Open Science Indicators Column Descriptions

This file was originally prepared on 8-3-2022 by Allegra Pearce and Lauren Cadwallader and was updated on 12-12-2023 by Lauren Cadwallader. It is part of the dataset:
Public Library of Science (2022) PLOS Open Science Indicators. Figshare. Dataset (version 5).
https://doi.org/10.6084/m9.figshare.21687686.

Named contact Information
Name: Iain Hrynaszkiewicz
ORCID: 0000-0002-9673-5559
Institution: Public Library of Science
Email: ihrynaszkiewicz@plos.org / plos@plos.org

Alternate Contact Information
Name: Lauren Cadwallader
ORCID: 0000-0002-7571-3502
Institution: Public Library of Science
Email: lcadwallader@plos.org / plos@plos.org

Column descriptions for data provided in PLOS-Dataset_v5_Dec23.csv and Comparator-Dataset_v5_Dec23.csv. The Method column indicates how data is gathered, whether from XML file data or found by DataSeer's assessment.

Column Name	Column description	Method	Additional Notes	Present in comparator set? Additional notes.
Publication_Day	Day of publication for each article	XML	"epub" date is used	
Publication_Month	Month of publication for each article	XML	"epub" date is used	
Publication_Year	Year of publication for each article	XML	"epub" date is used	

DOI	Digital object identifier (DOI) for each article	XML		Yes
Discipline	A list of all disciplines each article is identified as.	XML		Yes. Disciplinary information taken from the Major MeSH terms for each article
Country	Country of the listed affiliation of the first listed corresponding author	XML	If multiple affiliations are listed the first one is used. If multiple corresponding authors are listed, the first listed corresponding author is used.	Yes (only present for ~3% of sample)
DA_data	The article's data availability statement	XML		Yes. It comprises any Availability type statements (e.g. Data Availability or Code Availability) found in the article. If multiple statements are found they are separated by the "|" character.
Data_Section_Text_Generated	Which sections of the research article indicated data was generated when assessed by DataSeer algorithm.	DataSeer	Labeled N/A if no data/code generation detected. Multiple sections are separated by "|" (e.g. Methods|Supplementary) If XML section tags are not available this will be labeled	Yes

			"Full Text"	
Data_Generated	Was data generated in this text as assessed by DataSeer algorithm.	DataSeer	Labeled "Yes" or "No"	Yes
Data_Shared	Was data shared in this text as assessed by DataSeer algorithm.	DataSeer	Labeled "Yes" or "No"	Yes
Data_location	Location of data that was indicated as shared.	DataSeer	Labeled "Online" when a repository was used. Multiple sections are separated by "|" (e.g. Online|Supplementary)	Yes
Repositories_data	A list of standard repositories that were identified by DataSeer as having been used in the sharing of data/code in each article. Repositories are identified from URL/DOI/Accession IDs where possible. Only repositories from a standardized list will be identified. Domains will be provided for unidentified URLs where possible.	DataSeer	Multiple repositories separated by "|" (e.g. Github|Dryad)	Yes
URL_data	A list of any URLs provided that host data or code online that were detected by DataSeer algorithm.	DataSeer	Multiple URLs separated by ","	Yes
Accessions	A list of Accession IDs from identified repositories that host data or code online that were detected by DataSeer	DataSeer	Multiple Accession IDs separated by ","	Yes

	algorithm.			
Data_DOIs	A list of DOIs provided that host data or code online that were detected by DataSeer algorithm.		Multiple DOIs separated by ","	Yes
Preprint_Match	Whether a preprint was identified for this published article when searching common DOI registration agencies.	DataSeer	"Yes" or "No" indicating whether a match was found across DOI registration agencies. Match is completed using title and author information. Queries Crossref and Datacite.	Yes
Preprint_DOI	DOI of identified preprint match.	Preprint Metadata	Gathered from DOI agency metadata.	Yes
Preprint_Day	Day of preprint DOI registration, usually same date as publication.	Preprint Metadata	Note: Datacite DOI metadata does not always include a detailed date (will sometimes only include year level).	Yes
Preprint_Month	Month of preprint DOI registration, usually same date as publication.	Preprint Metadata	Note: Datacite DOI metadata does not always include a detailed date (will sometimes only include year level).	Yes
Preprint_Year	Year of preprint DOI registration, usually same date as publication.	Preprint Metadata	Note: Datacite DOI metadata does not always include a detailed date (will sometimes only include year level).	Yes
Preprint_URL	URL to access top preprint match.	Preprint Metadata	Provided where available, dependent on metadata.	Yes

Preprint_Server	Server top preprint match is hosted on.	Preprint Metadata	If server name is not provided in preprint metadata DOI prefix is used to identify server when possible.	Yes
Code_Section_Text_Generated	Which sections of the research article indicated code was generated when assessed by DataSeer algorithm.	DataSeer	Labeled N/A if no code generation detected. Multiple sections are separated by "|" (e.g. Methods|Supplementary) If XML section tags are not available this will be labeled "Full Text"	Yes
Code_Generated	Was code generated in this text as assessed by DataSeer algorithm.	DataSeer	Labeled "Yes" or "No"	Yes
Code_Shared	Was code shared in this text as assessed by DataSeer algorithm.	DataSeer	Labeled "Yes" or "No" Note: Shared *code is only detected if it is executable* (e.g. shared as a .R or .py file format), code shared in pdf of other image formats is not labeled as shared.	Yes
Code_Location	Location of code that was indicated as shared.	DataSeer	Labeled "Online" when a repository was used. Multiple sections are separated by "|" (e.g. Online|Supplementary)	Yes

Quarter	Quarter of publication inferred from Publication_Month and Publication_Year	DataSeer	Format is "Year Quarter", e.g. "2020 Q1" or "2018 Q4"	Yes
ANZSRC	Fields of Research categories as assigned by Dimensions	Dimensions	Multiple categories are separated by ";"	Yes