2021 State of Open at the University of Colorado Boulder

An Update on Open Access Practices Based on Data from 2020

September 9, 2021

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

Ryan Caillet, *Institutional Repository Program Manager*
Melissa H. Cantrell, *Scholarly Communication Librarian*
Andrew Johnson, *Lead, Data and Scholarly Communication Services Section*
Aditya Ranganath, *Data Librarian*
Jordan Wrigley, *Data Librarian*

University Libraries
UNIVERSITY OF COLORADO **BOULDER**

TABLE OF CONTENTS

I. Executive Summary

Using data from 2020, this report is the third 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). Due to the availability of new data sources, a repository software migration, and effects of the COVID-19 pandemic, many of the numbers in this report look quite different from past years. Explanations of these differences are provided in each respective section of the report. Data used to produce this report can be found here: https://doi.org/10.25810/x8nq-tb51

Key findings from this report include:

- 60.13% of articles published in 2020 by CU Boulder authors are available via some type of OA (Gold, Green, Hybrid, or Bronze);
- In 2020, the CU Boulder Libraries OA Fund funded author fees totaling $57,768.76 for 34 journal articles published by CU Boulder authors in full OA journals (reasons for the decline from $100,588.42 for 65 articles in 2019 are explained in detail in the report);
- At the end of 2020, there were 11,810 OA items in the CU Scholar institutional repository, and these items were downloaded a total of 43,236 times in 2020;
- In the annual Faculty Report of Professional Activities (FRPA), faculty reported 65 published data sets in 2020 (same as 2019) with 50 of these citations including Digital Object Identifiers (DOIs) (up from 44 in 2019) and 54 citations identifying a formal data repository (down slightly from 56 in 2019);
- The Libraries and its partners registered 320 DataCite DOIs for published data sets in 2020 (up from 197 in 2019).

II. Open Access Articles by CU Boulder Faculty

In a significant change from past years, this 2021 report leveraged 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 broader 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 previous State of Open reports, which only included articles published in full OA journals that were indexed in the Directory of Open Access Journals (DOAJ).[3] As a result, we will no longer report the data solely from DOAJ in future reports, and we will instead update this more comprehensive Unpaywall/CUBE data going forward. Because this is the first year using this approach, we also provided data for each year dating back to 2012 to align with the start date of our initial baseline State of Open Report.[4]

[1] Unpaywall: https://unpaywall.org/
[2] CU Boulder Elements: https://www.colorado.edu/fis/CUBE
[3] Directory of Open Access Journals: https://doaj.org/
[4] "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

Table 1. Types of Open Access Content

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

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

Table 2. Articles Published by CU Boulder Faculty by Open Access Type, 2012-2020

Year	Closed (n)	Gold (n)	Green (n)	Hybrid (n)	Bronze (n)	Total OA (n)	Total (n)
2020	1609	873	764	509	281	2427	4036
2019	1303	816	786	661	349	2612	3915
2018	1218	820	765	705	451	2741	3959
2017	1234	727	775	631	416	2549	3783
2016	1251	750	637	594	612	2593	3844
2015	1391	563	716	346	531	2156	3547
2014	1452	440	844	233	429	1946	3398
2013	1474	306	762	291	514	1873	3347
2012	1450	283	663	300	471	1717	3167

Table 2 provides the total number of articles published by CU Boulder faculty each year from 2012 to 2020 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 totals are given for each of these categories over time as well. These totals reveal overall shifts in OA article publishing practices at CU Boulder from 2012 to 2020. 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, there are recent signs that this trend might be plateauing around roughly 2500 total OA articles per year.

Figure 1.

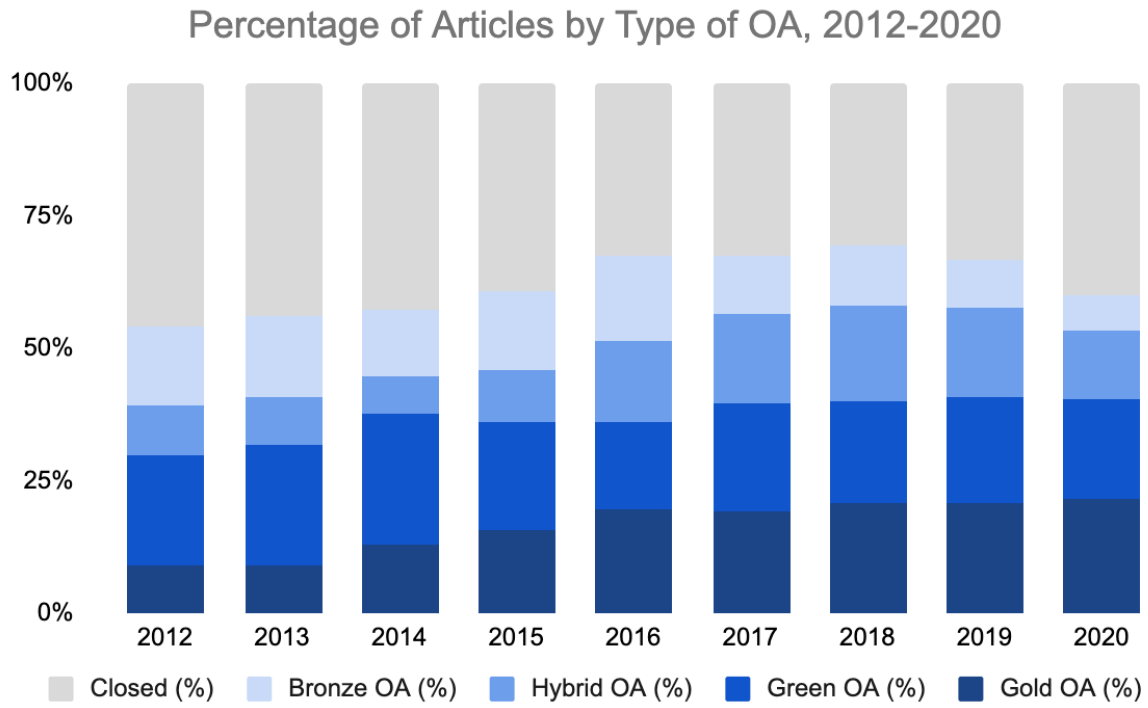
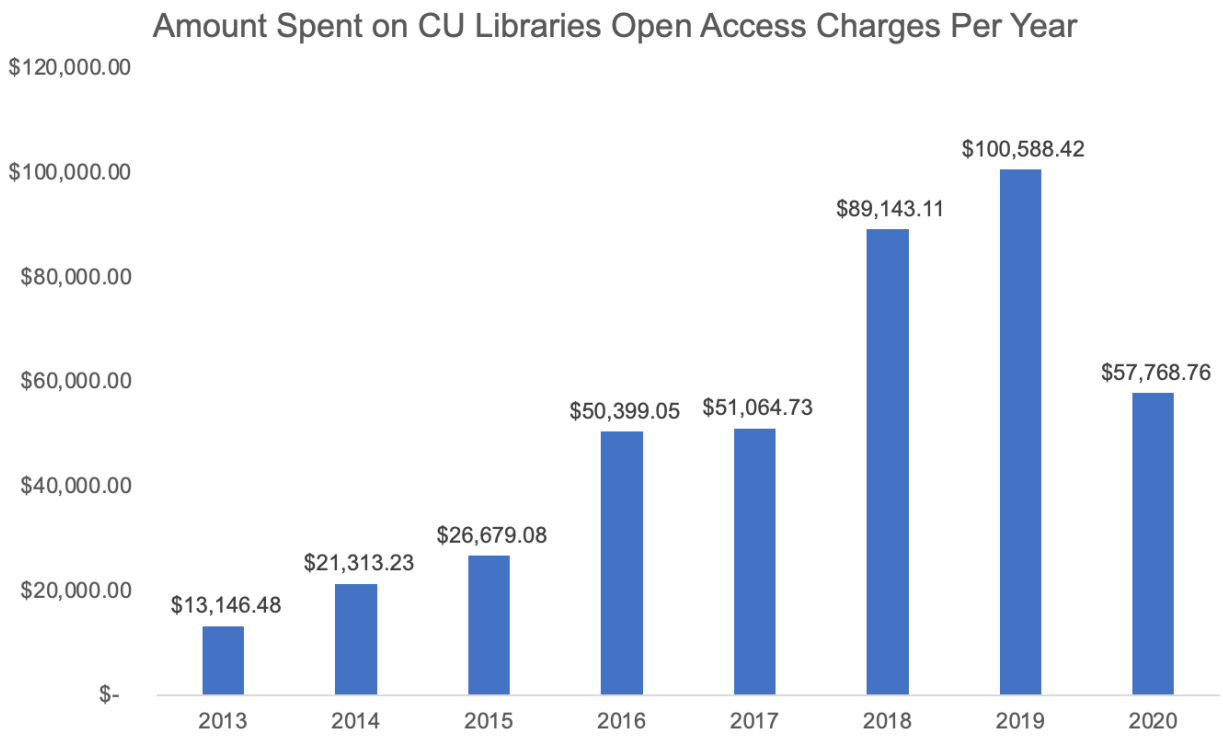


Figure 1 shows the trends in percentage of type of OA and closed articles over time. The total percentage of OA articles consistently increased from a low of 54.21% in 2012 to a high of 69.23% in 2018. There was a slight decrease to 66.72% in 2019, and a further decrease to 60.13% in 2020. In future years, it will be interesting to observe if this is the beginning of a trend, an anomaly, or evidence of some other phenomenon. 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 8.93% of all articles published in 2012 to a high of 21.63% of articles published in 2020. Green and Hybrid articles experienced weaker trends. Green fluctuated between 20.93% and 16.57% while Hybrid ranged from a low of 8.69% to a high of 17.80% during the years evaluated. Bronze OA ranged from a high of 15.92% in 2016 to a low of 6.96% in 2020.

III. CU Boulder Libraries Open Access Fund

Figure 2.[5]



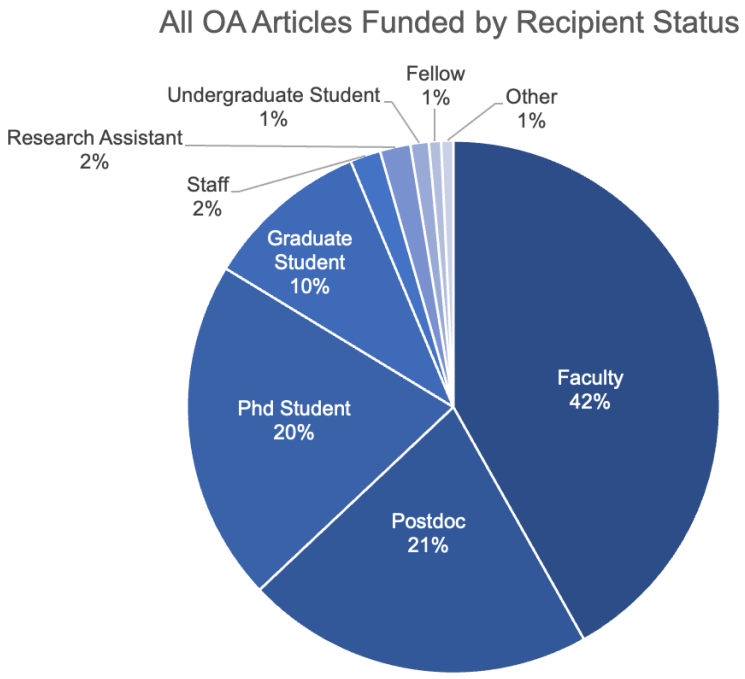
CU Boulder Libraries spending on article processing charges (APCs) in fully OA journals for faculty, staff, and students declined by almost half in 2020, for a total spend of $57,768.76. The reasons for this apparent reduction in funding are two-fold. First, the final payment of the fiscal year occurred in early April 2020.[6] Second, due to budget uncertainties created by the COVID-19 pandemic, payments in fall 2020 did not resume until late September 2020. While there appears to have been a dramatic reduction in funding, there was in fact only an overall 5% reduction in total funding available for this service from one fiscal year to the next, and discrepancies in funding by calendar year are primarily due to an approximate five-month gap in payments.

Thirty-four articles were funded in 2020, for an average APC cost of $1,699.08, which represents a nearly 10% increase in average APC cost from the previous year.

[5] There were slight errors in the calculation of the amount spent on OA charges for 2017 and 2018 in previous years' reports, which has been corrected in this report.
[6] Similar to previous years, several additional requests for funding were made in Spring 2020, but could not be completed due to depletion of funds.

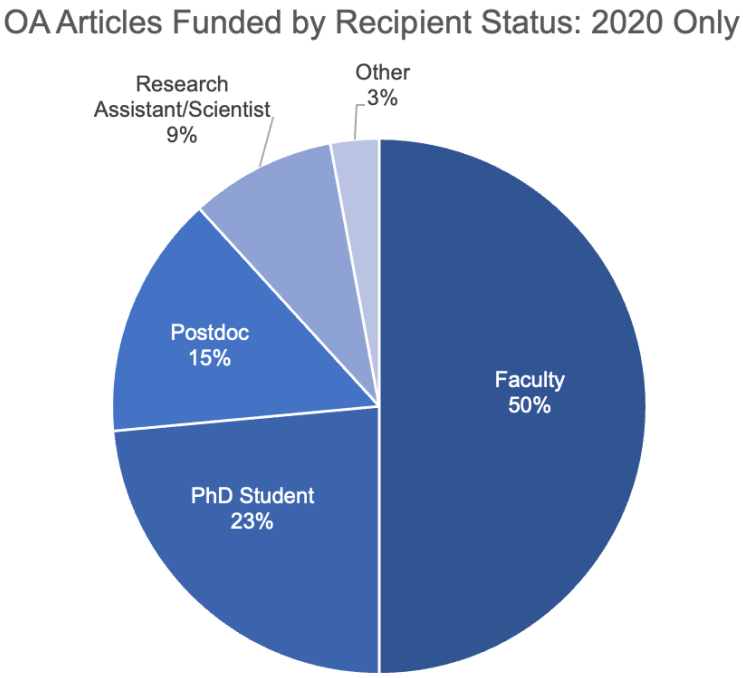
Cumulatively, the CU Boulder OA Fund has helped authors publish 270 fully OA articles in 111 unique journal titles. Of the 34 journal articles funded in 2020, seven of them were published in journal titles not previously funded through the CU Boulder OA Fund.

Figure 3.



The proportion of funded authors by university status over the lifetime of the OA Fund remains largely unchanged, as faculty still represent just over two-fifths of funded articles (42%) and graduate students and postdoctoral researchers represented a little over half (51%) of the articles funded. Research Assistants, Staff, Fellows, Undergraduate Students and Other affiliations collectively represent 7% of the total articles funded since the inception of the OA Fund in 2013.

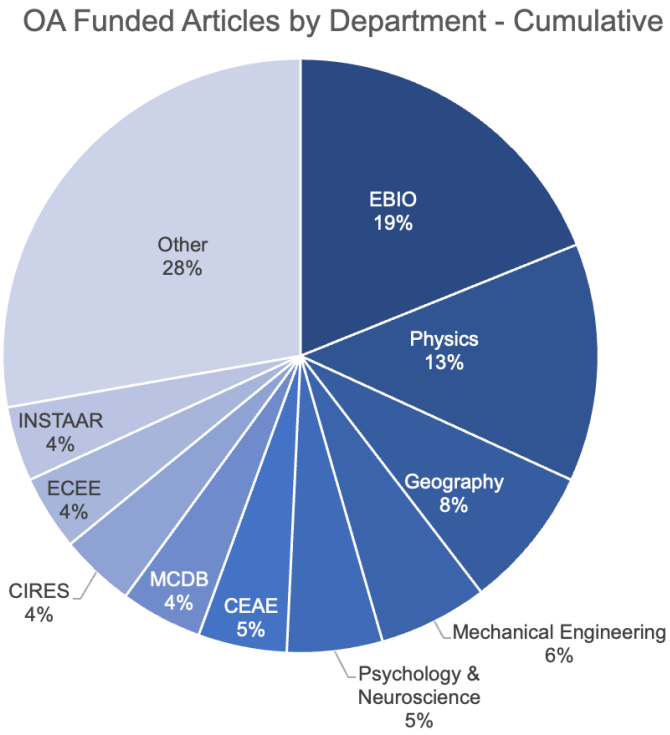
Figure 4.



The proportions of recipients of funding by status in 2020 show some notable differences from previous years. While the fewest number of faculty since 2018 received funding (17 in 2020 compared to 25 in 2019 and 22 in 2018), they nevertheless represented the highest share (50%) of total funding recipients since 2014. PhD students dropped from 29% of funding recipients in 2019 to 23% in 2020, and the proportion of Postdocs receiving funding fell even more significantly from 26% in 2019 to 15% in 2020. A trend of decreasing numbers of Graduate students applying for and receiving funding, noted in previous reports, continued in 2020. No recipients identified their status as “Graduate student” for this year. It is possible that a contributing factor to these changes in 2020 funding trends might be the COVID-19 pandemic’s well-known impact on research productivity and the outsized effects on early career researchers.[7]

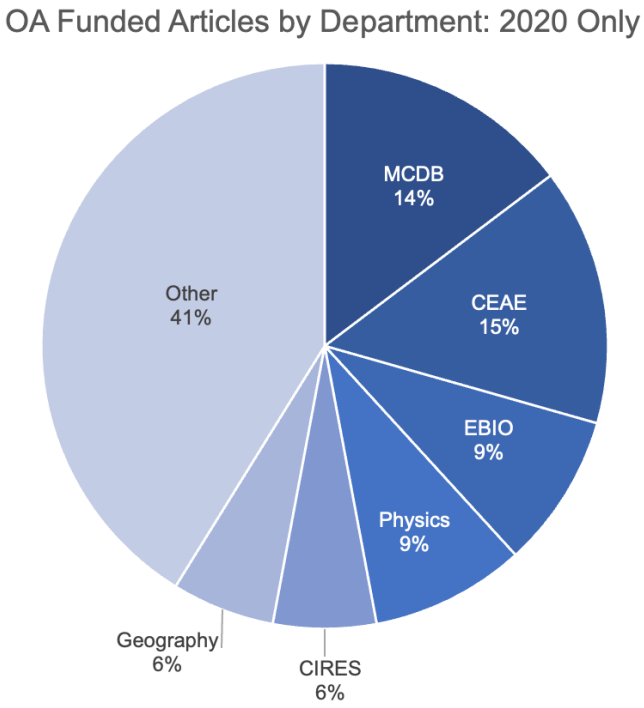
[7] Radecki, J., & Schonfeld, R. C. (2020, October 26). The Impacts of COVID-19 on the Research Enterprise: A Landscape Review. https://doi.org/10.18665/sr.314247

Figure 5.



The cumulative data for articles funded by department show no significant changes. EBIO and Physics represent a marginally smaller proportion of funded articles from past years (EBIO: 24% in 2018, 21% in 2019, 19% in 2020; Physics: 14% in 2018 and 2019, 13% in 2020), following previously noted trends. Molecular, Cellular, and Developmental Biology (MCDB) is the only new addition in the top ten funded departments or institutes, joining ECEE in edging out Environmental Studies and Integrated Physiology. The proportion of “Other” departments having articles funded continued to rise, representing 28% of funded articles in 2020. Two departments or programs had individuals receiving funding for the first time in 2020: The Museum of Natural History and the Technology, Cybersecurity, and Policy program (part of the Computer Science department).

Figure 6.

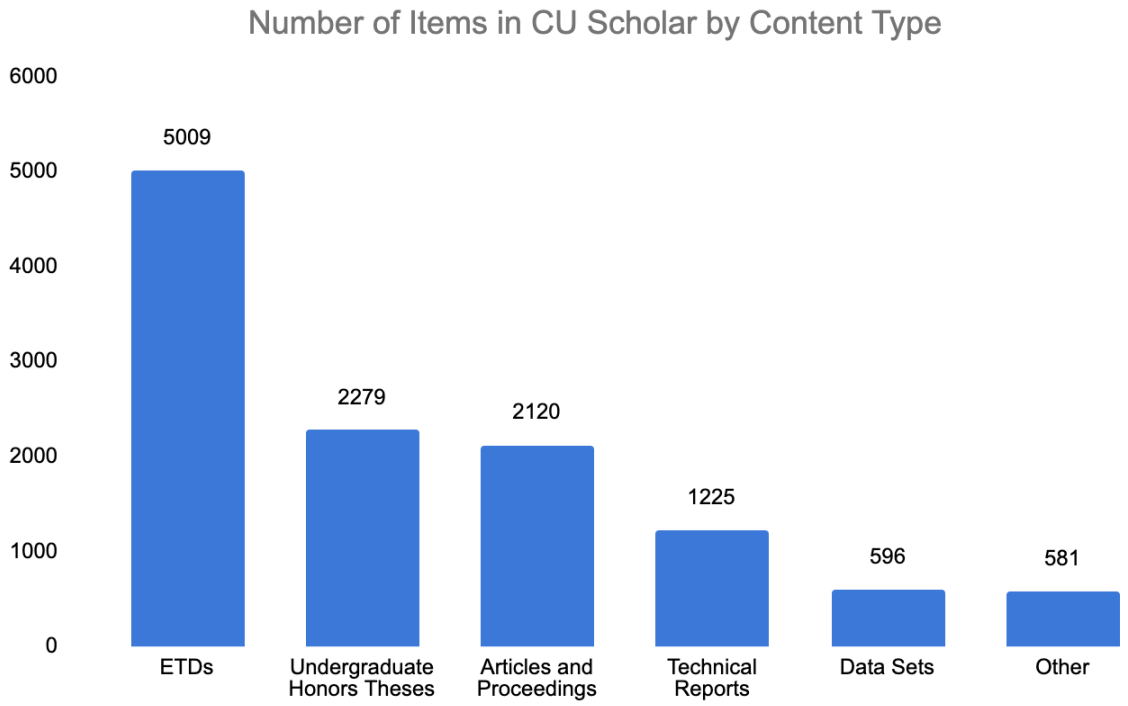


When isolating the 2020 articles funded by the department of the recipient, there are a few notable changes from 2019. While MCDB represented only 5% of funded articles in 2019, in 2020 this department represented 14% of the total articles funded. CEAE's share of funded articles similarly rose from 6% in 2019 to 15% in 2020. Conversely, the proportion of articles funded by authors in the Physics department fell from 17% in 2019 to 9% in 2020. After CIRES and Geography, every other department represented in the 2020 data was only funded for a single article, explaining the ballooning of the "Other" department category from 29% in 2019 to 41% in 2020. While this difference may be accounted for by the fewer total articles funded, and thus fewer departments represented in 2020 (i.e. only six departments were funded more than once in 2020), it may also suggest further diversification of authors across the university utilizing the Open Access Fund.

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. This edition of the State of Open Report Update is the first year content in CU Scholar will be reported on the calendar year, which we will continue to do going forward. 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 in past years. As such, we are not confident in the comparability of these numbers, and we will begin reporting the new numbers from the Samvera platform this year with the intent to provide comparable updates to these numbers each year going forward.

Figure 7.

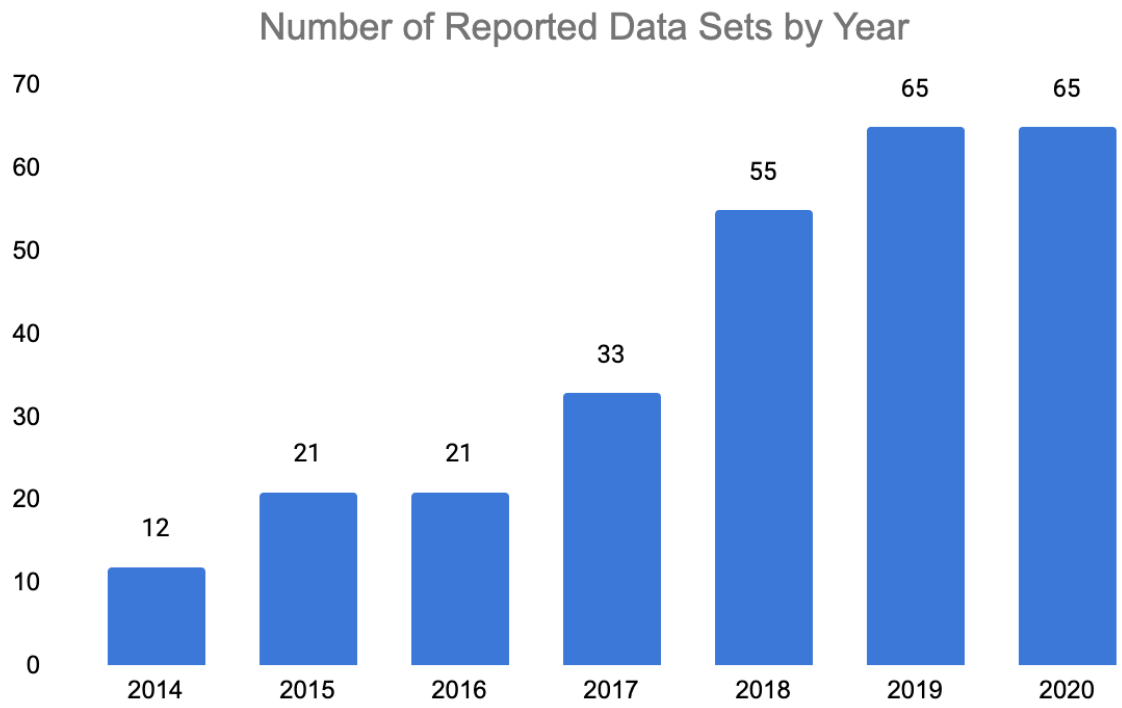


At the end of 2020, CU Scholar contained 11,810 items including journal articles, data sets, graduate theses and dissertations, undergraduate honors theses, conference materials, books, and book chapters. Content on CU Scholar was downloaded 43,236 times according to data exported from Google Analytics. Graduate theses and dissertations made up 42.41% of all content in the repository. Undergraduate honors theses accounted for 19.30% while articles and proceedings comprised 17.95% of repository items. Technical reports represented 10.37% of the repository with most

reports belonging to a computer science technical reports collection that is not actively growing. Data sets accounted for 5.05% of the repository contents.

V. Open Data at CU Boulder

Figure 8.



The number of published data sets reported on the annual Faculty Report of Professional Activities (FRPA) remained steady from 2019 to 2020, with 65 data sets published in each of the past two years. The absence of a pandemic-related decline in the number of published datasets in 2020 is perhaps surprising, but once again points to a growing acceptance of data sets as research outputs worth reporting for the purposes of annual faculty evaluations. As might be expected, the departments/units with the largest number of reported data sets (n>10) are primarily in disciplines where funding largely comes from funders with data sharing requirements (e.g., the National Science Foundation) and where journal publisher policies for data availability are increasingly common (e.g., American Geophysical Union).

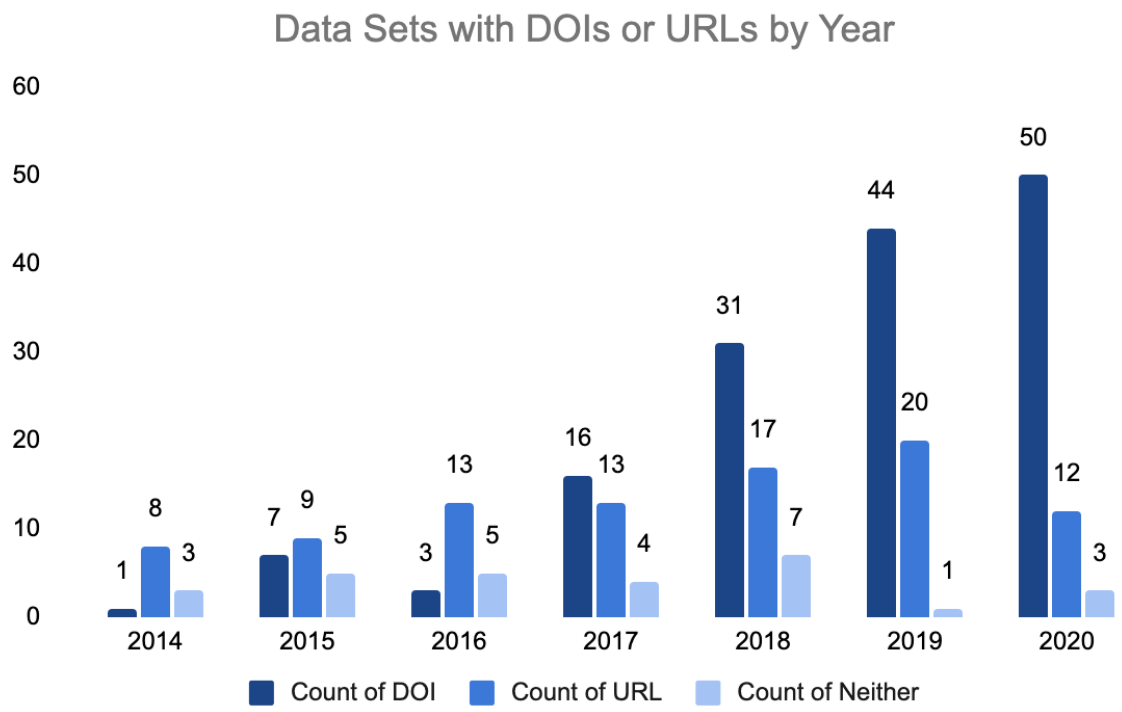
Table 3. Reported Data Sets by Department/Unit, 2014-2020 (n>1)

Department/Unit	Number of Reported Datasets
Cooperative Institute for Research in Environmental Sciences	53
Atmospheric and Oceanic Sciences	29
Ecology and Evolutionary Biology	25
Institute for Arctic and Alpine Research	25
Geography	13
Geological Sciences	13
Astrophysical and Planetary Sciences	12
Computer Science	10
Linguistics	10
Libraries	8
Laboratory for Atmospheric and Space Physics	7
Business	6
Chemical and Biological Engineering	6
Classics	5
Civil, Environmental and Architectural Engineering	6
Environmental Studies	4
Chemistry	3
Education	3
Environmental Design	3
History	3
Information Science	3
Media Studies	3
Physics	3
Sociology	3

Ethnic Studies	2
Journalism	2
Natural History Museum	2
Speech, Language, and Hearing Sciences	2
Aerospace Engineering Sciences	2
Asian Languages and Civilizations	2

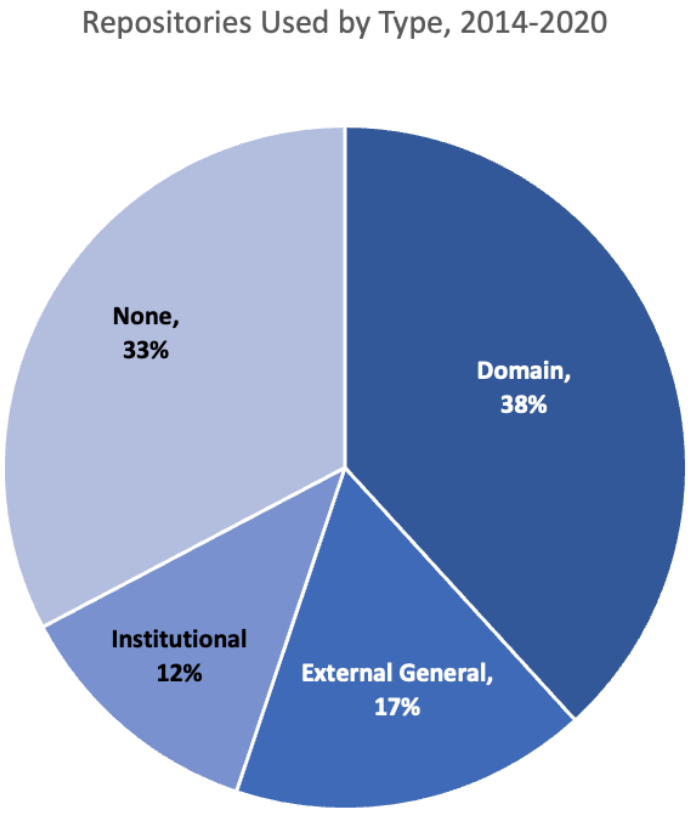
Several social science departments reported published data sets for the first time in 2020: Sociology, Political Science, and Ethnic Studies. In addition to these departments, 2020 also saw first-time data set publications from the Chemistry and Integrative Physiology departments. It is interesting to note that among departments that published data sets for the first time in 2020, several published more than one; in particular, both Chemistry and Sociology published three data sets, while Ethnic Studies published two. Whether this is a coincidence, or part of a larger trend in these disciplines (perhaps driven by changing mandates or norms) is a question worth exploring in future research. Table 3 provides the total number of published data sets reported in the FRPA for all departments with more than one data set since 2014.

Figure 9.



The use of DOIs for reported data sets increased from 44 in 2019 to 50 in 2020. The percentage of data sets with DOIs also increased to a new annual high of 76.9% (compared to the previous high, in 2019, of 67.6%). These results are consistent with the trend we noted in our previous report, namely, that data set citations (including a means for accessing data) are becoming a more common and consistent practice for faculty.

Figure 10.



In our previous report, we included data on repository types for the first time. Figure 10 presents an updated summary of repository types with data from 2020 included. Last year, we observed that a plurality (38%) of the reported published data sets dating back to 2014 did not use a formal repository. This year, with the inclusion of data from 2020, this is no longer the case, with the percentage of data sets that do not use a formal repository falling to 33%. A plurality of published data sets (38%) now 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 17% of the reported published data sets in the FRPA since 2014. Institutional repositories account for 12% of the reported published data sets. Four of the data sets published in institutional repositories appear in the institutional repositories of other institutions (two at Stanford, one at UCSD, and one at William and Mary) with the remainder appearing in CU Scholar.

Figure 11.

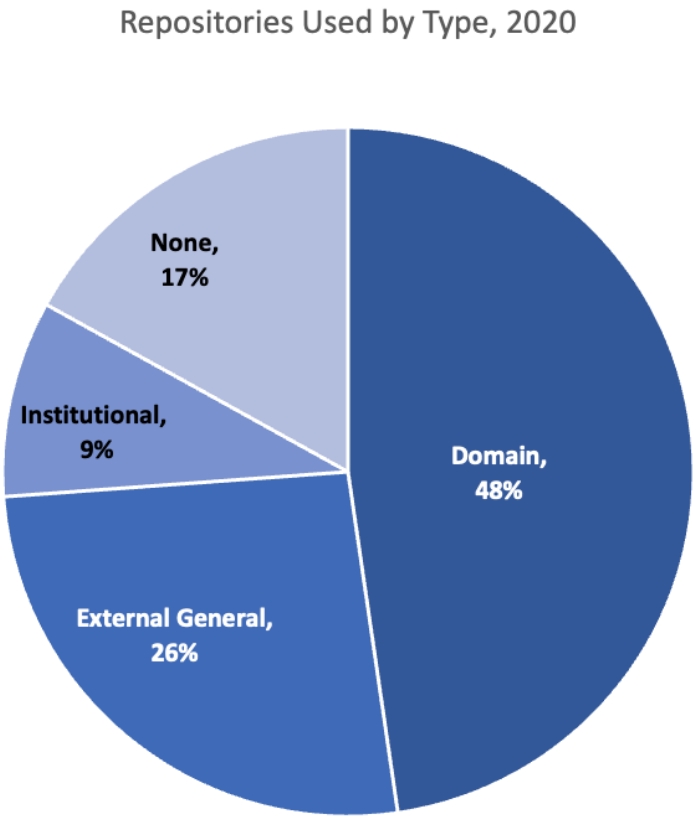


Figure 11 presents a breakdown of repository destinations exclusively for data sets published in 2020. As in 2019, data sets in 2020 were overwhelmingly published in formal repositories (with a total of 83% appearing in a formal repository of some kind). Nearly half (48%) of 2020's published data sets were published in domain repositories. External general repositories and institutional repositories accounted for 26% and 9%, respectively. In 2019, a virtually identical share of data sets were published in domain repositories (49%). However, compared to 2019, this year saw a relatively lower share of data sets published in institutional repositories (23% vs 9%) and a relatively higher share published in external general repositories (14% vs 26%). While this compositional shift in the institutional and external general categories from 2019 to 2020 is interesting to note, it is still too early to assess whether or not it is part of a meaningful 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. In addition, the Libraries provide DataCite DOI registration capabilities to a small number of campus partners through formal agreements. In 2020, the Libraries and its partners registered 320 DOIs for published data sets (up substantially from 197 in 2019). The disconnect between this larger number of DOIs and the 65 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 DOIs registered by the Libraries and its partners increasing by 64.1% in 2020 from 2019 again demonstrates the steady growth in demand for data publishing services at CU Boulder. As of the end of 2020, the Libraries and its partners had registered a total of 629 DOIs since the DataCite DOI registration service began at CU Boulder in 2018.