2025 State of Open at the University of Colorado Boulder

An Update on Open Access Practices Based on Data from 2024

October 1, 2025

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 2024, this report is the seventh 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 and monograph 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/v3bn-sq70.

Key findings from this report include:

- 74% of articles published in 2024 by CU Boulder authors are available via some type of OA (Gold, Green, Hybrid, or Bronze) (nearly the same as 75% in 2023);
- In 2024, the CU Boulder Libraries OA Fund funded author fees totaling $120,193 for 66 journal articles published by CU Boulder authors in full OA journals (up from $86,928 for 47 journal articles in 2023);
- The estimated total cost of APCs for all 2024 OA articles with a CU Boulder author is nearly $6.9 million (up from nearly $6.6 million in 2023);
- For the first time this year, this report provides data on the 11 open monographs funded through the CU Boulder Open Monograph Fund since 2021, which will be updated annually going forward;
- At the end of 2024, there were 19,817 OA items in the CU Scholar institutional repository (up from 17,448 in 2023), and these items were downloaded a total of 35,923 times in 2024;
- In the annual Faculty Report of Professional Activities (FRPA), faculty reported 75 published data sets in 2024 (down slightly from 83 in 2023) with 89.33% of these citations including Digital Object Identifiers (DOIs) (up slightly from 87.95% in 2023) and 96.00% of these citations identifying a formal data repository (nearly the same as 95.20% in 2023);
- The Libraries and its partners registered 544 DataCite DOIs for published data sets in 2024 (up significantly from 309 in 2023);

II. Open Access Articles by CU Boulder Faculty

Continuing with a change that was first implemented in 2021, this 2025 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, 2015-2024

Year	Closed (n)	Gold (n)	Green (n)	Hybrid (n)	Bronze (n)	Total OA (n)	Total (n)
2024	1164	1198	245	902	70	2415	3579
2023	986	1177	355	1029	288	2849	3835
2022	1069	1140	396	1006	322	2864	3933
2021	1194	984	614	1018	473	3089	4283
2020	1358	890	666	800	629	2985	4343
2019	1269	773	541	645	685	2644	3913
2018	1402	768	535	611	705	2619	4021
2017	1385	705	566	484	639	2394	3779
2016	1409	725	605	477	670	2477	3886
2015	1474	561	602	322	623	2108	3582

Table 2 provides the total number of articles published by CU Boulder faculty each year from the last 10 years (2015 to 2024) 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 2015 to 2024. 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 five years indicate that this might be plateauing in the 3000 OA articles per year range. Interestingly, the snapshot for this year shows a much different picture with regard to types of OA than we have seen in past reports. Notably, there are much lower Green OA counts this year than we have seen in the data presented previously. Since these decreases appear to be across all years studied, we suspect it is due to a change in our data sources (possibly Unpaywall). For example, this could be due to a different set of Green OA repositories indexed by Unpaywall this year in comparison to previous years, or it could be due to a change in how Unpaywall categorizes articles as Green OA. In past reports, Green OA and Gold OA were typically the top two types of OA each year, but this has shifted this year due to the decreases in Green OA as well as overall increases in Hybrid OA. As a result, Hybrid OA is now the second most prevalent type of OA (behind Gold OA) in four of the most recent five years. Similar to the Green OA data for this year, this could also indicate a change in how Hybrid OA is treated in our data sources (e.g., Unpaywall adding new data sources for Hybrid OA or changing the way they categorize existing Hybrid OA articles in their database). This year's Gold OA numbers do appear to be consistent with

those in past reports, and this year's counts continue to show a steady increase in Gold OA as the most common type of OA. While Green OA numbers are lower overall as noted, we do still see the expected "repository lag" effect that we have seen in past reports where Green OA numbers tend to be lower in the most recent year or two due to some journals' embargo periods that restrict deposit of articles into Green OA repositories until a certain time after publication (e.g., one year). This is another trend we continually monitor since there are currently changes taking place around "zero embargo" public access policies at federal funding agencies that could impact OA publishing practices in future years. The most prominent example of such policies that has been enacted so far is the National Institutes of Health's public access policy, which went into effect July 1, 2025.[4] All federal funding agencies are expected to have similar policies in place by the end of 2025, so it will be important to see what effect these "zero embargo" policies have on OA publishing practices in future years.

Figure 1.

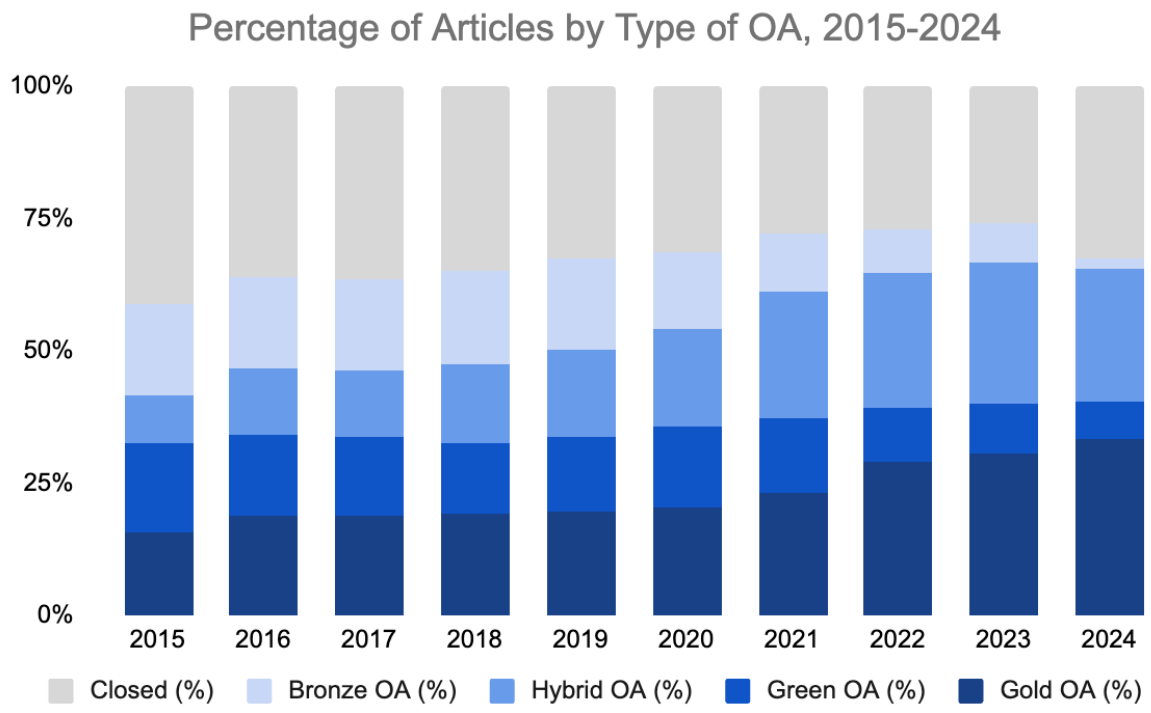


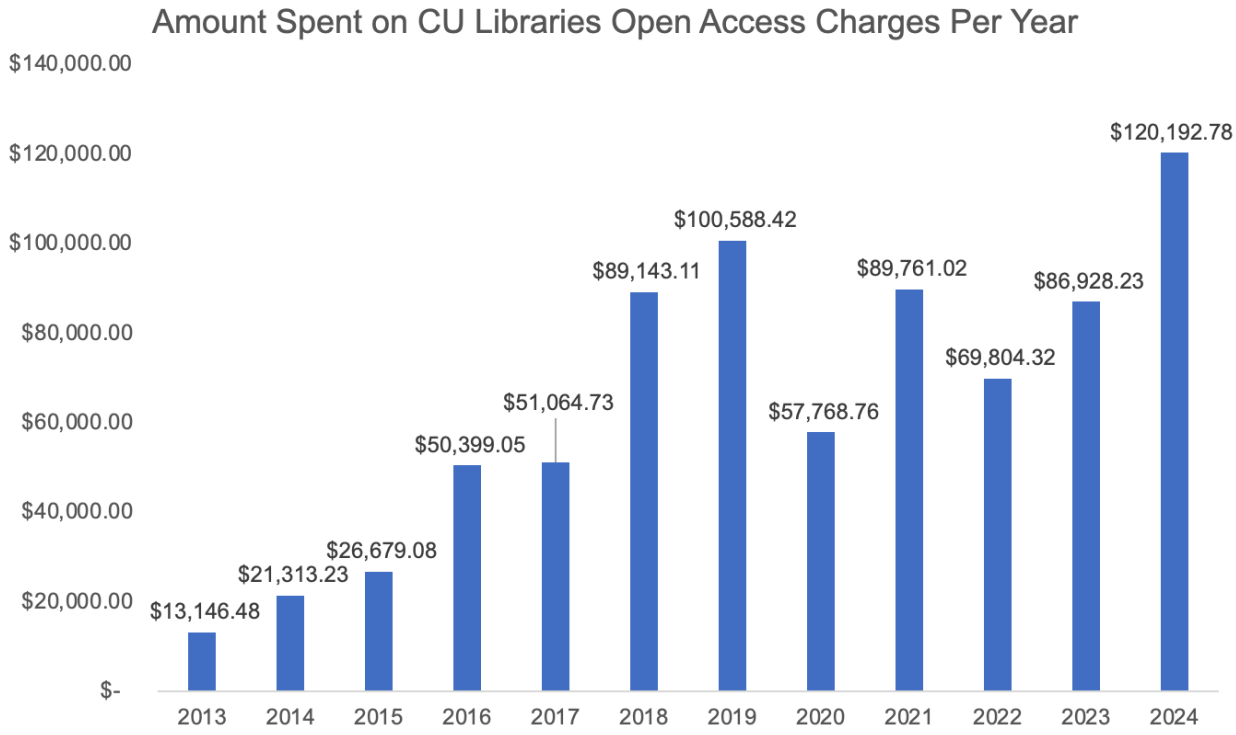
Figure 1 shows the trends in percentage of type of OA and closed articles over time. Overall, the total percentage of OA articles increased from a low of 58.85% in 2015 to a high of 74.29% in 2024. These numbers are very similar to those in last year's report, so any potential changes to our data sources regarding counts for types of OA

[4] NIH Public Access Policy: https://grants.nih.gov/policy-and-compliance/policy-topics/public-access

did not seem to impact these overall percentages. As expected given the differences in this year's snapshot described in the previous discussion of Table 2, the percentages of Gold, Green, Bronze, and Hybrid OA do look different from past reports while still providing further insights into the overall increases in OA publishing activities at CU Boulder. Over the period studied, Gold OA articles showed the strongest trend and greatest increase from a low of 15.66% of all articles published in 2015 to a high of 33.47% of articles in 2024. The percentage of Hybrid OA articles also generally increased over time leading to a high of 26.83% in 2023. This general trend toward an increasing number of Hybrid OA articles should be monitored closely since this type of article is associated with higher article processing charges (APCs) and problematic publisher business models. As noted in the previous discussion of Table 2, Green OA percentages are lower across the board in this report with a high of 16.81% in 2015 and a low of 6.85% in 2024. This most recent snapshot shows Green OA mostly hovering around 14-15% before beginning to decline in the three most recent years. As noted in the previous discussion, some of 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, but this will be important to monitor as we investigate the reasons for the large shifts in Green and Hybrid OA that we saw for the first time in this year's snapshot. According to this year's data, the percentage of Bronze OA articles appears to be in steady decline since 2019 with the percentage dropping each year and culminating in an overall low of 1.96% in 2024. This continues to suggest that Bronze OA will be a smaller part of the overall OA picture going forward.

III. CU Boulder Libraries Open Access Fund

Figure 2.

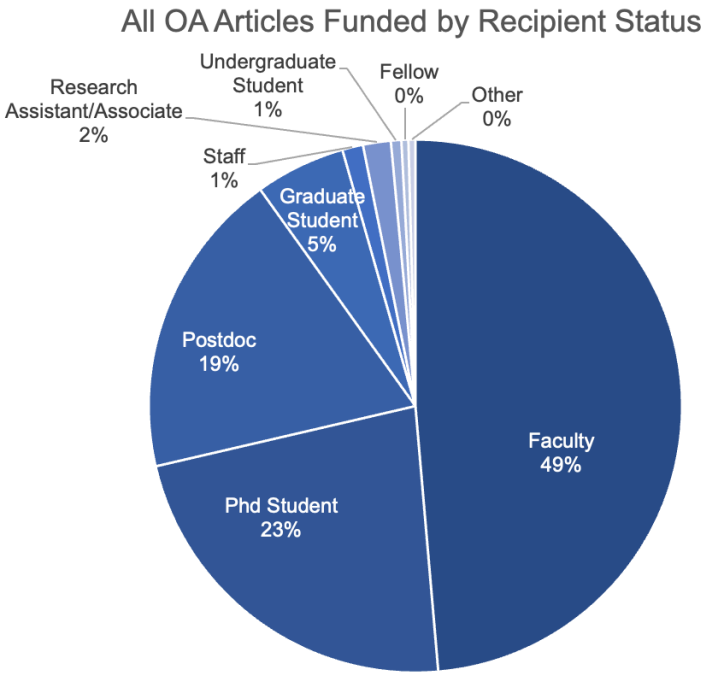


CU Boulder Libraries spent $120,193 on open access article processing charges for fully open access journal publication in 2024. In 2023, the CU Boulder Libraries Open Access Guild recommended capping the general OA Fund at $100k per fiscal year. Spending on the OA Fund by calendar year, as reflected in this report, may exceed $100k since the calendar year spans across two fiscal years.

66 articles were funded in 2024, with an average APC cost to the Libraries of $1,821.10. Individual article costs are capped at $2,000, but the total invoice often exceeds this cap. The least an author was invoiced in total was $525, and the most an author was invoiced in total was $6,790. The average total invoice was $2,514.

Cumulatively, the CU Boulder OA Fund has helped authors publish 477 fully OA articles in 207 unique journal titles.

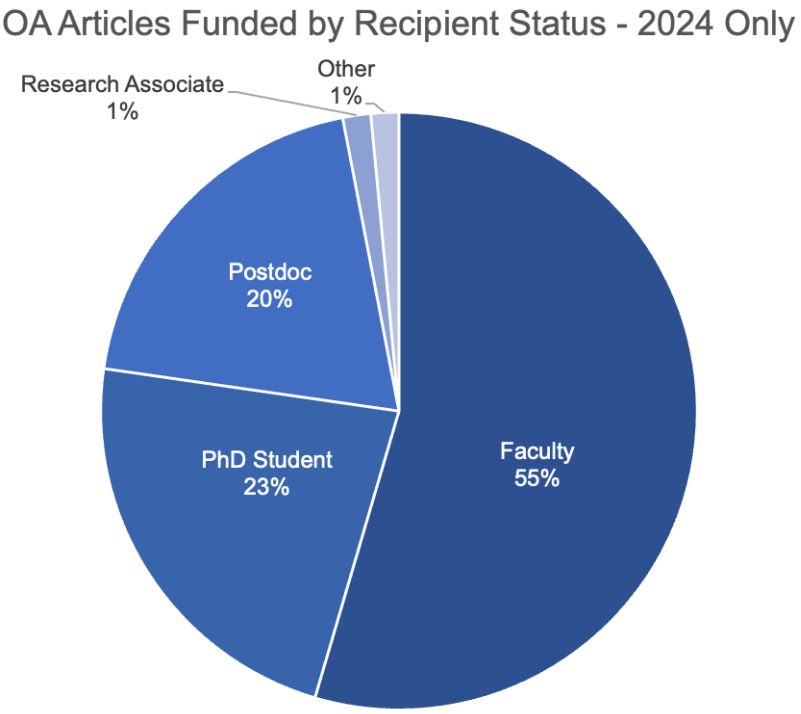
Figure 3.



As with previous reports, 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 (49%) of the total cumulative awards.

Graduate students and postdoctoral researchers represent just under half (47%) 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 2024 was awarded to faculty, followed by PhD students and postdocs. However, the share of faculty awards declined to 55% in 2024 from 68% in 2023, while PhD students and postdocs collectively saw a 15% increase in share of OA funding awards from the previous year. This rebounding of awards to PhD students and postdocs brings proportions closer to what they resembled several years ago, after dipping significantly in 2022 and 2023.

Table 3. Cumulative OA fund Awards, 2013-2024

Department/Unit	Count of OA Fund Awards (Cumulative)
EBIO	81
Physics	51
Geography	31
Mechanical Engineering	30
CIRES	25
CEAE	24
Psychology & Neuroscience	23
MCDB	23
INSTAAR	20
ECEE	19
Environmental Studies	17
Integrative Physiology	16
Geological Sciences	15
School of Education	7
Atmospheric and Oceanic Sciences	6
Institute of Behavioral Sciences	6
Chemistry and Biochemistry	5
Computer Science	5
Museum of Natural History and Geological Sciences	5
Applied Math	5

Chemical and Biological Engineering	5
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
Information Science	3
Anthropology	3
Biology	2
Aerospace Engineering	2
Political Science	2
Institute for Behavioral Genetics	2
Biofrontiers Institute	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
College of Music	1
CU Population Center	1
Business	1
Integrated Design Engineering	1
Space Weather Technology, Research and Education Center	1

The cumulative data for articles funded by department show no significant changes in 2024. 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 and 2024); Physics: 13% in 2020, 11% in 2021, 2022, 2023, and 2024) with no change in the top seven departments from the previous year. The only unit to jump significantly up the list of cumulative awards is the School of Education, which received 4 awards in 2024.

Departments with individuals receiving funding for the first time in 2024 included Integrated Design Engineering and the Space Weather Technology, Research and Education Center.

Figure 5.

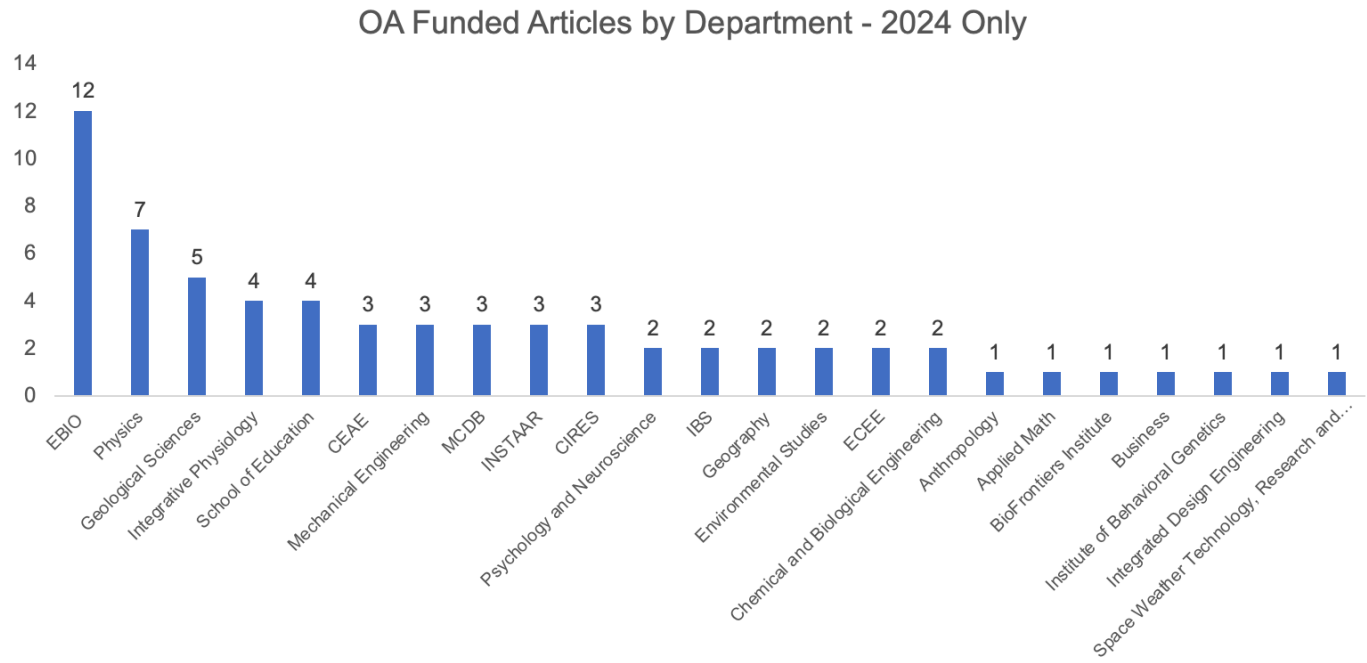


Figure 5 shows all of the representative departments for articles that received funding in 2024. Top departments receiving funding in 2024 continue to mirror the top departments cumulatively. However, a trend of declining funding for EBIO continued in 2024 (15% in 2021, 12% in 2022, 11% in 2023, and 8% in 2024). IPHY rose significantly in awards for 2024, receiving four OA fund grants compared to only one in 2023.

There were 23 unique departments that had individuals who received funding in 2024, up from 19 in 2023 and 21 in 2022.

IV. Article Processing Charges from CU Boulder Publications

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 fairly 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. For this reason, OpenAlex is currently the best source for estimating APC prices across all articles for CU Boulder.

OpenAlex was able to provide an APC price for 1,146 Gold articles and 349 Hybrid articles that were published in 2024. There were an additional 69 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 568 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.

The lowest hybrid APC price paid was $1023, and the highest was $12,512. $4,219 was the average hybrid APC price paid according to OpenAlex, which was the price used for the 568 articles which did not list a price.

Figure 6.

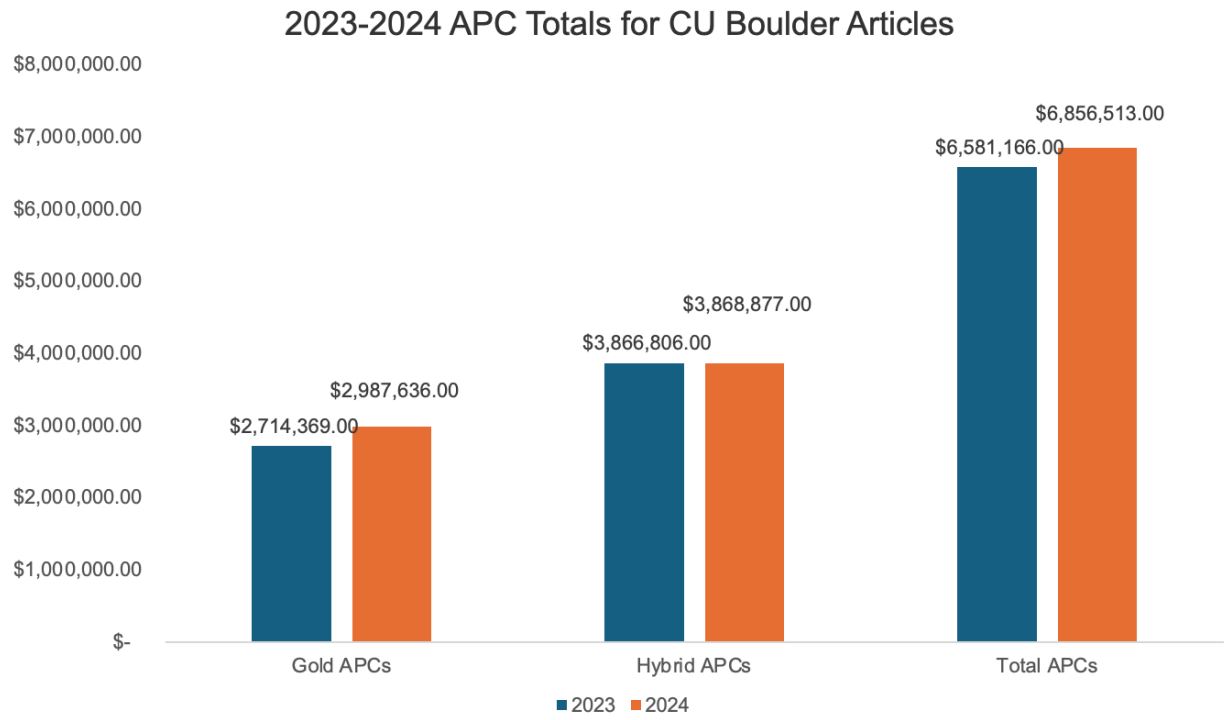


Figure 6 displays the total estimated APCs paid across Gold and Hybrid journals from 2023-2024. The estimated total APC spend in 2024 for CU Boulder open access articles is nearly $6.9 million, up from nearly $6.6 million last year. Most of this increase over the previous year is from an increase in Gold APC costs, since the estimated Hybrid cost for 2024 was nearly the same as in 2023.

We must provide the caveat that this estimate does not guarantee CU Boulder authors are responsible for all $6.9 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 institutions.

Figure 7.

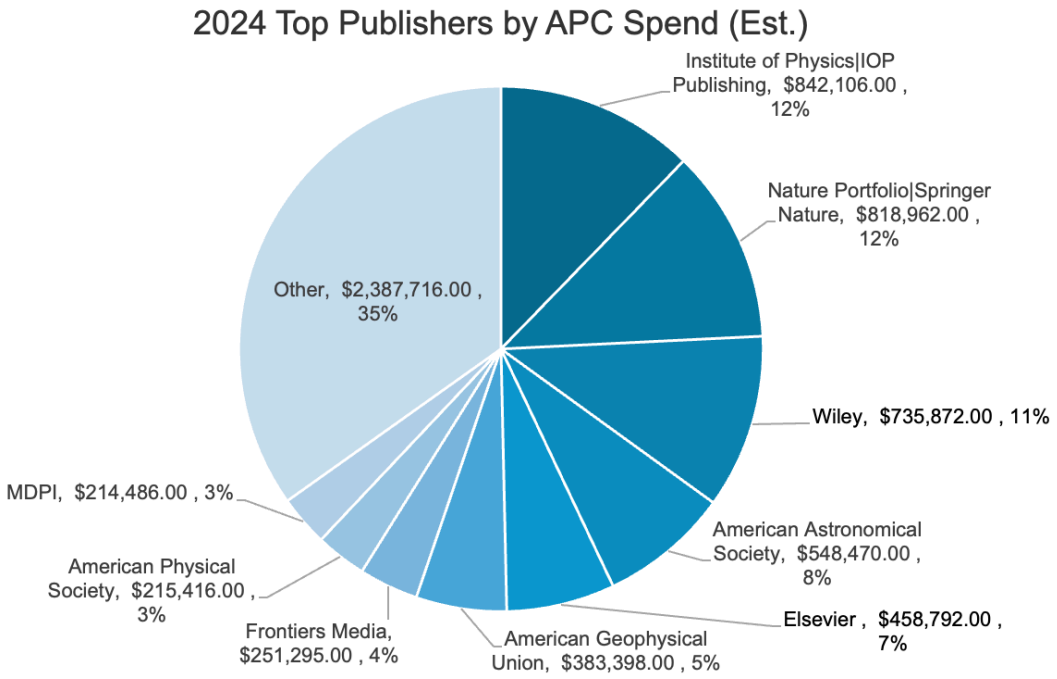


Figure 7 shows the top publishers by the estimated total amount paid to each for APCs. The top nine publishers account for 65% of all APCs, with IOP Publishing, SpringerNature, and Wiley-Blackwell alone accounting for over a third (35%) of all estimated APCs paid by CU Boulder authors. The remaining 35% of APCs ($2.3 million) went to over 100 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 9 publishers (all of which account for over $200k in APC payments), the Libraries has a transformative agreement with IOP Publishing, an agreement with Elsevier providing a 10% discount on select journals, a membership account with Frontiers to which we currently contribute around $60k, and a membership with MDPI which provides 10% discounts to CU Boulder affiliated publications.

V. CU Boulder Open Monograph Fund

The CU Boulder Open Monograph Fund began in 2021 as a pilot project under the TOME initiative, which ran from 2017-2022.[5] Participation in TOME was somewhat limited to participating institutions and university presses. The first four monographs funded by the CU Boulder Libraries were initiated under TOME.

After the sunsetting of the TOME initiative in 2023, the CU Boulder Libraries transitioned these activities to the CU Boulder Open Monograph Fund. We retained many of the same criteria for funding as were outlined under TOME. However, we chose to be somewhat more flexible, specifically with regard to the types of publishers and types of publications eligible for funding. Because of this change, we have funded two monographs with a large, commercial publisher (Taylor & Francis), and one edited volume since 2023 (all of which would not have met the initial TOME criteria).

Since the inception of the fund in 2021, the CU Boulder Libraries has funded 11 monographs across a wide range of disciplines in the social sciences and humanities. Table 4 provides details of each of these monographs.

Table 4. List of Monographs Funded by the CU Boulder Open Monograph Fund

Press	Author	Author Department	Title	Date Published
University of Illinois Press	Vilja Hulden	History	The Bosses' Union: How Employers Organized to Fight Labor before the New Deal	1/24/2023
University of Michigan Press	Srinivas Parinandi	Political Science	Following in Footsteps or Marching Alone? How Institutional Differences Influence Renewable Energy Policy	2/2023
University of California Press	Nathan Schneider	Media Studies	Governable Spaces: Democratic Design for Online Life	2/2024
Duke University Press	Brianne Cohen	Art & Art History	Don't Look Away: Art, Nonviolence, and Preventive Publics in Contemporary Europe	5/2023

[5] TOME: https://www.openmonographs.org/

Cambridge University Press	Jillian Heydt-Stevenson	English	Embodied Experience in British and French Literature, 1788-1814: Women and Belonging	12/2024
Routledge (Taylor and Francis)	Kelly A. Parkes	Music Education	The Applied Studio Model in Higher Music Education	11/2024
Routledge (Taylor and Francis)	Chris Vargo	ARPD	The Computational Content Analyst: Using Machine Learning to Classify Media Messages	12/2024
Fordham University Press	Lauren Stone	Germanic and Slavic Languages and Literatures	The Small Worlds of Childhood: Philosophy, Poetics, and Queer Temporalities of Early Life	5/2025
Edinburgh University Press	Dianne Mitchell (ed.)	English	Literary Form After Matter 1550-1700: Experiments in Close Reading	Expected 4/2026
The University of Michigan Press	Isabel Koster	Classics	Stealing from the Gods: Temple Robbery in the Roman Imagination	Expected 2/2026
Name: University of Minnesota Press	Brianne Cohen	Art & Art History	Empathic Lens: Art, Animism, and Ecology in Contemporary Southeast Asia	Expected 12/2026

VI. 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.

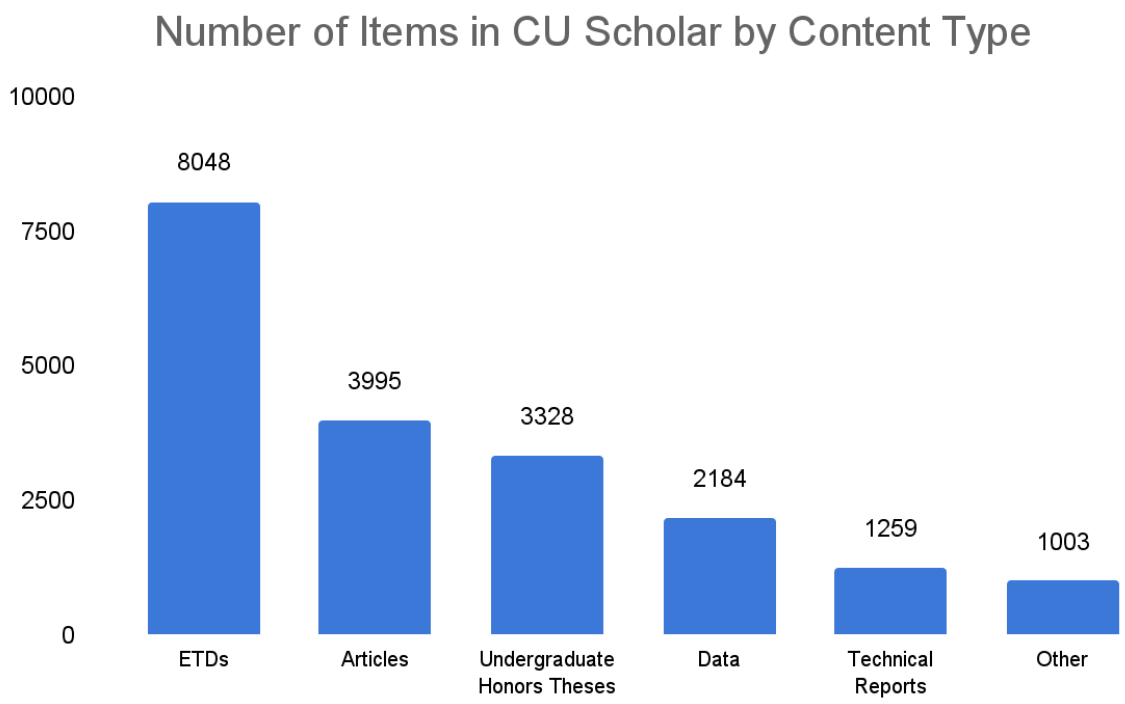


Figure 8 shows the total number of items in CU Scholar at the end of 2024 by content type. At the end of 2024, CU Scholar contained 19,817 items including journal articles, data sets, graduate theses and dissertations, undergraduate honors theses, conference materials, books, and book chapters. This represents an increase of 13.58% from the end of 2023. Content in CU Scholar was downloaded 35,923 times in 2024 according to data exported from Google Analytics. Total downloads were nearly the same as in 2023 when there were 36,926.

In 2024, graduate theses and dissertations made up 40.61% of all content in the repository (down slightly from 41.11% in 2023). Undergraduate honors theses

accounted for 16.79% of CU Scholar content (down slightly from 17.62% in 2023) while articles and proceedings accounted for 20.16%. Technical reports represented 6.35% of the repository with most items belonging to a computer science technical reports collection that is not actively growing. Data sets now account for 11.02% of CU Scholar content (up from 9.41% in 2023), representing one of the fastest growing content types in the repository. Other works consisting largely of books, working papers, presentations, and reviews account for 5.06% of the repository's content.

VII. Open Data at CU Boulder

Figure 9.

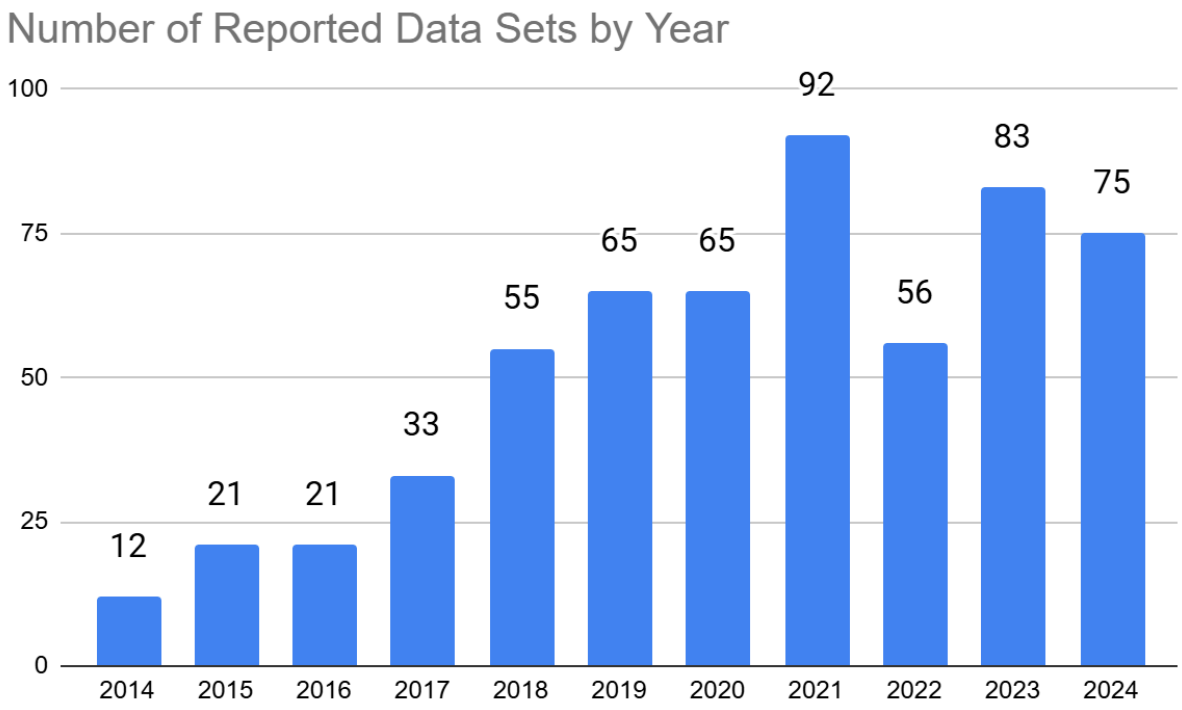


Figure 9 shows the number of published data sets reported in the annual Faculty Report of Professional Activities (FRPA) in 2024. This year the number of data sets declined slightly to 75. It is important to keep this year-over-year decrease in perspective; note, for example, that the number of datasets published in 2024 is the third highest on record, while the number published in 2023 was the second highest.

Table 5. Reported Data Sets by Department/Unit, 2014-2024 (n>1)[6]

Department/Unit	Number of Reported Data Sets (2014-2023)	Number of Reported Data Sets (2014-2024)	New Data Sets (2024)
Environmental Studies	58	64	6
Atmospheric and Oceanic Sciences	54	57	3
Civil, Environmental and Architectural Engineering	43	51	8
Geography	32	39	7
Geological Sciences	25	38	13
Ecology and Evolutionary Biology	34	36	2
INSTAAR	28	33	5
Sociology	19	27	8
Astrophysical and Planetary Sciences	18*	21	3
Computer Science	14	17	3
Linguistics	15	15	0
Chemistry	9	14	5
Libraries	10	12	2
Chemical and Biological Engineering	6	8	2
LASP	8	8	0
Business	7	7	0
Education	5	7	2
Aerospace Engineering Sciences	6	6	0
Classics	6	6	0

[6] Due to an error, last year's version of this table included separate entries for "Astrophysical and Planetary Sciences" and "Astrophysical & Planetary Sciences", as well as for "Molecular, Cellular, and Developmental Biology" and "Molecular, Cellular, & Developmental Biology". In both cases, the separate entries actually referred to the same department. The current table corrects the error by consolidating the data for each of these departments into a single entry.

History	6	6	0
Integrative Physiology	5	6	1
Environmental Design	5	5	0
Molecular, Cellular and Developmental Biology	5**	5	0
Electrical, Computer & Energy Engineering	3	4	1
Information Science	4	4	0
Physics	3	4	1
Media Studies	3	3	0
Asian Languages and Civilizations	2	2	0
English	0	2	2
Ethnic Studies	2	2	0
Journalism	2	2	0
Mechanical Engineering	1	2	1
Natural History Museum	2	2	0
Speech, Language, and Hearing Sciences	2	2	0

Table 5 provides updated information on the distribution of data sets published from 2014 through 2024 across disciplines.

Geological Sciences published the most new datasets in 2024 (13), followed by Civil, Environmental and Architectural Engineering (8). Environmental Studies, which published the most datasets in 2023 (with 18), published six new datasets in 2024; it remains the department with the most total published datasets (64).

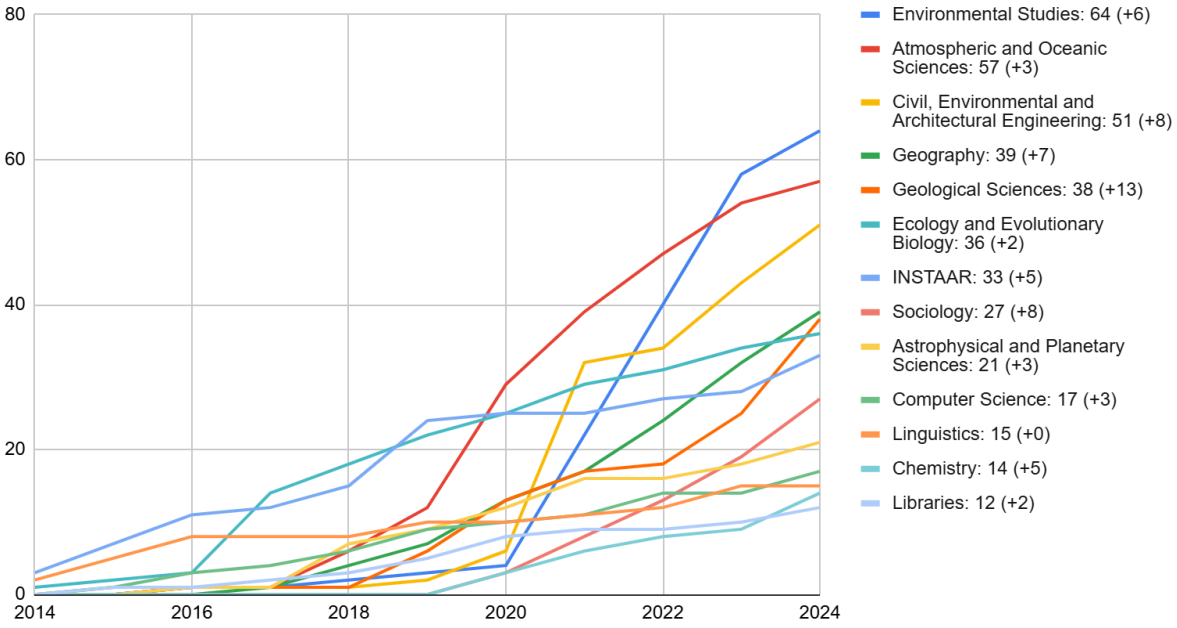
Other departments that published a significant number of new datasets in 2024 include Sociology (8), Geography (7), Chemistry (5), and the Institute of Arctic and Alpine Research (INSTAAR) (5). Many of the departments that published a large number of datasets in 2024 have an interdisciplinary orientation, which could point to a growing role for interdisciplinary data products (especially those focused on the environment) in CU Boulder's research output (though it is too early to draw firm conclusions).

Of the 19 departments that deposited data sets in 2024, the only new department (which had not previously deposited data) was English, which published two datasets.

In previous reports, we noted a robust increase in the number of datasets published by Sociology since 2020; over a five year span, the cumulative number of datasets published by that department has grown from 3 (in 2020) to 27 at the end of 2024, which perhaps reflects an increased awareness of reproducible and open research practices in the data-intensive social sciences. It will be interesting to observe whether English follows a similar growth trajectory in the next few years, as data-intensive fields of study in the digital humanities further mature.

Compared to 2023, when there were a total of 32 departments that had published more than one dataset[7], there are now 34. The new entrants to the set of departments with more than one data publication are English (which, as we noted above, published two data sets in 2024) and Mechanical Engineering (which had published one data set as of 2023, and published another one in 2024). Towards the top of the distribution, there are now 13 departments that have published 10 or more data sets (see Figure 10). Of particular note is tChemistry, which is the only new entrant to the set of departments with at least 10 cumulative data publications.

Figure 10. Growth of published data sets by department/unit



[7] This is after correcting for the error described in Footnote 6.

Figure 11.

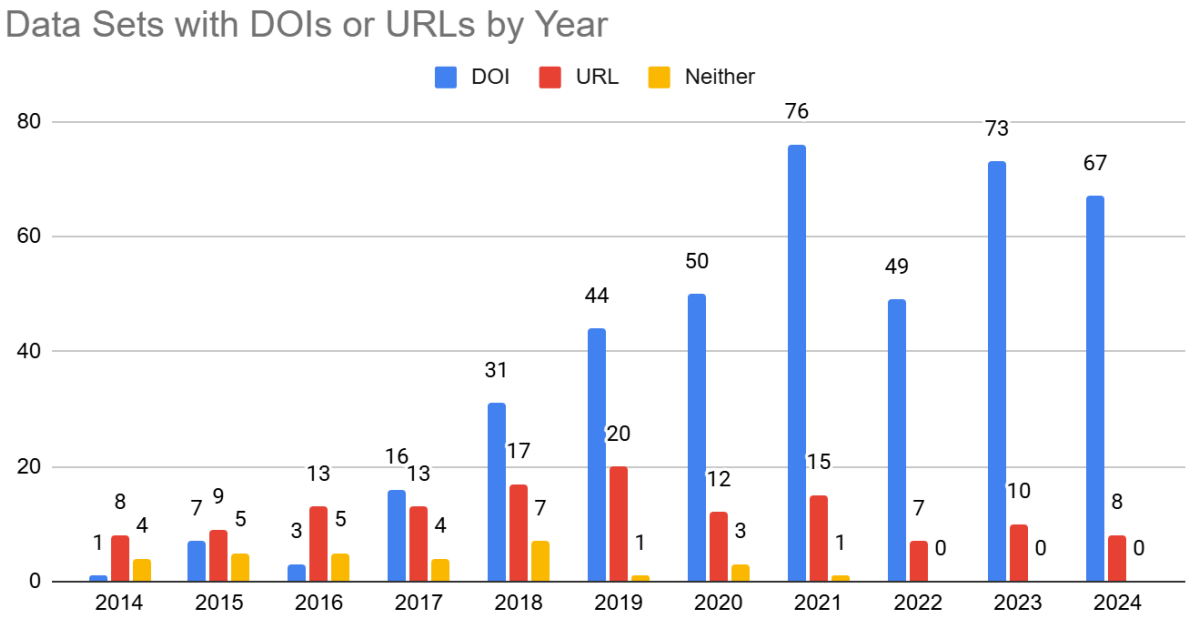


Figure 12.

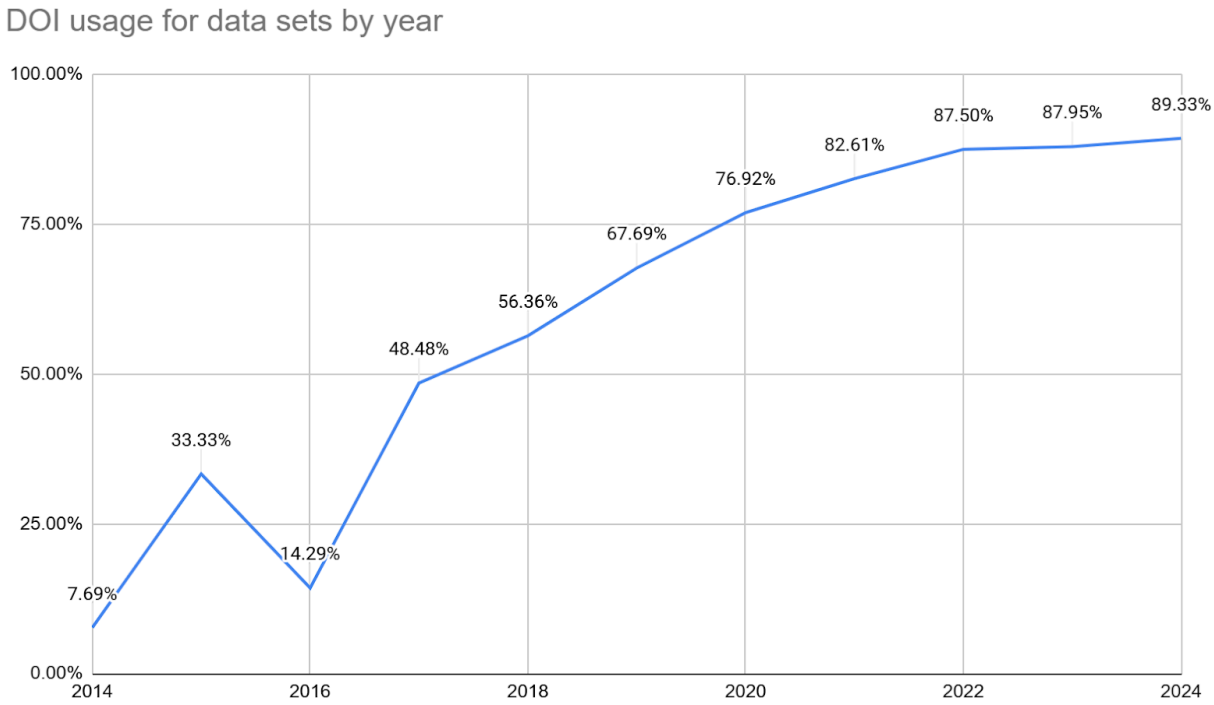


Figure 11 shows the use of DOIs or URLs over time, while Figure 12 visualizes the percentage of data sets with a DOI over time. The use of DOIs for reported data sets stayed relatively steady as 67 out of 75 (89.33%) data sets included DOIs, compared to 73 out of 83 datasets (87.95%) last year. The trend 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; the robustness of this trend is indicated not only by the growing use of DOIs, but also by the fact that 2024 is the third year in a row in which no data sets were published without a corresponding DOI or URL.

Figure 13.

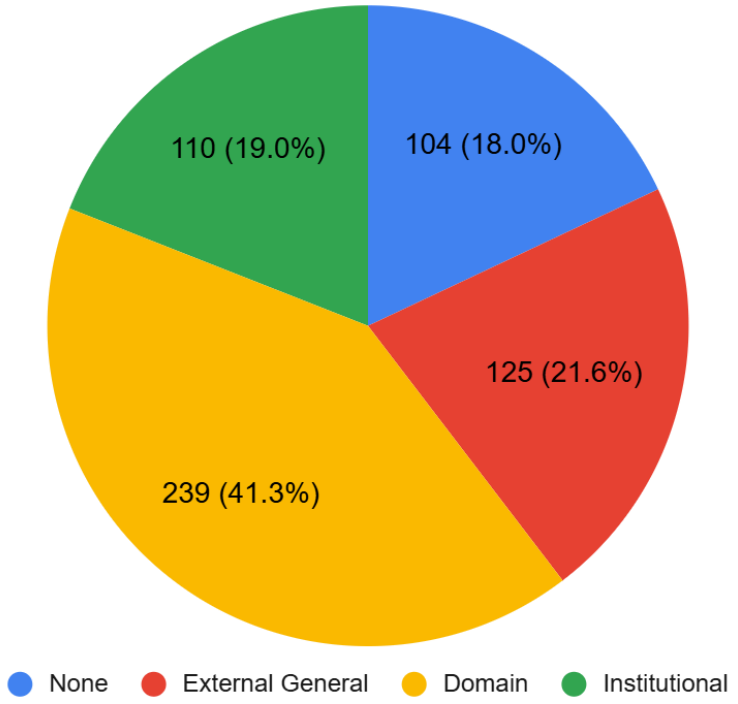
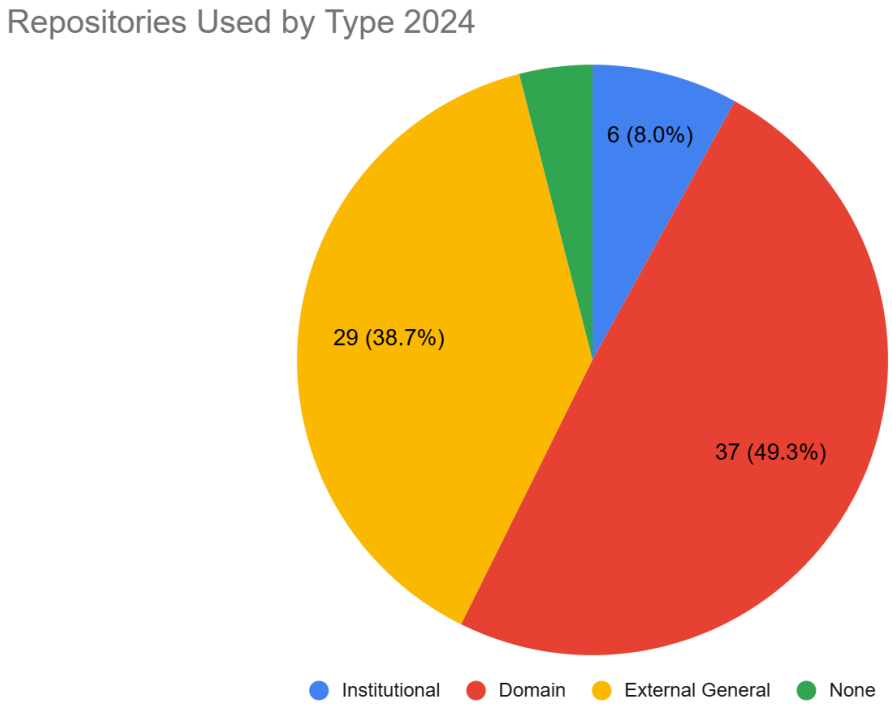


Figure 13 presents an updated summary of repository types, with 2024 data included. The cumulative distribution of data deposits across repository types at the end of 2024 is comparable to the distribution at the end of last year. At the end of 2024, 41.3% of all data sets had been deposited in domain repositories (compared to 40.1% at the end of 2023), 21.6% had been deposited in external general repositories (compared to 19% at the end of 2023), 19% had been deposited in institutional repositories (compared to 20.6% at the end of 2023), and 18% had not been deposited in a formal repository (compared to 20% at the end of 2023).

Figure 14.



Despite the year-over-year stability in the overall distribution of repository destinations (which is expected, given the accumulated weight of data from previous years), it is worth noting that on a more granular level, the distribution of published data sets across repository types in 2024 diverged significantly from the corresponding distribution in 2023. Last year, data sets were allocated fairly evenly between domain-specific (33.7%), external general (32.5%), and institutional repositories (28.9%). In contrast, data sets in 2024 (see Figure 14) were concentrated in domain (49.3%) and external general (38.7%) repositories; only 8% of data sets were deposited in institutional repositories, which represents a substantial decline in the relative use of institutional repositories compared to last year. However, the percentage of data sets in 2024 that were not published in a formal repository of any kind (4%) was comparable to last year (4.8%).

Figure 15.

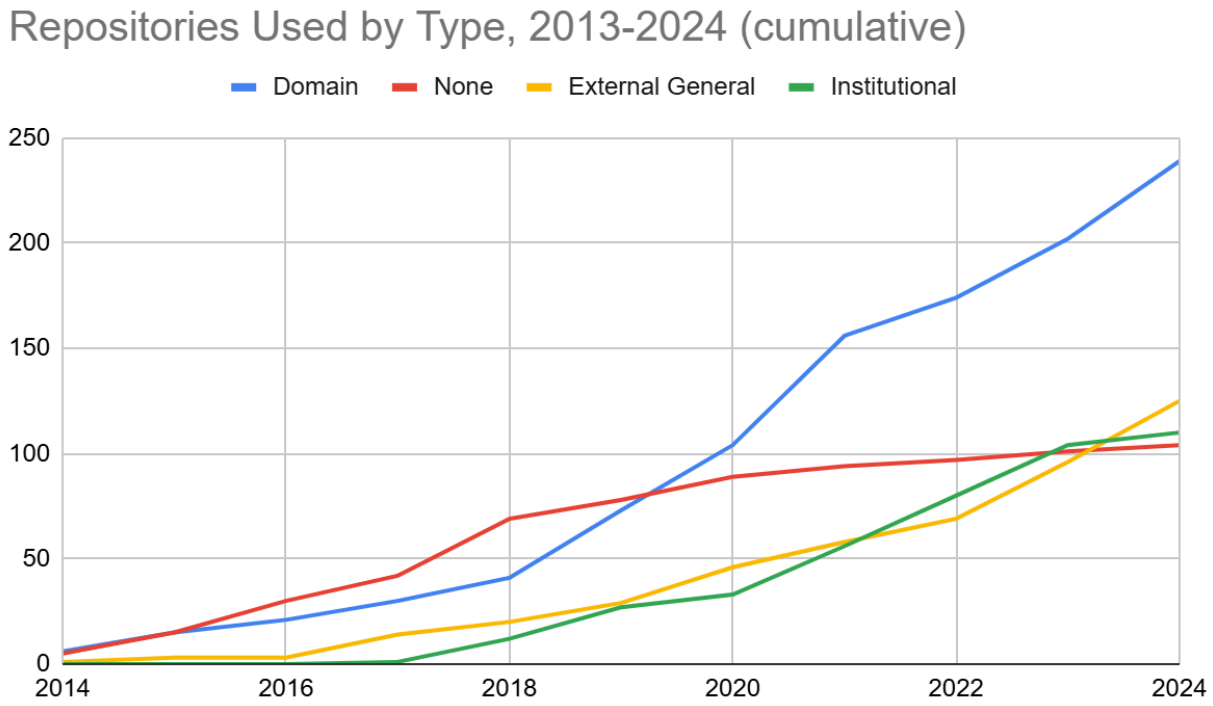


Figure 15 shows the cumulative growth of data sets across institutional, domain, and external general repositories from 2014 onwards. Last year, domain repositories were the most widely used repositories (202) and that continues to be the case this year (239). Given the slowing growth in the use of institutional repositories over the past year, cumulative deposits in external general repositories (125) now exceed those in institutional repositories (110). This represents a reversal compared to the end of 2023, when the number of data publications in institutional repositories (104) still exceeded those in external general repositories (96). It remains to be seen whether the slowing relative growth in the use of institutional repositories over the past year (compared to domain and external general repositories) is the start of a trend.

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 2024, the Libraries and its partners registered 544 DOIs for published data sets (up significantly from 309 in 2023). This is the largest single year total since the DOI registration service began (previous high was 420 DOIs in 2021). The disconnect between this much larger number of DOIs and the 75 published data sets reported by faculty in the preceding discussion in this section of the report 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 approximately 76% from 2023 to 2024 is notable as it is a major reversal of two years of modest decreases that might have indicated the start of a downward trend or plateau for the number of data sets published in CU Scholar each year. It is unclear what is driving the significant increase in DOI registrations from 2023 to 2024, but future years could help explain if this is the start of a new trend or simply an outlier. As of the end of 2024, the Libraries and its partners had registered a total of 2,243 DOIs for data sets since the DataCite DOI registration service began at CU Boulder.