Smithsonian Research Online
2021 Annual Report

Smithsonian
Libraries and Archives

Table of Contents

Program Overview

- Smithsonian Research Online (SRO) collects and manages information describing research conducted at the Smithsonian.

- SRO is managed by staff of the Digital Programs and Initiatives division but relies on support from staff across the Smithsonian Libraries and Archives (SLA).

- The information collected is stored centrally and shared widely via regular and custom reporting as well as Smithsonian systems including Smithsonian Profiles.

- SLA staff rely on the researchers for notification of new additions and verification of the accuracy of existing data.

Smithsonian Research Online (SRO) program documents the research activity of Smithsonian scholars and makes the information available to the Smithsonian community. The service, administered by the SLA allows Smithsonian leadership, scholars, and the public to identify research products created by Smithsonian staff and appointees.

The Program collects and stores information about publications, datasets, grants, areas of expertise and other demonstrations of Smithsonian scholarship. It is standardized and reused widely at the Institution in the form of regular and custom reports as well as via unit websites and other online systems including on Smithsonian Profiles, the Institution's expertise locator. Profiles indexes research information of federal and trust staff and leverages this to identify experts at the Institution who can then be contacted as needed by internal or external parties.

This annual report is designed to highlight Program accomplishments, notable research activity and to provide examples of possible analysis. Using SRO to highlight research impact is a key use-case for the program and one that is core to its mission[1].

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

---

[1] NOTE: The 2021 publications count, and subsequent metrics *may* be lower than in years past, likely due to disruptions to research activities due to COVID-19 shutdowns beginning the previous year.

2021 Program Highlights

During 2021 the SRO program . . .

- Performed a modest overhaul on the underlying technology systems for the Program which resulted in the automation of certain tasks. Among other things, these alterations further automated the movement of information between systems.

- Improved the API output including a uniform, formatted citation as the main component for reusing data by all systems and web sites, simplifying information delivery.

- Began a pilot service to associate research facilities with published research. Initially this will allow reporting on research from STRI research stations and selected instruments.

- Began including all research output types in Smithsonian Profiles (previously formats were limited to traditional publications/works).

- Migrated to a new platform for documenting workflow, tasks, user support and other processes. While there is no discernible benefit to SRO program users, this resulted in greater efficiency in program management.

- Hosted a LEADING[2] fellow who worked on a project to do text mining on Smithsonian-authored publications.

- Began systematically listing research data sets that have been deposited in the Institution's FigShare repository; over 400 have been included.

- Suspended work with a contractor on a new SRO system for managing information. Alternative solutions will be explored in the future.

- Standardized the updating of Smithsonian scholars' ORCID records with both published journal articles an institutional affiliation using industry-standard persistent identifiers.

- Enlisted the help of SLA staff who normally work with physical collections but who have been unable to because of COVID-related museum closures.

- Contributed to the revision of the Institution's plan for Public Access to federally funded research materials including a change in policy for archiving materials.

---

[2] LIS Education And Data Science Integrated Network Group, an IMLS supported program

2022 Goals

For 2022 the Program has the following high-level goals:

- Incorporate open access (OA) status for all SRO works with a DOI (digital object identifier). Among other things, this will enhance compliance reporting with the SI's Public Access Plan and allow all data from SRO to link to a reliable OA copy where applicable.

- Monitor published literature which refers to research resources (e.g., scientific instruments) for which there is a standard identifier registered and evaluate the usefulness of registering additional Smithsonian research instruments or facilities based on results.

- Work with EDAN to send research information regularly and uniformly between systems. A primary objective would be to make it much easier for existing units to reuse this information on individual curator/scholar web pages, particularly those without a dedicated webmaster.

A Note on Persistent Identifier Usage

Unique, persistent identifiers such as DOIs and ORCID ids enable interoperability of data from a variety of systems and multiply the value of research information. In 2021, SRO staff increased the use of these identifiers recognized by information and research professionals including:

- The large-scale updating of Smithsonian scholar ORCID records with publication and affiliation data from SRO. Over 250 scholars have opted in to have their ORCID profile updated automatically and without manual intervention.

- The registration of Smithsonian instruments (STRI Canopy Cranes; High Performance Computing Cluster) and facilities (STRI research stations) with recognized identifier schemes. These identifiers are published on instrument and facility web pages and users are encouraged to cite these resources (using the identifier) in publications resulting from use.

- The creation of instructions to assist National Science Foundation grant applicants at the Institution to generate the required SciENcv biosketch using data from their ORCID and circumvent the need to manually reproduce that information.

- Participation on the persistent identifier working group of the OSTP/NSTC Subcommittee on Open Science which promotes the use of identifiers in federal science research information systems.

Research Output at a Glance

Data collected March 1, 2022

Research Productivity

- Smithsonian Research Online has tracked **2,784** research outputs published in 2021.
- **2,376** (85.3%) of these outputs are journal articles, published in 668 different journals, representing a 3.3% increase over the 2020 reported number (2,299).
- **83** are book chapters and **37** are books, which is 14.4% and 33.9% less than 2020, respectively.
- **288** other outputs were collected including datasets, reports, exhibitions, and other formats.

Smithsonian-Affiliated Scholars

- There are **908** identified individuals who have Smithsonian affiliation and have authored, edited, or otherwise created outputs listed in Smithsonian Research Online in 2021.

Collaborations

- Over **83.5%** of scholarly outputs in SRO represent a collaboration with scholars at one of thousands of outside research organizations. **104** outputs include collaborations among Smithsonian units.
- Research Online attributes each research output to **48** different unit and departmental tags as appropriate, representing organizations across the Smithsonian Institution.

Open Access Publishing

- **1,646** (68.5%) of journal articles in Research Online published in 2021 are openly accessible, either directly from the publisher or through a repository. This does not include currently restricted publications that will become open after embargo expiration.

Impact

- Smithsonian publications from 2021 that have DOIs and are indexed in Web of Science Core Collection have already been cited **5,017** times.
- Smithsonian research published in 2021 has been mentioned on Twitter **33,000** times, in news articles **3,326** times, on Facebook **308** times and have been cited in **265** Wikipedia articles.

Individuals and Smithsonian Units

Every research output added to Research Online is the output of at least one individual (author, editor, etc.) who is either employed by or has an academic appointment at the Institution.

Likewise, 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. To identify affiliation, the SRO program relies on both Institutional systems (Outlook/PRISM, the Office of Fellowships) and the publication itself to verify or confirm Smithsonian affiliation.

Notable:
- Smithsonian scholars collaborated with scholars representing over **2,000** different academic, corporate, government, and museum organizations in 2021.
- In 2021 journal articles with multiple authors, Smithsonian scholars were listed as the lead author **658 times**
- **25** publications in 2021 included **100 or more co-authors**, while 30% of all 2021 publications included ten or more co-authors.
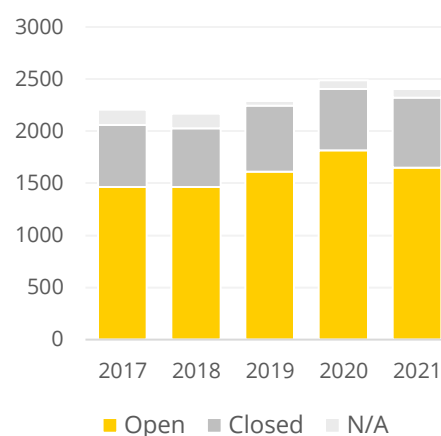
Open Access Publishing at the Smithsonian

The trend toward Open Access (OA) publishing is evident in Smithsonian research outputs and many Smithsonian scholars publish their papers in open access journals or where public access is otherwise ensured. Reports on the open access status of Smithsonian scholarship is limited to journal articles only due to availability of standard data sources and minimum publication elements.

OA Growth, 2017-2021

Figure 1: 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.

Notable:
During 2021, the Smithsonian Institution Scholarly Press began using a new system to capture information for the Institution's Public Access plan. SRO staff met regularly with Scholarly Press staff to refine the system and the plan to reflect attainable outcomes.

- In 2021 Smithsonian scholars were listed as author or co-author on **446** articles in open access journals which assess an article processing charge. (Data on whether Smithsonian authors paid any, some, or all the fees are not currently accessible, however the OA fees for these papers likely exceeds $930,000 for the year).
- In addition to OA publications freely available from the publisher's website, the SRO program operates its own repository containing over 20,000 scholarly works authored by Smithsonian staff and appointees. (https://repository.si.edu)

Journals

Publishing research in journals is still the primary outlet for scholarly communication, especially in the sciences. Smithsonian Research Online includes **2,394** journal articles published in **668** different journals during 2021. This includes 13 articles in *Science*, 16 articles in *Nature*, and 33 articles in *Proceedings of the National Academy of Sciences*.

Impact and Metrics

Measuring the impact of scholarly research has traditionally been done using citation metrics. Smithsonian Research Online relies on licensed commercial resources for analysis. The Smithsonian Institution includes analytics as part of its subscription although the data is limited to the sciences, where available data is much closer to complete.

Citation Counts and Journal Impact Factor

Calculating the average citation rate by year is valuable, however the typical research output takes years before its impact can be demonstrated in citation analysis. Meaningful data on the impact of 2021 research is therefore not yet available.

While comparison of journals by impact factor must take into account disciplines, SLA currently licenses tools which provide ranking of most mainstream journals' impact factor and discipline.

Despite known limitations, measurements about journals and journal articles do allow for many different types of analysis to be run. (See *Selected Analysis* for more information)

Notable:
The Web of Science includes the classifications, "Highly Cited Papers" and "Hot Papers". Among Smithsonian papers indexed in 2021, 29 have been indicated as Highly Cited, placing them in the top 1% and 2 were included in Hot Papers (top 0.1%), when compared to other articles published in the same field in the same time period.

SRO data shows that in 2021, more than half (66.4%) of all Smithsonian-authored articles are published in journals in the top quartile of their field, based on the impact factor of the journal.

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.

During 2021, Smithsonian research was mentioned in:

- Faculty 1000, in 10 Faculty Opinions (F1000),

- Wikipedia, with hundreds of mentions including in articles on Lyme Disease; 2021 Year in Science; Land Use; and Climate Change and Infectious Diseases,
- The World Economic Forum and international news sources, such as Jerusalem Post and South Africa Today, as well as local (U.S.) public radio stations from Iowa, Hawaii, Maine, Alaska, and many other states in the U.S.

While Altmetric mentions tend to be concentrated around the sciences, the outputs that garnered the greatest attention among the humanities were books *The Age of Acrimony (*Jon Grinspan, NMAH*)* and *American Independent Inventors in an Era of Corporate R&D* (Eric S. Hintz, NMAH).

In addition to media mentions, Altmetric Explorer contains information on institutional collaboration, research sponsorship and Open Access status. For example, in 2021 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.
- Were funded in part by 300 distinct funding agencies (although Smithsonian scholars were not necessarily listed as principal investigators on any/all of these awards).

Selected Analysis

Reports and visualizations can be customized to meet user demands. The options and possibilities are too numerous to list but below are several examples that are possible with the SRO data and commonly available tools. Smithsonian research managers are welcome to discuss specific needs for analysis with program staff.

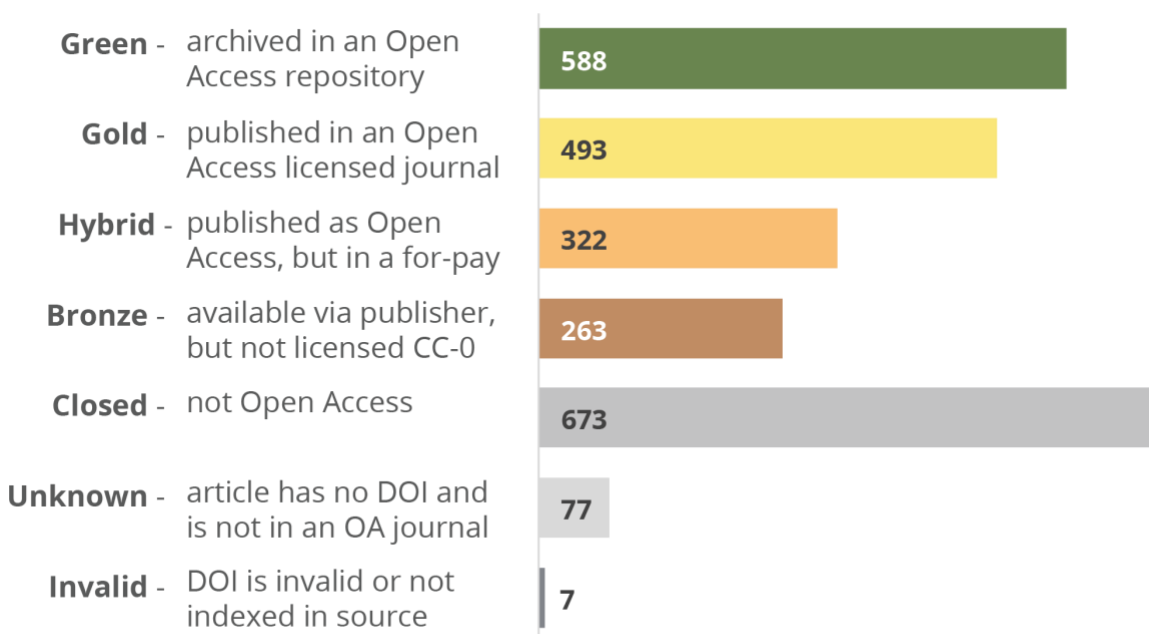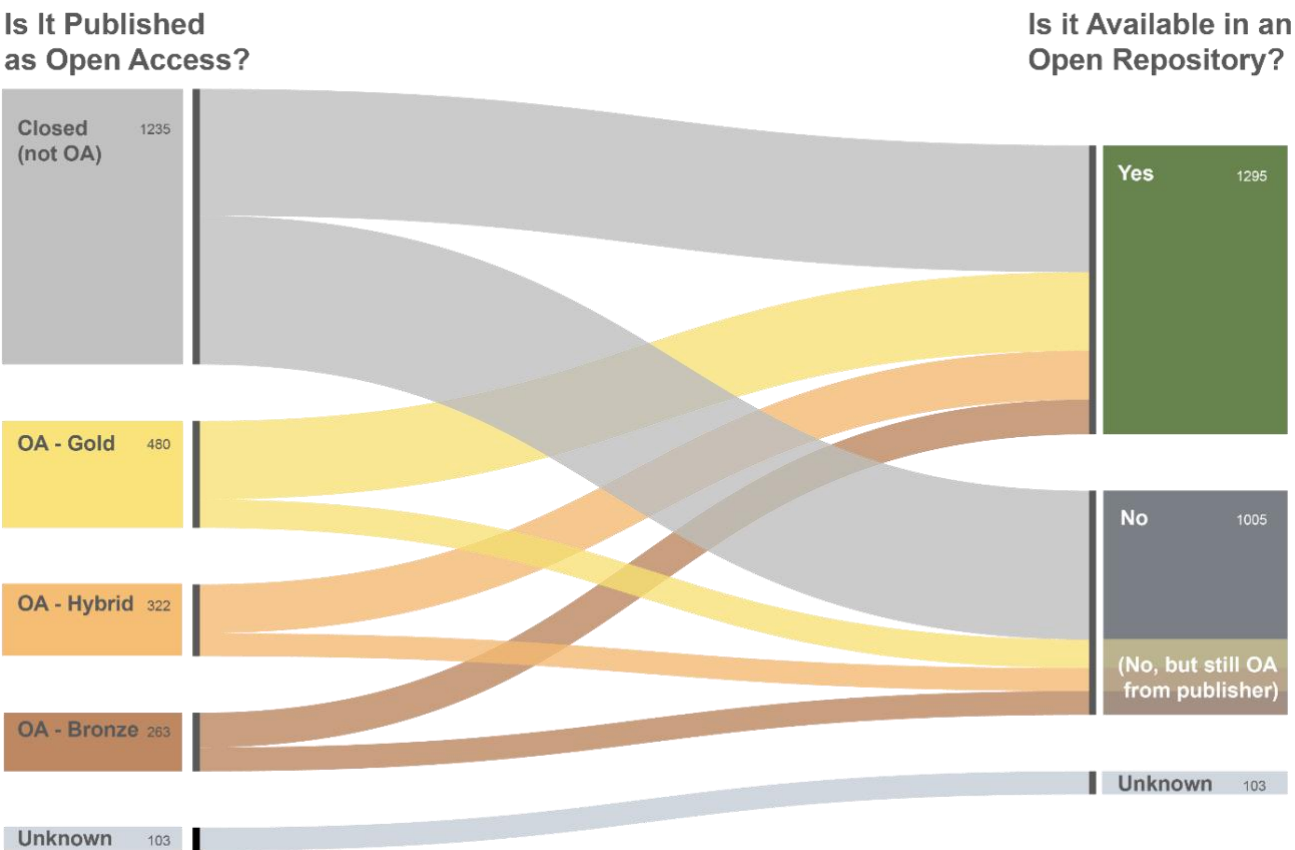### "Best" Open Access Status of 2021 Journal Articles

Figure 2: 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 available 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-

Open Access Status of 2021 Journal Articles by Publisher and Repository

Status

Figure 3: Sankey Chart showing the open access status of articles by the publisher on the left, and by whether it has been placed into an open access repository at the time of analysis on the right. More than half the articles in repositories were already open via the publisher, while half of the closed ones from the publisher still have not been placed in an open repository and are only accessible behind a paywall.

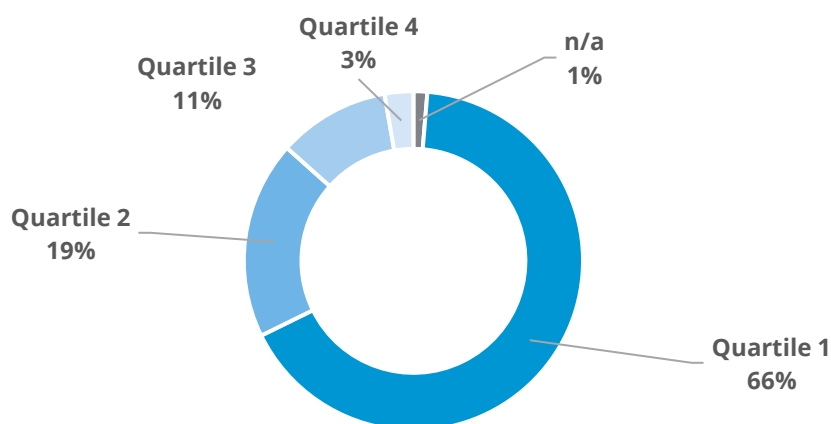# Open Access Status of 2021 Articles in Journal by Quartile

Figure 4: 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.

Top Ten Journals by Eigenfactor with Count of Articles in 2021
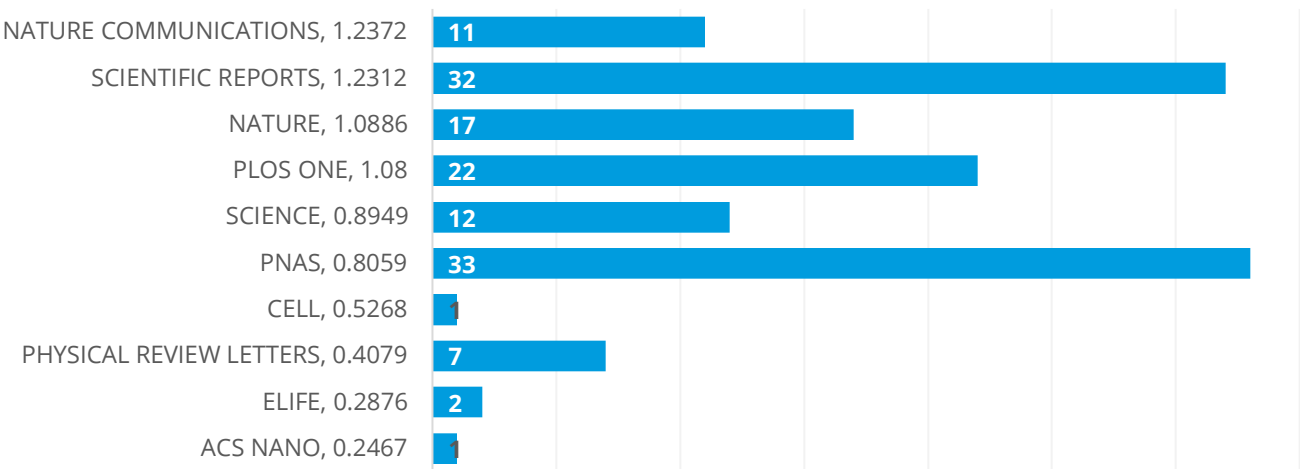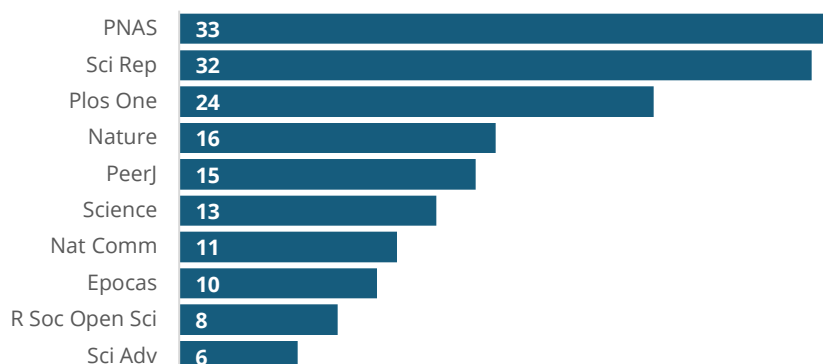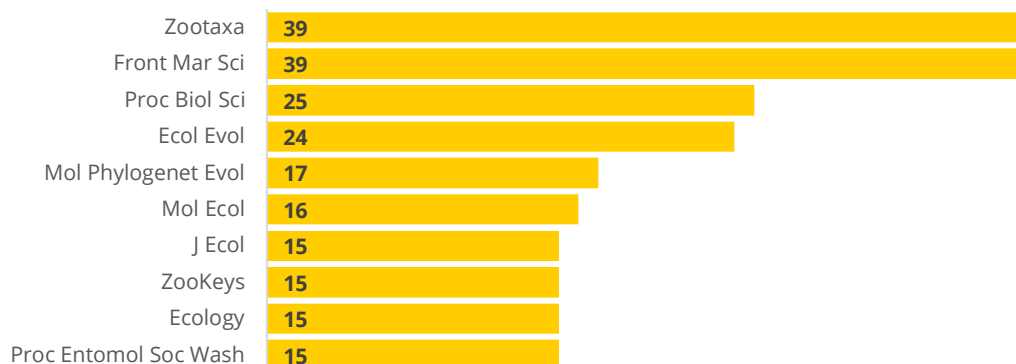

Figure 5: Number of articles in the top ten journals ranked by Eigenfactor. The Eigenfactor score is one way of evaluating journals and is based on the incoming citations to that journal while ranking incoming citations from more influential journals heavier. It purports to be a measure of a journal's total importance to the scientific community.

Journals with the Most Smithsonian-Authored Articles in 2021 by Category

Articles in Multidisciplinary Journals in 2021

Journal	Count
PNAS	33
Sci Rep	32
Plos One	24
Nature	16
PeerJ	15
Science	13
Nat Comm	11
Epocas	10
R Soc Open Sci	8
Sci Adv	6

Articles in Life Sciences Journals in 2021

Journal	Count
Zootaxa	39
Front Mar Sci	39
Proc Biol Sci	25
Ecol Evol	24
Mol Phylogenet Evol	17
Mol Ecol	16
J Ecol	15
ZooKeys	15
Ecology	15
Proc Entomol Soc Wash	15

Articles in Physical Sciences Journals in 2021

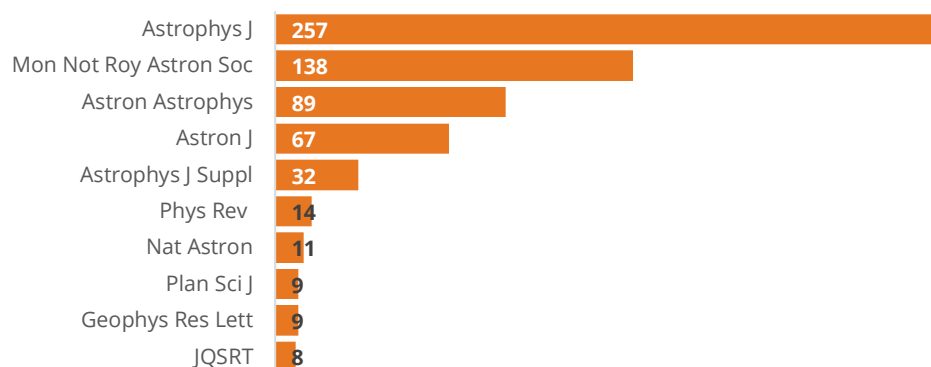| Journal	Count
Astrophys J	257
Mon Not Roy Astron Soc	138
Astron Astrophys	89
Astron J	67
Astrophys J Suppl	32
Phys Rev	14
Nat Astron	11
Plan Sci J	9
Geophys Res Lett	9
JQSRT	8

Figure 6: Each bar chart shows number of Smithsonian-authored contributions for the top ten journals based on the Web of Science classification for that journal, for the publication year 2021. Journal names are standard abbreviations via http://journalseek.net

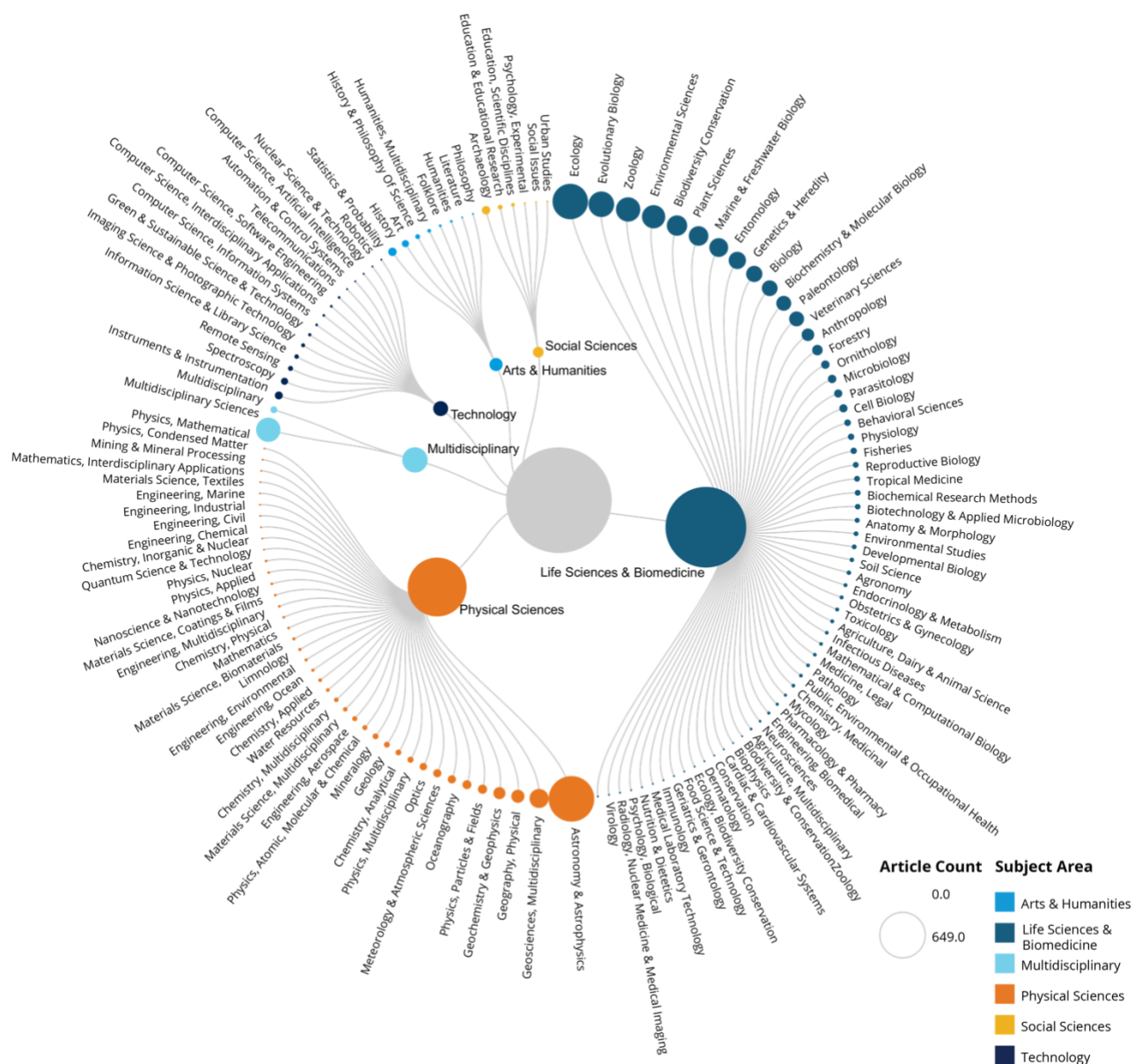# Journal Subject Categories by Count of Articles in 2021

Figure 7: Circular Dendrogram showing the number of articles published in each subject based on the journal's classification in the Web of Science. (Journals not indexed in the Web of Science were assigned subject categories manually.)

Top Five Articles by Altmetric Score in 2021

A possible planet candidate in an external galaxy detected through X-ray transit
Article in ***Nature Astronomy***, October 2021

People have shaped most of terrestrial nature for at least 12,000 years
Article in ***Proceedings of the National Academy of Sciences***, April 2021

The Per-Tau Shell: A Giant Star-forming Spherical Shell Revealed by 3D Dust…
Article in ***The Astrophysical Journal Letters***, September 2021

Baleen whale prey consumption based on high-resolution foraging measurements
Article in ***Nature***, November 2021

The electron-capture origin of supernova 2018zd
Article in ***Nature Astronomy***, June 2021

Figure 8: Top five articles by Altmetric Score for 2021. Each score is a weighted indicator based on the number of mentions in various attention sources including social media, Wikipedia, news media, Mendeley, policy documents, and more.

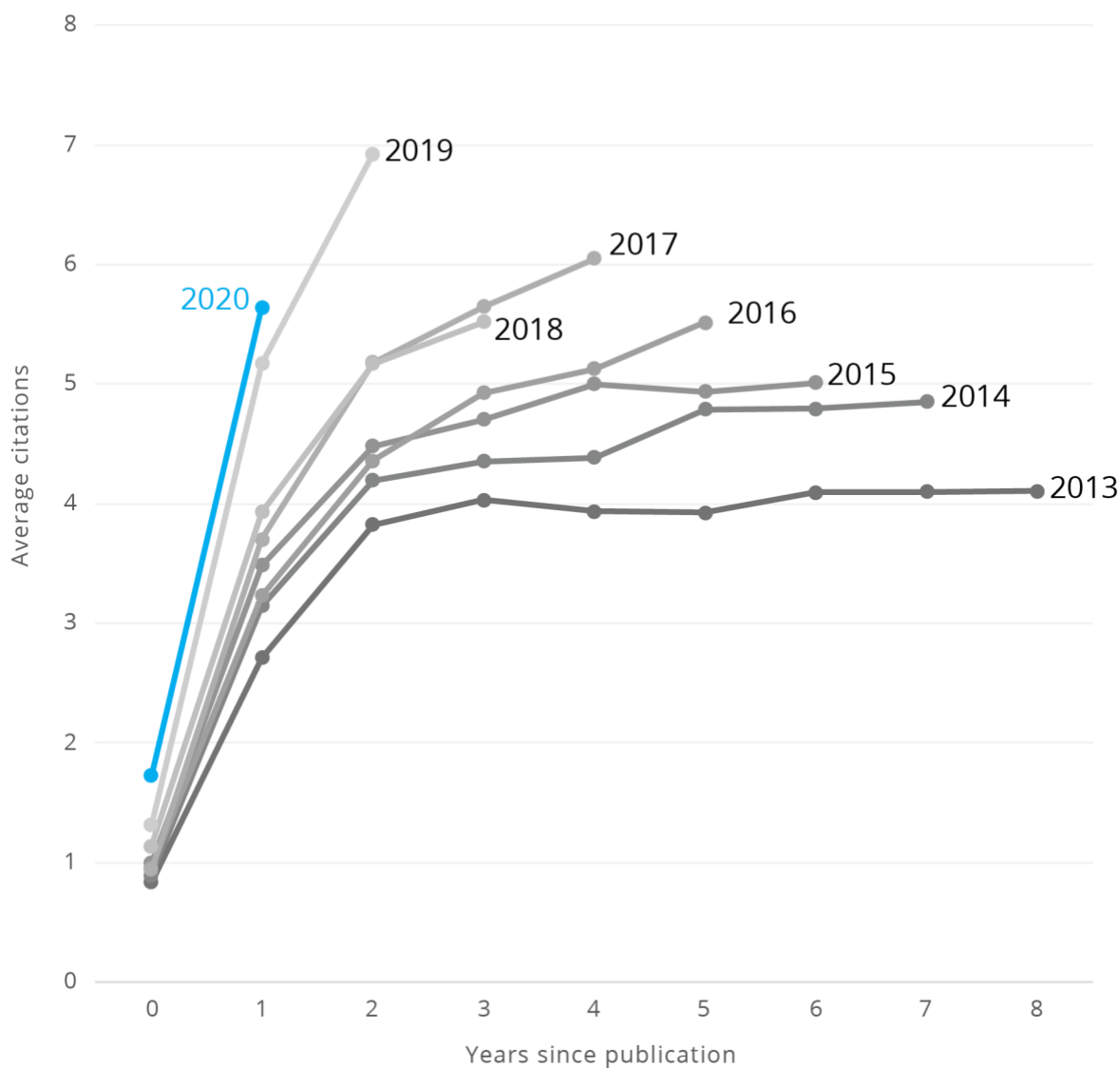# Average Citations for Articles in a Given Year by Years Since Publication

Figure 9: Total number of citations received by publications published each 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. The specific dataset for this analysis includes Smithsonian-authored publications as findable in the Web of Science Core Collection.

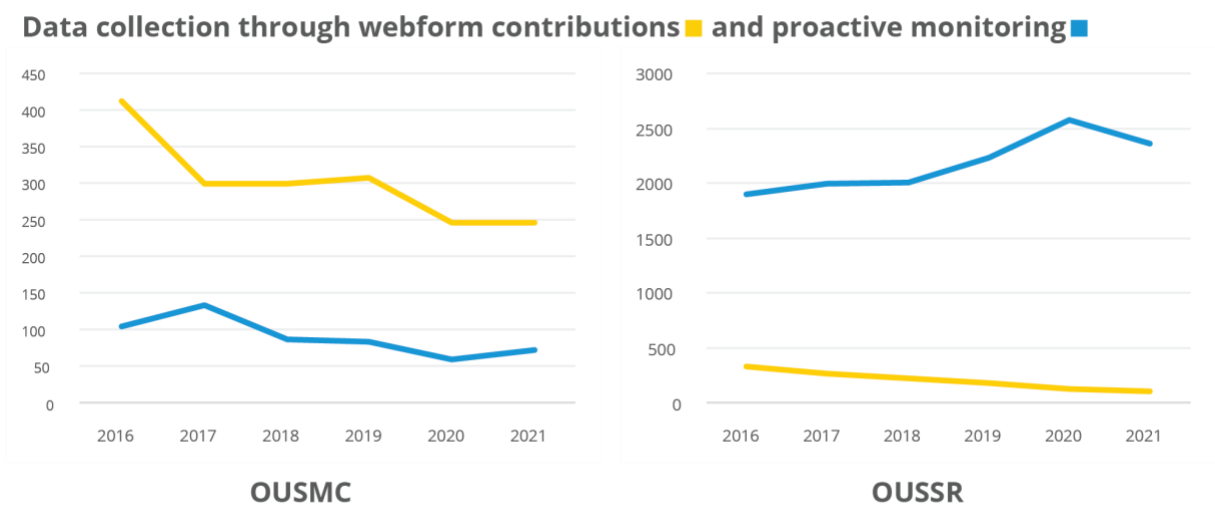# Data Collection by Smithsonian Research Online

Figure 10: Comparison of how research outputs have been ingested into Smithsonian Research Online between the Office of the Under Secretary of Museums and Culture (OUSMC) and the Office of the Under Secretary of Science and Research (OUSSR). Yellow represents outputs that are collected by a webform made available on the internal staff portal to Research Online, while blue represents outputs captured through publication alerts and other methods deployed by Smithsonian Libraries and Archives staff.

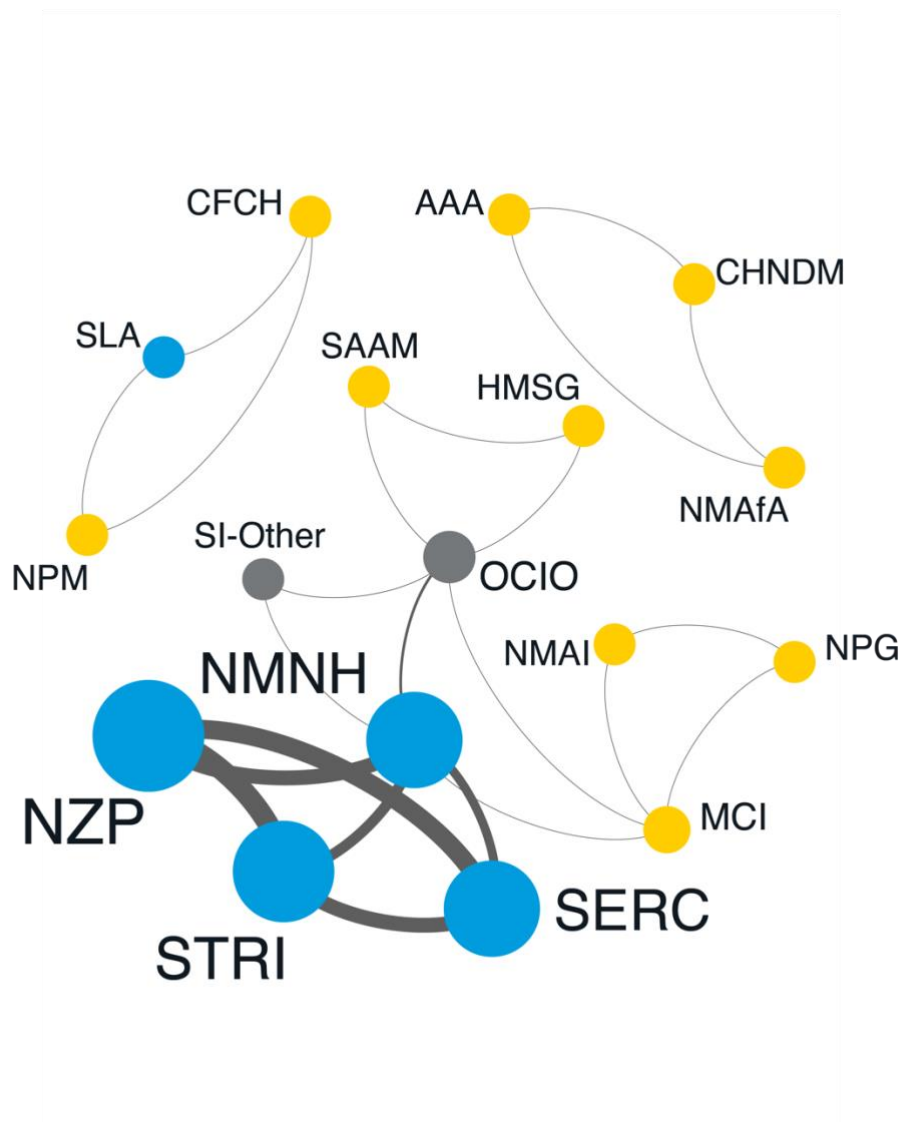# Network Graph of Internal Smithsonian Collaborations

Figure 11: Network showing collaboration among Smithsonian units in 2021. Circle size represents the number of publications with intra-unit collaboration, while lines connecting the circles represent how many times these units have collaborated. In many cases, unit collaboration is due to a researcher holding dual affiliation.

About this Report

The Smithsonian Research Online annual report was created by staff of the Smithsonian Libraries and Archives and released on March 24, 2022. This report uses data captured on March 1, 2022, 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.

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

Suggested Citation:

Hutchinson, Alvin R. and Naples, Richard M. 2022. *Smithsonian Research Online Annual Report 2021.* (Washington, D.C.: Smithsonian Libraries and Archives).