Smithsonian Research Online
2020 Annual Report

Smithsonian

Table of Contents

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 publications, 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.

This report is designed to raise awareness of SRO services 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. SRO strives to report on the scholarly output of the Smithsonian Institution. Using SRO to determine research impact is a key use-case for the program and one that is core to its mission. Use of SRO as a source for individual researcher metrics or evaluation is valuable only insofar as the data are complete and accurate—something which we pursue and for which we depend on continued feedback, updates and corrections.

SRO information is reused widely at the Institution including on Smithsonian Profiles, the Institution's expertise locator. Profiles indexes publications of federal and trust staff and uses this to identify experts at the Institution who can then be contacted with questions from either internal or external parties.

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.

2020 Program Accomplishments

During 2020 SRO program staff began . . .

- synchronizing Smithsonian research information with scholars' ORCID records. This will enable better tracking of research activity and reduce redundant data entry among all research staff.
- including rights and permissions statements in the Repository of full text documents corresponding to many research publications. These rights statements are in both human and machine-readable form, making the information more versatile.
- identifying scholars who have separated from the Institution and instead of removing them from Smithsonian Profiles, marking them as 'inactive' so that a historical record of research activity is retained while concealing their profiles from public view.
- identifying publications related to research done at Barro Colorado Island and incorporating this into workflows. This service may be expanded to other research facilities at STRI and the Institution at large but a special exception was made for BCI which will be celebrating its 100th anniversary soon.
- working in earnest with an OCIO contractor to develop a new system to manage research information and enable wider sharing and distributed inputs to the processes. This work is expected to culminate in a revamped tool which will streamline workflows and allow further development of services.

Comments and praise in 2020:

"Really great work. Thank you so much for what you do" . . . . NMNH Director

"I really appreciate you copying me on these [SRO reports]—they're always very helpful for me to see" . . . . Science Press Secretary

2021 Aspirations

During 2021 the Program has articulated the following high level goals:

- Move workflow to a new system which will allow for wider participation by Smithsonian staff, more timely updates and easier reuse of the information on SI websites and systems
- Provide a method for units to include information on Smithsonian research resources such as labs and research stations and to associate related research outputs
- Complete and establish a routine for synchronization of research information management with ORCID accounts of Smithsonian scholars
- Provide links to a broader range of publicly accessible works by Smithsonian scholars

2020 Summary Report

Data collected March 22, 2021

Research Productivity

- Smithsonian Research Online has tracked **2,804** research outputs published in 2020.
- **2,299** (82%) of these outputs are journal articles, published in 621 different journals, representing a 7% increase over the 2019 reported number (2,511)
- **51** (2%) are books, **80** (3%) are book chapters, which is 34% and 7% less than 2019, respectively.
- **374** other output types are indexed, including datasets, reports, posters, and exhibitions
- With more widespread usage of the Institution's Figshare account, there is a large increase in datasets reported in SRO (160%).

Smithsonian-Affiliated Scholars

- There are **1,191** identified individuals who have Smithsonian affiliation and have authored, edited, or otherwise created outputs listed in Smithsonian Research Online in 2020.

Collaborations

- **Over 85%** of scholarly outputs in SRO represent a collaboration with scholars at one of hundreds of outside research organizations. **102** outputs include collaborations among Smithsonian units.
- Research Online associates each research output to **49** different unit and departmental tags as appropriate, representing organizations across the Smithsonian Institution.

Open Access Publishing

- **1,501** (65%) journal articles in Research Online published in 2020 are openly accessible, either directly from the publisher or through a repository. This does not include currently embargoed publications that will become open after a set period of time.

Impact

- Smithsonian publications from 2020 that have DOIs and are indexed in Web of Science Core Collection have already been cited **5,661** times.
- Smithsonian research published in 2020 has been mentioned on Twitter **75,666** times, in news articles **8,410** times, on Facebook **1,051** times and have been cited in **484** Wikipedia articles.

Program Governance

- SRO is managed by staff of the Digital Programs and Initiatives division of the Smithsonian Libraries and Archives.
- An SRO Advisory Group meets several times per year to discuss issues and policies related to scope and services.
- The SRO Advisory Group includes representatives from the Libraries and Archives, Smithsonian Institution Scholarly Press, Office of Fellowships and the Office of the Under Secretary for Science and Research.
- Additional working relationships exist between SRO and both OCIO and the Office of Sponsored Projects, among other units.

Gathering Data

Reference Types

Of the 20+ types of research outputs in SRO, the great majority are books, articles and chapters. Other included formats such as websites, dissertations, and blog posts serve to illustrate the knowledge and expertise that Smithsonian staff and appointees have. While SRO staff endeavor to collect every book, chapter, and article, the addition of other research outputs is at the discretion of the unit and individual researcher.

Smithsonian Scholars

Every research artifact added to Research Online has at least one individual (author, editor, etc.) who is either employed by or has an academic appointment at the Institution. This generally can include federal or trust staff, post-doc/fellows, research associates, or in some cases visiting researchers. Scholars are considered active if they have authored a scholarly work dated 2020 and SRO staff is aware of their Smithsonian affiliation. There are over 1,000 active authors at the Institution who are known to Smithsonian Research Online.

A major challenge in working with bibliographic data is managing name variations for individual researchers. There are over 20,000 distinct name strings associated with publications SRO collected in 2020. Of these, approximately 2,000 are associated with slightly more than a thousand Smithsonian-identified authors.

Smithsonian Research Online staff add a unique identifier to associate name variations for individual Smithsonian scholars. Author information also includes duty station, ORCID, and other data. In 2020, SRO staff supplemented data on 3,171 name strings.

Smithsonian Units

Smithsonian Research Online aims to accurately track the research output for the nearly 30 museums, research centers, and the National Zoo. 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 organizations tracked includes over 35 different administrative organizations of the Institution (for example, STRI, NPG, etc.).

Proactive Data Collection

While collection of data on behalf of Smithsonian scholars is the aim of SRO program staff, a portion of research outputs are not easily captured by program tools—especially ones that are not available in digital format. For units covered by the Office of the Under Secretary of Science and Research (OUSSR), **96%** of outputs in 2020 were proactively added by the program's monitoring of research alerts and scraping metadata, in contrast to **21%** of outputs from units covered by the Office of the Under Secretary for Museums and Culture (OUSMC). A great majority of OUSMC research output in 2020 was manually added via the program's web form in the same year (79%).

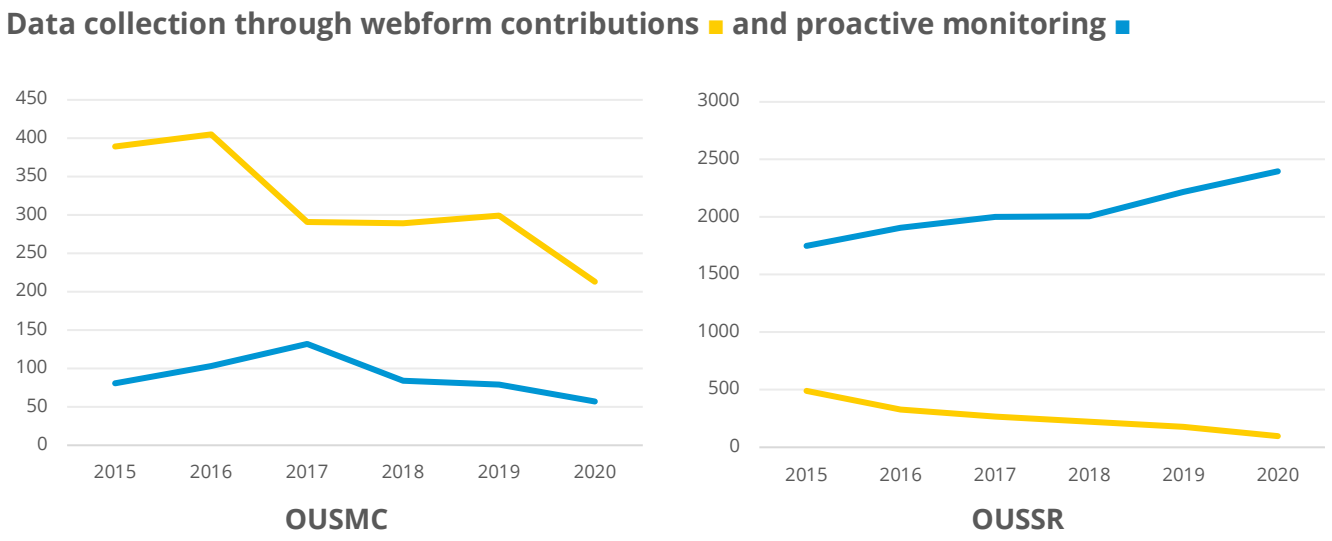

Figure 1: Number of contributions to Smithsonian Research Online per year for units under OUSMC and OUSSR by the method of addition, whether from monitoring research alerts from publishers and citation indexes, or contributions through the Smithsonian Research Online webform at http://staff.research.si.edu/contribute/.

A Note about Dates

Because a growing portion of publication information is supplied by authors, the exact date of publication often cannot be easily determined. Publications contributed to the SRO database may be added months or years after the copyright date listed. Reported numbers can change as research outputs continue to be added and updated well after the date they are published. However, these manually-supplied items also tend to arrive well after the items are published—sometimes months later. For this reason, a report showing total publications by publication date is subject to change over time, and a report by fiscal year is useful only to demonstrate level of effort in contributing to the database—not necessarily research productivity during the period. Previously under-represented items and those from non-science units may be growing.

Discussion

Unit directors are ultimately responsible for deciding what each unit considers as a scholarly output fit for inclusion to Smithsonian 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 units 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 each sub-discipline.

To identify affiliation, the SRO program relies on other Institutional systems—Outlook/PRISM, the Office of Fellowships—but also on the publication itself which should list a recognizable Smithsonian affiliation. The absence of coordination across data sources means that information about Smithsonian scholars must be manually maintained and is subject to being incomplete and/or out of date.

Additionally, 102 publications published in 2020 included co-authors from multiple units. Reports from Smithsonian Research Online use whole or full counting and not fractional counting, which means that those 102 publications are each counted once toward an organization's total.

Internal Collaborations

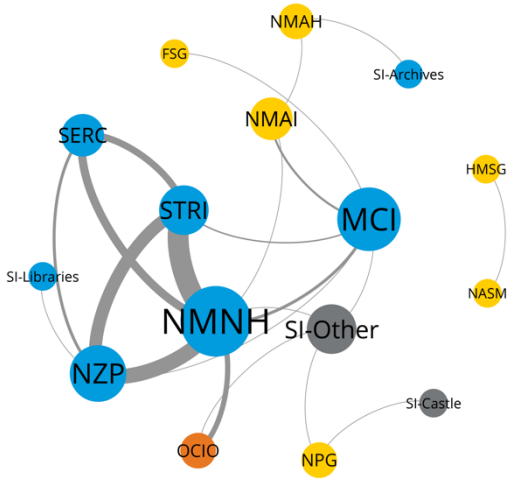

Figure 2: Network graph of collaborations between units on 2020 publications. Node size is based on the number of outputs with collaboration. Connecting line size indicate number of outputs worked on together.

Notable:

- Smithsonian scholars collaborated with upwards of **4,000** different academic, corporate, government, and museum organizations in 2020.
- In 2020 journal articles with multiple authors, Smithsonian scholars were listed as the lead author **610 times**.
- **33** publications in 2020 included **100 or more co-authors**, while 33% of all 2020 publications included ten or more co-authors.
- Among the almost **94,000 publications** in SRO are items co-authored by notables such as biologist, E.O. Wilson, science historian, Derek de Solla Price, journalist, Pete Hamill and ichthyologist, Gerald R. Allen.
- One article on bird genetic sampling contained co-authors from **four different units**—NMNH, NZP, OCIO, and STRI:
  Feng, Shaohong, et al. 2020. "Dense sampling of bird diversity increases power of comparative genomics." Nature 587 (7833): 252– 257. https://doi.org/10.1038/s41586-020-2873-9

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 less-than-reliable, many Smithsonian scholars publish their papers in open access journals or in journals which ensure public access.

The information presented here on the open access status of Smithsonian scholarship is limited to journal articles only. This limited focus is due to reliance on available standard data sources and minimum publication elements.

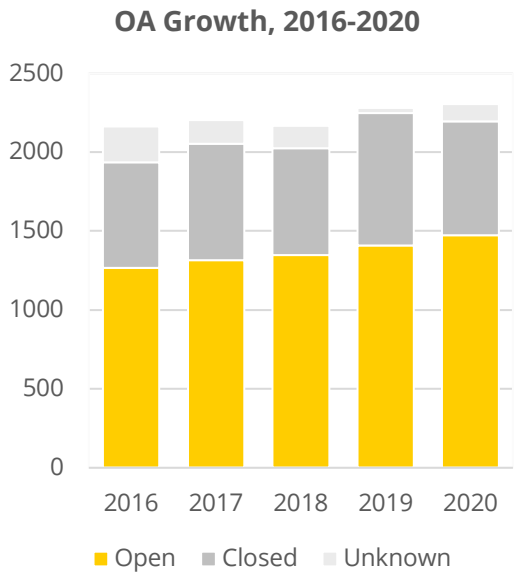

Figure 3: Open Access status of Smithsonian-authored journal articles. Data sources: Unpaywall.org and DOAJ.org. "Unknown" are either DOIs not covered by Unpaywall or were erroneous at the time of analysis.

Public Access Mandate

SRO has worked with the Smithsonian Institution Scholarly Press to implement the Institution's Public Access Plan for Federally Funded Research Materials, ensuring that over **400 works** created by Smithsonian scholars are available for reading and download by the general public. This is in addition to the many publications issued under open licenses.

Types of Open Access

Publications can be made freely available from the publisher's website. This can be a journal-wide policy or one that makes specific articles open while the rest of the journal remains behind a paywall. For either condition, the policy may or may not involve payments by authors or their institutions to grant open access.

Publications can also be made available through repositories. These repositories can be subject-specific, country-specific, or institution-specific. The SRO program operates its own repository containing over 20,000 scholarly works authored by Smithsonian staff and appointees, including 57 articles from 2020.

The source and licensing variations of open access policies are often conveyed by associating each condition with a color. Breaking down open access this way helps to reveal trends and explore the impact of changing models of scholarly publishing.

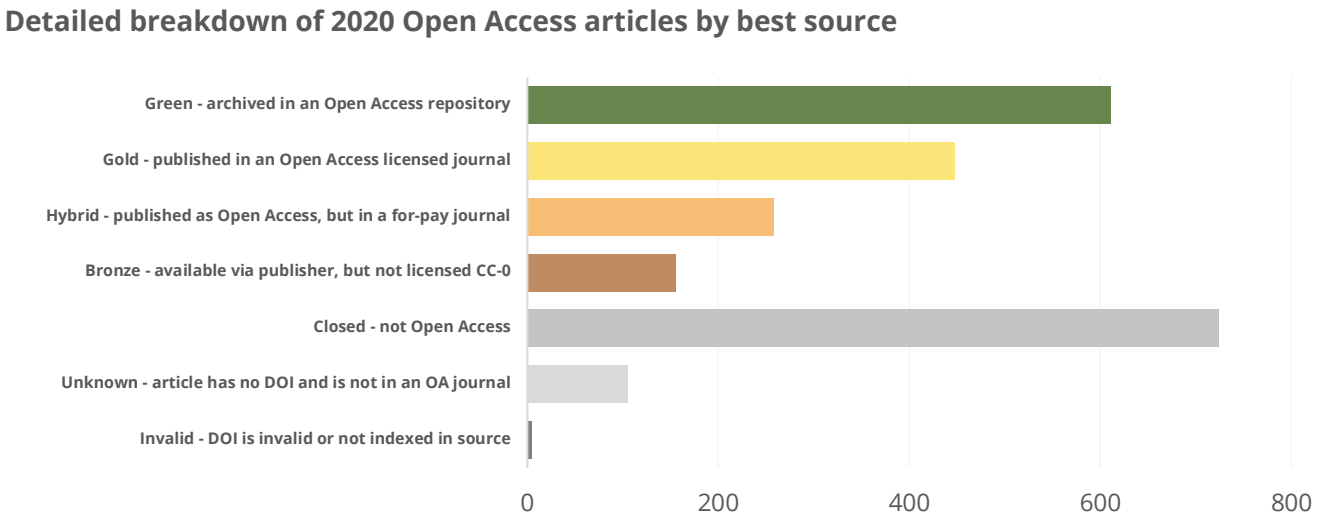

Figure 4: Number of journal articles by the category of Open Access status based on Unpaywall.org definitions. Data is for the "best" OA status, so articles are only counted once—even if they are OA by multiple means. See the following for more information: https://support.unpaywall.org/support/solutions/articles/44001777288-what-do-the-types-of-oa-status-green-gold-hybrid-and-bronze-mean-

Notable:

- In 2020 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 2020 exceeded $500,000.
- The Smithsonian Institution Scholarly Press Contributions Series and Open Monographs are examples of **platinum open access,** used to indicate when a publisher pays to make research freely available and does not charge authors.

Journals

Publishing research in journals is still the primary outlet for scholarly communication, especially for scientific research. Smithsonian Research Online includes **2,299** journal articles published in 2020, which represents **82%** of all research outputs for the year. These articles are found in **684** different journals, including 24 articles in *Science*, 20 articles in *Nature*, and 27 articles in *Proceedings of the National Academy of Sciences*.

Top Journals

There are **684** different journals in which Smithsonian authors have published in 2020. Most of these journals can be found in Clarivate's Web of Science with a specific category and a broader category. These tables show top journals by count of articles in journals in three areas: multidisciplinary journals, life sciences journals, and physical sciences journals.

Journals with the most Smithsonian-authored contributions for 2020, by category

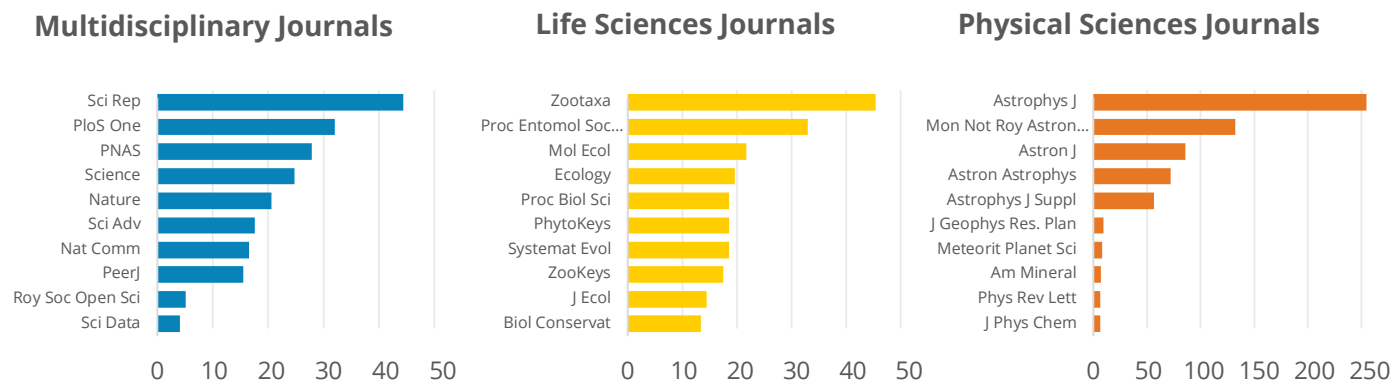

Figure 5: Each bar chart shows number of Smithsonian-authored contributions for the top 10 journals based on the Web of Science classification for that journal, for the publication year 2020. Journal names are standard abbreviations via http://journalseek.net

Journal Subject Categories

Journal subject classification provides an opportunity to explore the types of research produced by the Smithsonian Institution. Clarivate's Web of Science is a widely used bibliometric tool that assigns each journal to at least one subject category. This subject classification is one way to demonstrate the focus of research in 2020. The circular dendrograph below represents the count of articles based on the subject classification of the journal in which it is contained, demonstrating the breadth of research across the Institution.

Count of articles by subject category of journal

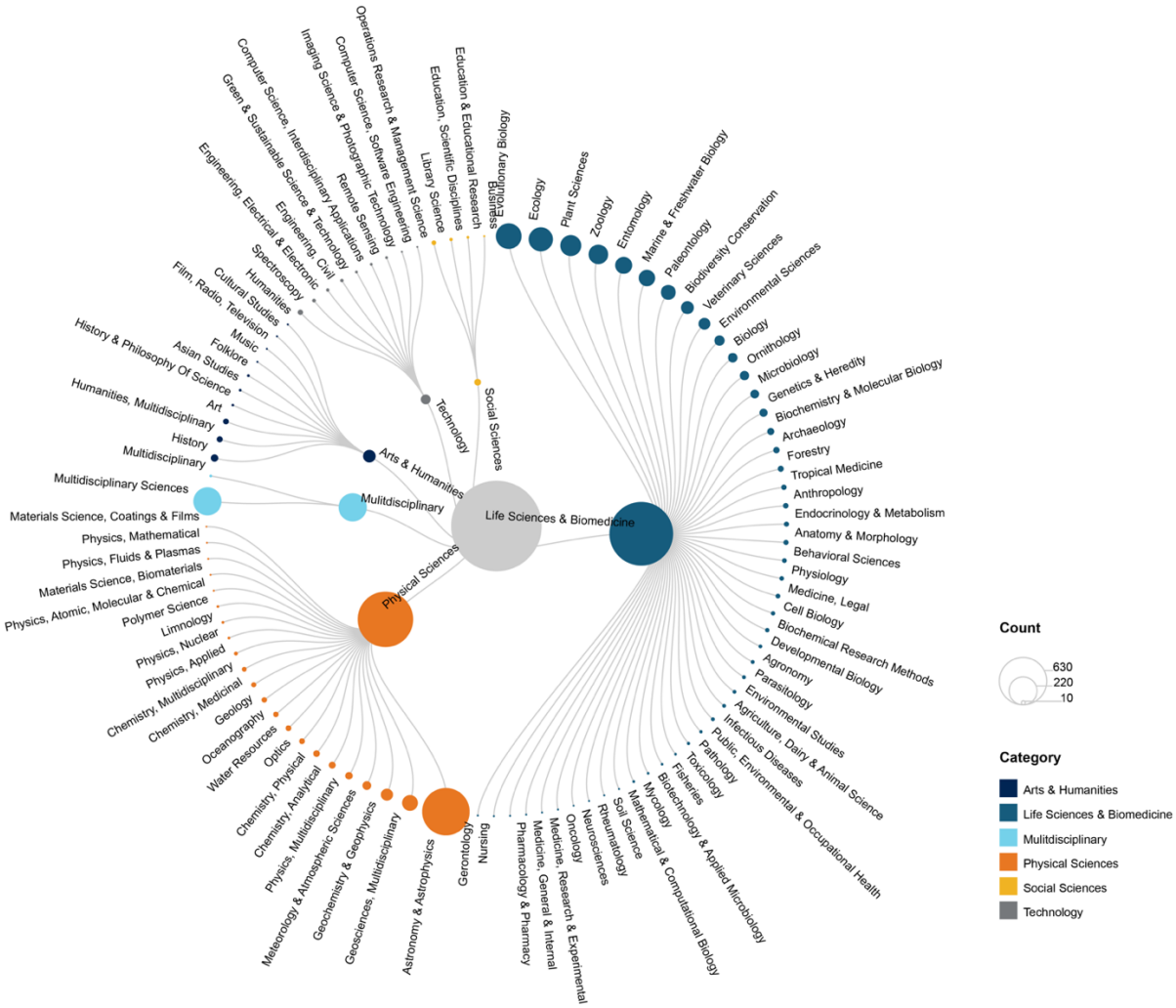

Figure 6: Circular Dendrograph showing Web of Science subject classification of journals (color) and the number of articles in journals with that category (circle size). Note that journals often have multiple categories, so articles can count towards more than one subject.

Impact and Metrics

Measuring the impact of scholarly research has traditionally been done using citation metrics. 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.

Citation Counts

Tools like Web of Science do allow for exploration of citation data, such as calculating the average citation rate by year. This is exhibited in figure 7 for Smithsonian research 2013-2019. However, the typical research output takes years before its impact can be demonstrated in citation analysis. Meaningful data on the impact of 2020 research is therefore not yet available.

The Web of Science does make two indicators available for recent publications: Highly Cited Papers and Hot Papers. 39 papers have been indicated 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. Six publications are noted as Hot Papers, entering 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. Smithsonian publications from 2020 that have DOIs and are indexed in Web of Science Core Collection have already been cited 5,661 times.

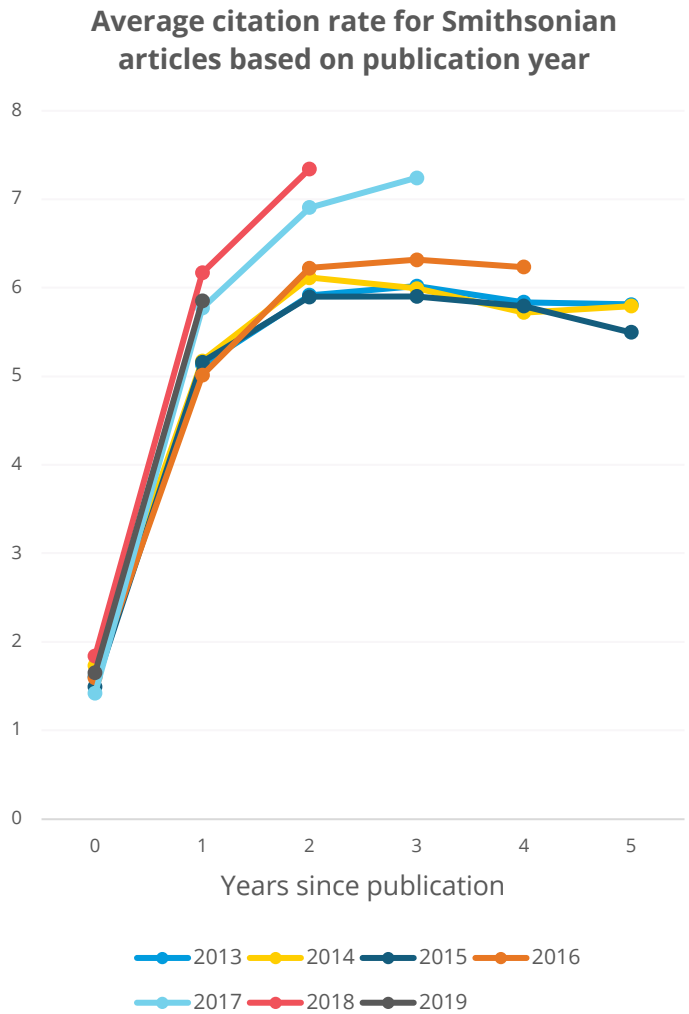

Figure 7: Total number of citations received by publications published in a given year divided by the total number of publications, giving the average number of citations each year across the number of years since it was published. Data source: Web of Science Core Collection.

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. While comparison of journals by impact factor must take into account disciplines, SLA has access to tools which provides quartile ranking of most mainstream journals impact factor and discipline. For example, SRO data shows that in 2020, more than half of all Smithsonian-authored articles are published in journals in the top quartile of their field (59%), based on the impact factor of the journal.

Despite known limitations, measurements about journals and journal articles do allow for many different types of analysis to be run. 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.

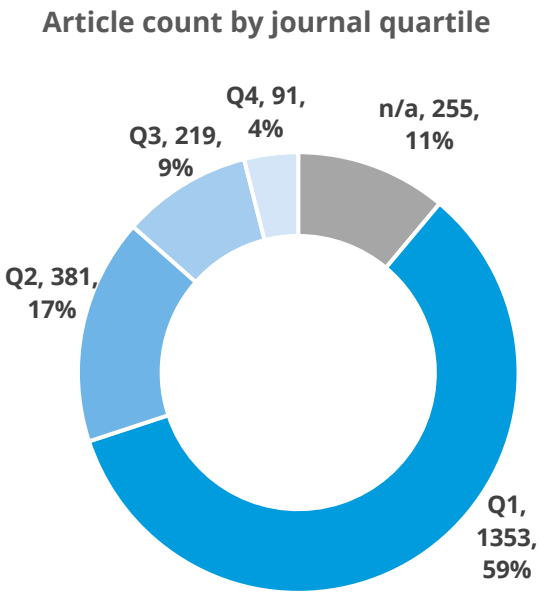

Figure 8: Number of articles in journals based on quartile rank of impact factor by subject in the Web of Science Journal Citation Report. Note that impact factor does not necessarily reflect the quality of the articles within the journal directly.

Altmetrics

The Institution's research publications are regularly mentioned in a variety of media platforms. Altmetrics show online attention paid to scholarly works either through social media or other digital communication channels. Altmetric data is often expressed in the form of a badge/icon which includes a numeric score along with colors representing different media.

Figure 9: Example of an Altmetric badge

During 2020, Smithsonian research was mentioned in:

- Policy Documents, including publications of the WHO, IUCN and FAO
- Faculty 1000, in 17 Faculty Opinions (F1000)
- Wikipedia, with hundreds of mentions including in articles on the Origin of the Domestic Dog; Uma Thurman; and 40 mentions in "2020 year in . . ." articles
- News sources, such as BBC, Al-Jazeera, CNN, New York Times and many others

Two of the articles mentioned in these policy documents relate to research on SARS-CoV-2, relating to vertebrate viral vectors[1] and filtration of aerosols by cloth masks.[2]

Top 5 articles by Altmetric Score for 2020

Rebuilding marine life
Article in ***Nature***, April 2020

Origins and genetic legacy of prehistoric dogs
Article in ***Science***, October 2020

Asynchronous carbon sink saturation in African and Amazonian tropical forests
Article in ***Nature***, March 2020

Body dimensions of the extinct giant shark Otodus megalodon: a 2D reconstruction
Article in ***Scientific Reports***, September 2020

An extremely energetic supernova from a very massive star in a dense medium
Article in ***Nature Astronomy***, April 2020

While Altmetric mentions tend to be concentrated around journal articles—and therefore, the sciences—among the humanities outputs from the Institution that garnered widespread attention were two books: *Health Design Thinking* (Ellen Lupton, CHDNM) and *Operation Moonglow* (Teasel Muir-Harmony, NASM)

The Smithsonian Libraries and Archives maintains a subscription to Altmetric Explorer which yields data based on SRO-listed publications. In addition to media mentions, Altmetric Explorer contains information on institutional collaboration, research sponsorship and Open Access status.

For example, in 2020 there were over 1,300 Smithsonian research outputs which:

- Listed Smithsonian collaborations with over 1,500 research institutions including 19 Max Planck institutes and 10 universities in the University of California system.
- Funded in part by 300 distinct funding agencies (although Smithsonian scholars were not necessarily listed as principal investigators on any/all of these awards).

---

[1] Damas, Joana, Hughes, Graham M., Keough, Kathleen C., Painter, Corrie A., Persky, Nicole S., Corbo, Marco, Hiller, Michael, Koepfli, Klaus-Peter, et al. 2020. "Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates." *Proceedings of the National Academy of Sciences of the United States of America* 117 (36): 22311– 22322. https://doi.org/10.1073/pnas.2010146117

[2] Zangmeister, Christopher D., Radney, James G., Vicenzi, Edward P., and Weaver, Jamie Lynn. 2020. "Filtration Efficiencies of Nanoscale Aerosol by Cloth Mask Materials Used to Slow the Spread of SARS CoV-2." *ACS Nano* 14 (7): 9188– 9200. https://doi.org/10.1021/acsnano.0c05025

Appendix

Reporting Categories

Science and Research	Museums and Culture	Education	Administration
Museum Conservation Institute National Museum of Natural History National Zoological Park | Smithsonian Conservation Biology Institute Harvard-Smithsonian Center for Astrophysics Smithsonian Environmental Research Center Smithsonian Institution Scholarly Press Smithsonian Libraries and Archives Smithsonian Tropical Research Institute	Archives of American Art Anacostia Community Museum Asian Pacific American Center Center for Folklife and Cultural History Cooper Hewitt, Smithsonian Design Museum Freer Gallery of Art | Arthur M. Sackler Gallery Hirshhorn Museum and Sculpture Garden National Museum of African American History & Culture National Museum of African Art National Museum of American History National Museum of the American Indian National Portrait Gallery National Postal Museum Smithsonian American Art Museum Smithsonian Latino Center	Smithsonian Center for Learning and Digital Access Smithsonian Science Education Center	Smithsonian Organization and Audience Research Smithsonian Gardens

Reference Type and Count of Items in Smithsonian Research Online

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Abstract	0	1	0	0	0	0	2	2	3	1	3
Book, Section	249	254	275	469	278	245	266	205	162	189	80
Book, Whole	126	115	110	122	97	76	84	78	83	69	51
Computer Program	0	0	3	0	0	0	0	0	0	0	2
Conference Proceedings	6	7	7	46	40	9	10	5	14	6	5
Dataset	0	1	3	2	13	6	25	33	27	77	200
Dissertation/Thesis	9	12	13	8	9	10	16	10	15	4	1
Exhibition	1	1	0	4	3	1	5	2	4	28	7
Generic	1	0	2	0	1	0	0	2	3	8	0
Journal Article	1,754	1,959	2,153	2,048	1,970	2,220	2,164	2,205	2,165	2,280	2,304
Journal, Electronic	3	8	8	6	3	4	2	1	1	0	0
Magazine Article	2	5	3	5	8	10	6	6	12	10	16
Map	0	0	0	1	1	0	0	0	0	0	0
Newspaper Article	1	0	0	2	0	1	4	0	6	0	14
Online Discussion Forum/Blogs	23	32	72	92	116	138	131	117	105	74	51
Patent	0	0	0	0	0	0	0	1	0	0	0
Poster	0	0	0	0	1	3	1	1	0	1	0
Presentation	9	13	9	8	11	23	33	26	21	53	55
Report	1	1	1	0	1	1	8	3	9	5	8
Sound Recording	1	0	0	0	0	0	0	0	0	0	0
Video/DVD	2	1	2	1	1	1	1	1	0	1	1
Web Page	15	9	17	17	23	17	36	35	19	20	11

Program Information

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 at research.si.edu. 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.

Please direct correspondence to:
Smithsonian Research Online
research-online@si.edu

About this Report

The Smithsonian Research Online annual report was created by staff of the Smithsonian Libraries and Archives and released on April 13, 2021. This report uses data captured on Monday, March 22, 2021 from Smithsonian Research Online, Clarivate's Web of Science, Altmetric.com, Unpaywall.org, and The Directory of Open Access Journals. Data are available upon request.

Suggested Citation:

Hutchinson, Alvin R. and Naples, Richard M. 2021. *Smithsonian Research Online Annual Report 2020.* (Washington, D.C.: Smithsonian Libraries and Archives).