2024 State of Open at the University of Colorado Boulder

An Update on Open Access Practices Based on Data from 2023

October 2, 2024

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 2023, this report is the sixth 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/wzkx-pn69.

Key findings from this report include:

- 75% of articles published in 2023 by CU Boulder authors are available via some type of OA (Gold, Green, Hybrid, or Bronze) (up from 72% in 2022);
- In 2023, the CU Boulder Libraries OA Fund funded author fees totaling $86,928 for 47 journal articles published by CU Boulder authors in full OA journals (up from $69,804 for 41 journal articles in 2022);
- As noted in a new section added this year, the estimated total cost of APCs for all 2023 OA articles with a CU Boulder author is nearly $6.6 million;
- At the end of 2023, there were 17,448 OA items in the CU Scholar institutional repository (up from 16,090 in 2022), and these items were downloaded a total of 36,926 times in 2022 (nearly the same as 36,730 in 2022);
- In the annual Faculty Report of Professional Activities (FRPA), faculty reported 83 published data sets in 2023 (up from 56 in 2022) with 87.95% of these citations including Digital Object Identifiers (DOIs) (nearly the same as 87.50% in 2022) and 95.20% of these citations identifying a formal data repository (nearly the same as 95.00% in 2022);
- The Libraries and its partners registered 309 DataCite DOIs for published data sets in 2023 (down from 335 in 2022);

II. Open Access Articles by CU Boulder Faculty

Continuing with a change that was first implemented in 2021, this 2024 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, 2014-2023

Year	Closed (n)	Gold (n)	Green (n)	Hybrid (n)	Bronze (n)	Total OA (n)	Total (n)
2023	963	1163	610	707	355	2835	3798
2022	1003	1093	861	488	485	2927	3930
2021	1216	949	1054	452	702	3157	4373
2020	1320	900	948	389	856	3093	4413
2019	1160	827	767	347	947	2888	4048
2018	1241	825	687	336	1094	2942	4183
2017	1256	735	734	286	903	2658	3914
2016	1332	763	625	306	990	2684	4016
2015	1457	598	637	225	840	2300	3757
2014	1528	464	649	179	649	1941	3469

Table 2 provides the total number of articles published by CU Boulder faculty each year from the last 10 years (2014 to 2023) 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 2014 to 2023. 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 be plateauing in the 3000-3100 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.

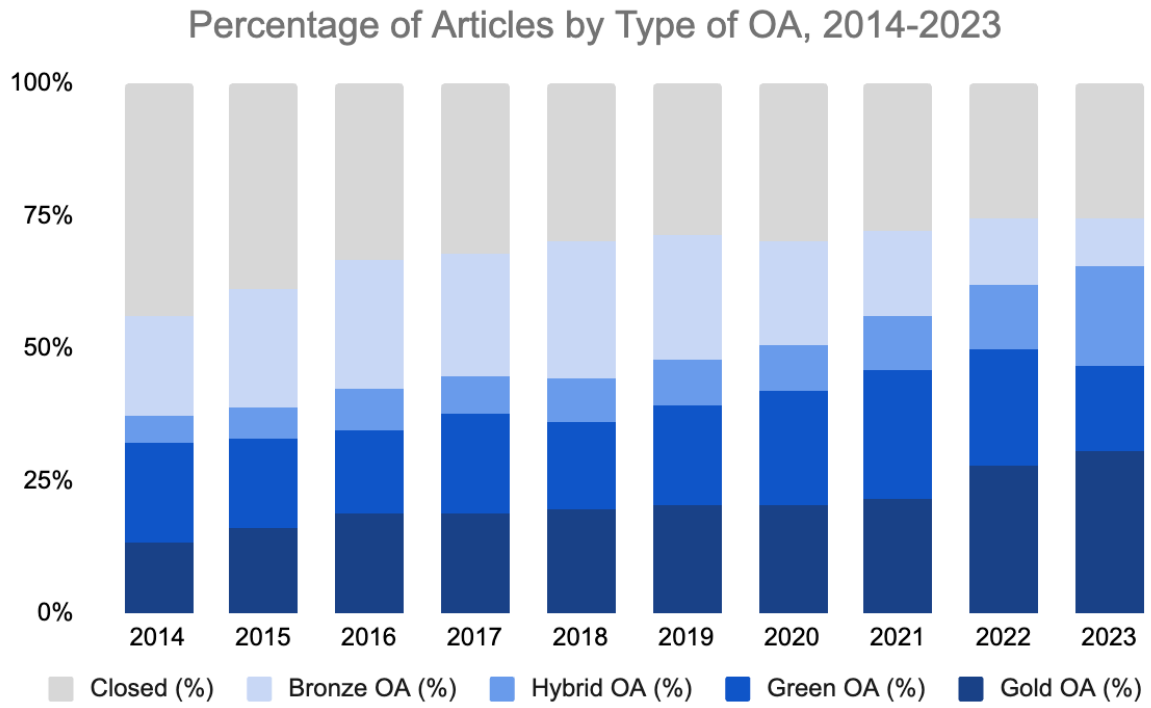
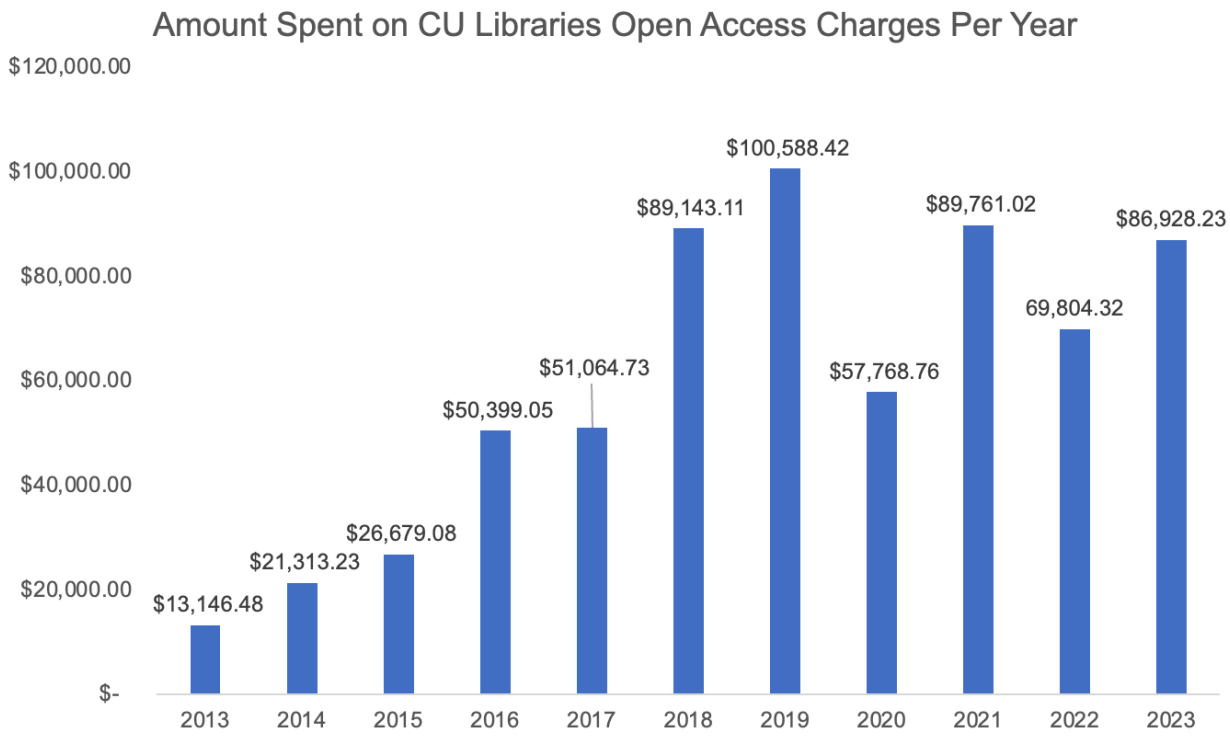


Figure 1 shows the trends in percentage of type of OA and closed articles over time. Overall, the total percentage of OA articles increased from a low of 55.95% in 2014 to a high of 74.64% in 2023. 2020 is the only year where the total percentage of OA articles did not increase from the previous year, and this was only a slight decrease from 71.34% in 2019 to 70.09% in 2020 before increasing again to 72.19% in 2021. 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 13.38% of all articles published in 2014 to a high of 30.62% of articles in 2023. The percentage of Green OA articles also generally increased over time leading to a high of 24.10% in 2021 before decreasing somewhat in 2022 to 21.91% and then to 16.06% in 2023. As noted in the previous discussion of Table 2, this decrease in the most recent years is expected since there is a lag time with Green OA articles becoming publicly available in repositories due to embargo periods and other factors. The percentage of Hybrid OA articles steadily increased from 5.16% in 2014 to 18.62% in 2023. While this still accounts for a smaller percentage of total OA articles over time than either Gold OA or Green OA, the general trend toward an increasing number of Hybrid OA articles should be monitored since this type of article is associated with higher article processing charges (APCs) and problematic publisher business models. The percentage of Bronze OA articles appears to be in steady decline since 2018 with the percentage dropping

3-4% each year culminating in an overall low of 9.35% in 2023. This could suggest that publishers' licensing practices are becoming increasingly consistent over time.

III. CU Boulder Libraries Open Access Fund

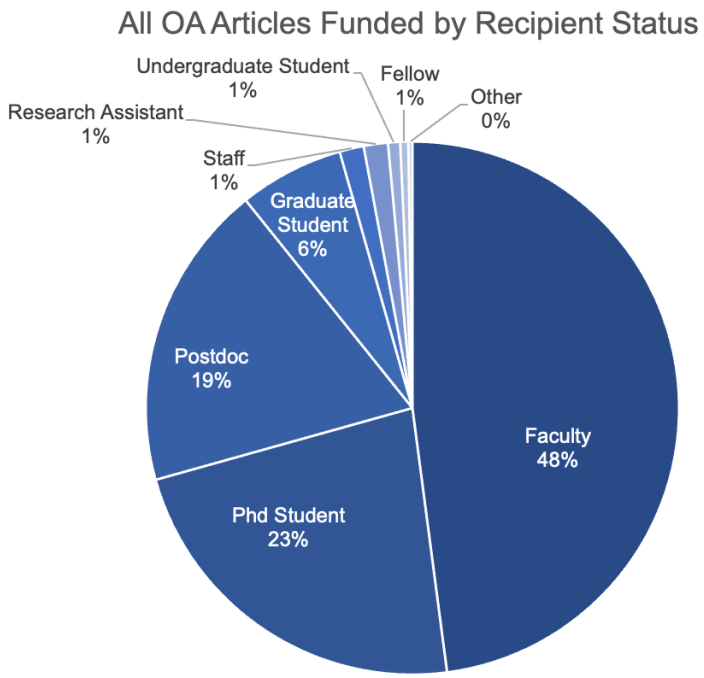
Figure 2.



CU Boulder Libraries spent $86,928.23 on open access article processing charges for fully open access journal publication in 2023. In 2023, the CU Boulder Libraries Open Access Guild recommended capping the general OA Fund at $100k, so as of 2023 total spending per year on this fund should never exceed that amount. Rather, it is the hope of the Libraries to increase funding for direct publisher memberships (e.g. Frontiers, for which the Libraries paid nearly $80k in 2023 in APC fees) as well as transformative agreements.

47 articles were funded in 2023, with an average APC cost of $1,849.54. This average jumped $147 from the average APC cost in 2022, representing an 8.6% inflation increase. Cumulatively, the CU Boulder OA Fund has helped authors publish 411 fully OA articles in 175 unique journal titles.

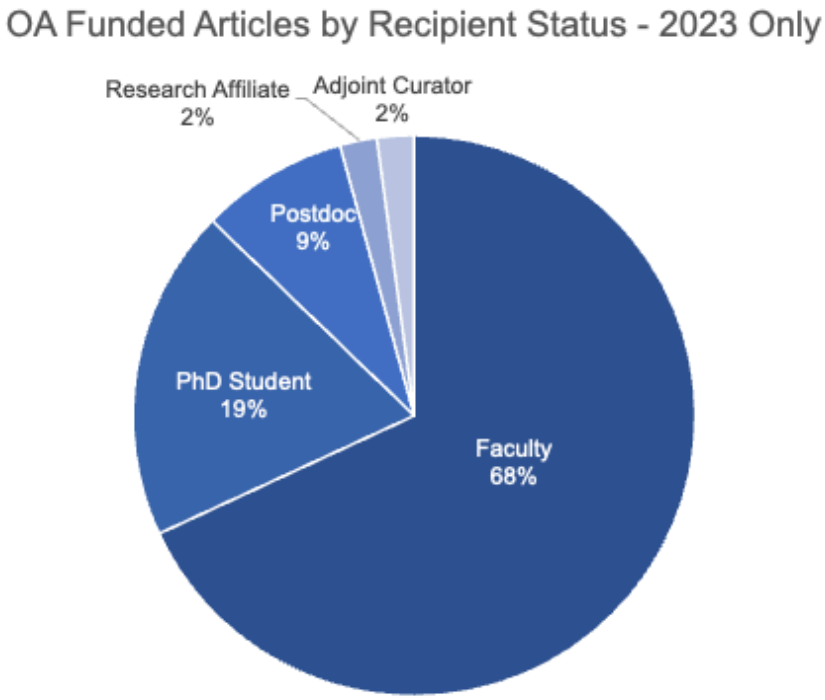
Figure 3.



As usual, the proportion of funded authors by university status over the lifetime of the OA Fund remains largely unchanged. Faculty awards continue to represent a larger portion of total awards, now representing nearly half (48%) of the total cumulative awards.

Graduate students and postdoctoral researchers represent just under half (48%) of the total articles funded. Research Assistants, Staff, Fellows, Undergraduate Students and Other affiliations collectively represent 4% of the total articles funded since the inception of the OA Fund in 2013.

Figure 4.



Similar to previous years, the majority of funding for 2023 was awarded to faculty, PhD students, and postdocs. Faculty alone represented a little over two-thirds (68%) of awards in 2023. One award each was provided to a Research Affiliate and an Adjoint Curator in 2023.

While in the previous year the proportion of PhD awardees had rebounded to 30% of recipients, this demographic fell back to an all time low of 19% of recipients in 2023. Postdoc recipients continued a trend of declining awards in 2023, representing 9% of awardees in 2023 (down slightly from 10% in 2022, and a high of 21% in 2021).

Table 3. Cumulative OA fund Awards, 2013-2023

Department/Unit	Count of OA Fund Awards (Cumulative)
EBIO	69
Physics	44
Geography	29
Mechanical Engineering	27
CIRES	22
Psychology & Neuroscience	21
CEAE	21
MCDB	20
INSTAAR	17
ECEE	17
Environmental Studies	15
Integrative Physiology	12
Geological Sciences	10
Atmospheric and Oceanic Sciences	6
Chemistry and Biochemistry	5
Computer Science	5
Museum of Natural History and Geological Sciences	5
Applied Math	4
Institute of Behavioral Sciences	4
Materials Science and Engineering Program	4

Speech, Language, and Hearing Sciences	4
Astrophysical & Planetary Sciences	3
Institute of Cognitive Science	3
Journalism	3
Sociology	3
School of Education	3
Chemical and Biological Engineering	3
Information Science	3
Anthropology	2
Biology	2
Aerospace Engineering	2
Political Science	2
Center for STEM Learning	1
Environmental Engineering	1
Latin American Studies Center (LASC)	1
Mathematics	1
Media Research and Practice	1
Microbiology	1
Program in Environmental Design	1
Silicon Flatirons Center at the law school	1
STEM Center and Physics	1
Student Academic Success Center	1
Sustainability Innovation Lab at Colorado	1

Technology, Cybersecurity, and Policy	1
Wardenberg Medical Services	1
Advertising, Public Relations, and Media Design	1
Libraries	1
Theatre & Dance	1
JILA	1
Institute for Behavioral Genetics	1
College of Music	1
CU Population Center	1
Biofrontiers Institute	1

The cumulative data for articles funded by department show no significant changes in 2023.

EBIO and Physics continue to represent the largest share of overall awards (EBIO: 19% in 2020, 18% in 2021 and 2022, and 17% in 2023; Physics: 13% in 2020, 11% in 2021, 2022, and 2023) with no change in the top seven departments from the previous year. Departments with individuals receiving funding for the first time in 2023 included the CU Population Center and Biofrontiers Institute.

Figure 5.

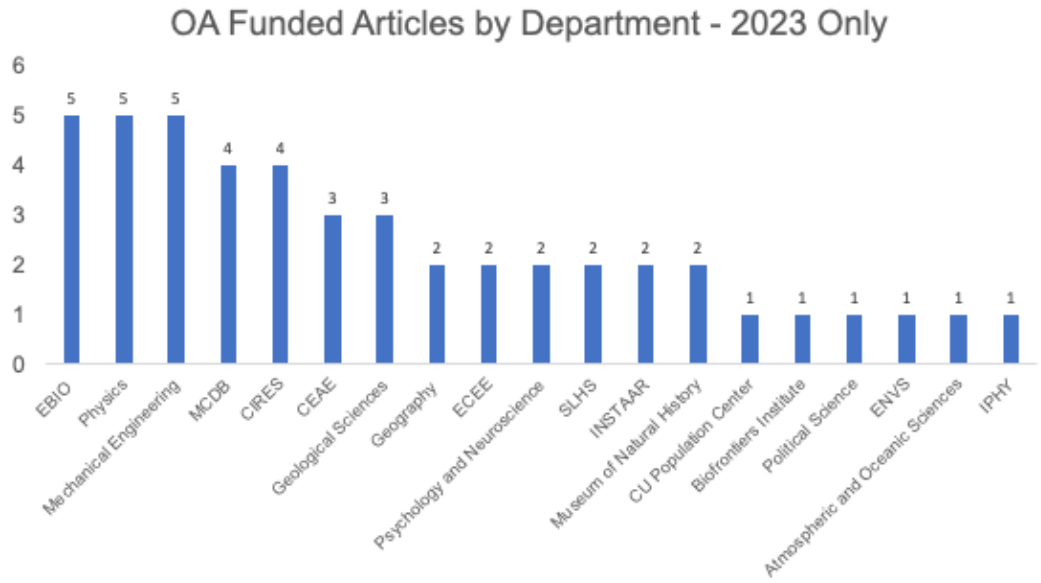


Figure 6 represents all of the representative departments for articles that received funding in 2023. Top departments receiving funding in 2023 continue to mirror the top departments cumulatively. However, a trend of declining funding for EBIO continued in 2023 (15% in 2021, 12% in 2022, and 11% in 2023). Affiliates from Mechanical Engineering, however, continue to use the OA Fund slightly more every year (with 2 awards in 2021, 4 in 2022, and 5 in 2023). There were 19 unique departments that had individuals who received funding in 2023, down slightly from 21 departments in 2022.

IV. Article Processing Charges from CU Boulder Publications

This section, new for this 2024 report, is a follow-up to a special report published in 2021 estimating the total cost of article processing charges (APCs) for CU Boulder publications from the year 2020.[5] In 2021, our report used Dimensions as a primary data set to identify open access articles by type, and then manually cross-listed journal titles with APC prices in the Directory of Open Access Journals (DOAJ). For hybrid journals without prices listed, scholarly literature was used to determine an average APC, which was then extrapolated to all hybrid articles reported in 2020. The report ultimately estimated that a little over $2.2 million was spent on APCs for articles authored by CU Boulder faculty members.

For this report, we were able to employ an updated and likely more accurate methodology for estimating APCs. OpenAlex, a bibliographic catalog of scientific papers, authors, and institutions which collates data from a variety of open sources, launched in 2022 and provides a much more rigorous snapshot of research outputs, including APCs. OpenAlex primarily deploys the DOAJ API to determine Gold article APC prices, but also creates a discrepancy between the APC “list price” and “paid price” based on available discounts, and includes some prices of Hybrid articles when they are available.

OpenAlex was able to provide an APC price for 1,039 Gold articles and 363 Hybrid articles that were published in 2023. There were an additional 87 Gold articles without a price listed, but since many Gold journals do not require an APC, these were left blank. There were also an additional 591 Hybrid articles without an APC price listed. Since all Hybrid journals charge an APC, we calculated an average APC price paid for the 363 articles that did list a price, and used that average for all Hybrid articles without a price listed. This average should be a more accurate representation of Hybrid APCs paid by CU Boulder authors because instead of relying on averages across the entire spectrum of Hybrid literature, it is derived only from titles in which CU Boulder faculty are actually publishing.

[5] https://scholar.colorado.edu/concern/reports/zc77sr262

Figure 6.

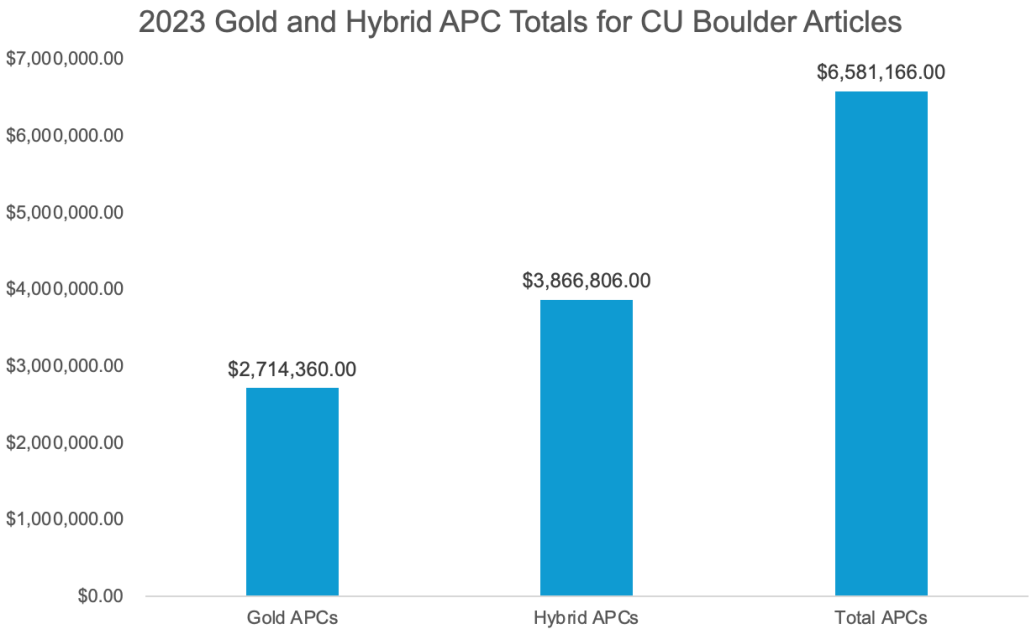


Figure 6 displays the total estimated APCs paid across Gold and Hybrid journals in 2023. The estimated total APC spend in 2023 for CU Boulder open access articles is nearly $6.6 million. Notably, this figure is nearly three times the estimated total APCs in the 2021 report. The significant jump between these analyses is likely due to a combination of factors, including increased uptake of open access as well as a more accurate snapshot from new data sources in this most current report.

Similarly to the 2021 report, we must provide the caveat that this estimate does not guarantee CU Boulder authors are responsible for all $6.6 million in APC funding. The costs estimated here do not account for the vast array of available funding sources, which may include waivers, grants, and departmental funds as well as out-of-pocket costs. Additionally, some proportion of these articles may be paid by co-authors from other universities or private sectors.

Figure 7.

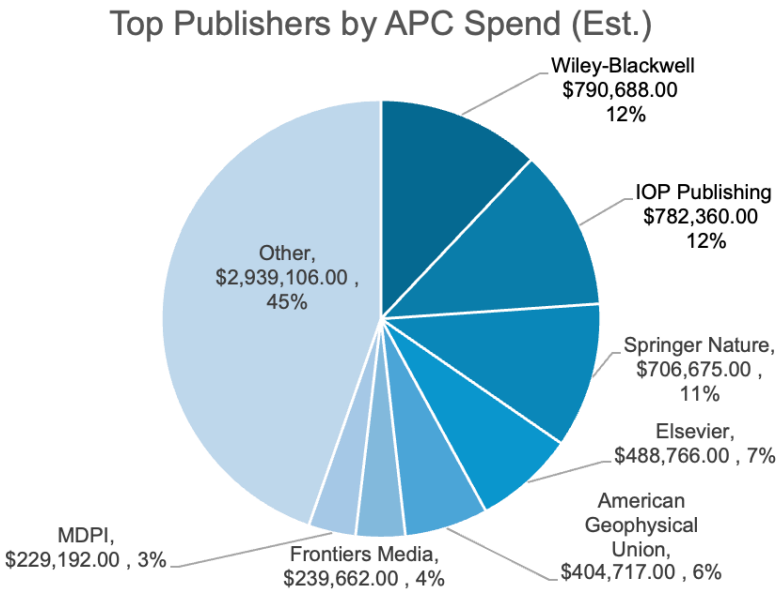


Figure 7 shows the top publishers by the amount paid to them for APCs. The top seven publishers account for 55% of all APCs, with Wiley-Blackwell and IOP Publishing alone accounting for nearly a quarter of all APCs paid by CU Boulder authors. The remaining 45% of APCs (just under $3 million) went to at least 110 other publishers. It was not possible to determine the exact remaining number of publishers since in some cases the publisher field was left blank in the OpenAlex data.

Of the top 7 publishers (all which account for over $200k in APC payments), the Libraries established a new agreement with IOP Publishing in 2024, has a membership account with Frontiers to which it currently contributes around $60k, and has a membership with MDPI which provides 10% discounts to CU Boulder affiliated publications.

V. 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 2024 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 different methods (Google Analytics and GA4) 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 8.

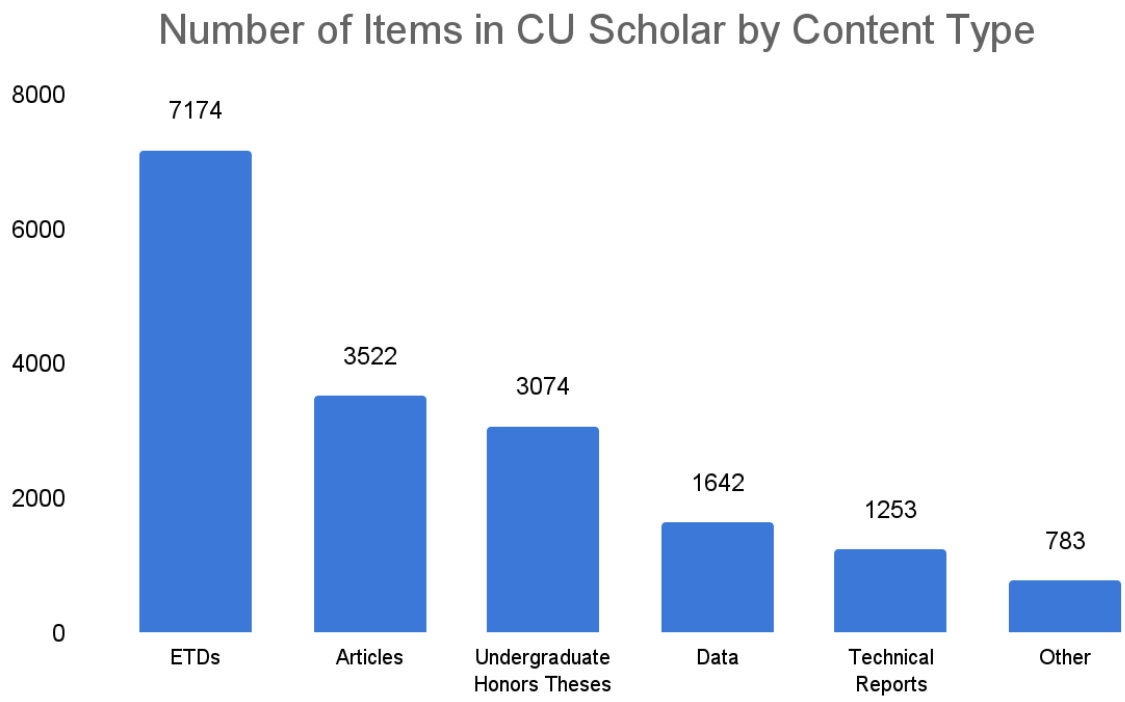


Figure 8 shows the total number of items in CU Scholar at the end of 2023 by content type. At the end of 2023, CU Scholar contained 17,448 items including journal articles, data sets, graduate theses and dissertations, undergraduate honors theses, conference materials, books, and book chapters. This represents an increase of 8.44% from the end of 2022. Content in CU Scholar was downloaded 36,926 times in 2023 according to data exported from Google Analytics. Total downloads were nearly the same as in 2022 when there were 36,730.

In 2023, graduate theses and dissertations made up 41.11% of all content in the repository (down slightly from 43.80% in 2022). Undergraduate honors theses

accounted for 17.62% of CU Scholar content (nearly the same as 17.29% in 2022) while articles and proceedings jumped to 20.19% from 18.93% between 2022 and 2023. Technical reports represented 7.18% of the repository with most items belonging to a computer science technical reports collection that is not actively growing. Data sets now account for 9.41% of CU Scholar contents (up from 8.29% in 2022), representing one of the fastest growing content types in the repository. Other works consisting largely of books, working papers, presentations, and reviews account for 4.49% of the repository's content.

VI. Open Data at CU Boulder

Figure 9.

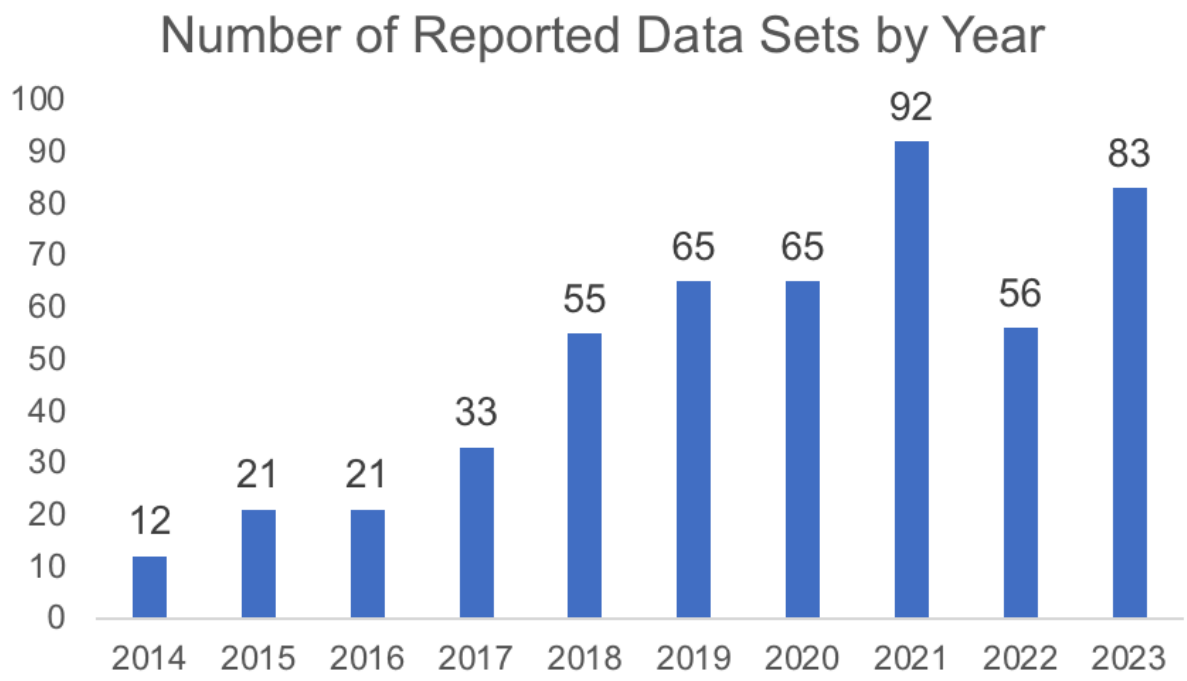


Figure 9 shows the number of published data sets reported in the annual Faculty Report of Professional Activities (FRPA) in 2023. This year the number of data sets increased to 83, which is the second highest number so far and is comparable to the 92 datasets in 2021.

Table 4. Reported Data Sets by Department/Unit, 2014-2022 (n>1)

Department/Unit	Number of Reported Data Sets (2014-2022)	Number of Reported Data Sets (2014-2023)	New Data Sets (2023)
Environmental Studies	40	58	18
Atmospheric and Oceanic Sciences	47	54	7
Civil, Environmental and Architectural Engineering	34	43	9
Ecology and Evolutionary Biology	31	34	3
Geography	24	32	8
INSTAAR	27	28	1
Geological Sciences	18	25	7
Sociology	13	19	6
Astrophysical and Planetary Sciences	16	16	0
Linguistics	12	15	3
Computer Science	14	14	0
Libraries	9	10	1
Chemistry	8	9	1
LASP	8	8	0
Business	7	7	0
Aerospace Engineering Sciences	5	6	1
Chemical and Biological Engineering	6	6	0
Classics	6	6	0
History	3	6	3
Education	4	5	1
Environmental Design	5	5	0
Integrative Physiology	1	5	4
Information Science	4	4	0
Electrical, Computer & Energy Engineering	0	3	3

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

Table 4 provides updated information on the distribution of data sets published from 2014 to 2023, across disciplines.

For the second year in a row the department of Environmental Studies has the most new published datasets (18) and this year becomes the department with the most total published datasets (58).

Other departments that continue to have larger numbers of data sets published include Atmospheric and Oceanic Studies (7 new data sets, 54 total), Civil, Environmental and Architectural Engineering (9 new data sets, 43 total), Geography (8 new data sets, 32 total), and Geological Sciences (7 new data sets, 25 total). The Sociology department continued its growth, starting with only three data sets in 2020, it added an additional 6 this year to reach 19 total; this sustained increase seems to reflect a growing emphasis on data publication in the social sciences.

Four new departments were added to the list of those that have published at least two data sets since 2014. They are Integrative Physiology (4 data sets in 2023, 5 total), Electrical, Computer & Energy (3 data sets in 2023, 3 total), Astrophysical & Planetary Sciences (2 data sets in 2023, 2 total), and Molecular, Cellular, and Developmental Biology (2 data sets in 2023, 2 total). There are now a total of 43 departments that have published at least one data set and 34 with at least two data sets. There are 12 departments that have published 10 or more data sets (see Figure 10).

Figure 10.

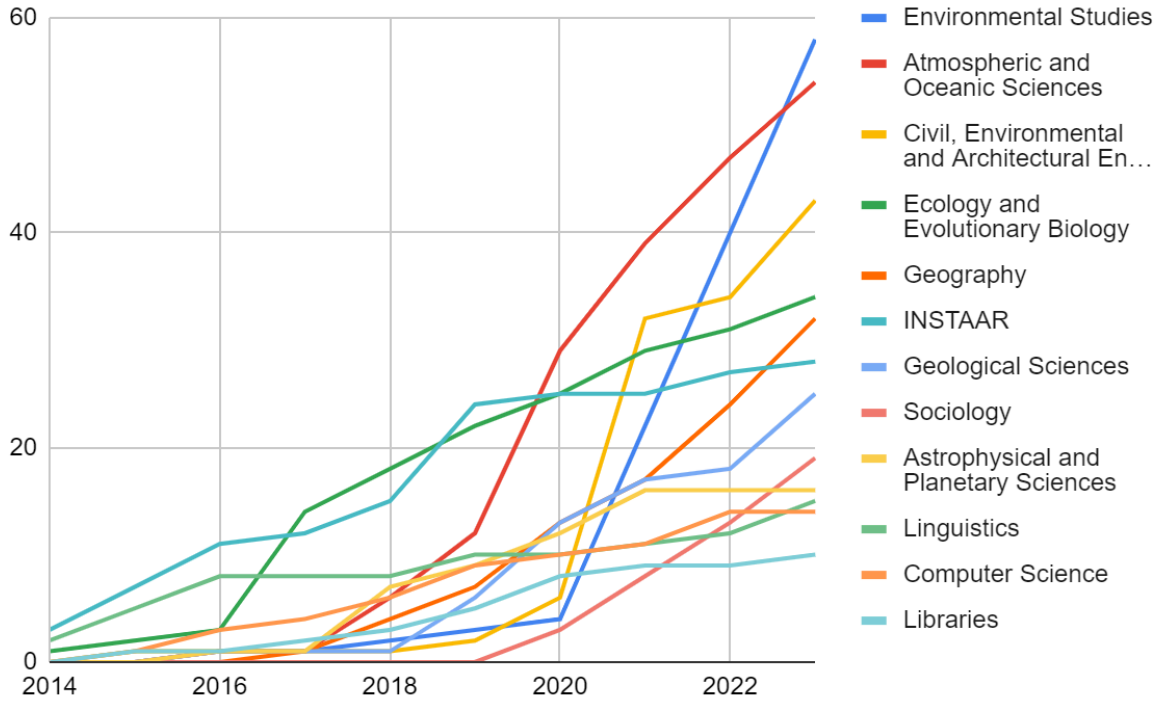


Figure 10 shows the growth of published data sets from the 12 departments that have published at least 10 total data sets. Of note is Civil, Environmental, and Architectural Engineering, who published 26 data sets in 2021, only 2 datasets in 2022, and 9 in 2023.

Figure 11.

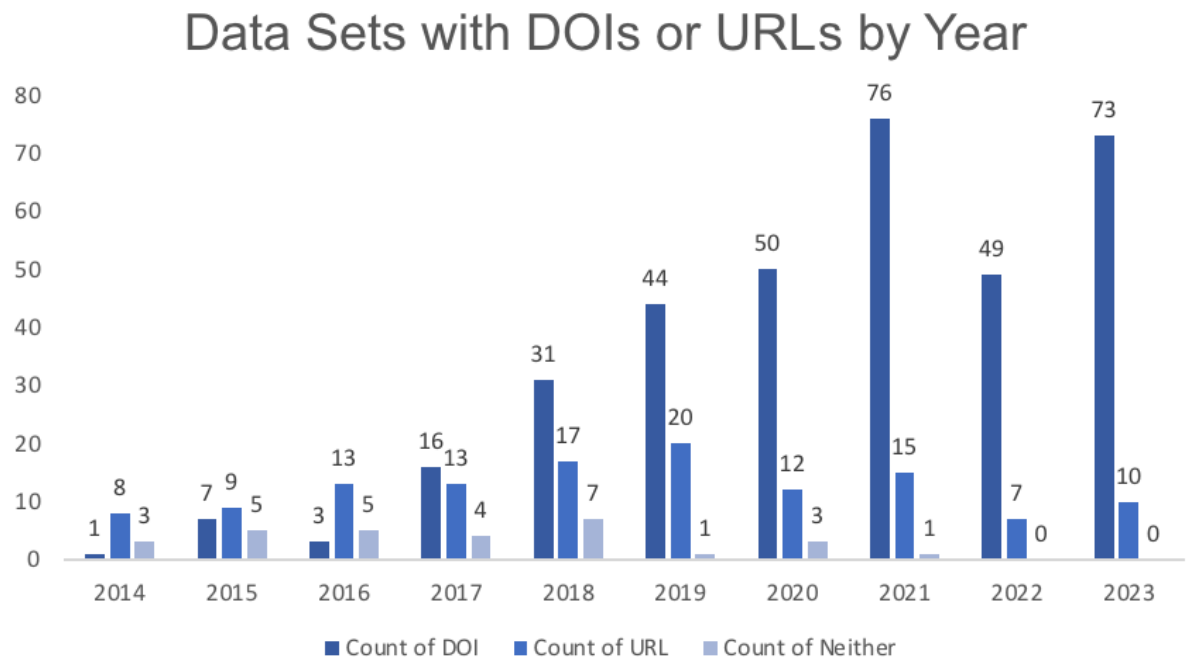


Figure 12.

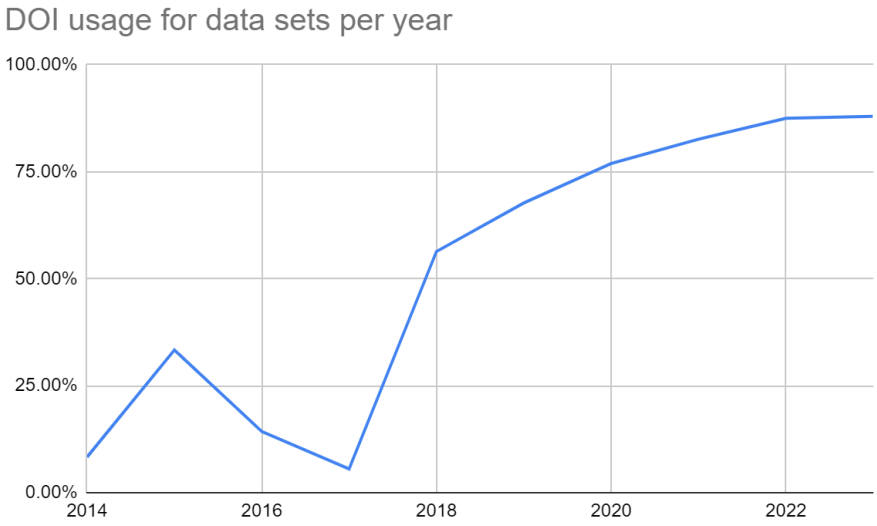


Figure 11 shows the use of DOIs or URLs over time while Figure 12 visualizes the percentage of data sets with a DOI over time. The use of DOIs for reported data sets stayed steady as 73 out of 83 (87.95%) data sets included DOIs, a slight increase over last year (87.50%) and a new all-time high. The 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 to be continuing.

Figure 13.

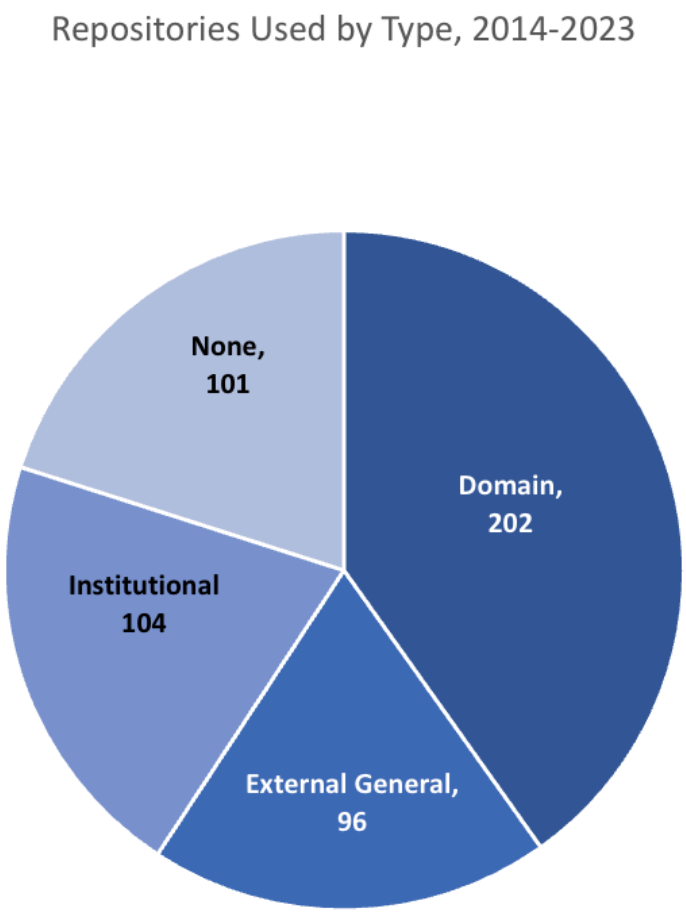


Figure 13 presents an updated summary of repository types, with 2023 data included. This year, the total number of data sets in institutional data repositories surpassed those in informal repositories for the first time (see below). Of the 104 data sets published in institutional repositories, 93 appear in CU Scholar. Of the remaining 11, four are in ScholarWorks@UNO, four are in the Stanford IR, and there is one in each of the UCSD IR, the Virginia Tech Data Repository, and the William and Mary IR. General repositories that are external to CU Boulder and cover a wide range of disciplines and data types (e.g., figshare, Dryad, Zenodo, etc.) are also increasingly being used.

A plurality of published data sets 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.).

Figure 14.

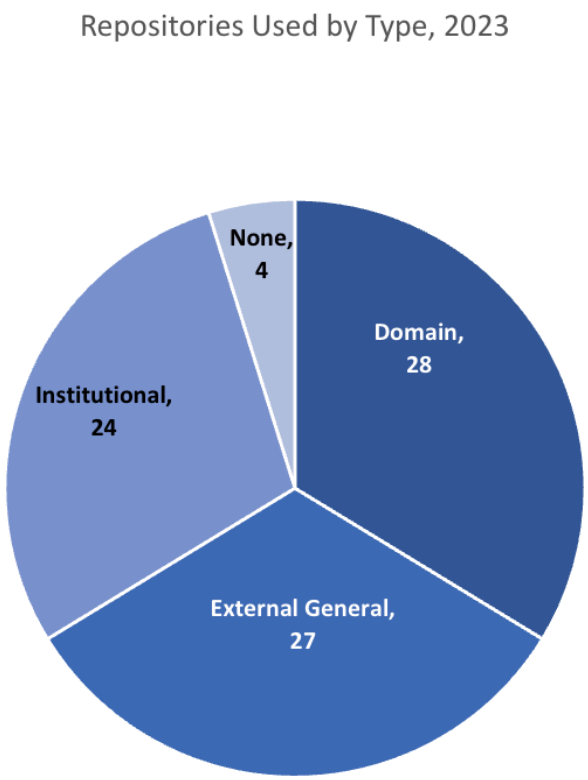


Figure 14 presents a breakdown of repository destinations exclusively for data sets published in 2023. The percentage of data sets that were not published in a formal repository of any kind continued to be low and over 95% of the data sets published in 2023 were published in a formal repository of some kind.

The type of repository used for publishing data sets continues to fluctuate, and this year that was fairly evenly divided between domain-specific (33.7%, slightly up from 32% in 2022), external general repositories (32.5%, up from 20% in 2022), and institutional repositories (28.9%, down from 43% in 2022).

Figure 15.

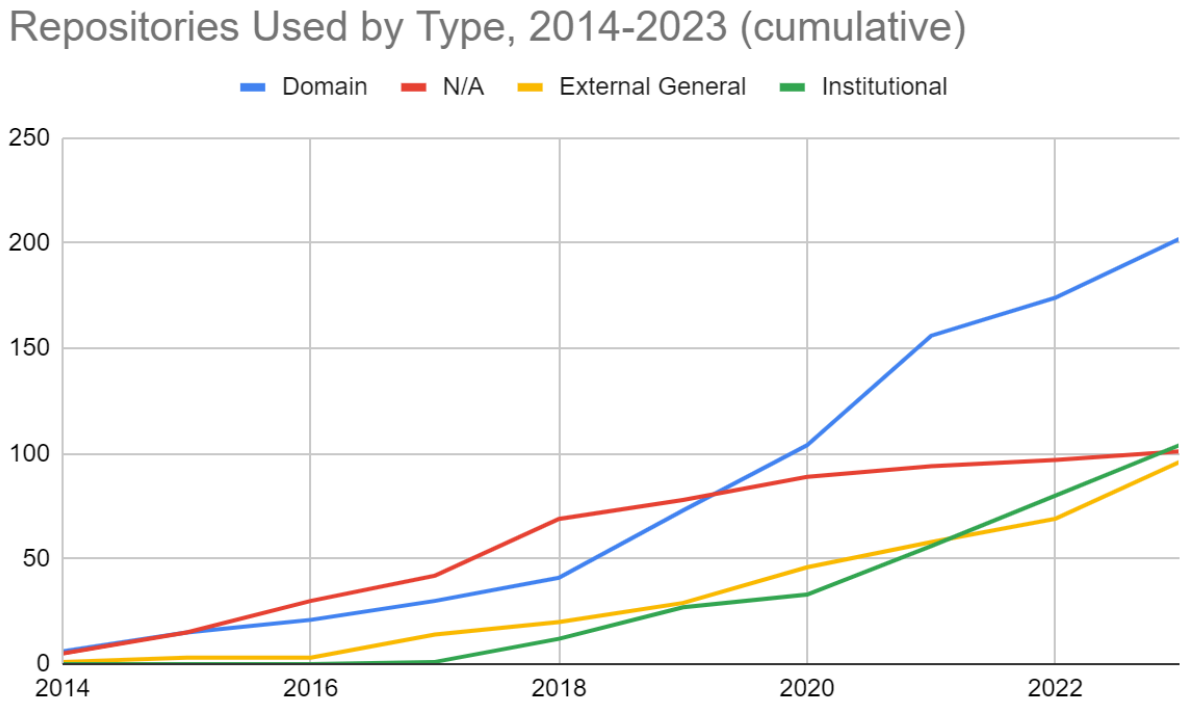


Figure 15 shows the cumulative growth of data sets across institutional, domain, and external general repositories. Domain-specific repositories continue to be the largest (202 data sets, up from 174 in 2022), but external general repositories (96 data sets, up from 69 in 2022) and institutional repositories (104 data sets, up from 80 in 2022) continue to grow while “unofficial” or “informal” repositories stay fairly flat (101 data sets, up from 97 in 2022).

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 2023, the Libraries and its partners registered 309 DOIs for published data sets (down from 335 in 2022). The disconnect between this larger number of DOIs and the 83 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 approximately 8% from 2022 to 2023 is notable as the second year in a row that there has been a decrease in the number of DOIs registered since the service began in 2018. That said, the decrease from 2022 to 2023 was smaller than the first decrease noted from 2021 to 2022 (a decrease of approximately 20%). Since 2022 was the first year there had ever been a decrease, this will be an important data point to keep monitoring to see if the downward trend continues or plateaus with regard to the number of data sets published in CU Scholar each year. As of the end of 2023, the Libraries and its partners had registered a total of 1,697 DOIs for data sets since the DataCite DOI registration service began at CU Boulder.