

2022 State of Open at the University of Colorado Boulder

An Update on Open Access Practices Based on Data from 2021

September 29, 2022

Prepared by the Data and Scholarly Communication Services Section, Center for Research Data and Digital Scholarship, University Libraries, University of Colorado Boulder

Ryan Caillet, *Institutional Repository Program Manager*

Melissa H. Cantrell, *Scholarly Communication Librarian*

Andrew Johnson, *Lead, Data and Scholarly Communication Services Section*

Aditya Ranganath, *Data Librarian*



University Libraries

UNIVERSITY OF COLORADO **BOULDER**

TABLE OF CONTENTS

I. Executive Summary	2
II. Open Access Articles by CU Boulder Faculty	3
III. CU Boulder Libraries Open Access Fund	7
IV. Open Access Content in CU Scholar	12
V. Open Data at CU Boulder	14

I. Executive Summary

Using data from 2021, this report is the fourth annual update to the “State of Open at the University of Colorado Boulder: A Baseline Analysis of Open Access Practices from 2012 to 2018”: <https://doi.org/10.25810/vprn-v113>. It includes analyses of open access (OA) article publishing activities, OA repository usage, and data publishing practices by researchers at the University of Colorado Boulder (CU Boulder). Data used to produce this report can be found here: <https://doi.org/10.25810/tt4b-9v88>

Key findings from this report include:

- 62% of articles published in 2021 by CU Boulder authors are available via some type of OA (Gold, Green, Hybrid, or Bronze) (up from 60% at the time of the 2020 report);
- In 2021, the CU Boulder Libraries OA Fund funded author fees totaling \$89,761 for 53 journal articles published by CU Boulder authors in full OA journals (up from \$57,769 for 34 journal articles in 2020);
- At the end of 2021, there were 13,791 OA items in the CU Scholar institutional repository (up from 11,810 in 2020), and these items were downloaded a total of 39,393 times in 2021 (down from 43,236 in 2020);
- In the annual Faculty Report of Professional Activities (FRPA), faculty reported 92 published data sets in 2021 (up from 65 in 2020) with 76 of these citations including Digital Object Identifiers (DOIs) (up from 50 in 2020) and 68 citations identifying a formal data repository (up from 54 in 2020);
- The Libraries and its partners registered 416 DataCite DOIs for published data sets in 2021 (up from 320 in 2020).

II. Open Access Articles by CU Boulder Faculty

Continuing with a change that was first implemented in 2021, this 2022 report leverages data on types of open access (OA) publishing from Unpaywall¹ matched against data on articles authored by CU Boulder faculty from CU Boulder Elements (CUBE)² in order to gain broad insight into the extent of OA publishing practices at CU Boulder. This approach allows for a more complete picture of all types of OA (e.g., Green, Gold, Hybrid, etc.) than the data provided in State of Open reports prior to 2021, which only included articles published in full OA journals that were indexed in the Directory of Open Access Journals (DOAJ).³ It should be noted that both of these current data sources are dynamic in nature. While CUBE data is updated on an annual basis, it is possible for both recently published and older articles to be added each year. In addition, articles that were included in CUBE in a previous year, may be removed for a variety of reasons by the time of the next annual data release. Unpaywall data is continuously updated with new articles, and information about OA status for any article in the database evolves over time as well. For example, an article previously included in Unpaywall as “closed” could be deposited in a Green OA repository at any time, which would change its OA status as a result. The dynamic nature of these sources means that the data presented in this section of the report should be treated as an annual snapshot rather than providing directly comparable data points with regard to what was included in the previous 2021 State of Open at CU Boulder report.

¹ Unpaywall: <https://unpaywall.org/>

² CU Boulder Elements: <https://www.colorado.edu/fis/CUBE>

³ Directory of Open Access Journals: <https://doaj.org/>

Table 1. Types of Open Access Content

Type	Description
Green Open Access	This content is made OA when a version of a closed access or subscription article is posted to a repository (institutional, subject, etc.)
Gold Open Access	This content is made OA through a journal that exclusively publishes OA articles. An APC sometimes but not always applies.
Hybrid Open Access	This content is made OA through a journal that offers the author(s) a choice to publish an article OA or via the closed/subscription model. An APC always applies if the OA option is selected.
Bronze Open Access	This content is free to read on a publisher's website but lacks a clearly identifiable license, typically making the article unavailable for reuse.

Table 1 provides descriptions of the different types of OA content that Unpaywall identifies: Green, Gold, Hybrid, and Bronze. In addition to these four types of OA content, Unpaywall also identifies when an article is “Closed,” which means that the content is not freely or openly available under any type of OA.

Table 2. Articles Published by CU Boulder Faculty by Open Access Type, 2012-2021

Year	Closed (n)	Gold (n)	Green (n)	Hybrid (n)	Bronze (n)	Total OA (n)	Total (n)
2021	1631	942	837	495	420	2694	4325
2020	1560	935	1043	508	592	3078	4638
2019	1547	863	926	501	643	2933	4480
2018	1532	891	890	514	691	2986	4518
2017	1670	790	955	461	598	2804	4474
2016	1670	807	785	466	750	2808	4478
2015	1734	677	844	364	514	2399	4133
2014	1859	520	931	274	379	2104	3963
2013	1834	330	830	313	460	1933	3767
2012	1795	291	749	307	397	1744	3539

Table 2 provides the total number of articles published by CU Boulder faculty each year from 2012 to 2021 that are included in both the CUBE and Unpaywall data sources. Inclusion in both data sources allows each article published by CU Boulder faculty to be categorized by type of OA, and the total number of OA articles is provided for each year as well. These totals reveal overall shifts in OA article publishing practices at CU Boulder from 2012 to 2021. With slight variations from year to year, there has been a general trend toward an increase in OA articles and a decrease in closed access articles published by CU Boulder faculty over the period studied; however, data from the last six years indicate that this might have plateaued in the 2800-3000 OA articles per year range. It will be interesting to see if recent developments like the White House Office of Science and Technology Policy’s memorandum on “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research”⁴ change the trajectory of this recent trend in the coming years.

⁴ White House Office of Science and Technology Policy (2022). “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research”: <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

Figure 1.

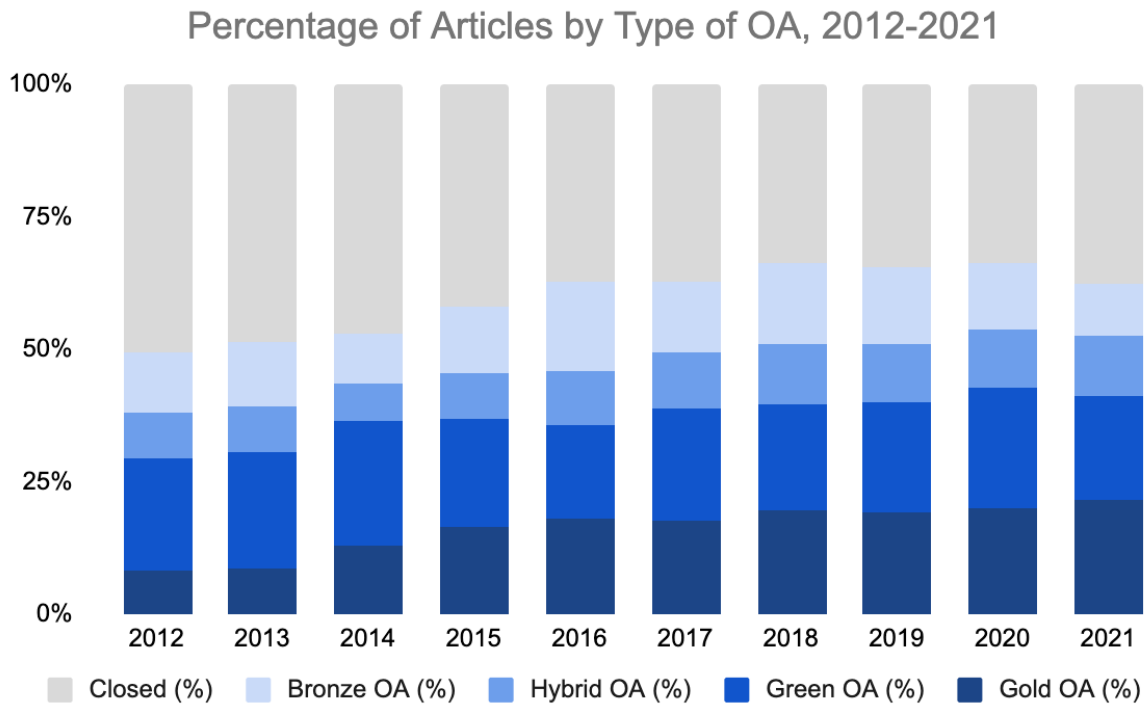
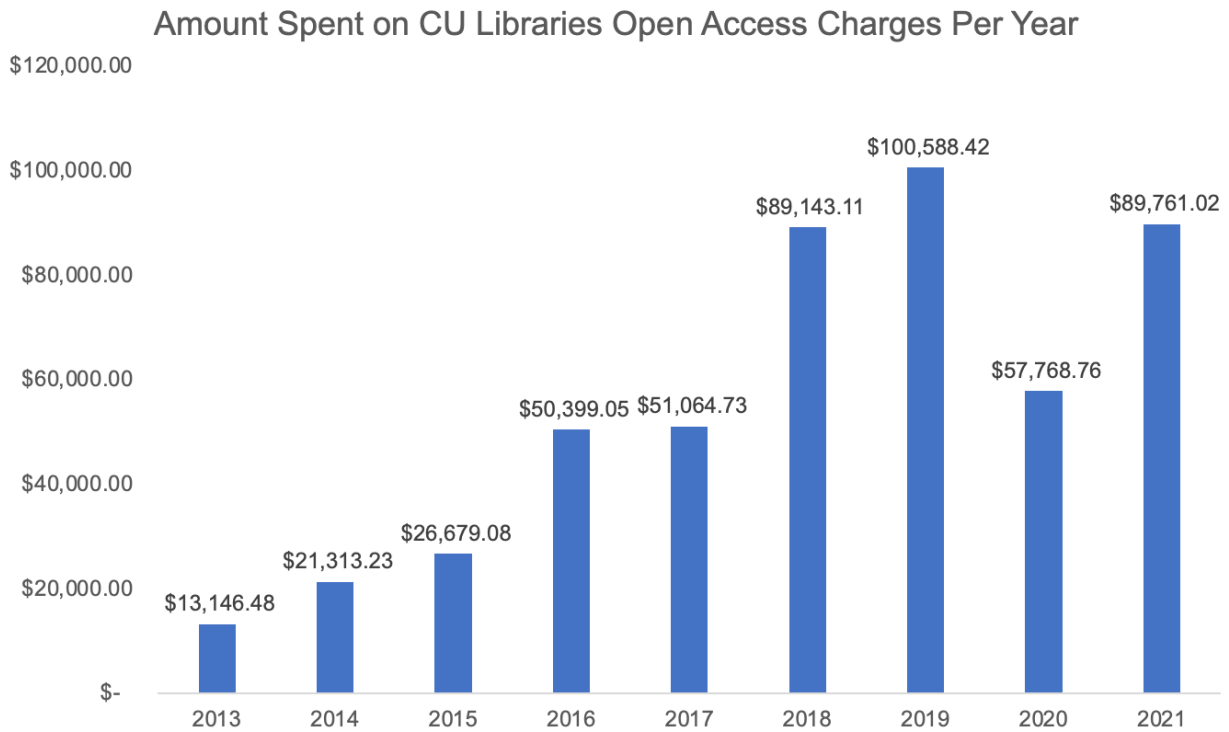


Figure 1 shows the trends in percentage of type of OA and closed articles over time. The total percentage of OA articles consistently increased from a low of 49.28% in 2012 to a high of 66.36% in 2020. Distinguishing between Gold, Green, Bronze, and Hybrid OA provides additional insight into the general trend of increasing OA activities at CU Boulder. Over the period studied, Gold OA articles showed the strongest trend and greatest increase from a low of 8.22% of all articles published in 2012 to a high of 21.78% of articles published in 2021. The percentage of Green OA articles has been relatively stable, fluctuating from year to year between 17.53% and 23.49% without a clear trend of continual increase or decrease. The percentage of Hybrid OA articles showed a somewhat steady increase from year to year over a smaller range than Gold OA (from 8.67% in 2012 to 11.45% in 2021). The growth or decline in percentage of Bronze OA articles has not shown a clear trend, but has ranged from a low of 9.56% in 2014 to a high of 16.75% in 2016.

III. CU Boulder Libraries Open Access Fund

Figure 2.

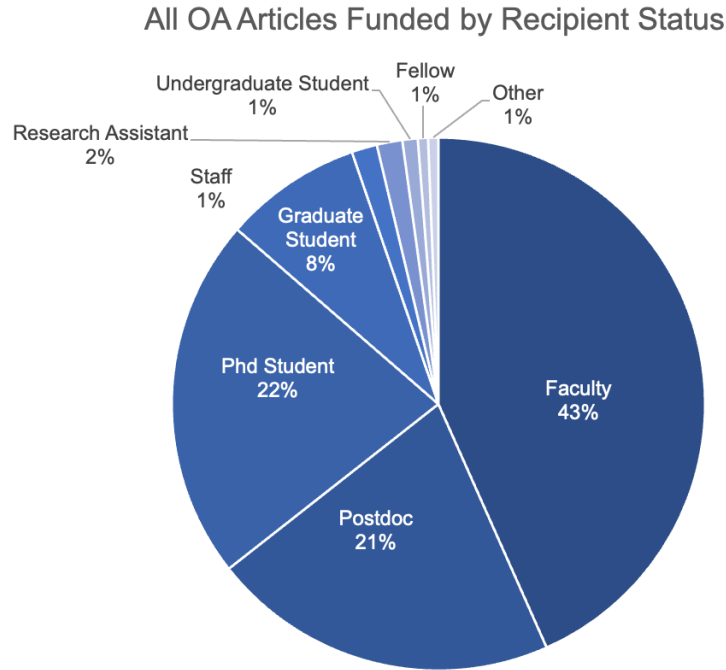


CU Libraries spending on article processing charges (APCs) for fully OA journal articles by faculty, staff, and students rebounded in 2021 to levels similar to 2018, with a total spend of \$89,761.02. This year represented the second highest expenditures on the OA Fund since the inception of the fund in 2013. The funds from fiscal year 2020/2021 were depleted in May 2021, so there was only a two-month gap before funding was renewed for fiscal year 2021/2022.⁵

53 articles were funded in 2021, with an average APC cost of \$1,693.60, which is comparable to the APC average from 2020. Cumulatively, the CU Boulder OA Fund has helped authors publish 323 fully OA articles in 136 unique journal titles.

⁵ Similar to previous years, several additional requests for funding were made in Spring 2021, but could not be completed due to depletion of funds.

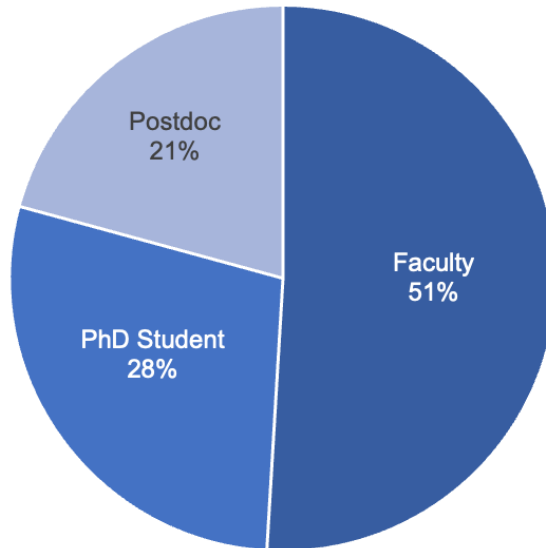
Figure 3.



The proportion of funded authors by university status over the lifetime of the OA Fund remains largely unchanged, as faculty still represent a little over two-fifths of funded articles (43%) and graduate students and postdoctoral researchers represented just over half (51%) of the articles funded. Research Assistants, Staff, Fellows, Undergraduate Students and Other affiliations collectively represent 6% of the total articles funded since the inception of the OA Fund in 2013.

Figure 4.

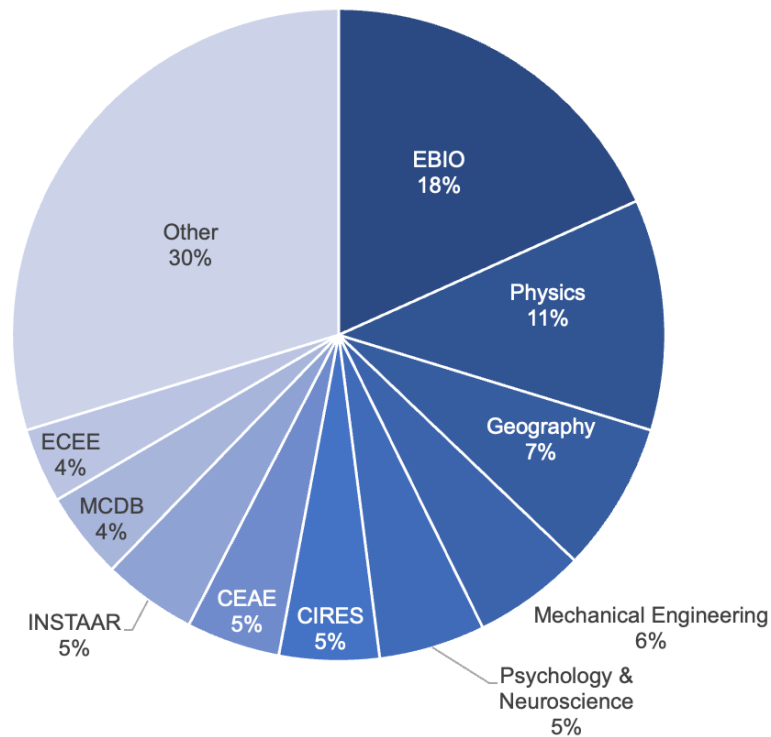
OA Articles Funded by Recipient Status: 2021 Only



Of note for 2021 is the fact that all funding for this year was awarded to faculty, PhD students, and postdocs. Each year since 2013, a small proportion of funding has gone to staff, research assistants, and other CU Boulder affiliates, and 2021 was the first year in which no funding went directly to an applicant claiming any of these statuses. The proportion of PhD recipients rebounded to near 2019 levels (from 23% in 2020 to 28% in 2021). The number of postdoc recipients rose from 15% in 2020 to 21% in 2021, but was still less than 2019 levels. The proportion of faculty receiving funding continued to hover around half of all OA awards (51%). By sheer numbers, the most faculty ever (26) were awarded funding in 2021.

Figure 5.

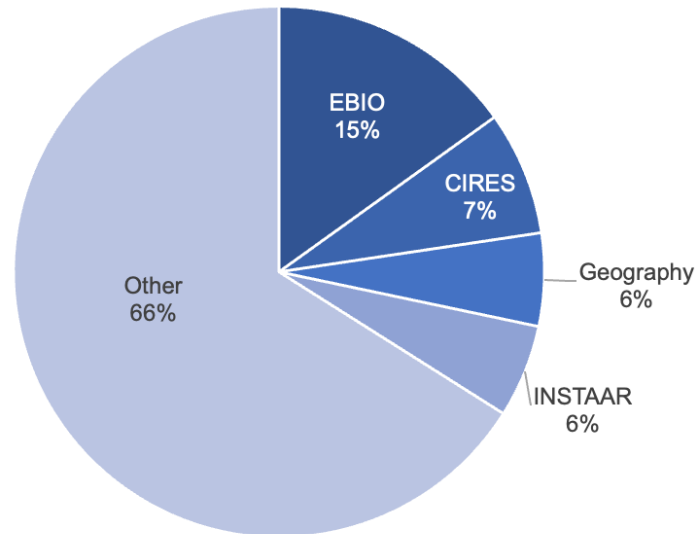
OA Funded Articles by Department - Cumulative



The cumulative data for articles funded by department show no significant changes. EBIO and Physics continue to fall as an overall share of the total awards (EBIO: 24% in 2018, 21% in 2019, 19% in 2020, and 18% in 2021; Physics: 14% in 2018 and 2019, 13% in 2020, 11% in 2021), following previously noted trends. There was also no change in the top ten departments, solidifying those listed in Figure 5 as the top users of the OA Fund. As noted in 2020, the proportion of “Other” departments continues to grow at a gradual rate (28% in 2020, 30% in 2021). We can conclude from this that while the fund has a core base of users, the disciplinary affiliations of authors seeking funding also continues to diversify. Several departments had individuals receiving funding for the first time in 2021, including Information Science, Aerospace Engineering, Advertising, Public Relations & Media Design, University Libraries, Political Science, and Theatre & Dance.

Figure 6.

OA Funded Articles by Department: 2021 Only



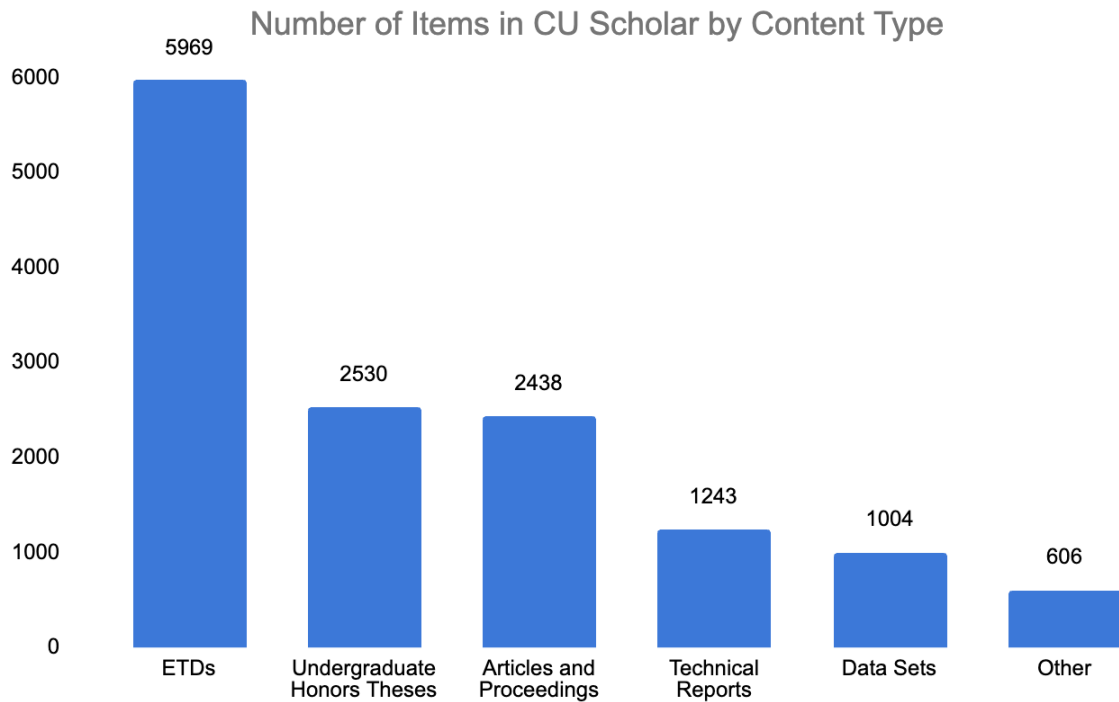
While there is little cumulative change in the departments seeking and being awarded open access funding, there continues to be somewhat substantial changes from year to year. For example, while MCDB and CEAE represented 29% of the funding awarded in 2020 (10 articles funded), these departments only accounted for 7% of the funding in 2021 (4 articles funded). Funding for EBIO rose slightly in 2021 and funding for Geography and CIRES remained at similar levels to 2020. This year, far more departments were designated into the “Other” category because they had 2 or fewer open access funding awards. In 2020, 16 departments had two or fewer awards, while 27 departments had two or fewer awards in 2021. This serves to reiterate the fund’s diversification, possibly due to better awareness as well as more authors from different departments publishing open access.

It is also worth noting this year that the University Libraries began entering into more formal agreements with some open access publishers that typically publish articles by CU Boulder authors paid through the OA Fund. In particular, the Libraries has a membership with Frontiers and a three-year agreement with PLOS. These deals are important because they not only extend the annual bandwidth of the general OA Fund but also may alter the makeup of departmental funding and titles funded in future years since these prevalent publishers will now be covered by other funding sources.

IV. Open Access Content in CU Scholar

At the beginning of 2020, CU Scholar migrated from the hosted bepress Digital Commons platform to the open source Samvera repository software. As with last year's 2021 report, this 2022 edition includes data reported on the calendar year unlike updates published before 2020. In addition, usage of repository content as measured by download counts is now being tracked using a different method (Google Analytics) than the proprietary download data provided by Digital Commons used prior to 2020. As such, we are not confident in the comparability of the pre-2020 and post-2020 numbers.

Figure 7.



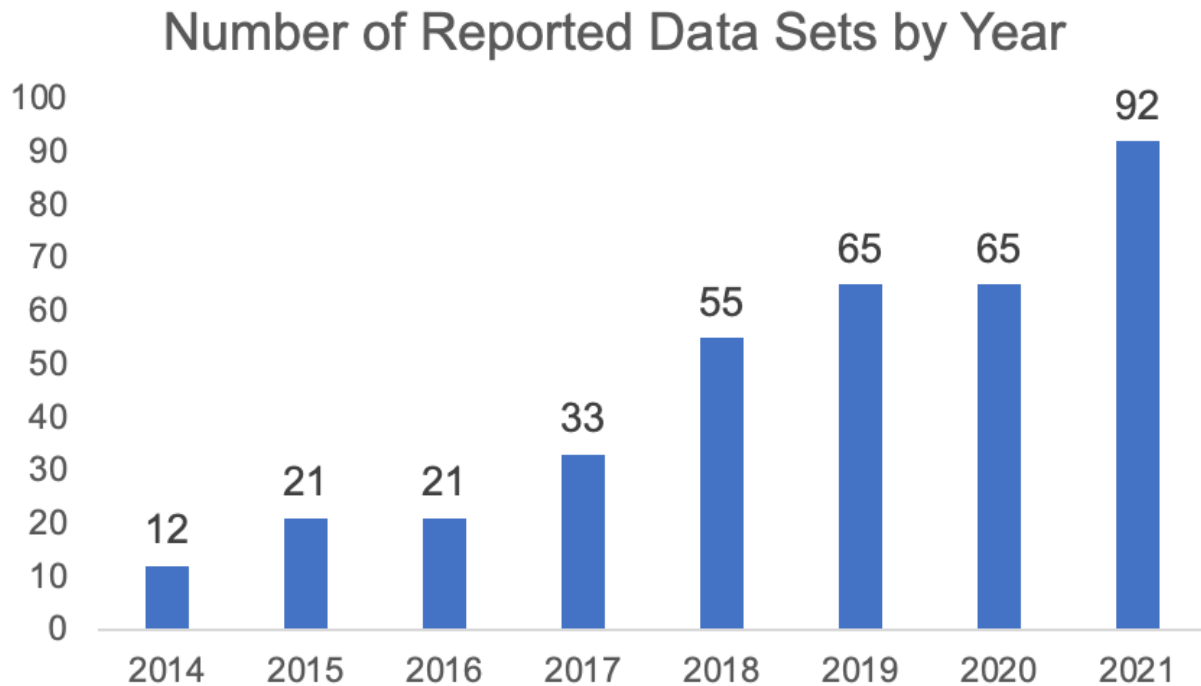
At the end of 2021, CU Scholar contained 13,791 items including journal articles, data sets, graduate theses and dissertations, undergraduate honors theses, conference materials, books, and book chapters. This represents an increase of 16.77% from 11,810 total items at the end of 2020. Content on CU Scholar was downloaded 39,393 times in 2021 according to data exported from Google Analytics. Total downloads decreased slightly by 8.89% from 43,236 in 2020. As this is the first year of post-migration download data to compare with 2020, it will be interesting to monitor this trend going forward. It is possible that 2020 was an unusually high year for downloads as the most downloaded CU Scholar content by far were articles related to the

COVID-19 pandemic that received widespread attention across both scholarly and popular media outlets.

In 2021, Graduate theses and dissertations made up 43.28% of all content in the repository (up slightly from 42.41% in 2020). Undergraduate honors theses accounted for 18.34% of CU Scholar content (down slightly from 19.30% in 2020) while articles and proceedings comprised 17.68% of repository items (virtually identical to 17.95% in 2020). Technical reports represented 9.01% of the repository with most items belonging to a computer science technical reports collection that is not actively growing, and the overall percentage decreased from 10.37% in 2020 as other content types continue to increase in relative total size in the repository. Data sets now account for 7.28% of CU Scholar contents (up from 5.05% in 2020), representing the fastest growing segment of the repository in terms of content types.

V. Open Data at CU Boulder

Figure 8.



After remaining steady at 65 published data sets in the years 2019 and 2020, the number of published data sets reported on the annual Faculty Report of Professional Activities (FRPA) in 2021 increased substantially to 92 data sets. This is the highest number of data sets that have been published since we began gathering data, and represents a 41.5% increase in the number of reported data sets relative to 2020. This is the second highest relative annual increase on record (the highest relative increase was from 2017 to 2018, which saw a 66% increase in the number of published data sets), and the highest absolute increase.

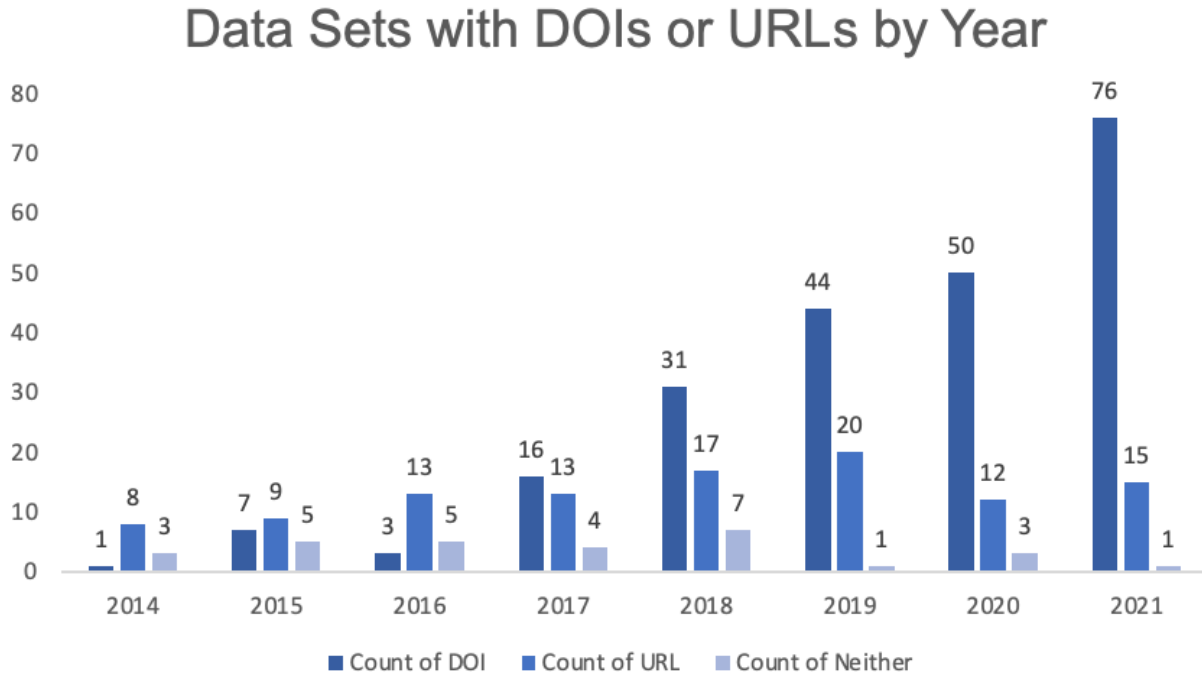
Table 3. Reported Data Sets by Department/Unit, 2014-2021 (n>1)

Department/Unit	Number of Reported Data Sets
Cooperative Institute for Research in Environmental Sciences	54
Atmospheric and Oceanic Sciences	39
Civil, Environmental and Architectural Engineering	32
Ecology and Evolutionary Biology	29
Institute for Arctic and Alpine Research	25
Environmental Studies	22
Geography	17
Geological Sciences	17
Astrophysical and Planetary Sciences	16
Computer Science	11
Linguistics	11
Libraries	9
Sociology	8
Laboratory for Atmospheric and Space Physics	7
Business	6
Chemical and Biological Engineering	6
Chemistry	6
Classics	6
Aerospace Engineering Sciences	5
Environmental Design	5
Education	3
History	3
Information Science	3

Media Studies	3
Molecular, Cellular & Developmental Biology	3
Physics	3
Asian Languages and Civilizations	2
Ethnic Studies	2
Journalism	2
Natural History Museum	2
Speech, Language, and Hearing Sciences	2

Table 3 provides updated information on the distribution of data sets published from 2014 to 2021, across disciplines. The top of the distribution looks fairly similar to the distribution reported in our previous report (which included data through the year 2020), with a few notable exceptions. In particular, the number of data sets reported by the department of Civil, Environmental, and Architectural Engineering, and the department of Environmental Studies, dramatically increased from 2020 to 2021. The cumulative number of data sets reported by the former increased from 6 in 2020 to 32 in 2021, while the cumulative number of data sets reported by the latter rose from 4 data sets in 2020 to 22 data sets in 2021. The only new entrant to this catalog of departments that have published at least two data sets since 2014 is the department of Molecular, Cellular and Developmental Biology; this department did not have any data sets at the end of 2020, but has since registered three through the year 2021. Finally, in our previous report, we highlighted data set publications from the Sociology department, which contributed data sets for the first time in 2020 (three in total). The Sociology department published an additional five data sets in 2021; this sustained increase may reflect a growing emphasis on data publication in the social sciences, but additional data is required before we can conclude that there is a meaningful upward trend in data publication practices of CU Boulder sociologists.

Figure 9.



The use of DOIs for reported data sets increased from 50 in 2020 to 76 in 2021, a 52% increase. In 2021, the overall percentage of data sets with a DOI was 82.6%, which represents an all-time high (the previous high was 76.9%, which was achieved in 2020). The trend of data set citations (including a means for accessing data) becoming a more common and consistent practice for faculty (which we have noted in previous reports) appears not simply to be continuing, but accelerating.

Figure 10.

Repositories Used by Type, 2014-2021

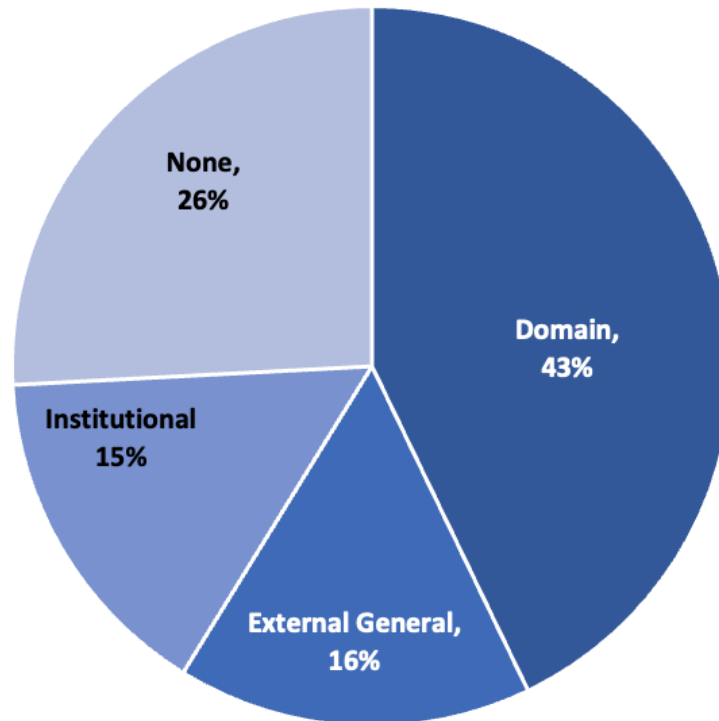


Figure 10 presents an updated summary of repository types, with 2021 data included. The percentage of data sets that were not deposited in a formal repository continues to steadily decline, from 38% in 2019, to 33% in 2020, to 26% in 2021. Similar to last year, a plurality of published data sets (43%) use domain repositories that provide access to data from particular disciplines and/or to specific types of data (e.g., ICPSR, NSF Arctic Data Center, Protein Data Bank, etc.). General repositories that are external to CU Boulder and cover a wide range of disciplines and data types (e.g., figshare, Dryad, Zenodo, etc.) provide access to 16% of the reported published data sets in the FRPA since 2014. Institutional repositories account for 15% of the reported published data sets. Four of the data sets published in institutional repositories appear in the institutional repositories of other institutions (two at Stanford, one at UCSD, and one at William and Mary) with the remainder appearing in CU Scholar.

Figure 11.

Repositories Used by Type, 2021

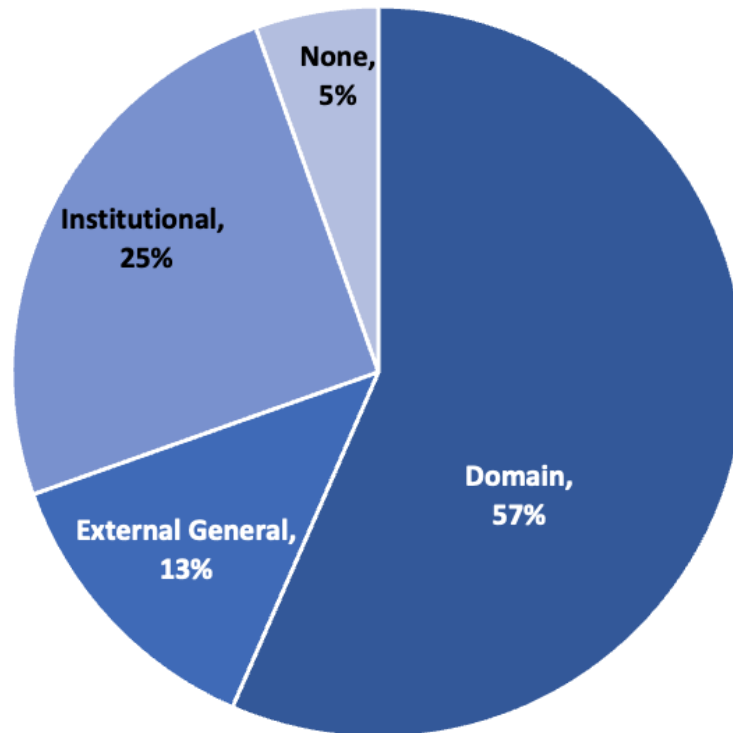


Figure 11 presents a breakdown of repository destinations exclusively for data sets published in 2021. The percentage of data sets that were not published in a formal repository of any kind was only 5% in 2021, compared to 17% in 2020. Fully 95% of the data sets released in 2021 were therefore published in a formal repository of some kind. A majority of the data sets published in 2021 (57%) were published in a domain repository, which represents an increase from 2020, when 48% were published in domain repositories. A quarter of 2021 data sets (25%) were published in institutional repositories, and 13% were published in external general repositories. This represents a departure from the composition of repositories in 2020, when only 9% of the data sets were submitted to institutional repositories, and 26% were published in external general repositories. The distribution of data sets across institutional repositories and external general repositories in 2021 is more similar to 2019 (when 23% of data sets were published in institutional repositories and 14% were in external general repositories)

than 2020, which suggests that the compositional shift from 2019 to 2020 may have been an anomaly.

Finally, in 2018, the Libraries began actively curating data sets in the CU Scholar institutional repository, including registering DataCite DOIs for every published data set housed there. In addition, the Libraries provide DataCite DOI registration capabilities to a small number of campus partners through formal agreements. In 2021, the Libraries and its partners registered 416 DOIs for published data sets (up from 320 in 2020). The disconnect between this larger number of DOIs and the 92 published data sets reported by faculty above could be due to a number of factors. Many of the data sets published by the Libraries are recurring data sets that receive a new DOI for every update but might only be reported as a single data set for the purposes of annual faculty reports. Also, some data sets published by the Libraries and its partners were created by individuals other than faculty (e.g., graduate students or staff). It is also possible that some data sets might not be considered appropriate for faculty annual reports for a number of reasons. For example, data sets supporting journal articles might be seen as duplicative when the journal article is already reported. The overall finding of DataCite DOI registration increasing 30% between 2021 and 2020 again demonstrates the steady growth in demand for data publishing services at CU Boulder. As of the end of 2021, the Libraries and its partners had registered a total of 1,021 DOIs for data sets since the DataCite DOI registration service began at CU Boulder in 2018.