



## Open Impacts Workgroup Report

*John Dove, Jason Hoyt, Rebecca Kennison, Natalia Manola, Trevor Owens*

### Abstract / OSI2016 Workgroup Question

This report summarizes the discussion of the Open Impacts workgroup held at the Open Scholarship Initiative meeting in Fairfax, VA, April 19-22, 2016. The workgroup tackled the following questions at OSI2016: How fast is open access growing? Is this fast enough? Why or why not? What are the impacts of currently evolving open systems? For instance, are overall costs being reduced for scholarly libraries? Is global access to scholarly information increasing? What about in the Global South? What is the impact in this region of increasing adoption of the author-pays system?

---

Society would be better off if the future were one of 100 percent open scholarship, including immediate open access to publications and to data. That hope, which is shared by most, must be allowed to shape how we act today if we are going to achieve such ambition. We must feel empowered to create change, but such change will require setting targets and developing the strategies, studies, and reflection to get there. A house cannot be built without first understanding the soil that it will rest upon and how the environment will be impacted.

Foundational, then, is that we have accurate and ample data in order to support the lofty goals of open scholarship. We believe such data is necessary to understand the effects that any strategies and targets for open scholarship will have upon all stakeholders.

Each discipline and stakeholder will have its own idiosyncratic needs and reflexes, thus there must be a specific data-gathering approach by and for each field to

understand the impact of moving toward open scholarship.

To create the open impacts measures that we believe are needed to achieve such change, we wish to target reward and evaluation systems, the transparency of academic publishing costs, support for new scholarly communication mediums, and improved text and data mining.

There are already numerous case studies on these issues. We lack milestones, goals, and widely accepted routine measures. In short, we need to know where we've been and where we are in order to go where we want to be.

We know the "why" of data capturing, but to arrive at the who, how, and what requires study and debate between and among all stakeholders. Therefore, we have aimed to set forth a launch pad with this document for others to build upon in determining the impact(s) of becoming open. We set out three high-level areas in a framework to create an agenda for going

forward: measuring openness, measuring the use of open, and understanding the economic impacts of open. Finally, we conclude with a target to fill the knowledge gaps.

As the initial workgroup for “Open Impacts,” we have gone in circles, found our way out again, only to then paint ourselves back into a corner once or twice. A look

back on where we’ve already been may be helpful to future groups looking to extend this work.

The initial question we asked ourselves was “what should we be working on or solving?” Preliminary discussions of the workgroup’s remit fell into four loosely defined categories:

<p><b>We should set specific targets</b></p>	<ul style="list-style-type: none"> <li>• By [X year] scholars should be able to incorporate 2x as much research in their work as they do now.</li> <li>• Businesses and jobs based on access to scholarly literature should measurably increase year on year.</li> <li>• Discipline by discipline we should arrive at a point where a broad consensus exists by [X year] that scholars and institutions anywhere in world can get by with only what’s available for free on the Internet.</li> <li>• By [X year] at least 60% of academic libraries have “flipped” from pay-wall to open access.</li> <li>• The majority of scholarly output will be allowed to be machine-readable by 2020.</li> </ul>
<p><b>We should define the stakeholders &amp; impacts</b></p>	<ul style="list-style-type: none"> <li>• Map out who the stakeholders are and explore the kinds of impacts that should be better understood for each of those groups.</li> <li>• Is there a set of measures and recommendations we can come to that might apply to all stakeholders?</li> </ul>
<p><b>We should gather existing data</b></p>	<ul style="list-style-type: none"> <li>• What % of literature is currently open access?</li> <li>• What are the costs of open scholarship?</li> <li>• Which fields are already open enough?</li> <li>• What are the impacts of openness on libraries, the public, current systems, etc.?</li> <li>• Where and how can we get the data in a consistent/systematic way to know where we are and how to measure progress?</li> </ul>
<p><b>We should determine which data to capture</b></p>	<ul style="list-style-type: none"> <li>• Desire for good measures.</li> <li>• Approaches to measurements.</li> <li>• How do we measure progress in open scholarship?</li> </ul>

Table 1. Proposed categories discussed for what should be the initial workgroup’s remit and outputs.

Given our short timeframe, we determined in the end that specific targets (Table 1) related to open scholarship were

out of scope for our workgroup. We also concluded that an attempt to gather existing data within the workgroup’s limited

hours would be of minimal value. Such an exercise could, however, be part of a future workgroup or more likely a task for organizations created or enlisted and devoted to such.

Similarly, even defining who all of the stakeholders are and the impacts upon them seemed to be both too granular and too expansive for an initial workgroup gathering. It seemed neither realistic nor productive to develop consensus on measures since we were working in the abstract. We will not be able to come up with a concrete plan for measurement without a budget, timeline and other real world constraints.

Therefore, we decided to focus on developing a general framework for understanding what data to gather on the impacts of open scholarship.

### Three areas for an Open Impacts framework

The team ended up with three different high-level areas of foci to create the agenda for understanding open impact (Fig. 1):

1. Measuring openness
2. Utilization measures
3. Understanding economic impacts of open

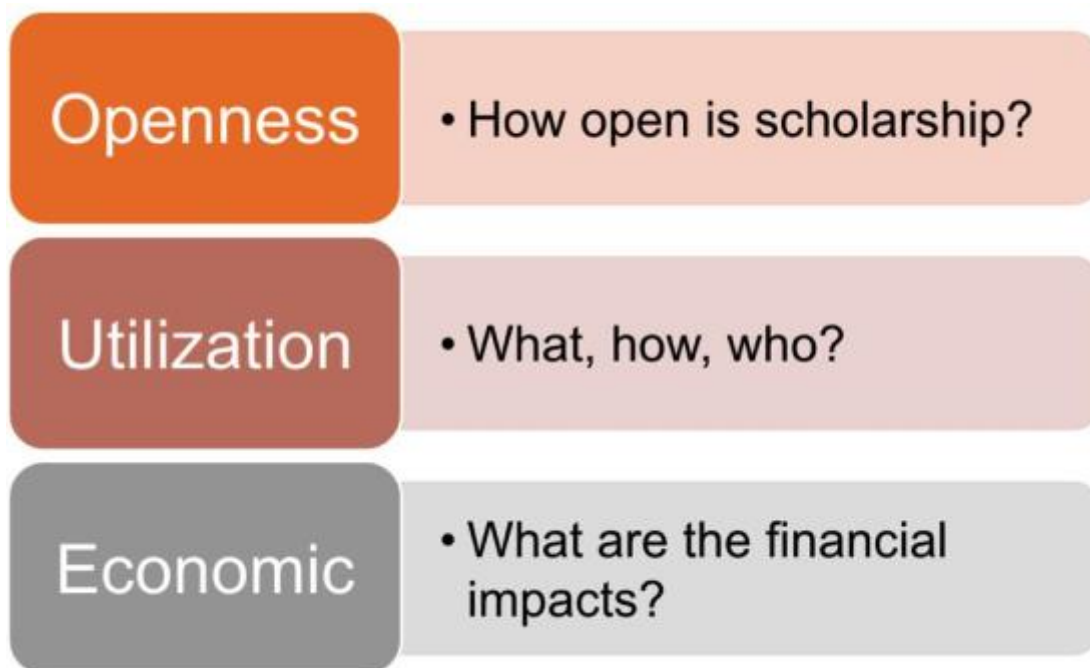


Figure 1: Three areas of the Open Impact framework

Measuring openness is a near-term opportunity to establish key baseline measures for how open scholarship is deployed in different fields and at different institutions. It would also be useful for policy makers (e.g. funders, institutions) to have insight into the effectiveness of an open

access policy. Finally, such measures can help all stakeholders (publishers, funders, scientists, institutions, etc.) measure their progress towards openness in their field.

## Measuring Openness

To determine how open scholarship may be for particular fields we found it helpful to define the possible measures and group them by domain (Table 2). Challenges for future workgroups abound:

- There is a need to firm up common measures for what will be counted and

how. This will require stakeholder consensus.

- Who will do this work? Existing projects need to be coordinated on a global scale, resources secured, data collected and curated, and so on.
- Which incentives can be recognized and developed to drive participation?
- How will any pilot projects scale?

Products	License measures	Availability measures	Permanence measures	Format measures
<ul style="list-style-type: none"> <li>• Artides</li> <li>• Monographs</li> <li>• Data</li> <li>• Software etc</li> </ul>	<ul style="list-style-type: none"> <li>• Creative Commons</li> <li>• Free to read</li> <li>• Free to mine</li> <li>• Embargoed and embargo length</li> <li>• Pay-walled</li> </ul>	Different ways to measure availability [needs more research]. Examples: <ul style="list-style-type: none"> <li>• Metadata quality,</li> <li>• Discoverability,</li> <li>• Crawling</li> <li>• Machine readability</li> <li>• Links to other resources</li> <li>• Public access to usage data</li> </ul>	Official certification <ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> <li>• No but committed to long-term preservation</li> </ul>	Per file formats (e.g., PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml)

Table 2: Measuring Openness: metrics devise to address openness.

**Recommendation:** Work and organizations already exist, that can be leveraged to coordinate this work. Examples include, but are in no way limited to VIVO, SHARE Notify, OpenAIRE, SPARC, ORCID, CrossRef, and USUS.

Our recommendation in this area is to create an “Openness Score” based upon Table 2. This score could apply to specific research objects, aggregated by object to disciplines, funders, institutions, etc. An Openness Score Summit could be organized. Such a summit would involve all

stakeholders, including those collecting data. This meeting’s initial focus would be to establish a plan for developing the infrastructure, funding, and sustainability needed. It would also focus on aligning with existing infrastructures. The key output should be a strategic, organizational, and technical plan for how the data above would be collected, organized, coordinated and managed.

## Utilization measures

This is the “who” and “what” for the various types of use. Leveraging the available information, a key consideration is

whether a common approach can be established. We defined categories of current uses of open scholarship and a limited set of possible metrics for each (Table 3).

Current use	Possible Metrics
To grow global and national economies	New job sector growth, invention-to-product time
To increase literature access	Page reads, downloads, citations, derivative works (translations, reuse of figures), number of publications, machine usage, time on site, altmetrics, geo-location, etc.
To engage a broader audience	Page reads (number, distribution, reader ID), downloads, references in non-academic sources, increased diversity in readers, crowd sourcing/funding
To gain access to more data	Access frequency, distribution, user ID, citations, crowd sourcing/funding.
To support education (K-12, higher ed)	Access by learners, references in educational material, programs, lesson plans, syllabi
To accelerate innovation	Use in patent applications, number of patents arising, investment growth
To facilitate archiving and curation	Frequency of use, growth of archives

Table 3: Utilization measures.

**Recommendation:** We were split in our recommendation. Some delegates suggested that several working groups should be created in order to tackle this effort in coordination. Such working groups could include the National Information Standards Organization (NISO) and the Research Data Alliance (RDA) to propose standards for open access usage metrics to be adopted by the community at large that would include published material provided via publishers, content in institutional repositories, domain specific repositories, and funder repositories like PubMed, and other sources of research output. Following this path we would strongly encourage that any U.S. initiative be aligned with existing initiatives: Impact Story for alt-

metrics, IRUS UK, OpenAIRE analytics, DE OA Statistik, and COAR working group on repository usage stats.

Other delegates would prefer to entrust others to develop the metrics we need rather than going through the extra step of creating standards for metrics first.

## Measuring Economic Impact

What are the economic impacts of open Scholarship, and what is considered to be worth paying to provide access in the long term (i.e., to preserve)? Much like the early discussion of the workgroup’s scope, many concrete measurements were considered: increased economic growth,

democratization of innovation, increase in cross domain research, change of peer-review, disruption to the academic publishing industry, job placement at universities, institutional ranking, etc.

**Recommendation:** We agreed that economic impact is too broad to address through specific measures and needs to be analyzed via ongoing inquiries. We recommend that one or more funders explore this domain in the areas listed above and more. We recommend that as part of this effort the Open Scholarship Initiative

(OSI) establish a research agenda for open scholarship if an existing program does not yet exist somewhere else in the scholarly community. Subsequent OSI meetings could include a program aiming to address the economic impact. Finally, we suggest gathering a diverse stakeholder group to define reasonable metrics for the publishing industry, as well as identify the socioeconomic impacts of open access policies.

## **OSI2016 Open Impacts Workgroup**

**Jean-Gabriel Bankier**, President, bepress

**John Dove**, library and publishing consultant

**Jason Hoyt**, CEO, PeerJ

**Rebecca Kennison**, Principal, K|N Consultants

**Natalia Manola**, Director, OpenAIRE

**Trevor Owens**, Senior Program Officer, Institute of Museum and Library Services (IMLS)

**Christopher Thomas**, Administrator, Defense Technical Information Center (DTIC), US Department of Defense

**Jack Schultz**, Director, Christopher S. Bond Life Science Center, University of Missouri

**Neil Thakur**, Special Assistant to the Deputy Director for Extramural Research, NIH, and program manager for the NIH Public Access Policy