Code for Research Data
Community Consultation

For public release 9 May 2024

Background

First released in 2017, the COUNTER Code of Practice for Research Data (CRD) was developed in parallel with Release 5 of the main Code of Practice. The full text of the CRD can be found on the COUNTER website at https://www.countermetrics.org/code-of-practice/. With this experience from the implementation of the initial release, and reflecting on changes in the repository space, the Make Data Count and COUNTER teams feel the time is right to look at an update to the Code of Practice for Research Data. This update provides an opportunity to revisit one of the original intentions of the Code of Practice for Research Data to explore alignment with the Release 5.1 of the COUNTER Code of Practice.

During development of Release 5.1 (R5.1), COUNTER's team aimed to maximise the usefulness of R5.1 to mixed-content repositories (i.e. repositories that contain various types of outputs, such as articles, datasets, multimedia files, software, and other outputs, e.g. IRUS), and to this end it incorporated an extended list of Data_Types including Dataset. Subsequent to publication of R5.1, COUNTER and Make Data Count have been working together to determine whether the CRD could be successfully merged into R5.1.

As key aspects of the CRD and R5.1 overlap, we view a merging of the codes as a relevant path forward. The elements of the codes that are the same are detailed in the section below labelled as ‘Matters which are comparable across the CRD and R5.1’. Other aspects of R5.1 are different from the CRD. Our goal with this consultation is to determine whether the similarities between the codes are close enough to allow us to merge the CRD into the main Code of Practice.

As with all significant COUNTER projects, this proposal is subject to community consultation and you are encouraged to get involved by responding to us in any of the ways we’ve made available to you:

Through our GitHub repository: https://github.com/Project-Counter/CodeRD/wiki

Through our Google form: https://forms.gle/4jZpnLi4qMexfMWj8

By emailing our Executive Director: tasha@countermetrics.org

Project rationale

The CRD was originally aimed at repositories hosting data, while R5.1 now accounts for the needs of mixed-content repositories. Maintaining two Codes requires duplication of effort and resources, and runs a risk of data usage counts being processed differently over time depending
on the Code the repository is applying. To facilitate consistency in usage reports for data and reduce duplication of efforts, we recommend merging the two Codes.

**How this consultation is structured**

You'll see that this consultation is split into sections. Within each section there is a description of the similarities and differences between R5.1 and the CRD, followed by the specific question(s) we have for our community.

**Consultation questions**

**Demographics**

To help us understand your answers, please tell us about yourself.

1. Are you: A repository; A librarian; A consortium manager; A publisher; An organisation providing COUNTER reports on behalf of others; A funder; Other (please specify)
2. Which country are you based in?
3. If you operate a repository, does your repository offer Code of Practice for Research Data usage metrics?

**Matters which are comparable across the CRD and R5.1**

As mentioned in the Background, there is a great deal of overlap between the CRD and R5.1. Five key aspects are identical:

- Both Codes include the Host_Types Data_Repository and Repository, with the same definitions.
- Both Codes include the Data_Type Dataset, with the same definition.
- The Year of Publication (YOP) attribute appears in both Codes, with the same definition.
- Both Codes include Access_Method Regular, with the same definitions.
- Both Codes handle zero usage and missing fields in the same way.

Two further aspects are functionally the same, but with slightly different naming conventions:

- CRD specifies four ‘Dataset’ Metric_Types which map to R5.1 ‘Item’ Metric_Types
  - Total_Dataset_Investigations = Total_Item_Investigations
  - Unique_Dataset_Investigations = Unique_Item_Investigations
  - Total_Dataset_Requests = Total_Item_Requests
  - Unique_Dataset_Requests = Unique_Item_Requests
- Access_Method Machine in the CRD is labelled as TDM in R5.1

**Questions for repositories**

4. Do you foresee any significant difficulties for repositories using the R5.1 Metric_Type nomenclature in place of that used in the CRD? Yes; No; I have a comment (specify)
5. Similarly, do you anticipate any difficulties for repositories using the Access_Method=TDM in place of Machine? Yes; No; I have a comment (specify)

**Reports**

**Required Reports**
R5.1 includes four COUNTER Reports and several Standard Views of COUNTER Reports. At present, report providers with Host_Type=Repository have to deliver a Platform Report and its associated Standard View, as well as the Item Report described in the next section. The Platform Report is a very high-level summary view that is often used for management reporting (note that the PR includes search metrics, for which we have a separate section under the Attributes heading).

**Questions for repositories**

6. As a repository manager, will delivering a Platform Report present difficulties for you or prevent you from using R5.1? Yes; No; I have a comment (specify)

**Mapping Item Reports to the Dataset Master Report**

Of the four COUNTER Reports the Item Report (IR) is most closely related to the Dataset Master Report defined by the CRD, but it accommodates reporting on multiple content types and therefore includes a more comprehensive selection of metadata fields. R5.1 allows for metadata fields to be left empty where there is no metadata (e.g. ISBN for non-book items).

Fields which have are required in R5.1 but do not appear in the CRD:

- Platform will be used to supply the repository name
- ISBN, Print_ISSN and Online_ISSN are required in R5.1 but not applicable to research data. In this case, these column headings will remain required however fields for each dataset may remain blank

Fields from the CRD which appear in R5.1 with a different name:

- Dataset_Title appears as Item
- Creators appears as Authors
- Dataset_Version appears as Article_Version
- Other_ID appears as Proprietary_ID

There are also a series of parent and component data fields. Component fields are not mandatory, but are included in the Item Report specification. These can be excluded from Item Reports, or left blank if they are included in tabular reports. Requests for parent fields must be supported, but if there is no relevant parent information the fields may be left blank.

**Questions for repositories**

7. As a repository manager, will including the extra R5.1 metadata field headings create difficulties for you in generating usage reports? Yes; No; I have a comment (specify)

**Questions for libraries and report consumers**

8. As a library or other consumer of usage data, will delivery of repository usage data in an IR format be helpful? Yes; No; I have a comment (specify)

**Components within the Item Report**

An aspect of R5.1 that we feel may be particularly helpful for repositories is Components, which allow for nested files. To use the example of https://arcticdata.io/catalog/view/doi:10.18739/A2P55DH4M, we would recommend handling the whole dataset (DOI 10.18739/A2P55DH4M) as the Item, while each file (e.g. Heatmap_Family.png)
would be a Component. Using this structure would allow repositories to report on usage of the dataset as a whole, or to be more granular and report usage of each Component.

Questions for repositories
9. As a repository manager, would Components be a helpful mechanism for managing nested content? Yes; No; I have a comment (specify)

Questions for libraries and report consumers
10. As a librarian, is reporting on Components valuable to you? Yes; Possibly; No; I have a comment (specify)

Report headers
R5.1 Reports include an additional piece of information in the headers linking out to the COUNTER Registry. Repositories which do not appear in the Registry may leave the field blank. You can find out more about the Registry at https://registry.projectcounter.org/.

Attributes
Access Types
The CRD has no attribute for Access_Type: we're recommending use of the standard R5.1 Access_Type definitions.

R5.1 includes three Access_Type: Controlled, Open and Free_To_Read (described at https://cop5.projectcounter.org/en/5.1/03-specifications/03-counter-report-common-attributes-and-elements.html#access-types). In the context of datasets, we propose utilising the following Access_Type elements:

- Controlled: describes materials in a controlled access repository - for example where registration is required.
- Open: describes open access materials, accessible without registration or subscription. Open applies where the report provider asserts that the content is open access, irrespective of the license (that is, a Creative Commons license is not essential), and the content may have been Open from the day of publication, or after expiry of an embargo, but it is not intended to return to Controlled status.
- Free_To_Read: describes content that is available to all users on this platform, regardless of authorization status, but not Open (e.g. it may return to Controlled status in the future).

Questions for repositories
11. Do you foresee any significant difficulties for repositories using the R5.1 Access_Type attribute? Yes; No; I have a comment (specify)
12. Will the inclusion of the Access_Type attribute help you to better report on and understand usage within the repository? Yes; No; I have a comment (specify)

Data Types
R5.1 includes many more Data Types than the CRD Dataset.

The CRD includes only one Data_Type, which is Dataset. For data-only repositories, Data_Type=Dataset remains the default. For mixed-content repositories, the full list of COUNTER Data Types as described in R5.1 at Section 3.3 is available. A mapping between COUNTER
Data_Types and the resource types within the DataCite metadata schema is provided below for quick reference, excluding those Data_Types used specifically for databases.

- Terms which are the same: Audiovisual, Book, Dataset, Image, Journal, Other, Report, Software, Sound, Standard
- Mapped terms (COUNTER Data_Type > DataCite resource type)
  - Article > JournalArticle
  - Book_Segment > BookChapter
  - Conference > ConferenceProceeding
  - Conference_Item > ConferencePaper (Posters could be logged under Other, recordings under Audiovisual)
  - Interactive_Resource > InteractiveResource
  - Multimedia > Other
  - News_Item > Text
  - Newspaper_or_Newsletter > Text
  - Patent > Other
  - Reference_Item > Text
  - Reference_Work > Text
  - Thesis_or_Dissertation > Dissertation
  - Unspecified > Other

Questions for repositories

13. Does the extended list of Data_Types present any problems for your repository? Yes; No; I have a comment (specify)

Filtering Reports using Attributes

R5.1 requires that report consumers be able to pre-filter Reports by Data_Type and Access_Type in addition to the CRD filters of YOP, Access_Method, and Metric_Type. Similarly, R5.1 allows report consumers to explicitly Include_Parent_Details and Include_Component_Details.

Questions for repositories

14. Would the additional filters present implementation difficulties for you as a repository manager? Yes; No; I have a comment (specify)

Search and Denial Metrics

R5.1 includes Search and Denial metrics in addition to the shared usage Metric_Types. Denials apply where there are access controls, either paywall or registration, while Search metrics for repositories would appear only in the Platform Report.

Questions for repositories

15. If your repository operates access controls, do you already track denials? Yes; No; I have a comment (specify)
16. If your repository does not track denials, would the requirement to track these create significant problems? Yes; No; I have a comment (specify)
17. If your repository has search functionality, do you already track search activity? Yes; No; I have a comment (specify)
18. If your repository does not track search activity, would the requirement to track these create significant problems? Yes; No; I have a comment (specify)
**COUNTER API and JSON**

The COUNTER API (previously known as SUSHI) was significantly updated for R5.1 to be more robust and easier to use. You can access the details of the API at https://countermetrics.stoplight.io/docs/counter-sushi-api.

As part of the API changes, we reduced the size, improved the readability and simplified the processing of JSON-format COUNTER reports.

Processed data usage reports can be sent to DataCite for aggregation. DataCite provides documentation for contributing usage reports: https://support.datacite.org/docs/contributing. Usage reports submitted to DataCite are made accessible via DataCite's API.

**Questions for repositories**

19. Do you have any concerns about implementing the updated API? Yes; No; I have a comment (specify)

20. Having viewed the sample Item Report in our Stoplight service, do you have any comments about the updated JSON schema? Yes; No; I have a comment (specify)