consistent, reliable, and interoperable usage statistics for research data from repositories and platform providers. Currently, it only addresses dataset-level usage statistics, with future plans to provide statistics for all components within datasets based on user needs and feedback. The specification defines statistical data elements, their definitions, report content and format, data processing requirements, and duplicate counting avoidance.

Developed and governed collaboratively by MakeDataCount and COUNTER, the specification follows the *COUNTER Code of Practice Release 5*, diverging

only when necessary. For instance, research data does not require institution-level usage statistics but should provide geographic distribution data. Another significant difference is versioning: the specification requires aggregated usage statistics across all versions of a dataset. Additionally, it does not require reporting by file format (e.g., separate counts for CSV vs. XLSX downloads).

3.4 Core Content of the Specification

The specification comprises eight sections: (1) Preface; (2) Overview; (3) Technical Implementation of Reports; (4) Reports; (5) Report Delivery; (6) Usage Data Collection Methods; (7) Data Processing; and (8) Automated Report Harvesting via SUSHI. Sections 3, 4, 5, 7, and 8 constitute the core implementation content.

3.4.1 Section 3: Technical Implementation of Reports This section describes mandatory reports, their common format, and report attributes. The primary report is the "Dataset Master Report," providing granular and personalized reporting of dataset-level usage behaviors with filters and configuration options. This report applies to both repositories (hosting multiple research output types, such as institutional repositories like Figshare) and data repositories (hosting only research data, such as disciplinary repositories like CDL Dash and Dryad).

Reports can be in tabular format or machine-readable JSON (JavaScript Object Notation). All reports share the same structure with a header containing ten elements: (1) Report_{name}; (2) Report_{ID}; (3) Version; (4) Metric_{types}; (5) Report_{filters}; (6) Report_{attributes}; (7) Exceptions; (8) Reporting_{period}; (9) Created; and (10) Created_{by}. Unlike COUNTER Release 5, research data usage reports exclude institution-related elements (Institution_{name} and Institution_{ID}).

The most important report element is the metric type. The specification adopts "Investigations" and "Requests" metrics from COUNTER Release 5. These newly introduced metrics indicate: Investigations represent user access to information about a content item (e.g., abstract or descriptive metadata) or the item itself (e.g., full text); Requests represent user requests for the full content item, typically through viewing, downloading, emailing, or printing. Figure 1 [Figure 1: see original paper] illustrates that Requests are a subset of Investigations. For research data, any user action applicable to a dataset (including metadata) constitutes an "Investigation," while "Requests" specifically represent retrieval or viewing of the dataset itself. Both metrics have "Total" and "Unique" variants (see Table 1).

3.4.2 Section 4: Reports This section provides detailed specifications for each report and its elements. The specification currently offers only one report: the Dataset Report. Standard dataset reports must include all ten header elements in strict order with precise spelling and capitalization. Beyond header

elements, reports must include dataset name, publisher, creator, publication date, version, DOI, URL, and other information when available. Tabular reports must include DOI or other ID information or URL.

3.4.3 Section 5: Report Delivery This section specifies mandatory information for user access to reports. Requirements include: (1) Reports must be available in TSV (tab-separated values) format for easy, error-free import into Excel, or JSON format following the Research Data SUSHI API specification; (2) Reports must be deliverable as single files for automated processing; (3) Tabular reports must be uploadable to a password-controlled website with email alerts for updates, providing filters and options for users to browse all standard reports; (4) Reports must be provided at least monthly, updated within one month of the previous report's release, typically processed by full month but allowing partial data; (5) Usage data must be retained for the current year and two previous years; and (6) Reports must be harvestable via the SUSHI protocol.

3.4.4 Section 7: Underlying Data Processing Principles This section addresses data collection and processing principles, including return codes, duplicate click filtering, and robot/crawler handling. The *COUNTER Code of Practice Release 5* emphasizes human usage, filtering known web crawlers and spiders while allowing legitimate machine retrieval via scripting languages (e.g., Python, curl, wget, Java). The specification similarly permits legitimate machine retrieval, distinguished by the "Access_{method}" report attribute ("Regular" vs. "Machine"), but excludes web crawlers or robots through blacklisting (see COUNTER's robot/crawler list).

Duplicate click filtering specifies that double-clicks on a link within 30 seconds count as a single click. For example, clicks at 10:01:00 and 10:01:29 are considered duplicates (recorded once), while clicks at 10:01:00 and 10:01:35 are separate (recorded twice). The first request is discarded, retaining the second. Four methods (in descending reliability) identify same-user clicks: (1) username from login; (2) user cookie; (3) session cookie; and (4) IP address and browser user agent.

3.4.5 Section 8: Automated Report Harvesting via SUSHI This section provides detailed SUSHI support. The SUSHI standard, introduced in COUNTER Release 3 (2008) and a requirement since Release 5, enables libraries to efficiently collect COUNTER-compliant usage data. The specification similarly requires content providers to support automated harvesting via SUSHI and has developed a "Research Data SUSHI API specification."

4 Unique Aspects of Research Data Usage Statistics

Analysis reveals several data processing and output methods unique to research data. The specification emphasizes four distinctive features:

First, usage is not divided by institution. Unlike subscribed electronic journals and books, subscription-based access is uncommon for research data, making institutional breakdowns less meaningful. To meet geographic distribution needs, reports are provided by country rather than institution, better facilitating public sharing of usage data.

Second, reports are provided by geographic location (country) rather than IP address. Country-level statistics are more meaningful than IP-based division for understanding dataset usage across locations. For large countries like the United States, state- or province-level statistics can be provided.

Third, usage statistics are provided by version. Versioning is common and complex in research data. The specification emphasizes providing usage reports for each specific version and aggregating usage across all versions.

Fourth, usage is reported by data volume rather than format. Unlike text-based resources, research data can be retrieved in multiple file formats. The specification does not separate requests by format (e.g., CSV vs. XLSX downloads) but includes requested data volume as a report component, a more meaningful variable for research data that helps compare data packaging differences across repositories and usage statistics at different granularity levels.

5 Application Case Studies

The specification primarily applies to institutional and data repositories, which must complete five steps for implementation [26]: (1) read and understand the specification; (2) process usage logs according to the standard; (3) send standardized logs to an open hub (currently DataCite Hub); (4) extract usage and citation metrics from the hub; and (5) display standardized metrics on repository platforms.

5.1 California Digital Library Dash and DataONE

Since the specification's development, the project team's two repositories—Dash (CDL) and DataONE—have implemented standardized usage and citation metrics. Usage metrics include views and downloads. Processed logs are sent to DataCite Hub for public use and integration. Citation information comes from CrossRef Event Data.

CDL Dash provides researchers with a repository to describe, upload, manage, and share research data. Following the DataCite Metadata Schema, Dash assigns DOIs to datasets and collects usage date/time, requesting IP address, session cache ID, user cache ID, username/user ID, requested URL, dataset DOI, and data volume (for requests only). Dash distinguishes machine vs. human usage and divides usage by country according to Section 7 principles [27].

Figure 2 [Figure 2: see original paper] shows Dash's interface displaying usage metrics (views and downloads) and citation metrics. Views correspond to the specification's "Investigations," while downloads correspond to "Requests." The

interface also provides data volume, version, related publications, and standard citation formats. DataONE has provided usage and citation metrics since July 2018, showing citations, downloads, and views for each dataset [28]. Figure 3 [Figure 3: see original paper] displays DataONE's metrics interface.

5.2 Zenodo

Beyond the project team repositories, Zenodo provides more comprehensive usage statistics. Managed by CERN since May 2013 and funded by the EU's OpenAIRE project, Zenodo supports Europe's open access and open data movements. Since July 2018, Zenodo has tracked views and downloads according to COUNTER and the specification, updating statistics every three hours and tracking visitors, visitor types (human, machine, crawler), country, and referring domain [29]. Figure 4 [Figure 4: see original paper] shows Zenodo's metrics interface, which provides unique views, unique downloads, and data volume by version, though it lacks citation metrics.

5.3 Major Data Repositories

To investigate broader adoption, we examined seven major repositories: ICPSR, Figshare, GenBank, PANGAEA, UK Data Service, Research Data Australia, and Dryad (accessed January 2-5, 2019). As Table 2 shows, ICPSR, Figshare, Research Data Australia, and Dryad provide usage metrics, while others do not. Few provide citation or altmetrics, with only Figshare offering citations. Among those providing usage metrics, only Figshare uses standardized metrics; others use non-standard indicators (e.g., ICPSR's "Total Downloads," "Total Sessions," "Total Users"; Research Data Australia and Dryad's non-standard metrics). Thus, major repositories need to strengthen standard adoption.

6 Measures and Recommendations for Promoting Data Usage Metrics

Data usage metrics face challenges intertwined with general data metrics issues. This section addresses broader data metrics to advance data usage metrics specifically, requiring cooperation among stakeholders across data production, management, dissemination, and utilization.

6.1 Standards Organizations

Standards organizations play a crucial role in promoting adoption. MakeData-Count and COUNTER should: (1) recruit data repositories, starting with major providers already offering usage statistics (ICPSR, Figshare, Research Data Australia, Dryad) to achieve scale effects; (2) strengthen promotion through dedicated websites, articles, and conference presentations; (3) develop implementation guidelines (CDL has created a guide with tips and tools [30] and hosted a webinar); (4) promote coordinated use with citation and altmetrics

standards to build a comprehensive impact assessment system; and (5) ensure continuous maintenance and updates every 4-5 years. Future versions should address content provider auditing, as current requirements remain unclear.

6.2 Researchers

Researchers are the primary stakeholders. Without data sharing, data metrics are meaningless. Researchers should: (1) share data in trusted, sustainable repositories meeting criteria such as reputation, certification, community adoption, quality assessment, DOI provision, usage statistics collection, and indexing in DataCitation Index [31]; and (2) formally publish data as research output through data journals or supplementary data, leveraging existing usage, citation, and altmetrics indicators.

6.3 Institutional and Data Repositories

Repositories are the primary implementers. They must: (1) provide detailed dataset descriptions using the DataCite Metadata Schema [32] for cross-disciplinary discovery or domain-specific standards [33]; (2) provide display metadata enabling users to understand and reuse data, including detailed collection processes and application instructions (currently lacking in most repositories); and (3) provide standard citation formats to facilitate citation metrics and integrated data metrics.

6.4 Publishers (Data Journals)

Publishers should: (1) establish policies encouraging data deposition in stable repositories with permanent identifiers; (2) guide researchers to correctly cite data in reference lists; and (3) index data citations in CrossRef. CrossRef's infrastructure links publications with data and other resources, offering guidelines for publishers to deposit citation metadata in references or relationship fields [35].

6.5 Funding Agencies and Research Institutions

These organizations should provide incentive systems incorporating data metrics into hiring, tenure, and promotion decisions, either as standalone data metrics (usage, citation, altmetrics) or integrated into existing evaluation systems.

6.6 Libraries

Libraries, traditionally central to citation-based research evaluation, should: (1) promote data citation and guide proper citation practices; (2) provide data citation and altmetrics data using indexes and tools; (3) offer usage analysis reports following the specification for repositories they operate; (4) develop institutional data metrics; and (5) create discovery tools for repositories.

Conclusion

The *Code of Practice for Research Data Usage Metrics* standardizes the generation and delivery of research data usage statistics, providing new indicators and perspectives for data metrics and impact assessment. Its adoption by all stakeholders across the data lifecycle will establish a complete research data evaluation system, promoting data sharing and reuse.

References

[1] KRATZ JE, STRASSER C. Making data count. [EB/OL]. [2018-12-13]. https://doi.org/10.1038/sdata.2015.39.
[2] FENNER M, LOWENBERG D, JONE M, et al. Code of practice for research data usage metrics, release 1[EB/OL]. [2018-12-13]. https://www.projectcounter.org/code-of-practice-rd-sections/foreword/.
[3] NISO. Outputs of the NISO alternative assessment metrics project (NISO RP-25-2016)[EB/OL]. [2018-12-29]. http://www.niso.org/apps/group_{public}/download.php/17091/NISO-RP-25-2016-Outputs-of-the-NISO-Alternative-Assessment-Project.pdf.
[4] FEAR KM. Measuring and anticipating the impact of data reuse[EB/OL]. [2018-12-15]. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/102481/kfear_1.pdf?sequence=1.
[5] BOLLEN J, DESOMPHEL H V, RODRIGUEZ M A, et al. Towards usage-based impact metrics: first results from the MESUR project[C]//Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries. New York: ACM, 2008: 231-240.
[6] FORCE11. Joint declaration of data citation principles[EB/OL]. [2018-12-29]. https://www.force11.org/datacitationprinciples.
[7] The scholix initiative[EB/OL]. [2018-12-29]. http://www.scholix.org/.
[8] NISO. Alternative outputs in scholarly communications—data metrics (NISO-RP-25-201x-2A)[EB/OL]. [2018-12-29]. https://groups.niso.org/apps/group_{public}/download.php/125-201x-2A%20Alternative%20Outputs%20in%20Scholarly%20Communications-Data%20Metrics.pdf.
[9] Partnerships: PLOS, CDL, and DataONE launch pilot project to develop data-level metrics[EB/OL]. [2018-12-29]. http://www.infodocket.com/2014/10/07/partnerships-plos-cdl-and-dataone-launch-pilot-project-to-develop-data-level-metrics/.
[10] FEAR KM. The impact of data reuse: a pilot study of five measures[EB/OL]. [2018-12-30]. https://www.slideshare.net/asis_{org}/kfear-rdap.
[11] COSTAS R, MEIJER I, ZAHEDI Z, et al. The value of research data: metrics for datasets from a cultural and technical point of view[EB/OL]. [2018-12-30]. http://www.knowledge-exchange.info/event/value-research-data-metrics.
[12] BALL A, DUKE M. DCC how-to guides: how to track the impact of research data with metrics[EB/OL]. [2018-12-30]. http://www.dcc.ac.uk/resources/how-guides.
[13] CASRAI. Research dataset-level metrics[EB/OL]. [2018-12-30]. http://ref.casrai.org/Research_{Dataset}-Level_{Metrics}.

[14] Gu Liping. Data-level metrics: concept analysis and practice progress[J]. Journal of Library Science in China, 2015, 41(2): 56-71.
[15] Wang Yiping, Ma Jianling. Research progress on scientific data impact abroad[J]. Library and Information Service, 2017, 61(7): 118-126.
[16] Meng Yang, Qu Baoqiang. Comparative study between data metrics and bibliometrics[J]. Information Theory and Practice, 2017, 40(11): 139-144, 138.
[17] Ding Nan, Li Jiao, Li Wenyuze, et al. Research on scientific data evaluation based on citation[J]. Library and Information, 2014(5): 95-99.
[18] Xing Hongmei, Lü Xianjing, Liu Wenjun, et al. Analysis of sociology data impact based on DCI[J]. Library Theory and Practice, 2016(2): 43-46.
[19] INGWERSEN P, CHAVAN V. Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure[J]. BMC bioinformatics, 2011, 12(Suppl 15): S3.
[20] Ding Pei. Research on application and key processes of scientific data usage statistics[J]. Journal of Modern Information, 2017, 37(7): 116-122.
[21] Li Hong. Features and future development of the new COUNTER[J]. Journal of Library Science in China, 2012, 38(6): 29-37.
[22] DEKKERS A, ISAAC A. Data Catalog Vocabulary (DCAT) 1.1[EB/OL]. [2018-12-13]. https://w3c.github.io/dxwg/dcat/.
[23] COUNTER Code of Practice Release 5[EB/OL]. [2018-12-13]. https://www.projectcounter.org/code-of-practice-five-sections/foreword/.
[24] ANSI/NISO Z39.93-2014 SUSHI Protocol[EB/OL]. [2018-12-13]. http://www.niso.org/apps/group_{public}/download.php/14217/Z39-93-2014_{SUSHI}-1_7.pdf.
[25] BURTON A, FENNER M, HAAK W, et al. Scholix metadata schema for the exchange of scholarly communication links[EB/OL]. [2018-12-13]. https://doi.org/10.5281/zenodo.1120265.
[26] How to make your data count[EB/OL]. [2018-12-25]. https://makedatacount.files.wordpress.com/2018/07/l to-make-your-data-count-webinar-july-2018.pdf.
[27] Dash how-to guideline[EB/OL]. [2018-12-24]. https://github.com/CDLUC3/Make-Data-Count/blob/master/getting-started.md.
[28] DataONE implements new usage and citation metrics to make your data count[EB/OL]. [2018-12-25]. https://makedatacount.org/2018/10/24/dataone-implements-new-usage-and-citation-metrics-to-make-your-data-count/.
[29] Zenodo usage statistics launched[EB/OL]. [2018-12-27]. http://blog.zenodo.org/2018/07/18/2018-07-18-usage-statistics/.
[30] California Digital Library. Implementing the COUNTER code of practice for research data in repositories[EB/OL]. [2018-12-21]. https://github.com/CDLUC3/Make-Data-Count/blob/master/getting-started.md.
[31] WHYTE A. Where to keep your data: key considerations (check-list)[EB/OL]. [2018-12-21]. http://www.dcc.ac.uk/resources/how-guides.
[32] DataCite Metadata Working Group. DataCite metadata schema for the publication and citation of research data[EB/OL]. [2018-12-21]. http://doi.org/10.5438/0010.
[33] DCC disciplinary metadata standards[EB/OL]. [2018-12-21]. http://www.dcc.ac.uk/resources/metadata-standards.

[34] Publishers: make your data citations count[EB/OL]. [2018-12-25]. https://makedatacount.org/2018/05/29/publishers-make-your-data-citations-count/.
[35] Crossref data & software citation deposit guide for publishers[EB/OL]. [2018-12-25]. https://support.crossref.org/hc/en-us/articles/215787303-Crossref-Data-Software-Citation-Deposit-Guide-for-Publishers.

Author Contributions: Lin Weiming was responsible for writing the paper; Ye Lan was responsible for conceptualization and revision guidance.

Research on the Code of Practice for Research Data Usage Metrics and Its Implementation

Lin Weiming, Ye Lan
Shenzhen University Library, Shenzhen 518060

Note: Figure translations are in progress. See original paper for figures.