

NISO Altmetrics Working Group C "Data Quality" - Code of Conduct Self-Reporting Table
For data aggregator: Plum Analytics

Item	Description	Supports CoC Recommendation	Aggregator / Provider Submission*	Last update of self-reporting table**
#1	List all available data and metrics (providers and aggregators) and altmetric data providers from which data are collected (aggregators)	T1	<p>Plum Analytics has a suite of products called PlumX. A description of each PlumX product can be found on our product pages here: http://bit.ly/1WZ3kHB</p> <p>PlumX collects metrics data from many sources and groups them into 5 categories of metrics. Sources for each category are defined below:</p> <p>Usage – bepress, bit.ly, CABI, Dryad, DSpace, EBSCO, ePrints, Facebook, figshare, Forbes, Github, Institutional Repositories, OJS Journals, PLOS, PubMedCentral, Pure, RePEc, Slideshare, SSRN, WorldCat More info: http://bit.ly/1KtCO7a</p> <p>Captures – Delicious, EBSCO, GitHub, Goodreads, Mendeley, SlideShare, Vimeo, YouTube More info: http://bit.ly/1onkGBZ</p> <p>Mentions – Amazon, blogs, Facebook, GitHub, Goodreads, mainstream media, Reddit, Slideshare, SourceForge, StackExchange, Vimeo, YouTube, Wikipedia More info: http://bit.ly/1Szs1cd</p> <p>Social Media – Amazon, Facebook, Figshare, Google Plus, Goodreads, SourceForge, Reddit, Twitter, Vimeo, YouTube More info: http://bit.ly/1ZMb9C3</p> <p>Citations – CrossRef, PubMed Central, PubMed Central Europe, RePEc, Scopus (for mutual customers), SSRN, United States Patent and Trademark Office More info: http://bit.ly/1V7JNpg</p>	2016/03/31

#2	Provide a clear definition of each metric.	A1	<p>The PlumX Suite provides the raw usage, capture, mention, social media, or citation counts by source, e.g., the number of Wikipedia articles we have mined about a specific book or article. Raw counts can be viewed in the application, embedded in other sites through widgets, or exported.</p> <p>We strive to keep the naming of these metrics consistent with how the source we are harvesting them from. E.g., Mendeley “readers” and Delicious “bookmarks.”</p> <p>We have over 35 specific, granular metrics that we calculate. A complete list and definition of each can be found here: http://bit.ly/1iLQ0IH</p>	2016/03/31
#3	Describe the method(s) by which data are generated or collected and how data are maintained over time.	T1, T2, R1	<p>Data are collected via a range of methods, largely via data provider APIs, third-party provider APIs, FTP data transfers, OAI-PMH harvesting, web crawlers and RSS feeds.</p> <p>The data is maintained over time as described in section #7 below.</p>	2016/03/31
#4	Describe all known limitations of the data.	A3	<p>When PlumX begins utilizing a source of metrics, the amount of historic data from that source will vary.</p> <p>Our text mining for calculating mentions of artifacts often requires that the artifact is mentioned by URL or another scholarly identifier to associate the mention with the artifact.</p> <p>Links to original posts on third party blog and news sources may break or posts may be deleted.</p> <p>Our match and merge algorithms for combining and aggregating metrics from all the different online locations where it is published (as described in #6 below) depend upon a knowledge base of how to cross-walk different identifiers (like going from a DOI to a PubMed ID). If there are errors in this crosswalk data, it is possible to “over-merge” a record. Any examples of this can be reported to</p>	2016/03/31

support@plumanalytics.com. Similarly, if there is not enough data to automatically merge two preprints from two different services together, they may also need to be manually identified and merged by the PlumX staff.

We license twitter data in PlumX directly through Twitter/GNIP. We have a filtered view of all tweets based upon the domain names of the links in the tweets. Our historic twitter data begins on January 1, 2011. We accommodate URL shorteners and have match and merge technology for combining tweets from multiple, separate URLs into a single view for a given artifact. However, if the original artifact is published at a domain that we do not yet track, once identified and added by the Plum Analytics team, twitter mentions for that domain will only begin to be counted from the time the new domain is added.

#5	Provide a documented audit trail of how and when data generation and collection methods change over time and list all known effects of these changes. Documentation should note whether changes were applied historically or only from change date	R1, R2, R3	PlumX does not have an audit trail. Tracking of an audit trail for PlumX will begin in April 2016.	2016/03/31
----	--	------------	--	------------

forward.

#6	Describe how data are aggregated.	T2	<p>Each research output in PlumX is called an artifact. PlumX tracks over 40 different types of artifacts including books, book chapters, conference proceedings, journal articles, slide presentations, videos, etc. A full list of artifact types can be found here: http://bit.ly/1VUueC7</p> <p>Online events about different versions of the same artifact (Publisher + Green Open Access + Preprint + Aggregated versions + A&I) are collected and aggregated based on algorithms that examine matching identifiers (such as DOI, ISBN, or URI) across versions.</p> <p>Usage, Capture, Social Media, and Mention metrics counts are summed across all versions of each artifacts.</p> <p>Citation counts are not added together across different providers since this would result in double-counting the citations. Instead, we represent the cited by count for an artifact as the maximum value reported.</p> <p>Within PlumX Dashboards and PlumX +Grants, metrics are aggregated based on researcher, grant, or any other customer-defined group hierarchy for comparisons at the aggregate level. Group hierarchies are defined by each client and might include grouping by school or department, by geography or by journal issue or volume.</p> <p>Within PlumX Benchmarks, metrics are aggregated per institution, to allow comparisons between all institutions who have received NIH funding from 2012-2015. They are also aggregated at the NIH grant level, so that users can see the ROI on any NIH grant.</p>	2016/03/31
----	-----------------------------------	----	--	------------

#7	Detail how often data are updated.	T3	<p>Metrics data is kept up to date by re-harvesting on the frequency that the source of the metrics updates. For some data providers, like twitter, we license part of the twitter firehose, and we get the metrics in real time. For other sources, we get daily updates of metrics. For example, we update usage data from EBSCO on a daily basis. For other providers, they only give us their data on a weekly or monthly basis.</p> <p>Every 3-4 hours we refresh the entire PlumX index to have the most up to date metrics from all of our sources.</p>	2016/03/30
#8	Describe how data can be accessed.	T4	<p>Plum Analytics provides access to the data via end-user interfaces, widgets that customers can integrate to their site, free artifact widgets or via our open Application Programming Interface (API).</p> <p>Article level widgets can be accessed by the following identifier types:</p> <ul style="list-style-type: none"> • arxiv • cabi_abstract_id • doi • github_repo_id • isbn • nct_id • oclc • pmid • repo_url • slideshare_slideshow_id • sourceforge_repo_id • ssrn_id • us_patent_publication_id • vimeo_video_id • youtube_video_id <p>Author level widgets can be accessed by their PlumX user id. This user id can also be associated with both publicly available author identifiers such as ORCID or with institution specific unique author identifiers. Each customer of PlumX can decide if their PlumX user ids are public or private.</p>	2016/03/30

		<p>Group level widgets can be accessed by their PlumX group id. These group ids can be mapped to and associated with institution-specific group ids. Each customer of PlumX can decide if their PlumX group ids are public or private.</p> <p>Grant level widgets can be accessed by their PlumX grant id. These grant ids can be mapped to and associated with institution-specific or funder-specific grant ids. Each customer of PlumX can decide if their PlumX grant ids are public or private.</p> <p>Documentation about our widgets and API is available here: http://bit.ly/SXYTCd</p>
#9	<p>Confirm that data provided to different data aggregators and users at the same time are identical and, if not, how and why they differ.</p> <p>R4</p>	<p>All Plum Analytics applications are based on the same set of data. Users access the same data across each tool, except where data is restricted according to access level. Access level varies across products, but all products require a subscription to access all data. Artifact-level PlumX pages are free and publicly accessible; they provide access to all our article-level metrics.</p> <p>2016/03/31</p>
#10	<p>Confirm that all retrieval methods lead to the same data and, if not, how and why they differ.</p> <p>R4</p>	<p>We eat our own dog food at Plum Analytics, and the entire PlumX product suite is developed on top of the same API we expose to customers. Different retrieval methods will lead to the same data.</p> <p>2016/03/31</p>

#11	Describe the data-quality monitoring process.	T5, A2	<p>Data quality is monitored in a variety of ways. Some sources of data (such as the set of blogs that PlumX covers) are hand-curated to focus on research-oriented blogs. This set of blogs is created in conjunction with customer driven requests, and content managers on the Plum Analytics team facilitate this process.</p> <p>Outlier analysis is done on our data to identify and investigate potential gaming or erroneous metrics.</p> <p>Our match and merge technology for bibliographic data prioritizes high quality sources like CrossRef over unedited sources like when our crawlers harvest data off of the open web.</p> <p>Each new source of metrics goes through a rigorous data quality assurance cycle before being added to PlumX.</p>	2016/03/31
#12	Provide a process by which data can be independently verified (aggregators only).	R5 ww	<p>See Item #8 - all PlumX Suite tools and services use the API documented here: http://bit.ly/SXYTCd</p> <p>We consider all the metrics in all of the PlumX products to be fully auditable. For all third-party data providers like Facebook, Twitter, etc., metrics can be verified with those parties. For example, PlumX reports a Facebook likes count of 141 for this article: https://plu.mx/a/0ekxGOXPjTF3JmszhgRPI1yvt4F39t2sc54bCJ4BeHg/?display-tab=summary-content#LIKE_COUNT. A search of the Facebook API for that article by URL shows the same count.</p> <p>In all cases, we clearly break down metric counts and indicate the provider. Our data is fully transparent. We do not calculate any scores on top of the raw results returned by each data partner.</p>	2016/03/31
#13	Provide a process for reporting and correcting data or metrics that	A2	<p>Suspected inaccurate metrics or data can be reported to support@plumanalytics.com.</p>	2016/03/31

are
suspected to
be
inaccurate.