2023 State of Open
at the University of Colorado Boulder

An Update on Open Access Practices Based on Data from 2022

October 11, 2023

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, *Head, Data and Scholarly Communication Services Section*
Matthew Murray, *Data Librarian*
Aditya Ranganath, *Data Librarian*

University Libraries
UNIVERSITY OF COLORADO **BOULDER**

TABLE OF CONTENTS

I. Executive Summary

Using data from 2022, this report is the fifth 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/ktb4-ce48

Key findings from this report include:

- 72% of articles published in 2022 by CU Boulder authors are available via some type of OA (Gold, Green, Hybrid, or Bronze) (up from 62% at the time of the 2021 report);
- In 2022, the CU Boulder Libraries OA Fund funded author fees totaling $69,804 for 41 journal articles published by CU Boulder authors in full OA journals (down from $89,761 for 53 journal articles in 2021); however, these decreases have more to do with the OA Fund being exhausted earlier in the fiscal year than an actual decrease in funding;
- At the end of 2022, there were 16,090 OA items in the CU Scholar institutional repository (up from 13,791 in 2021), and these items were downloaded a total of 36,730 times in 2022 (down from 39,393 in 2021);
- In the annual Faculty Report of Professional Activities (FRPA), faculty reported 56 published data sets in 2022 (down from 92 in 2021) with 87.5% of these citations including Digital Object Identifiers (DOIs) (up from 82.6% in 2021) and 95% of these citations identifying a formal data repository (same as 95% in 2021);
- The Libraries and its partners registered 335 DataCite DOIs for published data sets in 2022 (down from 416 in 2021);
- This is the first year there has been a decrease in either the number of reported published data sets in FRPA or the number of DataCite DOIs registered for published data sets, so it will be important to monitor these numbers in the coming years to see if this is an anomaly or the start of a new trend.

II. Open Access Articles by CU Boulder Faculty

Continuing with a change that was first implemented in 2021, this 2023 report leverages data on types of open access (OA) publishing from Unpaywall[1] 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 previous State of Open at CU Boulder reports.

---

[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, 2013-2022

Year	Closed (n)	Gold (n)	Green (n)	Hybrid (n)	Bronze (n)	Total OA (n)	Total (n)
2022	1122	1082	903	596	250	2831	3953
2021	1283	995	1177	726	449	3347	4630
2020	1392	939	1118	699	612	3368	4760
2019	1174	850	905	688	660	3103	4277
2018	1170	852	827	695	774	3148	4318
2017	1210	759	852	651	600	2862	4072
2016	1309	800	732	672	659	2863	4172
2015	1425	622	762	466	599	2449	3874
2014	1544	483	787	298	543	2111	3655
2013	1526	308	743	363	597	2011	3537

Table 2 provides the total number of articles published by CU Boulder faculty each year from the last 10 years (2013 to 2022) 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 2013 to 2022. 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 3100-3300 OA articles per year range. It will be interesting to see if recent developments like implementation of federal agency policies in response to the White House Office of Science and Technology Policy (OSTP) memorandum on "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research"[4] change the trajectory of this recent trend in the coming years. In addition, recent trends show a steady increase in the number of Gold OA articles per year, which is typically eclipsed by the number of Green OA articles at least one year after publication. This is to be expected given many journal publishers' embargo policies for Green OA, but this trend might also shift given the "zero embargo" requirement included in the 2022 OSTP memo.

---

[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.

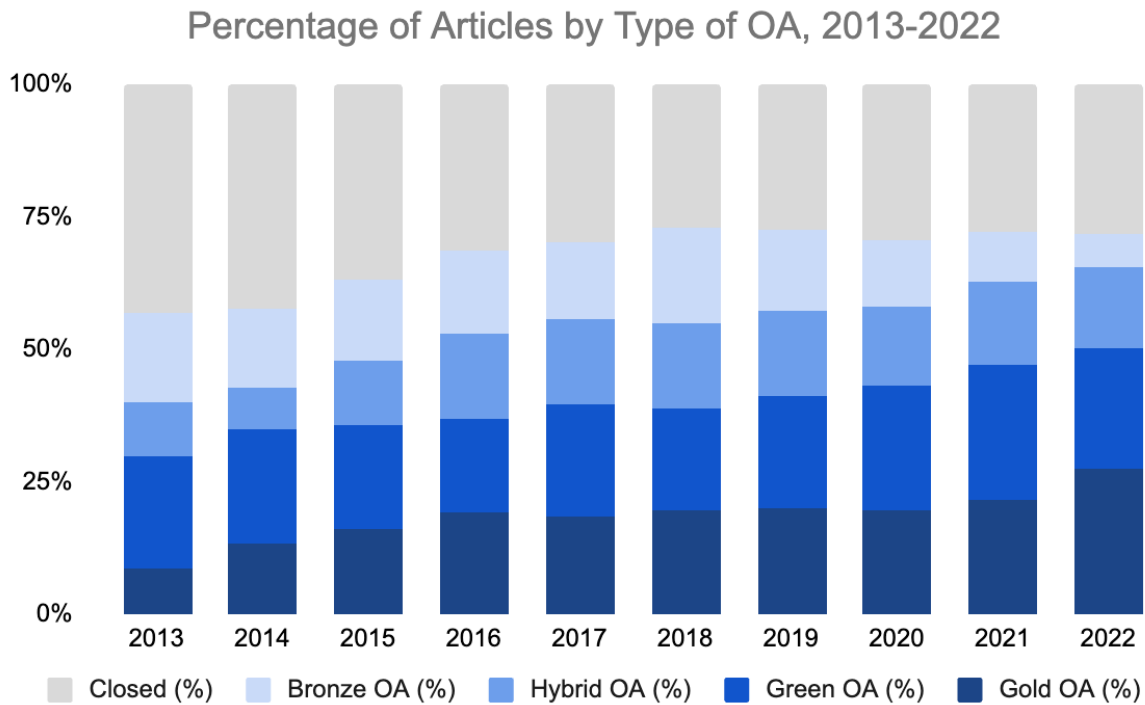Percentage of Articles by Type of OA, 2013-2022

        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 56.86% in
2013 to a high of 72.90% in 2018. From 2019-2022, the total percentage of OA articles
remained in the 71-72% range, suggesting a possible plateau. 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 8.71% of all articles
published in 2013 to a high of 27.37% of articles published in 2022. The percentage of
Green OA articles has been relatively stable, fluctuating from year to year between
17.55% and 25.42% without a clear trend of continual increase or decrease. The
percentage of Hybrid OA articles from 2016 to 2022 also remained quite stable (with all
percentages falling between a low of 14.68% in 2020 and a high of 16.11% in 2016).
The percentage of Bronze OA articles appears to be in steady decline since 2018 with
the percentage dropping 2-3% each year culminating in an overall low of 6.32% in 2022.
This could suggest that publishers' licensing practices are becoming increasingly
consistent over time.

## III. CU Boulder Libraries Open Access Fund

Figure 2.

Amount Spent on CU Libraries Open Access Charges Per Year
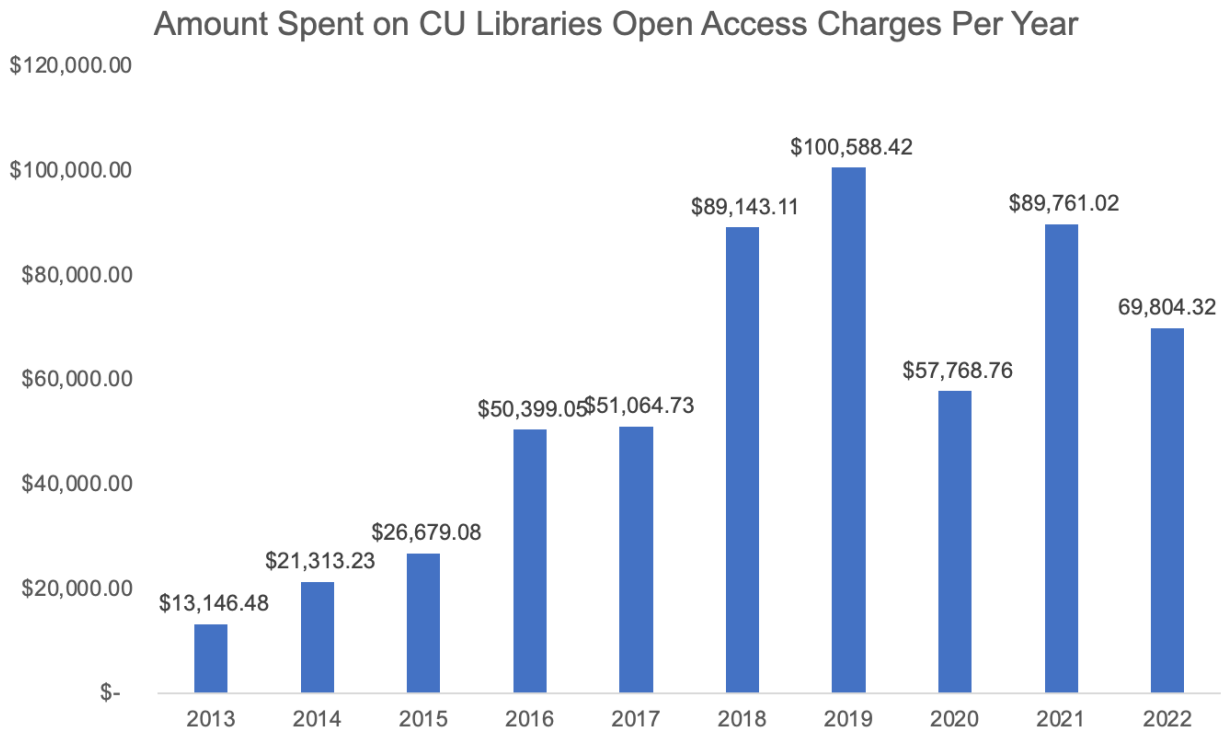
CU Boulder Libraries spending on article processing charges (APCs) for fully OA journal articles via the Open Access Fund appears much lower in 2022 compared to 2021, with a total spend of $69,804.32. However, this dip in spending is an illusion created by calculating the spend by calendar year as opposed to by fiscal year, which is how this support is funded. The amount dedicated by the Libraries to this fund in 2022 was similar to the level dedicated in 2021. Since funds allocated for FY22 ran out at the end of February 2022, there were no articles accepted between March-June 2022. Also, factoring in that $26,412 was spent on APCs for articles published in Frontiers via our CU Boulder institutional agreement, the actual spend for APCs across these funds was just over $96,000.

41 articles were funded in 2022, with an average APC cost of $1,702.54, which is about nine dollars higher than the average from 2021. Cumulatively, the CU Boulder OA Fund has helped authors publish 364 fully OA articles in 152 unique journal titles.

Figure 3.

All OA Articles Funded by Recipient Status
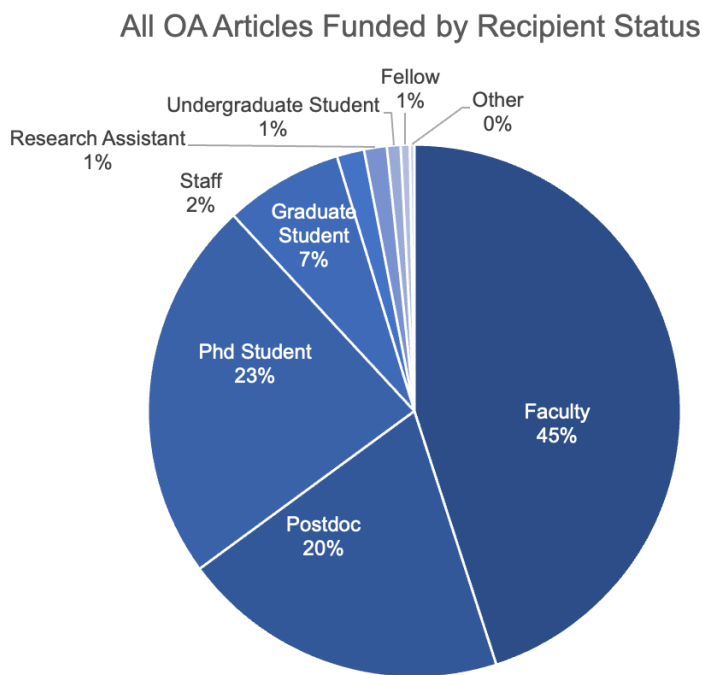
As usual, the proportion of funded authors by university status over the lifetime of the OA Fund remains largely unchanged. Faculty awards are edging closer to half (45%) of the total cumulative awards. Graduate students and postdoctoral researchers represent exactly half (50%) of the total articles funded. Research Assistants, Staff, Fellows, Undergraduate Students and Other affiliations collectively represent a little over 5% of the total articles funded since the inception of the OA Fund in 2013.

Figure 4.

OA Articles Funded by Recipient Status 2022 Only
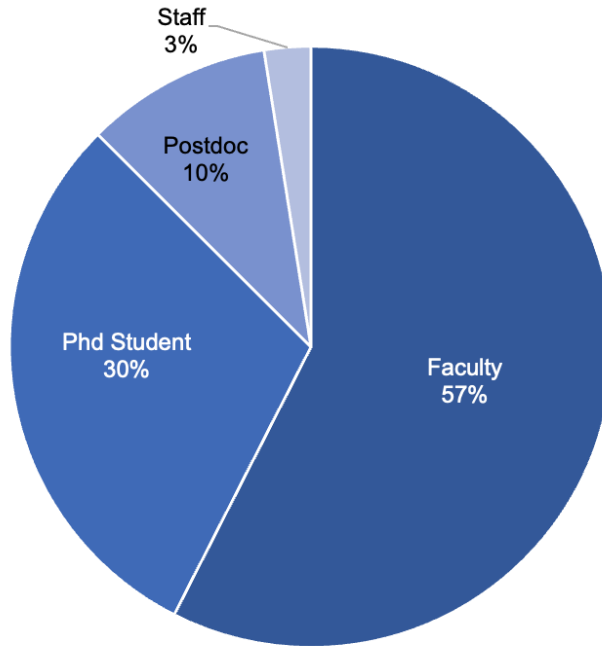
Similar to previous years, the majority of funding for 2022 was awarded to faculty, PhD students, and postdocs. One award was provided to a Staff member in 2022. The proportion of PhD recipients continued to rebound from a low of 23% in 2020 to nearly a third (30%) of recipients in 2022. The number of postdoc recipients was only half of the proportion from the previous year (21% in 2021 compared to 10% in 2022). The percentage of faculty awards rose from 51% in 2021 to 57% in 2022, though the total number of faculty recipients was actually slightly lower in 2022 (23) compared to 2021 (26) due to less awards given in 2022 overall.

Figure 5.

OA Funded Articles by Department - Cumulative
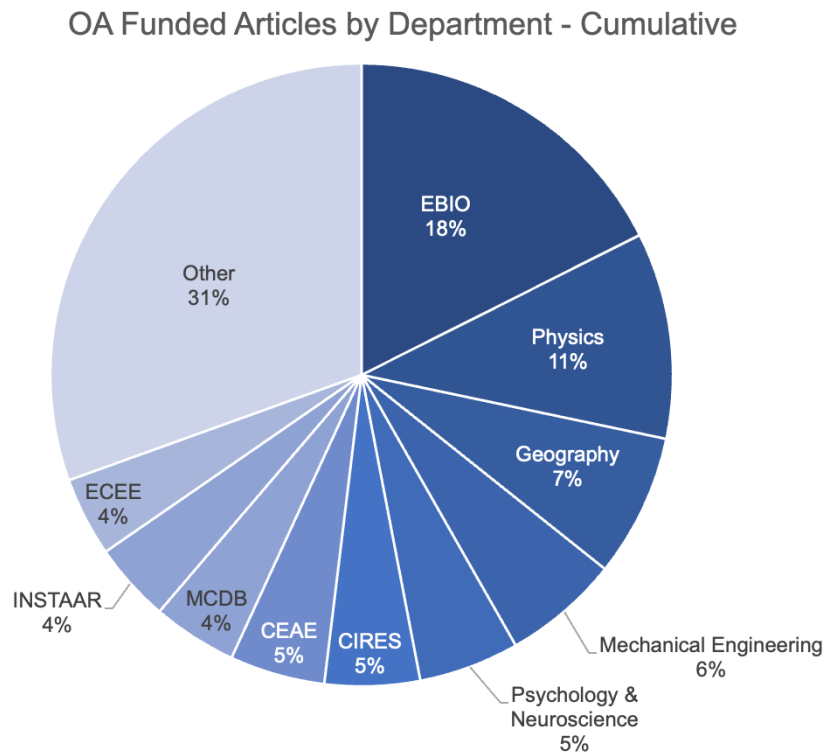
The cumulative data for articles funded by department show no significant changes in 2022. EBIO and Physics continue to represent the largest share of overall awards (EBIO: 21% in 2019, 19% in 2020, 18% in 2021 and 2022; Physics: 14% in 2019, 13% in 2020, 11% in 2021 and 2022) with no change in the top seven departments from the previous year. There was also no change in the top ten departments, solidifying those listed in Figure 5 as the top users overall of the OA Fund. Departments with individuals receiving funding for the first time in 2022 included JILA, the Institute for Behavioral Genetics, and the College of Music.

Figure 6.

### OA Funded Articles by Department: 2022 Only
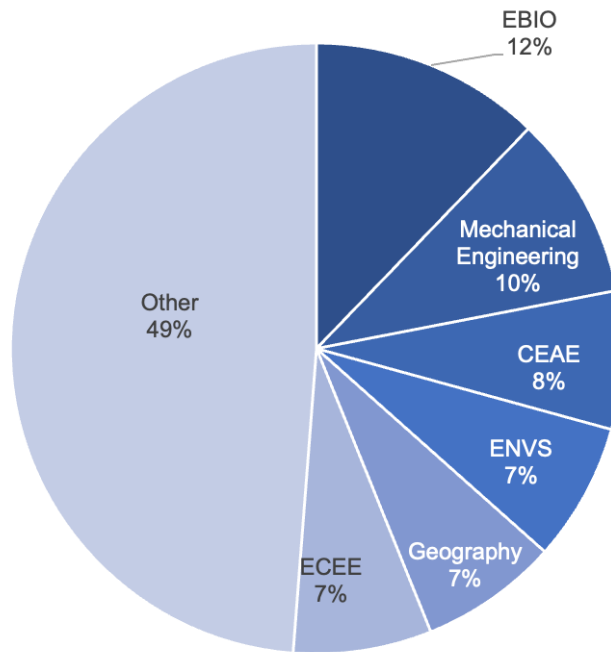
Top departments receiving funding in 2022 continue to mirror the top departments cumulatively, though there are some minor notable shifts. Funding for EBIO dipped slightly (from 15% in 2021 to 12% in 2022). Departments shown in the chart for 2022 above outside of the Other category represent all departments which received more than 2 awards for the year, and by this criteria CIRES and INSTAAR both fell out of the top departments for the year. Additionally, Mechanical Engineering doubled its awards in 2022 - going from two awards in 2021 to four awards in 2022 - making it the second highest awarded department for the year, whereas it wasn't included in the top departments in 2021 at all. Fifteen unique departments had two or fewer awards in 2022 (compared to 16 in 2020 and 27 in 2021). This drop can be attributed to fewer total awards being funded in 2022 compared to the previous year.

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 recent reports, this 2023 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.

Number of Items in CU Scholar by Content Type
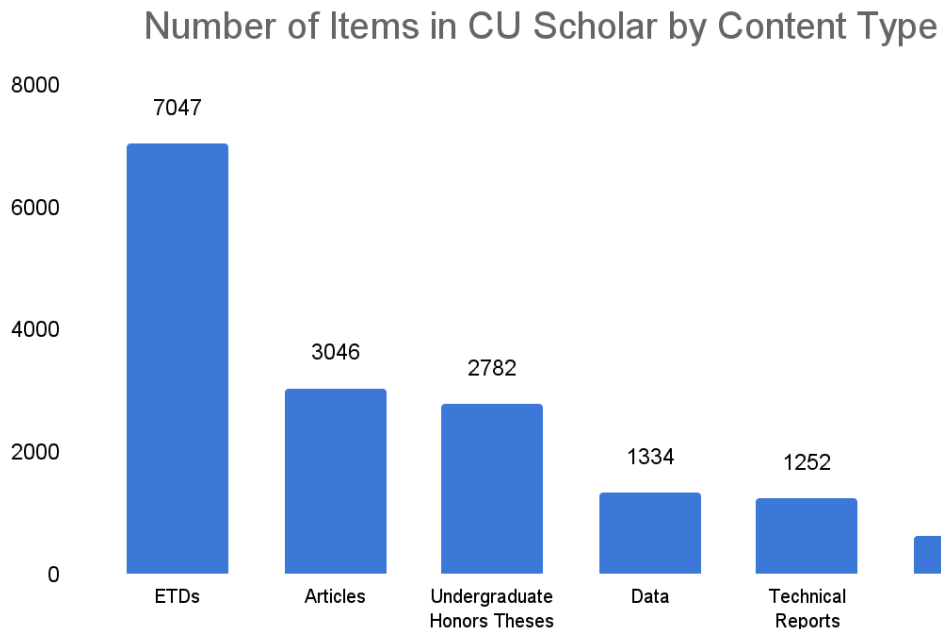
At the end of 2022, CU Scholar contained 16,090 items including journal articles, data sets, graduate theses and dissertations, undergraduate honors theses, conference materials, books, and book chapters. This represents an increase of 16.67% from 13,791 total items at the end of 2021. Content on CU Scholar was downloaded 36,730 times in 2022 according to data exported from Google Analytics. Total downloads decreased slightly by 6.76% compared to 2021.

In 2022, graduate theses and dissertations made up 43.80% of all content in the repository (up slightly from 43.28% in 2021). Undergraduate honors theses accounted for 17.29% of CU Scholar content (down slightly from 18.34% in 2020) while articles and proceedings comprised 18.93% of repository items. Technical reports represented 7.78% of the repository with most items belonging to a computer science technical

reports collection that is not actively growing. Data sets now account for 8.29% of CU Scholar contents (up from 7.28% in 2021), representing the fastest growing segment of the repository in terms of content types.

V. Open Data at CU Boulder

Figure 8.

Number of Reported Data Sets by Year
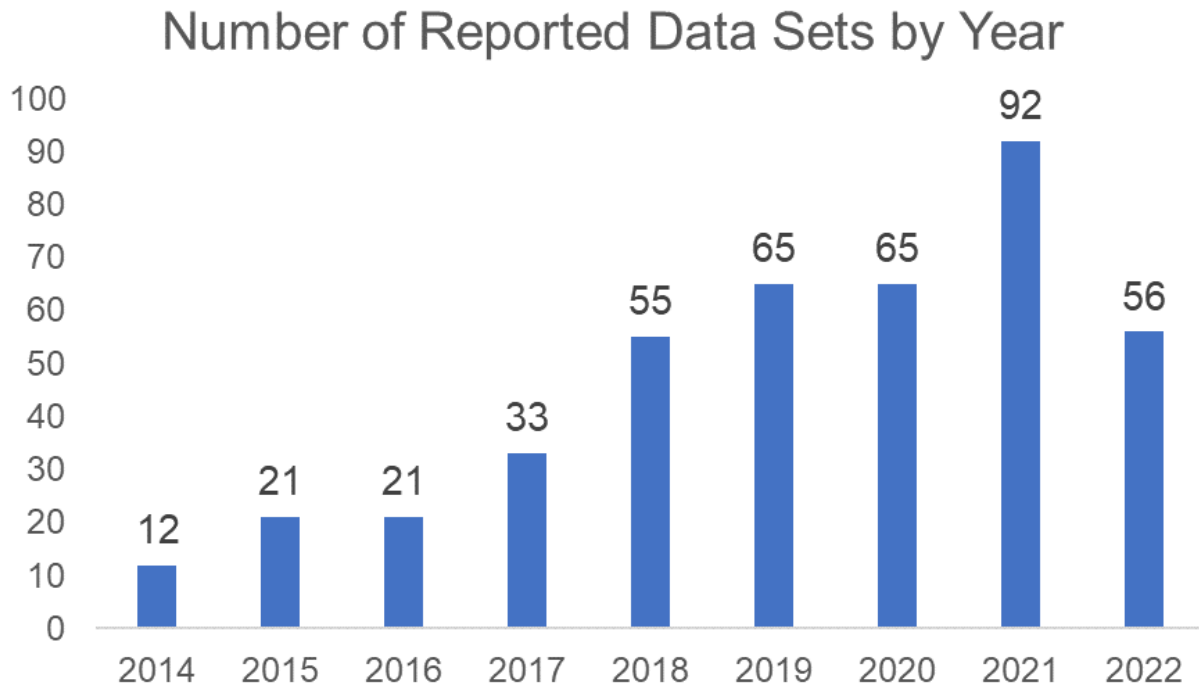
After a large increase in data sets in 2021, the number of published data sets reported on the annual Faculty Report of Professional Activities (FRPA) in 2022 decreased to 56. While this seems like a large drop, the number is comparable to those from 2018 (55), 2019 (65), and 2020 (65) and can mostly be traced to one department not depositing as many data sets this year (see discussion of Table 3). Since this is the first year there has been any decrease in the number of published data sets, it will be important to monitor this number next year for evidence of possible new trends.

Table 3. Reported Data Sets by Department/Unit, 2014-2022 (n>1)

Department/Unit	Number of Reported Data Sets (2014-2021)	Number of Reported Data Sets (2014-2022)	New Data Sets (2022)
Cooperative Institute for Research in Environmental Sciences	54	54	0
Atmospheric and Oceanic Sciences	39	47	8
Environmental Studies	22	40	18
Civil, Environmental and Architectural Engineering	32	34	2
Ecology and Evolutionary Biology	29	31	2
Institute for Arctic and Alpine Research	25	27	2
Geography	17	24	7
Geological Sciences	17	18	1
Astrophysical and Planetary Sciences	16	16	0
Computer Science	11	14	3
Sociology	8	13	5
Linguistics	11	12	1
Libraries	9	9	0
Chemistry	6	8	2
Laboratory for Atmospheric and Space Physics	7	8	1
Business	6	7	1
Chemical and Biological Engineering	6	6	0
Classics	6	6	0
Aerospace Engineering Sciences	5	5	0
Environmental Design	5	5	0
Education	3	4	1
Information Science	3	4	1
History	3	3	0

Media Studies	3	3	0
Molecular, Cellular & Developmental Biology	3	3	0
Physics	3	3	0
Asian Languages and Civilizations	2	2	0
Ethnic Studies	2	2	0
Journalism	2	2	0
Natural History Museum	2	2	0
Speech, Language, and Hearing Sciences	2	2	0

Table 3 provides updated information on the distribution of data sets published from 2014 to 2022, across disciplines. The top of the distribution looks fairly similar to the distribution reported in our previous report (which included data through the year 2021), with a few notable outliers. In particular, the number of data sets reported by the department of Environmental Studies and the department of Atmospheric and Oceanic Sciences featured the largest increases from 2021 to 2022. The cumulative number of data sets reported by the former increased from 22 in 2021 to 40 in 2022 (an increase of 18), while the cumulative number of data sets reported by the latter rose from 39 data sets in 2021 to 47 data sets in 2022 (an increase of 8). There were no new entrants to the catalog of departments that have published at least two data sets since 2014 this year. The Sociology department continued its recent growth, starting with three data sets in 2020 and adding five more in both 2021 and 2022; this sustained increase seems to 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. It is also worth mentioning the decrease in new data sets reported by the Civil, Environmental, and Architectural Engineering department. In 2021 they reported publishing 26 data sets, but that number fell to only two in 2022.

Figure 9.

Data Sets with DOIs or URLs by Year

The use of DOIs for reported data sets decreased from 76 in 2021 to 49 in 2022. In 2022, 49 out of 56 data sets included DOIs, leading to an overall percentage of 87.5%, which represents an all-time high (the previous high was 82.6%, which was achieved in 2021). 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.

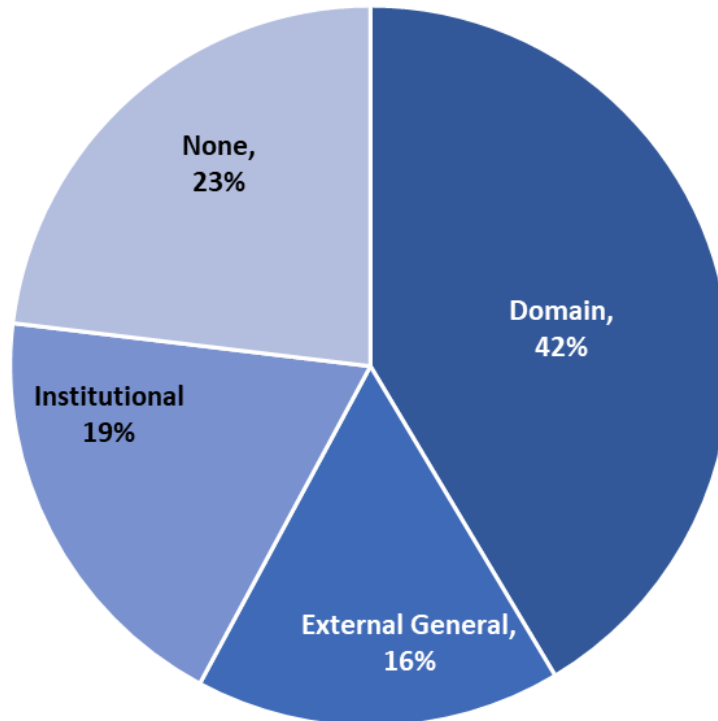Repositories Used by Type, 2014-2022

Figure 10 presents an updated summary of repository types, with 2022 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, to 23% in 2022. Similar to last year, a plurality of published data sets (42%) 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 19% of the reported published data sets. Five of the data sets published in institutional repositories appear in the institutional repositories of other institutions (three at Stanford, one at UCSD, and one at William and Mary) with the remainder appearing in CU Scholar.

Figure 11.

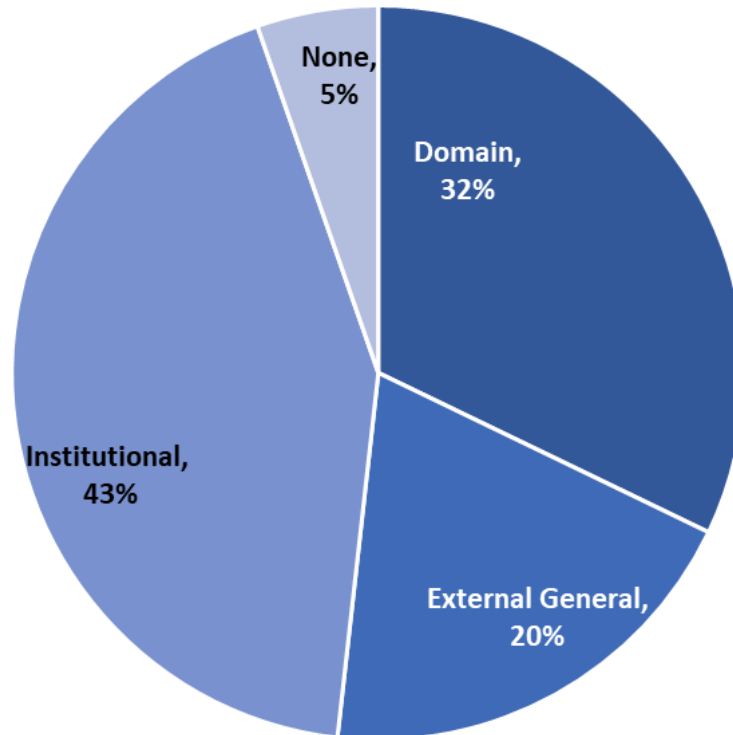### Repositories Used by Type, 2022

Figure 11 presents a breakdown of repository destinations exclusively for data sets published in 2022. The percentage of data sets that were not published in a formal repository of any kind stayed steady at 5% in 2022, compared to 5% in 2021 and 17% in 2020. Fully 95% of the data sets released in 2022 were therefore published in a formal repository of some kind. A plurality of the data sets published in 2022 (43%) were published in an institutional repository, which represents an increase from 2021 and 2020, when 25% and 9% respectively were published in institutional repositories. Nearly a third of 2022 data sets (32%) were published in domain repositories, compared with 57% in 2021 and 48% in 2020. One-fifth (20%) of data sets were published in external general repositories, compared with 13% in 2021 and 26% in 2020.

Figure 12
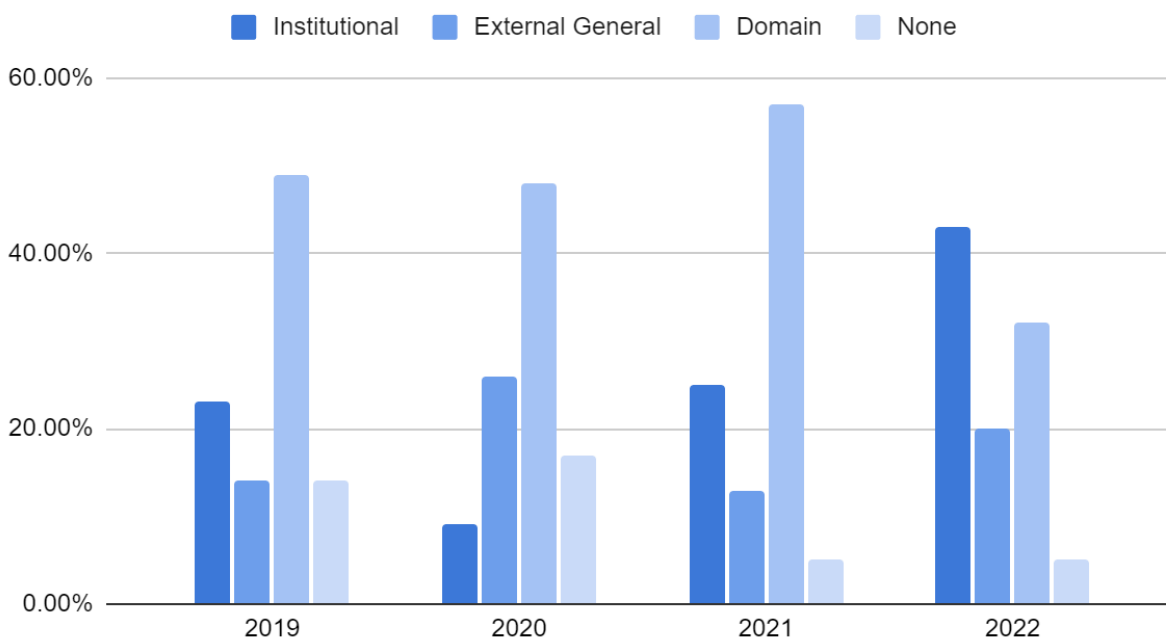
Repositories used by type (2019-2022)

Figure 12 shows that the distribution of data sets across institutional, domain, and external general repositories continues to vary year-by-year, making it hard to judge trends.

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 2022, the Libraries and its partners registered 335 DOIs for published data sets (down from 416 in 2021). The disconnect between this larger number of DOIs and the 56 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 decreasing 20% from 2021 to 2022 is the first time there has not been an increase in the number of DOIs registered since the service began in 2018. This will be an important data point to monitor in the coming years to see if it is the start of a

downward trend or plateau with regard to the number of data sets published in CU Scholar each year. As of the end of 2022, the Libraries and its partners had registered a total of 1,378 DOIs for data sets since the DataCite DOI registration service began at CU Boulder.