

NISO Altmetrics Standards Project White Paper

(Draft 4, June 6, 2014)

Table of Contents

- Summary 2
 - Potential Action Items*.....2
- Introduction..... 3
- Definitions 4
 - Potential Action Items*.....5
- Research Outputs 5
 - Potential Action Items*.....6
- Discovery 7
 - Potential Action Items*.....7
- Research Evaluation..... 7
 - Potential Action Items*.....9
- Data Quality and Gaming..... 9
 - Potential Action Items*.....10
- Grouping and Aggregation.....10
 - Potential Action Items*.....11
- Context12
 - Potential Action Items*.....12
- Stakeholders’ Perspectives.....12
 - Researchers*.....12
 - Institutions*13
 - Funders*14
 - Publishers*.....14
 - General Public*.....14
 - Potential Action Items*.....14
- Adoption.....15
 - Timing*15
 - Drivers of Adoption*.....15
 - Outreach*.....16
 - Potential Action Items*.....16
- Next Steps17

Summary

In the first phase of the NISO Alternative Assessment Metrics Project (July 2013 through June 2014), three in-person meetings were held and 30 in-person interviews conducted. The goal was to collect input from all relevant stakeholder groups, summarize the discussion in this white paper, and identify potential action items for further work in Phase II of the project.

Because of the open format used in the meetings and interviews, we were able to collect a broad range of input that touched on many aspects of metrics and assessment, which also includes input not directly related to standards or best practices. Overall there were very lively discussions with much consensus as to the areas that need further work and very little controversial discussion. These observations are a good indication that an evolving community cares about this topic and that we can expect productive work going forward. Overall, a total of 25 action items in 9 categories were identified, listed below and again within each category.

Potential Action Items

1. Develop specific definitions for alternative assessment metrics.
2. Agree on proper usage of the term “Altmetrics,” or on using a different term.
3. Define subcategories for alternative assessment metrics, as needed.
4. Identify research output types that are applicable to the use of metrics.
5. Define relationships between different research outputs and develop metrics for this aggregated model.
6. Define appropriate metrics and calculation methodologies for specific output types, such as software, datasets, or performances.
7. Agree on main use cases for alternative assessment metrics and develop a needs-assessment based on those use cases.
8. Develop statement about role of alternative assessment metrics in research evaluation.
9. Identify specific scenarios for the use of altmetrics in research evaluation (e.g., research data, social impact) and what gaps exist in data collection around these scenarios.
10. Promote and facilitate use of persistent identifiers in scholarly communications.
11. Research issues surrounding the reproducibility of metrics across providers.
12. Develop strategies to improve data quality through normalization of source data across providers.
13. Explore creation of standardized APIs or download or exchange formats to facilitate data gathering.
14. Develop strategies to increase trust, e.g., openly available data, audits, or a clearinghouse.
15. Study potential strategies for defining and identifying systematic gaming.
16. Identify best practices for grouping and aggregating multiple data sources.

17. Identify best practices for grouping and aggregation by journal, author, institution, and funder.
18. Define and promote the use of contributorship roles.
19. Establish a context and normalization strategy over time, by discipline, country, etc.
20. Describe how the main use cases apply to and are valuable to the different stakeholder groups.
21. Identify best practices for identifying contributor categories (e.g., scholars vs. general public).
22. Identify organizations to include in further discussions.
23. Identify existing standards that need to be applied in the context of further discussions.
24. Identify and prioritize further activities.
25. Clarify researcher strategy (e.g., driven by researcher uptake vs. mandates by funders and institutions).

Introduction

Assessment of scholarship is a critical component of the research process, impacting most elements of the process from which projects get funded to who gains promotion and tenure, and which publications gain prominence in fields of inquiry. As a metric, citation reference counts have historically been an exceptionally strong and rich source of accessible data upon which to draw conclusions about the quality of scholarship and will remain an important component of research assessment.

Unfortunately these traditional metrics, primarily based on print processes, have always had their faults but increasingly they are failing to keep pace with the expanded scope of forms and usage that are presently available. Online reader behavior, network interactions with content, social media references, and online content management—all important indicators of scholars' interaction with research outputs—are not reflected in today's measures. In addition, newer forms of network and researcher behavior analysis can provide a means to assess non-traditional scholarly outputs.

New forms of scholarly outputs, such as datasets posted in repositories, software tools shared in Github, or algorithms or molecular structures are now commonplace, but they are not easily—if at all—assessed by traditional citation metrics. For most of these outputs, there isn't yet a culture of citation around them, despite efforts to foster their development. Citation analysis also lacks ways to measure the newer and more prevalent ways that articles generate impact such as through social networking tools. These are among two of the many concerns the growing movement around alternative metrics, sometimes called "altmetrics," is trying to address.

With any new concept or methodology there exist limitations and gaps that need to be addressed in order to facilitate its adoption. The NISO Alternative Assessment Metrics Project tries to address the following issues through a process of community consensus:

1. What exactly gets measured?
2. How do we decide what the criteria are for assessing the quality of the measures?

3. At what granularity should these metrics be compiled and analyzed?
4. How long a period should altmetrics cover?
5. What is the role of social media in altmetrics?
6. What is the technical infrastructure necessary to exchange these data?
7. Which metrics will prove most valuable and how do we decide?
8. What types of assessment criteria could and should be applied to these new metrics to best assess the value of the analysis?
9. How do we ensure consistent quality across providers?

This was only a starting list of potential issues for discussion and underscores why it was important to have in-person meetings and other forms of interaction with the community to develop a comprehensive list of the issues and concerns.

In the first phase of the NISO Alternative Assessment Metrics Project, which began in July 2013, three in-person meetings were held and 30 in-person interviews conducted. The focus of the in-person meetings and the interviews was to collect input from a wide variety of stakeholders, including researchers, librarians, university administrators, scientific research funders, and publishers regarding their views on alternative assessment metrics, and to identify areas where best practices and standards might be helpful. The in-person meetings took place:

- October 9, 2013 in San Francisco
- December 11, 2013 in Washington, DC
- January 23, 2014 in Philadelphia

Between 30 and 50 people attended each of the meetings. In addition, each meeting was recorded and live-streamed to the broader information distribution community. Each of the three streaming sessions was viewed by 50-100 concurrent participants throughout the events, and interactive participation was made possible through twitter and e-mail exchanges. To allow the unbiased collection of ideas and to foster discussion, the in-person meetings were done in an open format with 5-10 short lightning talks followed by 1-2 discussion rounds in smaller breakout groups. The personal interviews were conducted in March and April 2014 using a semi-structured interview format with a list of predefined questions. This white paper summarizes the topics that were discussed, the consensus or lack of consensus in the group discussions, and the areas highlighted as needing further work. Complete information about the meetings, the discussion summaries, as well as the recordings of the videos are all posted on the NISO Alternative Assessment Initiative section of the NISO website at: http://www.niso.org/topics/tl/altmetrics_initiative/

Definitions

“Altmetrics” is the most widely used term to describe alternative assessment metrics. Coined by Jason Priem in 2010, the term usually describes metrics that are alternative to the established citation counts and usage stats—and/or metrics about alternative research outputs, as opposed to journal articles.

Alternative assessment metrics cover many varied areas, both in the artifacts that are assessed and in what metrics are collected. A question was raised whether there are

distinct types of altmetrics that differ with regard to how they measure impact, how open they are, and how they collect diverse research outputs. One related question is the association of the term “altmetrics” to related concepts, specifically article-level metrics, data citation, Snowball Metrics, and bibliometrics.

Some people expressed their discomfort with this term because:

- these metrics are no longer alternative, but have become mainstream;
- these metrics are viewed as *complementary* rather than as an alternative;
- the more focused term “social media metrics” would seem more appropriate; and/or
- altmetrics are not clearly defined, and may mean different things to different people.

There was no consensus whether one or more new definitions of appropriate terms should be attempted, but some people expressed concerns that the term altmetrics might not be helpful for wider acceptance and adoption of the new measures.

The discussions at the in-person meetings and in the interviews covered a wide range of topics that go beyond a narrow definition of altmetrics, including the discussion of usage stats; linking of research outputs to contributors, institutions, and funders; and general issues regarding research evaluation.

Potential Action Items

1. Develop specific definitions for alternative assessment metrics.
2. Agree on proper usage of the term “Altmetrics,” or on using a different term.
3. Define subcategories for alternative assessment metrics, as needed.

Research Outputs

The evaluation of research has traditionally focused on scholarly journal articles and—in the humanities and social sciences—books or book chapters. While the focus on these traditional outputs will likely remain critical into the foreseeable future, it is clear that the importance of other research outputs is increasingly recognized. Some of the research outputs mentioned in the discussions include:

- Research datasets
- Scientific software
- Posters and presentations at conferences, in particular in the medical sciences
- Electronic theses and dissertations (ETDs)
- Performances in film, theater, and music
- Blogs
- Lectures, online classes, and other teaching activities

Several organizations have already defined the types of scholarly artifacts that they deemed relevant for research evaluation, including EuroCRIS/CERIF, VIVO, OpenAIRE, and IRUS-UK. The National Science Foundation (NSF) has recently changed the biosketch requirements for proposal submissions; instead of requiring a list of five relevant

publications, it now requires five research products, which could include databases, software, etc.

There seems to be consensus that research datasets and scientific software should be included in the list of valuable research outputs, but inclusion is less clear for some of the other types of research outputs, and it remains an open question as to whether it is necessary to clearly define those artifacts at all. Peer review is seen by many as a useful mark of quality and it is unclear how peer review is presently applied to these other outputs, or how it could be in the future. Many libraries have taken an inclusive approach and offer to host any type of digital scholarly content in their institutional repositories. Nanopublications—in which assertions, or data, or discovery elements, are shared with minimal additional context—were also mentioned in the discussions, but they currently remain just a concept as no distribution infrastructure is developed for them, such as the purpose that journals serve for articles. Another concept that was discussed, and which could potentially replace the journal article or other single research output, is the aggregated model of related research outputs, although components of this system, such as micro-attribution have not yet developed.

Many workshop participants felt that it is important to use persistent identifiers such as DOIs to make it easier to track metrics associated with a research object. One concern raised by several people is that the techniques developed for citing and measuring the impact of scientific journal articles might not be appropriate for other types of research outputs; software and code was cited as an example of this potential misappropriation. The consensus of the group was that the infrastructure needed for data citation of these outputs isn't fully in place yet within the community, and that simply mirroring the infrastructure for research paper citation might not be the best option when building this new infrastructure.

The argument was made that different research outputs might be most appropriate for different scientists, depending greatly on their fields of study, their role in the scholarly processes, their particular types of scholarly output, or even their personality types. For example, not every scientist should be expected to set up a Twitter account or science blog. It may not be necessary to attach metrics to *all* researcher or faculty activities. The importance of presentations and posters at meetings was seen as the exchange of information that leads to inspiration and collaboration, rather than something that can be measured, and some people expressed similar views regarding blog posts.

Some participants worried about the *journalization of science*—the centrality of the published journal article in evaluation and advancement—and felt that it would be important to give more weight to other research outputs. One argument given for this perspective was that there exists a bias in the publication and review system where negative results are difficult to publish, and therefore not something people get credit for. This lack of credit reduces the incentives for publishing such research, which can be quite important, and hinders reproducibility. Research outputs better tailored to communicating negative results may fill this gap. However, the discussions also contained opinions that the scholarly article *is* the most important research output, and everything else is supplementary material.

Potential Action Items

1. Identify research output types that are applicable to the use of metrics.

2. Define relationships between different research outputs and develop metrics for this aggregated model.
3. Define appropriate metrics and calculation methodologies for specific output types, such as software, datasets, or performances.

Discovery

There are ideas percolating in the academy that are not given enough attention. Maybe a system that captured information earlier could accelerate the way we learn and innovate.

The conversations about what altmetrics can do have tended to conflate two use cases:

- Discovery
- Evaluation

Many participants felt that this distinction is important. Whereas altmetrics are particularly well suited for discovery because of their immediacy and variety, they play a much smaller role in evaluation, both because of what they try to measure and because they have not yet withstood the test of time.

Discovery fits well with the current trend of instant access, immediate discussion, and sharing. There was overall agreement that altmetrics can play an important role as a discovery tool, in particular for newly published content where citations are not yet available. To that extent an increasing amount of bibliometrics research has been done regarding the correlation of alternative assessment metrics with traditional citations, in particular addressing the question of the predictive value for future citations. It has been known for a long time that usage statistics don't correlate too closely with citations, as they reflect different behaviors—reading versus research for writing. From the newer metrics it has been observed that Mendeley reader numbers seem to correlate best with citations; however, that correlation is also not strong enough to serve as a strong predictor of future citations.

Another use under the theme of discovery encompasses social media reference. One challenge for social media is that the general acceptance of social media indicators is lagging behind the increasing use of social media in academia. There is still quite a lot of discussion of how much notice one should make of social interaction, whether it's buzz or notoriety of research, or the activity level of the researcher on social media platforms. Several participants also brought forward the notion that as most altmetrics services are currently implemented, the focus is very much on the author and less on the reader, still leaving much room for improvement as discovery tools.

Potential Action Items

1. Agree on main use cases for alternative assessment metrics and develop a needs-assessment based on those use cases.

Research Evaluation

It is easier to measure an activity than it is to measure quality because quality is a subjective judgment.

The speed with which one can get information and the kinds of information one can get is definitely changing, but for promotion, tenure, and awards, the tried and true processes, cultural issues, and barriers haven't changed. Many participants felt that altmetrics aren't going to change these social interactions any time soon, as they are still immature and even might cause more new problems than help solve existing problems.

Senior colleagues in committees remain the ones who make the tenure and promotion decisions, just as it is with grants and funding decisions. Often these decisions come down to how a researcher and her work are viewed in the field. These decisions almost never come down strictly to numbers or quantitative measures. While no one participating in the conversations suggested that these qualitative measures be removed from the process, it should be recognized that subjective assessment presents its own challenges. Particularly, two big challenges caused by this kind of peer evaluation are the substantial cost—especially in time and effort on the part of reviewers—and the subjective biases that a non-blind evaluation could introduce.

Bibliographic metrics don't work that well for evaluation. Part of the problem is that the methodology from which they are derived is often complex and poorly understood. The role traditional bibliometrics can play is to use a comparative view to determine if an output has exceeded a minimum level and after which a qualitative assessment can be applied to a smaller subset of outputs. Unfortunately, the scientific rigor applied to using these numbers for evaluation is often far below the rigor scholars use in their own scholarship; the misuse of the Journal Impact Factor to assess individual researchers is often cited as an example. As in any field, people are overwhelmed with work and seek quick and easy answers in evaluations, which seem to be provided by bibliometrics. Citations tend to be more important internationally in emerging research areas like Brazil and China, because they are seen as indicators that can help establish international reputation.

Many participants felt that issues of misapplying quantitative evaluation methods are exaggerated for altmetrics, not least because they have been studied for only a short time. Critics of alternative assessment metrics usually argue that they correlate poorly with scholarly impact and draw attention and resources in the wrong direction. Participants in the NISO initiative raised the concern that there is a lack of interaction between those who have studied research evaluation intensely and the emerging altmetrics community.

One aspect of evaluation where altmetrics *can* play a role is the assessment of non-academic impact, i.e., engagement with the social, cultural, and economic environment. This engagement is increasingly seen as important; for example, it represents 20% of the evaluation in the next version of the Research Excellence Framework (REF) in the UK.

The other aspect where altmetrics can play a role in evaluation is for the assessment of newer research outputs that have not traditionally been part of the evaluation, e.g., research data or scientific software. Another example that was mentioned is in the emergence of video as a significant non-traditional content format. The best practices for this assessment are still evolving, e.g., how best to track the impact of data as noted previously.

In the evaluation of medical research, there is increasing emphasis on systematic reviews to inform decisions for diagnostic tests and treatment as part of evidence-based medicine. This very involved process has become the standard approach for this type of evaluation.

Some participants felt that collaboration is an aspect of assessing research impact that is underutilized and needs more attention. Some funders and institutions combine the evaluation of research with open access, i.e., only research outputs made openly available in a digital repository are part of the evaluation.

Potential Action Items

1. Develop statement about role of alternative assessment metrics in research evaluation.
2. Identify specific scenarios for the use of altmetrics in research evaluation (e.g., research data, social impact) and what gaps exist in data collection around these scenarios.

Data Quality and Gaming

Alternative assessment metrics are relatively new, and the reliability and validity of the data has not yet been studied in detail. There was consensus that this is an important area of study and that further work is needed regarding data quality. Some data sources are more problematic than others because data can be pulled in many different ways—usage stats and Twitter are good examples. One question that was raised is how to evaluate what data sources are added or removed as altmetrics sources; there is a tendency to include what can be counted rather than to include what adds particular value. Another key problem identified is the failure of bloggers, tweeters, and traditional journalists to use persistent identifiers like DOIs when discussing scholarly works, which means that it is difficult to collect specific references outside traditional scholarly literature.

Making all altmetrics data openly available via a standardized API and/or download, a centralized altmetrics data clearinghouse, and audits for altmetrics data are some of the approaches discussed that could improve data quality. For broader acceptance, efforts are required to normalize the data across data providers.

One important aspect of data quality is the potential for gaming metrics, e.g., behavior that is meant to unfairly manipulate those metrics, generally for one's benefit. Many alternative assessment metrics are more prone to gaming compared to traditional citations. Before effects of gaming on alternative assessment metrics can be factored out in computations, the community needs consensus on what behaviors are considered cheating/gaming vs. what is considered acceptable promotion. A better understanding of the factors that make a metric more or less likely to be gamed is needed, such as:

- Provenance, i.e., linking to the original source (A metric is less likely to be gamed if the source is openly available, e.g., a citation or blog post.)
- Availability (Should metrics be hidden to prevent herd mentality?)
- Anti-gaming measures by the data provider (For example, cleaning up usage stats following COUNTER criteria.)
- Identifying the incentives for gaming

There are currently no best practices for detecting and preventing gaming of alternative assessment metrics, and there seems to be no consensus on data sources that are considered credible in terms of data quality or for their potential to prevent or mitigate

gaming. We discussed that looking at the correlation between two metrics known to be correlated in the absence of gaming can help to flag suspicious incidents.

One outcome of the discussions was that we need more bibliometrics research into data quality, e.g., looking into reproducibility across multiple providers. It also appears that many users could be overwhelmed by the nuances of data quality, and that trust into these metrics will need to be gained, perhaps over time.

Potential Action Items

1. Promote and facilitate use of persistent identifiers.
2. Research issues surrounding the reproducibility of metrics across providers.
3. Develop strategies to improve data quality through normalization of source data across providers.
4. Explore creation of standardized APIs or download or exchange formats to facilitate data gathering.
5. Develop strategies to increase trust, e.g., openly available data, audits, or a clearinghouse.
6. Study potential strategies for defining and identifying systematic gaming of new metrics.

Grouping and Aggregation

Alternative assessment metrics typically measure one particular metric for one particular research output, and these metrics are then usually grouped or aggregated in a variety of ways. Aggregation means that a single score is produced for a group of metrics of component elements that are related in some way. For a single research output, for example:

- Grouping or aggregation of several metrics produced by a single data source, e.g., HTML pageviews and PDF downloads for usage stats, or Facebook likes, shares, and comments
- Aggregation of metrics for a research output available from multiple locations, e.g., publisher, PubMed Central, and institutional repository
- Aggregation of metrics for multiple versions of the same research output
- Aggregation of metrics for multiple manifestations of the same research output, e.g., journal article, pre-print, and presentation about the same research
- Grouping of related metrics for the same artifact, e.g., Mendeley readers and CiteULike bookmarks
- Aggregation of all metrics for an artifact into a single score

These metrics for a single research output can then be further aggregated or grouped by:

- Journal
- Contributor
- Institution
- Funder

- Subject area
- Region or country

The above grouping and aggregation is common, e.g., aggregation of HTML and PDF usage stats, creating a single score for all altmetrics for a research output, journal-based metrics, or grouping of publications by contributor. Grouping can improve the understanding of a sometimes overwhelming number of metrics, but also introduces biases and carries the risk of grouping things together that should be kept separate, e.g., because different time scales apply.

Aggregation further exacerbates these risks of error or data quality issues, particularly if underlying metrics and algorithms are not made available. At the same time it is clear that, as described above, various stakeholders (including researchers) have a strong desire for an easy-to-understand metric that spreads across a variety of providers or data sources. There is also a potential of providing too much data, as behavioral economics research has shown that the ability to reach a decision declines after a certain amount of data.¹

There are currently no community best practices or standards for grouping and aggregation, and there is also no consensus that these mechanisms are needed. One concern in this realm is that aggregation combines metrics where we have a good understanding of their individual meaning and data quality with other metrics where these understandings may not exist.

The different aggregation mechanisms available can sometimes favor a particular group, e.g., the H index is more favorable to senior researchers. Another consideration is that different stakeholders might be interested in different groupings and aggregations, and that these numbers could be calculated on the fly, similar to how the OECD calculates the Better Life Index. One example is funders, who are particularly interested in aggregation at the author level. We might also want to allow authors to create a portfolio of the things they want credit for, e.g., via their ORCID profile.

Grouping by contributor, institution, or funder presents additional challenges related to unambiguously identifying those relationships with the research artifact, ideally using persistent identifiers such as ORCID, ISNI for individuals or institutions, and FundRef for funders. There was consensus that solving these relationship identification issues is as essential for wider adoption of alternative assessment metrics as it is for traditional citations. One related aspect is the as-yet-unsolved question of contributorship, i.e., capturing in a consistent way how an individual contributed to a research output created by a sometimes very large group of people. One suggested solution was to give the credit to the group of contributors as a whole.

Potential Action Items

1. Identify best practices for grouping and aggregating multiple data sources.
2. Identify best practices for grouping and aggregation by journal, author, institution, and funder.
3. Define and promote the use of contributorship roles.

¹ Iyengar, Sheena. *The Art of Choosing*. Grand Central Publishing, 2010. ISBN 978-0446504119

Context

It's important to use this data in context and intelligently rather than just using the numbers and getting a quantitative metric.

There are significant disciplinary and geographic differences in the use of various scholarly outputs and social media, as well as other alternative assessment metric data sources. Similar to the use of traditional citations, we have to be careful when comparing numbers from different fields. It might be necessary to explore new metrics that are better suited to particular subject areas and geographic regions (e.g., Weibo for China).

Many scholars in the humanities feel that traditional metrics do not apply to their discipline because of the absence of the monograph in computing these metrics and the low frequency of publishing and of citations in these fields generally. There are concerns that the workflow in the sciences dictates how metrics are applied and that the different workflows in the humanities and social sciences are not given enough consideration.

The focus in discussing alternative assessment metrics is often on aspects of quantitative data, but qualitative data can be as informative. For example, it is not only interesting how many times a paper has been discussed or cited, but also by whom, e.g., from which geographic location and by someone with what professional or influential background. The fact that policy organizations, members of the European Parliament, or front line organizations in health are tweeting about a paper often matters more to a funder than the absolute number of tweets. Network analysis can help show the most important publications in a field, e.g., by showing citations by influential researchers—although privacy concerns and policies obviously limit this kind of analysis and these data are not available for all metrics. Until now, little attention has been paid to the intent behind a citation, e.g., when a source is mentioned because the author is refuting it.

The target audience is also of relevance; it is quite possible that different communities and different stakeholder groups need different metrics.

Potential Action Items

1. Establish a context and normalization strategy over time, by discipline, country, etc.

Stakeholders' Perspectives

Researchers

Even though every researcher wants credit for the work s/he has done, it appears that many still remain quite unaware of the existence or potential value of altmetrics; most researchers still focus on journal articles—using journal name and Journal Impact Factor—as the primary means of evaluating impact assessment. One participant in the brainstorming meetings presented data that a mere 5% of researchers were familiar with the term Altmetrics.

Social media facilitate the more timely discussion of research results compared to citations in the scholarly literature. Researchers are also interested in seeing that their research makes a difference outside of academia, such as in policy decisions. Outside of the academy there is a variety of practitioners, who can easily access and interact with scholarship,

whether in medicine, engineering, teaching, or librarianship. Related to this is the public science element: scientists who are able to translate work for the average person.

The approach to metrics taken by researchers is often dependent on their career stage. Tenured faculty are often in easier positions to pursue their choice of scholarship, research and publication venue, and, unless grant funding is essential for ongoing research, evaluation of their work is less of a concern compared to researchers seeking tenure. For those seeking tenure, any strategy that can increase the likelihood of success is carefully considered, thus explaining the perhaps surprisingly conservative approach taken towards alternative assessment metrics by many early-career faculty in this position. These researchers are “trained” to concentrate on the more traditional aspects of tenure and promotion and to not spend time on mechanisms seen as more risky and distracting. On the other hand, use of altmetrics can be a great opportunity to strengthen the value of the tenure package, e.g., if a researcher has produced a lot of research data or other non-traditional outputs, or if he/she can demonstrate social outreach and societal impact.

A third group of researchers, and a group that is growing in size, are those researchers following non-tenure-based career tracks. The current evaluation system is very much focused on the first and last authors of publications, and is less concerned about properly evaluating all the others who make significant contributions to the scholarly effort. As science increasingly becomes much more collaborative, the community needs to reconsider how proper credit is given to all these contributions, and how alternative assessment metrics could play a role in this area by tracking or measuring credit for more non-traditional research outputs such as research data.

Institutions

Institutions are increasingly paying attention to impact assessment, including assessment of non-traditional outputs. Administrators at research institutions often have to respond to greater calls for accountability either from internal or external sources, and frequently view metrics as a way to help with that goal. However, disconnects between university administrators and faculty may be created in that faculty wish to maintain standards of their discipline, whereas administrators also value other aspects of the research, such as the creation and cultivation of a wider audience beyond the peer community.

Other motivations for administrator support of altmetrics—as tuition and other financial factors become increasingly important for some institutions—include attracting potential students and funding by showing a more complete picture of the institution’s talent and the broad impact of its work. The financial pressures felt at many academic institutions drive new pushes for grants and other third-party funding as well as public/private partnerships and entrepreneurship involvement by faculty. These activities can drive attention away from the scholarly literature as the main way by which scholars evaluate themselves.

Institutions recognize that their students and younger faculty have grown up with social media and are interested in leveraging this intertwining, often not so much for measurement of scholarship but rather as a means of connecting with the broader community.

Libraries are increasingly asked by administrations to provide metrics about their faculty, and both institutional repositories and research information systems play a central role in this activity. Many librarians are involved in discussions around metrics, including

alternative assessment metrics, and often serve as ambassadors for new approaches within an institution.

Funders

The initial reaction of some funders to altmetrics was to view them as a vanity factor without much relevance in the scholarly community. That view seems to be changing more recently, as funding agencies—in particular funding agencies using taxpayer money—have an inherent interest in funding science that meets the needs of the public. Funders also often see tremendous value in the general public understanding of publicly funded research projects and the scientific process. Use of alternative assessment metrics may help support needs in these areas.

Many funders are interested in other outcomes besides the advancement of knowledge, e.g., changes in clinical practice or economic impact. Alternative assessment metrics that look at citations in clinical practice guidelines and those tracking patents have started to address these questions. Similarly, funders are generally keen on pushing research boundaries and novel forms of output are encouraged, yet they lack verifiable sources of data on outcomes or performance.

Publishers

Some publishers have taken a proactive approach to altmetrics and see them as an alternative to promoting the Journal Impact Factor. As use of altmetrics and usage statistics become more attractive to potential authors, additional publishers are following suit in the incorporation of these tools into their products. In addition, the potential of new metrics in discovery tools is particularly of interest to publishers who are always interested in expanding markets interested in their content.

General Public

Social media inherently represents a greater democratization of communications, an equal mixing of the expert and inexpert.

Altmetrics have opened up the scholarly discourse to people who, under previous circumstances, wouldn't be part of these types of conversations. Anyone with Internet access can now tweet about a particular article or write a blog post about it, whereas in earlier days communications were typically limited to those who were publishing scholarly articles. Some altmetrics data providers are trying to distinguish between contributions from scholars vs. contributions from the general public, whereas others see this as a futile exercise. Regardless, determining the participating level of the general community might be a valuable reference point in scholarly communications.

Potential Action Items

1. Describe main use cases for the different stakeholder groups.
2. Identify best practices for identifying contributor categories (e.g., scholars vs. general public).

Adoption

Some of the discussions included ideas for strategies on how standards and best practices for alternative assessment metrics could be moved forward. It was clear from the comments that all stakeholder groups, including researchers, bibliometricians, librarians, administrators, funders, publishers, and third-party tool providers, should be included in these steps. Questions that need to be addressed in the area of uptake are how these activities would be funded in the long term, whether there are existing standards and organizations involved in these areas, and, if so, how best to coordinate with them.

Timing

As alternative assessment metrics are very much an evolving field, the best timing for developing best practices and standards was discussed in detail. We know that this process is one that may take several years in total duration, but we don't want the process to start too early and suppress innovation and experimentation, or to develop standards that only address the requirements of a too-narrow set of stakeholders.

Several people expressed their concern that it is still too early to consider altmetrics for impact assessment, and that wider adoption and more established standards are needed before these metrics can be used for tenure and promotion. Since assessment metrics are strong incentives for faculty to change their behaviors, any modifications in this area of use should be considered carefully. In contrast, it was felt by many participants that the time is right to move forward with best practices and standards for altmetrics for their use in the realm of information discovery. Some participants felt that community efforts should be encouraged to begin before commercial players define the field.

Many participants suggested a prioritization of activities. Some ideas that were identified as having a higher priority include:

- Unique identifiers for scholarly works and for contributors
- Standards for usage stats in the form of views and downloads
- Focus on building infrastructure rather than detailed metrics analysis

Drivers of Adoption

Many participants felt that authors should be the primary drivers of change: if altmetrics work for authors, then these users will put pressure on other participants in the scholarly ecosystem to adopt them. In particular, it was felt that publishers will likely adopt altmetrics if their incorporation makes their journals more desirable to authors. A particularly interesting subgroup to engage with are those researchers seeking tenure and promotion, as they will be more open to strategies that increase their likelihood to get these advancements.

Many participants suggested that the best platforms for education and wider adoption of altmetrics among various researchers are the discipline-related societies because of the communities they serve, focusing on the particular interests of their discipline. The San Francisco Declaration on Research Assessment (DORA) is one recent example.

It was recognized widely that the funder often drives policy change and no change usually takes place without changes in reward systems—places where funders have much say.

Several professional organizations were mentioned as partners for NISO in the standardization process, including the National Federation of Advanced Information Services (NFAIS) in the US, Consortia Advancing Standards in Research Administration Information (CASRAI) in Canada, and the Association of Research Managers and Administrators (ARMA) in the UK.

One further open question is whether, in addition to coordinating standards and best practices, the community also needs an organization to act as auditor and/or clearinghouse for these data used in altmetrics calculations; similar discussions in the realm of electronic usage statistics led to the creation of Project COUNTER. However, creation of such a central community effort raises further questions regarding financial support for it, and there is no consensus direction emerging yet. A related topic that was discussed is data ownership, i.e., who owns the data, and what are the conditions and licenses for reuse.

Outreach

Altmetrics are still largely unknown to researchers, and to a lesser extent other stakeholder groups. There was a feeling that we need to better understand what people in the community are currently doing with altmetrics, e.g., what early adopters want altmetrics to be used for and how altmetrics are used in other parts of the world. Some factors that hinder wider adoption of alternative assessment metrics were identified, e.g., the entrenched power of journal brands, the current tenure and promotion system, and, more generally, an academic culture focused on citations in particular journals.

The topic of how to promote alternative assessment metrics for wider use was discussed at length. Some ideas included:

- Building a compelling profile page for every author
- Incentivizing data sharing
- Creating a standard logo
- Creating well-documented examples
- Featuring articles in popular media
- Finding a Nobel Prize winner to promote altmetrics
- Building and supporting tools that make use of these data

To what extent these activities need to be coordinated, or targeted at specific stakeholder groups, was not discussed in detail.

Potential Action Items

1. Identify organizations to include in further discussions.
2. Identify existing standards to include in further discussions.
3. Prioritize further activities.
4. Clarify researcher strategy (e.g., driven by researcher uptake vs. mandates by funders and institutions).

Next Steps

This white paper describes the outcomes, recommendations, and potential action items identified during Phase I of the NISO Altmetrics Initiative. This paper will be circulated and vetted publicly by the research community, assessment experts, publishers, and vendors who we ask to help identify the recommended best practices and/or standards that need to be developed by NISO (possibly in conjunction with others) in order to speed wide adoption of altmetrics. The potential action items included in this white paper will be used as the basis for draft New Work Item Proposal(s) within NISO to commence its Phase II work to advance these outputs. Presuming the approval of the NISO leadership committees and the voting membership, the final output of this initiative will be one or more published standards or recommended practices regarding new assessment forms. With the draft and community input in hand, NISO will be positioned to leverage the community used in developing the paper to create the working groups needed to develop those practices and standards. There will also be training and resources produced about how to implement and adopt the standards and/or recommended practices, which will provide further participation opportunities for the community.