

Smithsonian Research Online
2019 Annual Report

Table of Contents

INTRODUCTION	2
2019 SUMMARY REPORT	4
RESEARCH PRODUCTIVITY	4
SMITHSONIAN-AFFILIATED	4
SCHOLARS	4
SMITHSONIAN UNITS	4
OPEN ACCESS PUBLISHING	4
IMPACT	4
SMITHSONIAN SCHOLARS	5
WHO IS A SMITHSONIAN-AFFILIATED SCHOLAR?	5
NEW COMPLEXITIES	5
ORCID: A PERSISTENT DIGITAL IDENTIFIER FOR RESEARCHERS	5
SMITHSONIAN UNITS	7
PRE-SMITHSONIAN	7
LEGACY OUTPUTS & MASTERS THESES	7
TYPES OF RESEARCH OUTPUT IN SMITHSONIAN RESEARCH ONLINE	9
EXPANDED REFERENCE TYPES	9
REPORTING: CALENDAR YEAR VS. FISCAL YEAR	10
FISCAL YEAR VS. CALENDAR YEAR	10
OPEN ACCESS PUBLISHING AT THE SMITHSONIAN	11
OPEN ACCESS VIA THE PUBLISHER	11
OPEN ACCESS VIA REPOSITORIES	11
TOP JOURNALS	13
MEASURING IMPACT	14
CITATION COUNTS	14
JOURNAL IMPACT FACTOR	15
ALTMETRICS	15

Introduction

Smithsonian Research Online (SRO) is part of a suite of scholarly communications services provided by Smithsonian Libraries and Archives (SLA) for the Smithsonian Institution. SRO serves as the key collection and dissemination tool for scholarly activity at the Institution for a wide variety of internal and external audiences. SRO provides a robust website where SI leadership, scholars, and the public alike can search for citations of publications and other scholarly outputs, research.si.edu. The majority of the data collected by the SRO program comes through automated collection of data from available sources. SRO relies on the Smithsonian research community (both the researchers as well as staff of SLA) for filling gaps and verifying the accuracy of the data.

The attached inaugural report is designed to raise awareness of the possibilities of additional reporting beyond a simple count of publications by explaining the source of the data and providing examples of analysis. At this point, the data measurements contained in this report are just that—measurements. Smithsonian Research Online will strive to report accurate measurements about the scholarly output of the Smithsonian Institution. Determining impact or metrics for scholar/researcher evaluation is a challenge faced by research institutions around the world. SRO is a key component for the continuing dialogue at the Smithsonian.

An Appendix at the end of this report provides definition of terms used throughout.

This report is intended to be internal to the Smithsonian Institution. Feedback from this report can provide guidance on new service development and desired outcomes from the program.

2019 Summary Report

Research Productivity

- Smithsonian Research Online has tracked **2,511** research outputs published in 2019.
- **2,154** (86%) of these outputs are journal articles, published in 621 different journals
- **55** (2%) are books, **122** (7%) are book chapters, and **180** are other output types including datasets, reports, posters, and exhibitions.

Smithsonian-Affiliated Scholars

- There are **811** identified scholars who have Smithsonian affiliation and have authored, edited, or otherwise created research outputs in Smithsonian Research Online in 2019.
- **Over 90%** of scholarly outputs in Research Online have more than one co-author.

Smithsonian Units

- Research Online adds one or more of **49** different unit and departmental tags to all research outputs, representing organizations across the Smithsonian Institution.

Open Access Publishing

- **1,272** (59%) journal articles in Research Online published in 2019 are open access, either directly from the publisher or through a repository. This does not include currently embargoed publications that will become open after a set embargo period.

Impact

- Smithsonian publications from 2019 that have DOIs and are indexed in Web of Science Core Collection have already been cited **3,668** times.
- Smithsonian research published in 2019 has been mentioned on Twitter **over 50,000** times, in news articles **4,583** times, and have been cited in **430** Wikipedia articles.

Smithsonian Scholars

Every research artifact added to Research Online must have at least one agent (author, editor, etc.) who is either employed by or has an academic appointment at the Institution. This generally means federal or trust staff, but can also include post-doc/fellows, research associates, or in some cases visiting researchers.

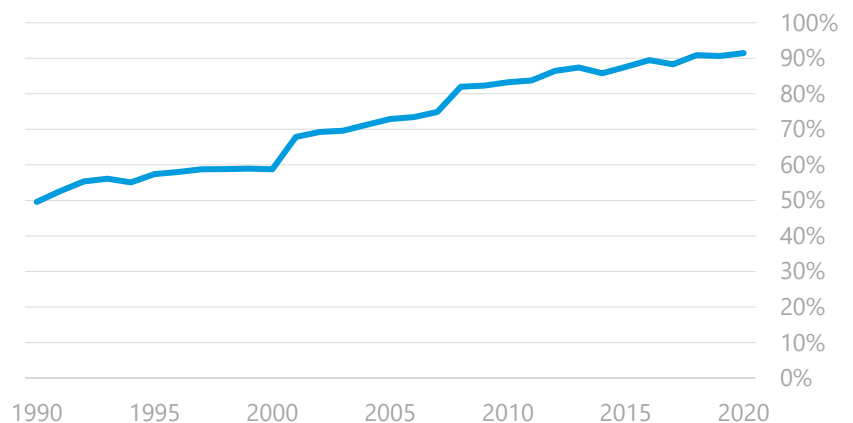
Who is a Smithsonian-Affiliated Scholar?

There are approximately 800 active scholars at the Institution who are known to Smithsonian Research Online. Scholars are considered active if they have authored a scholarly work dated 2019 and for whom SRO staff are aware of a Smithsonian affiliation. Put in perspective, this represents 516 federal or trust employees, 112 post-doc/fellows, 64 affiliates and 119 whose status is unknown.

New Complexities

The landscape of scholarly publishing is also continuing to change, adding complexity to how SRO accounts for affiliation. Over the past few decades, there has been a steady increase in the average number of co-authors on publications. Over 90% of journal articles in Research Online with a publication year of 2019 have more than one co-author. Looking at just the subset of Smithsonian journal articles indexed in Web of Science for 2019 (1,888), there are 3,200 different affiliations and variants listed, with Smithsonian authors listing their respective affiliation in 267 different ways.

Smithsonian-Authored Journal Articles with More Than One Author Have Become Nearly Universal



ORCID: A Persistent Digital Identifier for Researchers

The Smithsonian Libraries and Archives is a member of ORCID (Open Researcher and Contributor Identifier), an international effort to streamline, unify and standardize research activity including affiliation, output, and attribution. The SRO program encourages scholars to register at ORCID. To date, the program staff are aware of 1,092 scholars (past and present) who have registered with ORCID.

What this means: The SRO program relies on other Institutional services to identify affiliations—generally the Outlook/PRISM system and the Office of Fellowships and Internships database. But SRO staff also rely on the actual publication listing a recognizable Smithsonian affiliation. The absence of coordination across data sources means that information about Smithsonian scholars must be manually maintained and subject to being incomplete and/or out of date. Developments like ORCID are promising, and widespread adoption should help alleviate these challenges.

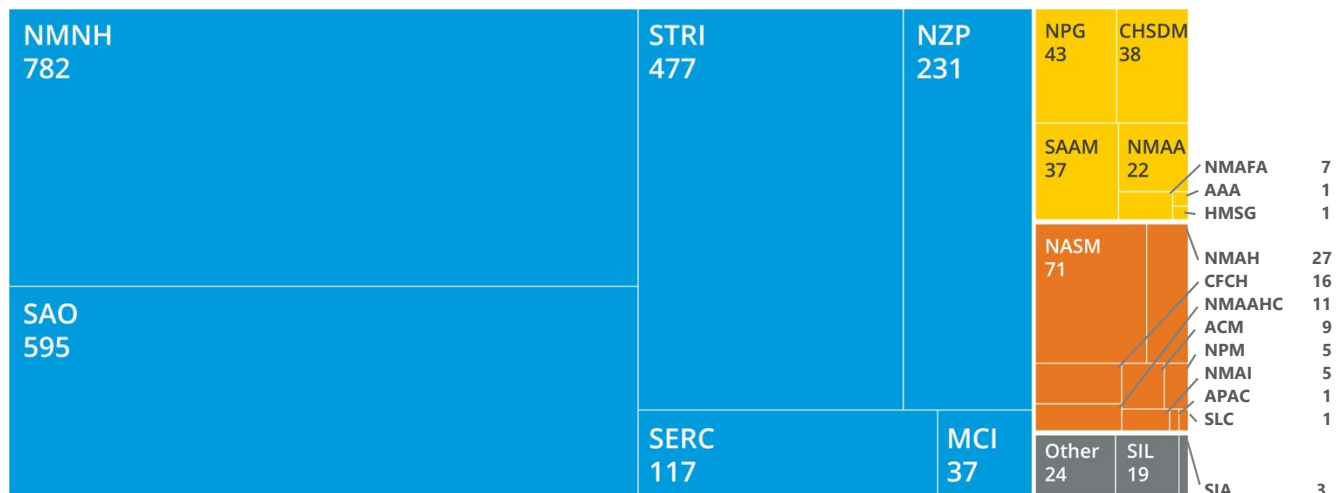
Notable:

- Each year Smithsonian scholars collaborate on publications with over **10,000 individuals** from outside organizations. (That number falls to **5,000** when omitting Smithsonian Astrophysical Observatory papers.)
- In 2019 publications with multiple authors, Smithsonian scholars were listed as the lead author more than **600 times**.
- **42** publications in 2019 included **100 or more co-authors**, while 25% of all publications included ten or more co-authors.
- Among the almost **90,000 publications** in SRO are papers co-authored by notables such as Noam Chomsky (*Manufacturing Consent*) and Paul Ehrlich (*The Population Bomb*). (And a frequent non-Smithsonian co-author with one of our Botany curators is D.H. Lorence (no relation))
- Collectively, Smithsonian scholars have received degrees from over **300 universities**, the most popular being Harvard, Cornell and George Washington University.

Smithsonian Units

With its 19 museums, 9 research centers, and the National Zoo, that span sciences, the arts, and history and culture, the Smithsonian is a complex organization. Research Online aims to accurately track the research output for these many organizations—not necessarily to be a tool for comparing this output between units. To track organizations in a way that is scalable, every research output added to Smithsonian Research Online must be attributed to at least one high-level organizational unit, or must have an accepted reason for being included despite this. The list of high-level organizations tracked for reporting includes over 24 different administrative organizations of the Institution (for example, STRI, NPG, etc.).

2019 Scholarly Outputs by Unit



Pre-Smithsonian

As a courtesy to newly-hired researchers, SRO includes outputs that were published before a researcher became affiliated with the Institution. These “Pre_SI” data populate web pages and Smithsonian Profiles for new staff. Currently there are over 900 (1%) “Pre_SI” items in the database. These may appear in searches of Research Online but are excluded from reporting of Smithsonian scholars’ outputs.

Legacy Outputs & Masters Theses

Smithsonian Research Online also includes legacy publications (identified broadly as U.S. National Museum) and master's theses produced from the Master of Arts in the History of Design and Curatorial Studies from Parsons School of Design & Cooper Hewitt, Smithsonian Design Museum, and the Master of Arts in Decorative Arts & Design History from Smithsonian Associates in partnership with the George Washington University Corcoran School of the Arts & Design programs.

What this means: Unit directors are ultimately responsible for deciding what each unit considers as a scholarly output fit for inclusion in Research Online. While SRO maintains a system that allows for many different research outputs to be included, this does not imply that all units are treating their content the same. Because disciplines have widely differing scholarly cultures, the research output by SI unit are not directly comparable to each other. Even within the sciences there is a wide variation in research outputs among units thanks to co-authorship rates, the inclusion of output from research associates, and the publishing customs of sub-disciplines.

Additionally, 75 publications published in 2019 included co-authors from multiple units. Reports from Research Online use whole or full counting and not fractional counting, which means that those 75 publications are counted in each organization's total. This type of counting ensures each unit knows how many outputs their authors are responsible for, but it is not conducive to proper comparison among those units.

Notable:

- One article, a letter to *Ecology Letters*, contained co-authors from **four different units**—NMNH, NZP, SERC, and STRI:
Chu, Chengjin, Lutz, James A., Kral, Kamil, Vrska, Tomas, Yin, Xue, Myers, Jonathan A., Abiem, Iveren, Alonso, Alfonso, Bourg, Norm, Burslem, David F. R. P., Cao, Min, Chapman, Hazel, Condit, Richard S., Fang, Suqin, Fischer, Gunter A., Gao, Lianming, Hao, Zhanqin, Hau, Billy C. H., He, Qing, Hector, Andrew, Hubbell, Stephen P., Jiang, Mingxi, Jin, Guangze, Kenfack, David, Lai, Jiangshan et al. 2019. "Direct and indirect effects of climate on richness drive the latitudinal diversity gradient in forest trees." *Ecology Letters* 22 (2):245–255. <https://doi.org/10.1111/ele.13175>.

Types of Research Output in Smithsonian Research Online

Smithsonian Research Online contains 21 types of research outputs. Many traditional outputs like journal articles are proactively collected by our staff. Research Online includes many formats such as websites, dissertations, and blog posts that go beyond traditional scholarly publications.

Expanded Reference Types

In 2019, SRO expanded to start collecting datasets, presentations, posters, and exhibitions. As a research organization with a strong public impact mission, providing the ability to include a multitude of research outputs best captures the true scholarly work of the Smithsonian.

Proactive Data Collection

Using online tools, Smithsonian Libraries and Archives staff identify and collect research outputs proactively on behalf of Smithsonian scholars, relieving them of the burden of administrative reporting. However, the information ecosystem that permits us to accomplish this was designed for scientific journal articles. Because of this discrepancy, webforms for submission are provided.

What this means: While proactive data gathering is one of the most important features for the success of Smithsonian Research Online, it does mean that there are most likely gaps in capturing other forms of scholarly output. This may result in an under-representation of items in other formats and those from non-science units, especially when looking at recent data. These manually-supplied items tend to arrive well after the calendar year is ended. Therefore, reports based on publication year or fiscal year can and do change over time, especially as total scholarly output data include these other output types. Improving the accuracy of reports that include these reference types requires the participation of the entire scholarly community and the encouragement of unit leaders.

Reference Type	Count (all years)	Count (2019)
Abstract	24	1
Chapter	11,244	122
Book	3,870	55
Computer Program	4	0
Conference Proceedings	464	4
Dataset	114	5
Dissertation/Thesis	500	4
Exhibition	57	23
Generic (Other Reference Types)	17	8
Journal Article	73,498	2,154
Magazine Article	131	9
Map	20	0
Newspaper Article	27	0
Blog	865	52
Patent	5	0
Poster	11	1
Oral Presentation/Lecture	235	48
Report	161	6
Sound Recording	8	0
Video/DVD	34	0
Web Page	301	19

Reporting: Calendar Year vs. Fiscal Year

While most reporting at the Smithsonian is done on a fiscal year basis, the data collected by Smithsonian Research Online on scholarly publishing is not conducive to this format. Until the infrastructure for collecting and storing publication metadata matures, any work-around to provide fiscal year data would be inconsistent and inaccurate.

Fiscal Year vs. Calendar Year

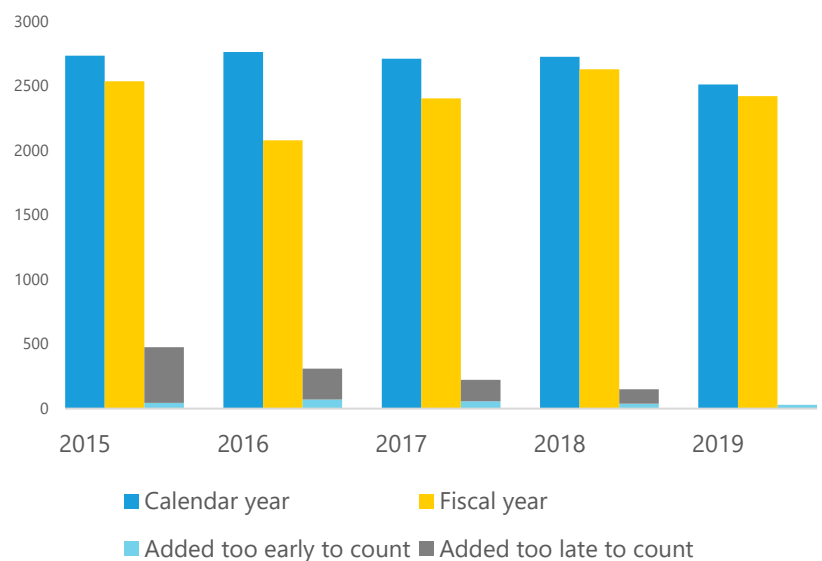
The date a publication is released is a data point that can be reliably captured. However, this is frequently limited to the four-digit year alone. SRO also records the date an item was added, but this can be well in advance or well past the publication year. It does not directly correlate to the publication year collected, and is therefore only useful as an administrative data point.

These data may imply that SRO can provide fiscal year data, but in truth, the only reliable reports that can be provided are those based on the calendar year in which an output was published. Even then, publication reports based on publication dates can and do change.

It can be argued that the publication year and the date added could approximate fiscal year data. However, items are added years before or after their publication date and are therefore never captured in fiscal year reporting, as the chart indicates. For example, of the 2,762 items published with a copyright year of 2016, 1,495 of them would count for FY 2016, 957 would count for FY 2017, while 70 were added too soon and 240 were added too late to count at all.

What this means: Reported numbers will frequently change as research outputs continue to be added and updated well after the date they are published. For this reason, a report showing total publications by publication date is subject to change over time, while a report by fiscal year is highly inaccurate and should be avoided.

Approximating the Fiscal Year Excludes 10% of Outputs



Open Access Publishing at the Smithsonian

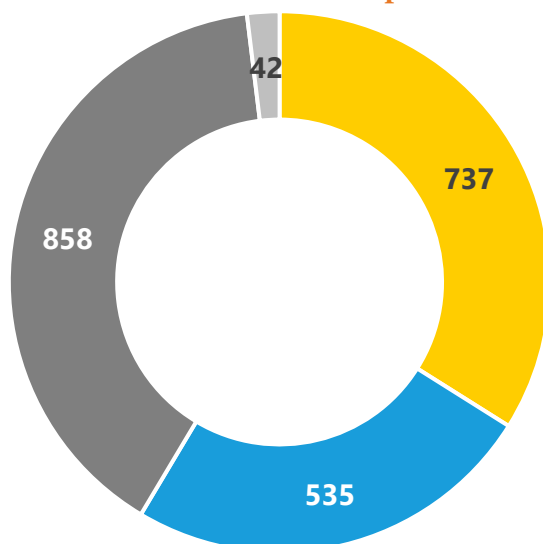
The Open Access publishing movement intends to make research accessible without cost to the reader. The trend towards this model of research distribution is evident in Smithsonian research outputs. Although public availability may depend on third parties (e.g. authors or their institutions) and are therefore not permanently accessible, many Smithsonian scholars publish their papers in open access journals or in journals which ensure public access.

The information on open access to Smithsonian scholarship presented here is based on reporting from certain standard sources and minimum publication elements. Because these sources focus on journal and article data, reporting here is limited to journal articles only.

Open Access via the Publisher

Publications can be made freely available from the publisher's website. This can be a journal-wide policy (gold or platinum open access) or one that makes specific articles open while the rest of the journal remains behind a paywall (hybrid or bronze open access). For either condition, the policy may or may not involve payments by authors or their institutions to grant open access. In 2019, 737 articles were made open access this way.

59% of 2019 Publications are Open Access



Open Access via Repositories

Publications can be made available through digital repositories (green open access). The SRO program operates a repository containing over 20,000 scholarly works authored by Smithsonian staff and appointees. Repositories can be subject-specific, country-specific, or institution-specific. SRO documents 535 articles in repositories that would otherwise be only available behind a paywall. There are certainly more, but articles in repositories that were already made open access via the publisher are included in that count.

1,272 Smithsonian journal articles from 2019 are openly available from publishers or in a repository.

	From Publisher	33.9% (737)
	In a Repository	24.6% (535)
	Behind Paywall	39.5% (858)
	Unknown	1.9% (42)

Notable:

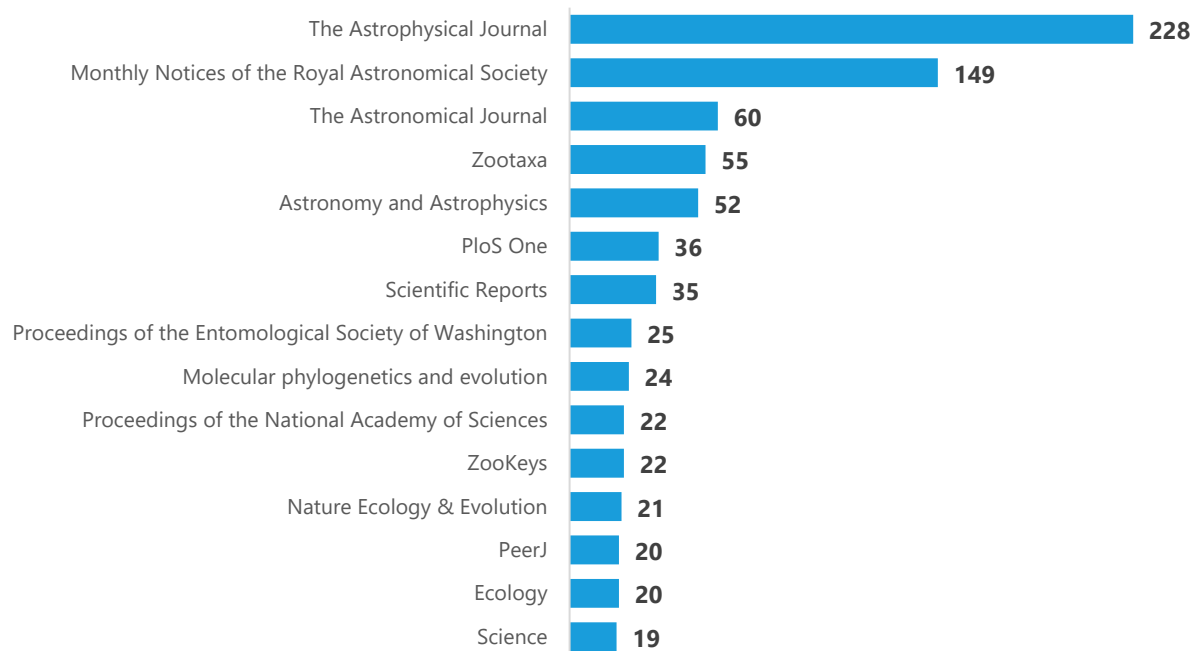
- In 2019 Smithsonian scholars were listed as author or co-author on **over 300 papers** in open access journals which operate an “author pays” model. Data on whether Smithsonian authors paid any, some, or all of the fees is not currently accessible. However, the total article processing charges of these papers for 2019 exceeded \$500,000.
- The Smithsonian Institution Scholarly Press Contributions series are an example of **platinum open access**, when an institution does not require payment by an author nor an institution in order to make research freely available.

Top Journals

Publishing research in journals is still the primary outlet for scholarly communication, especially for scientific research. Smithsonian Research Online includes 2,154 journal articles published in 2019, which represents 86% of all research outputs for the year. These articles are found in 642 different journals, including 19 articles in *Science*, nine articles in *Nature*, and 22 articles in *the Proceedings of the National Academy of Science*.

There are 642 different journals in which Smithsonian authors have published in 2019. Examining the 2019 journal articles in Research Online reveals a number of journals where Smithsonian authors are more likely to publish. This is probably reflective of disciplinary culture—four of the top five journals by number of articles are astronomy-related. Indeed, most units average one to two articles per journal while SAO has averaged 12.93 articles per journal in 2019.

27% of Journal Articles Published in 2019 are in These 15 Journals:



Measuring Impact

Citation metrics have set the standard for measuring the impact of scholarly research, generated for articles, journals, researchers, or even institutions. In recent years, alternative metrics have been developed to capture new ways that research is disseminated as digital infrastructure has expanded. Smithsonian Libraries and Archives makes available resources for determining the impact of research through subscriptions to Clarivate's Web of Science and Digital Science's Altmetric.com, in addition to freely available resources.

Citation Counts

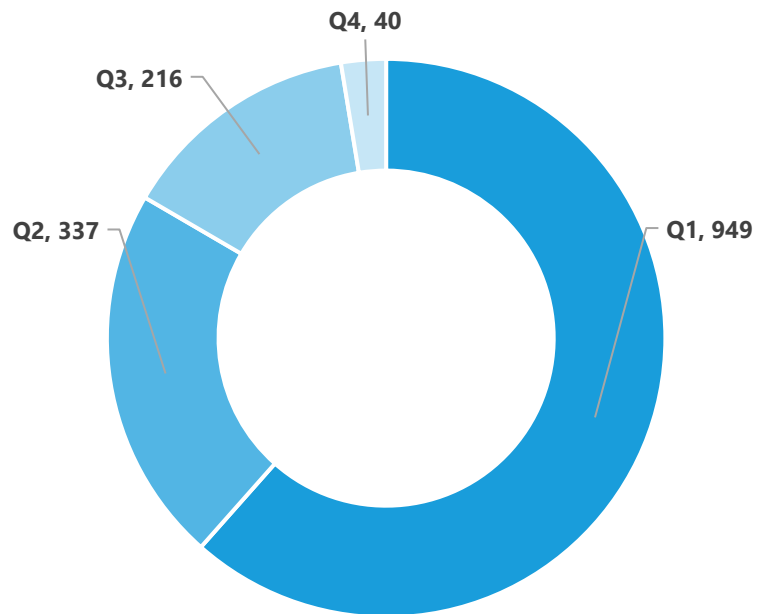
One of the fundamental measurements for research impact is the count of times other articles have cited a publication. In order to calculate these measurements, a source for citation data must be chosen. For institution-level analysis, these are typically the large citation index sources like Clarivate's Web of Science or Elsevier's Scopus. While available on a researcher level, Google Scholar is difficult to use as a source for the data mining necessary to analyze on an institutional level. Smithsonian Research Online relies on the Web of Science core collection for analysis, since the Smithsonian Institution includes analytics as part of its subscription to the Clarivate Web of Science package. Because of this, citation analysis herein must be limited to the sciences, where available data is much closer to complete.

The typical research output may take two to four years before its impact can be demonstrated in citation counts. This makes it too soon to have meaningful data on the impact of 2019 research. Sometimes, publications do amass citations more quickly. At the time of analysis, Smithsonian publications from 2019 that have DOIs and are indexed in Web of Science Core Collection have already been cited 3,668 times. Of these publications, 49 have been indicated by Web of Science as Highly Cited Papers, placing them in the top 1% of all articles based on the number of citations received by these articles when compared to other articles published in the same field in the same year. Web of Science also provides an indication for articles that meet their criteria for Hot Papers. Currently, there are six Hot Papers. A Hot Paper is in the top 0.1% of articles by citation count over the past two-month period as compared to other articles in the same field and added to Clarivate's database in the same period.

Journal Impact Factor

A journal's impact factor is a measurement of the average number of citations for articles published in the journal over the last two years. Journals with higher impact factors have articles that are more frequently cited in other articles. While comparison of journals by impact factor must take into account disciplines, Clarivate's Journal Citation Reports provides quartile ranking of all journals they index based on impact factor and field. With this data, SRO shows that more than half of all Smithsonian journal articles are published in journals in the top quartile of their field.

62% of Articles are in Top Quartile Journals



Altmetrics

Altmetrics show online attention paid to scholarly works either through social media or other digital communication channels. The Institution's research publications are regularly mentioned in a variety of platforms. Altmetric data from Altmetric.com are incorporated into Smithsonian Research Online search results in the form of a badge displayed. This badge includes a weighted score for outputs that have a digital object identifier (DOI) and have been indexed in Altmetric.com, along with a donut with colors representing different areas of attention.

What this means:

There are many opportunities for measuring the impact of research, and it is clear that metrics are important. It is also clear that metrics are open to many criticisms, from gaming the system to using a metric out of context. The best assessment takes into account both the quantitative and the qualitative when evaluating research. The Libraries and Archives has the source data that make many research metrics able to be calculated—the journals, databases, and services we subscribe to or compile are the main source data for measuring research impact.

While publishing an article in a specific journal, or counting how many times an article was cited cannot and should not be used to evaluate that article itself, these measurements can give an approximation of the importance of a journal or article. Measurements about journals and journal articles do allow for many different types of analysis to be run, from the h-index to Eigenfactor. Yet each of these metrics has strengths and weaknesses, and none should be considered outside of the context of that research—whether based on subject, age of the research, or aims of the researcher. While journals are the primary method for disseminating scientific research, this does not hold true for all disciplines. And as an institution with a varied mission, metrics and evaluation should reflect that varied mission.

Notable:

- Smithsonian research has accumulated over **two million** citations in the Web of Science.
- At the time of publication, Smithsonian journal articles indexed in the Web of Science have received **3,668** citations.
- Smithsonian research publications as far back as **2010** are still being mentioned on social media channels. This ten-year-old research output includes the following:
 - Schulte, Peter, Alegret, Laia, Arenillas, Ignacio, Arz, Jose A., Barton, P. J., Bown, Paul R., Bralower, Timothy J., Christeson, Gail L., Claeys, Philippe, Cockell, Charles S., Collins, Gareth S., Deutsch, Alexander, Goldin, Tamara J., Goto, Kazuhisa, Grajales-Nishimura, Jose, Grieve, Richard A. F., Gulick, Sean P. S., Johnson, Kirk R., Kiessling, Wolfgang, Koeberl, Christian, Kring, David A., MacLeod, Kenneth G., Matsui, Takafumi, Melosh, Jay, Montanari, Alessandro et al. 2010. "The Chicxulub Asteroid Impact and Mass Extinction at the Cretaceous-Paleogene Boundary." *Science* Mar 5;327(5970):1214–8. doi: 10.1126/science.1177265. https://www.altmetric.com/explorer/highlights?q=chicxulub&show_details=103997
 - Henry, Amanda G., Brooks, Alison S., and Piperno, Dolores R. 2011. "Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium)." *Proceedings of the National Academy of Sciences of the United States of America* 108 (2):486–491. <https://doi.org/10.1073/pnas.1016868108>. https://www.altmetric.com/explorer/highlights?q=neanderthal%20diets&show_details=100637

- Pimiento, Catalina, Ehret, Dana J., MacFadden, Bruce J., and Hubbell, Gordon. 2010. "Ancient Nursery Area for the Extinct Giant Shark Megalodon from the Miocene of Panama." *PLoS One* 5 (5): e10552. <https://doi.org/10.1371/journal.pone.0010552>.
https://www.altmetric.com/explorer/highlights?q=giant%20shark%20megalodon&show_details=104772
- Smithsonian research publications from 2019 have already been mentioned or cited in policy documents three times including documents from **The Analysis and Policy Observatory** (Australia), **UK Parliament Briefing Notes** (UK), and **Public Health England** (UK)

Appendix

Terms we use and why

Item vs. publication: SRO contains more than publications. It includes datasets, presentations, and now exhibitions.

Scholar vs. researcher: These terms are used interchangeably within this report, and reflect a broad definition of either.

Author vs. person: Research Online includes items that are not authored, but otherwise reflect intellectual work of Smithsonian-affiliated people or agents. These roles include editor, translator, data compiler, series editor, etc.

Full Counting vs. Fractional Counting: When a publication has authors with more than one Smithsonian organization or authors from different Smithsonian organizations, it can be counted with either full counting or fractional counting. If using full counting, such a publication would count as one publication for each organization, while fractional counting would assign a fraction of the paper to each organization.

Organization vs. Unit, Department, Museum, Facility, Research Center, etc.: The Smithsonian is a complex organization with many layers. Smithsonian Research Online has the capacity to track research only at a high organizational level. This is usually mutually exclusive administrative units at the museum or research center level. For NMNH and NASM, data is also collected on a departmental level.

Reporting Categories

Smithsonian Research Online reports data for the following categories. Additional administrative units are tracked by the program but not included in high-level reporting channels.

AAA – Archives of American Art

ACM – Anacostia Community Museum

APAC – Asian Pacific American Center

CFCH – Center for Folklife and Cultural History

CHSDM – Cooper Hewitt, Smithsonian Design Museum

HMSG – Hirshhorn Museum and Sculpture Garden

MCI – Museum Conservation Institute

NASM – National Air and Space Museum

NMAA – National Museum of Asian Art

NMAAHC – National Museum of African American History & Culture

NMAFA – National Museum of African Art

NMAH – National Museum of American History

NMAI – National Museum of the American Indian

NMNH – National Museum of Natural History

NPG – National Portrait Gallery

NPM – National Postal Museum

NZP/SCBI – National Zoological Park / Smithsonian Conservation Biology Institute

SAAM – Smithsonian American Art Museum

SAO – Smithsonian Astrophysical Observatory

SERC – Smithsonian Environmental Research Center

SIA – Smithsonian Institution Archives

SIL – Smithsonian Libraries

SLC – Smithsonian Latino Center

STRI – Smithsonian Tropical Research Institute

OP (+NCP/OFI/SDP/SISP) – Office of the Provost (National Collections Program, Office of Fellowships & Internships, Smithsonian Digitization Program, Smithsonian Institution Scholarly Press)

Smithsonian Research Online Program Information

Website: research.si.edu

Email: Research-Online@si.edu

Program Overview

Smithsonian Research Online (SRO) is a program which documents the research activity of Smithsonian scholars and makes the information available to the Institution and (with restrictions) to the general public. Information about publications, datasets, grants, areas of expertise and other demonstrations of Smithsonian scholarship are collected by the Smithsonian Libraries and Archives, standardized, and redistributed in the form of regular and custom reports as well as via unit websites, etc.

About this Report

The inaugural Smithsonian Research Online annual report was created by staff of the Smithsonian Libraries and Archives and released on April 22, 2020. This report uses data captured on Monday, March 23, 2020 from Smithsonian Research Online, Clarivate's Web of Science, and Altmetric.com. Data are available upon request. For questions and for feedback, please contact SRO staff at research-online@si.edu.