Beyond authorship: Analyzing disciplinary patterns of contribution statements using the CRediT taxonomy

Elvira González-Salmón[1], Victoria Di Césare[1*], Aoxia Xiao[2] and Nicolas Robinson-Garcia[1]

[1]Unit for Computational Humanities and Social Sciences (U-CHASS), University of Granada (Spain)
[2]School of Information Management, Wuhan University (China)
*Corresponding author: vdicesare@ugr.es

Abstract

In this research article we present the first cross-disciplinary descriptive analysis on the use of contribution statements. Our main objective is to obtain further insight on contributions by a variety of fields (Multidisciplinary, Health, Life, Physical and Social Sciences) from the largest dataset used up to now. We examine more than 700,000 articles published between 2017 and 2024 in Elsevier and PLOS journals, in combination with bibliometric data extracted from the Scopus database. The descriptive analysis of the dataset focuses on the overall coverage of the merged data, the distribution of authorship and disciplines at paper level, and the interactions between contribution statements, author order and disciplines. Our two main findings indicate that, on the one hand, looking at contributions and authorship order can enrich the way we understand science as a social endeavor. On the other hand, delving deeper into contributorship differences by field is key. We underscore the value of the CRediT taxonomy in unveiling nuanced research dynamics and offering a more equitable framework for evaluation.

Keywords: CRediT taxonomy; Authorship; Contributorship; Division of labor

Introduction

Team science is on the rise (Wuchty et al., 2007), challenging previous conceptions of what accounts for science and authorship (Birnholtz, 2006). This may require new norms that are prepared to deal with the new structure of science (Jabbehdari & Walsh, 2017) and has already led many to suggest replacing the concept of author with that of contributor, which acknowledges the distributed and collaborative nature of science as it is conducted in the 21st Century. Since their introduction in biomedical journals by the end of the 1990s (Rennie et al., 1997; Rennie et al., 2000), contribution

statements are becoming more widespread in academia, especially thanks to the launch of the Contributor Roles Taxonomy (CRediT) (Brand et al., 2015). This taxonomy was developed to homogenize contribution statements across publications and facilitate discussions about authorship that help avoid author disputes (Allen et al., 2019). Since its launch, the CRediT taxonomy is now used by more than 40 different publishers across a wide range of disciplines (https://credit.niso.org/). Their expansion not only has the potential of overcoming the limitations of using authorship as a "credit" in research assessment but can help us understand how science is done, how researchers distribute tasks, how this influences author order, and how it relates to their prestige as reflected in the author byline of publications (Haeussler & Sauermann, 2013).

Contributorship has been studied for some years now. Cronin et al. (2003) traced them even before they appeared in a special section within research articles by looking into the acknowledgements for Psychology and Philosophy throughout the 20th Century. Drawing on a multidisciplinary dataset of more than 80,000 documents, Larivière et al. (2016) examined the relationship between division of labor, contribution types and authors' seniority to provide evidence on the existence of conceptual contributions made by senior researchers and technical tasks performed by younger scholars. Later, Larivière et al. (2021) updated their analysis using the CRediT taxonomy for a set of more than 30,000 PLOS papers and studied the distribution of contributions across teams. Other relevant works have applied contributions to look into credit allocation (Ding et al., 2021; Sauermann & Haeussler, 2017), credit bias (Matheson, 2022), scientific trajectories (Robinson-Garcia et al., 2020), and gender differences (Macaluso et al., 2016; Sugimoto & Larivière, 2023).

However, much remains unknown since most analyses focus on specific disciplines and publications, mainly Biomedical Sciences and PLOS journals. In addition, from a bibliometric point of view, the possibilities of this still novel taxonomy are not fully explored. More than one decade since the birth of CRediT, we now have enough data to start getting a glimpse of its adoption, patterns and prospects. Here we present the first cross-disciplinary descriptive analysis on the use of contribution statements. The aim of this descriptive paper is to obtain further insight on contributions by a variety of fields, covering Health, Life and Physical Sciences as well as Multidisciplinary and Social Sciences, from the largest dataset used up to now, to the best of our knowledge. We analyze a set of over half a million research articles belonging to Elsevier and PLOS journals from all fields of science which we combine with bibliometric data extracted from the Scopus database. We look into disciplinary differences on the use of contributions, their overall coverage and their relationship with author position in papers. We conclude by discussing the potential of this data to open new venues of research on career trajectories, scientific impact and recognition, and the social organization of the sciences.

Data & Methods

In this study we examine a total 714,732 journal articles published between 2017 and 2024. The data was facilitated by the ICSR Lab from Elsevier as well as by PLOS. Elsevier was one of the first publishers adopting the CRediT taxonomy within their journal portfolio (Elsevier, 2024). Since then, they provide the option to include such information in their Editorial Manager and are working on offering it as well through the ScienceDirect journal platform (Genova, 2023). The ICSR Lab provided access to a total of 633,443 unique articles from 1,951 journals, authored by a list of 1,854,552 unique author profiles ranging from 2017 to 2024. We conducted some manual quality checks to ensure the reliability of the dataset which are available in the Appendix (see Table A1).

PLOS data was also facilitated by the publisher through a data use agreement. PLOS has stand out as one of the main drivers on the expansion of contribution statements, by liberating their bibliographic data and allowing the scientometric community to explore it to better understand team dynamics. Examples of such efforts are the studies conducted by Larivière et al. (2016, 2021), Sauermann and Haeussler (2017), Macaluso et al. (2016) and Robinson-Garcia et al. (2020) to name just a few. In this case they provided contribution data from their journal portfolio related to 97,819 publications for the 2018-2023 period, from which 83.2% belonged to PLOS One. Along with the author contribution statements assigned to every author, this dataset includes each article's Digital Object Identifier (DOI). After a manual validation check (See Appendix, Table A2), the dataset was merged into Scopus data via DOI and authors' given name and surname. A total of 81,289 records were correctly matched, corresponding to 415,014 disambiguated authors (based on Scopus Author Profile).

Our final dataset resulting from the merging of the ScienceDirect data and the PLOS data contains a total of 1,965 journals and 714,732 bibliographic records, authored by a total of 2,182,041 unique Scopus Author Profiles and published between 2017 and 2024. We have a minimum number of authors per paper of 1, and a maximum of 271. The full Python code developed to merge, compute and visualize the data is freely available in Di Césare and Xiao (2024) and the supplementary material can be accessed through González-Salmón et al. (2024). Next, we conduct a descriptive analysis of the dataset focusing on the overall coverage of the merged data, the distribution of authorship and disciplines at paper level, and the interactions between contribution statements, author order and disciplines.

Results

General overview

Figure 1 provides an overview of the magnitude of the dataset analyzed in comparison with the overall size of Scopus. The first three years of the period as well as 2024 present low coverage, but the

years 2020 to 2023 represent on average 6% of Scopus' content, with 2022 reaching the highest coverage point at 7.2% (208,207 articles). Table S1 (Supplementary material) shows this comparison by discipline and field. The list of fields, disciplines and their corresponding acronyms is included in the Appendix (Table A3).

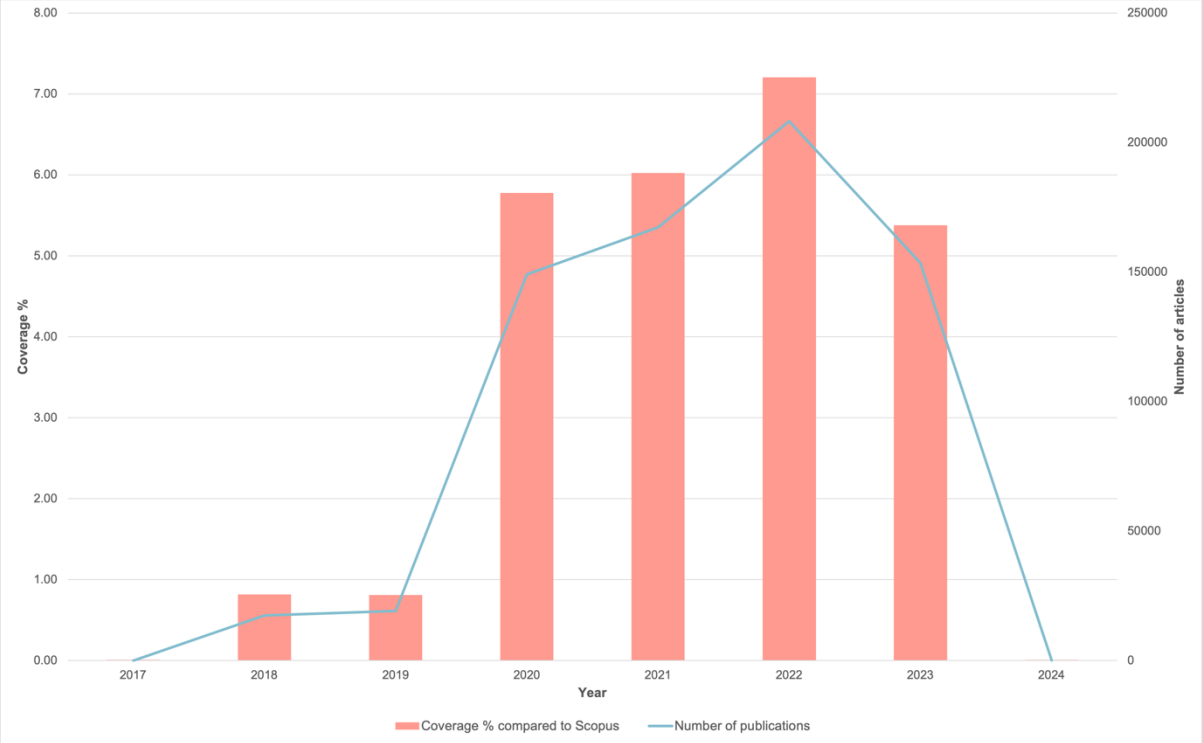

Figure 1. Coverage of our merged dataset compared to Scopus by year.

Both in absolute and relative terms, the most represented field in our dataset and compared to Scopus is Physical Sciences (503,145 articles, 5% of Scopus' content). Within it, the disciplines Engineering (198,231), Materials Science (161,105) and Chemistry (129,971) stand out with the highest numbers of articles for the whole period. Multidisciplinary as a field is also prominent because it presents the highest percentages of shared articles between our dataset and Scopus throughout the years (15% on average) (Supplementary material, Table S1). In addition, Table S2 (Supplementary material) lists the 1,965 journals that form our merged dataset with the number of articles each one contains per year. The journal covering the largest number of publications from our dataset is PLOS One, with almost 10% of the total articles.

Paper level analysis

Figure 2 shows the distribution of authors per paper overall and by major field. To ensure readability, we include a cut-off threshold of up to 50 authors per paper. The mode for the whole set is of 4 authors per paper (113,991 papers). When looking at number of authors across fields, we observe narrow differences. For Social Sciences, the most common number of authors is 3 (16,150 papers),

while for both Multidisciplinary (9,772 papers) and Physical Sciences (85,592 papers) it is 4. In Health Sciences and Life Sciences most papers are written by 5 authors (9,570 and 23,321 papers respectively).

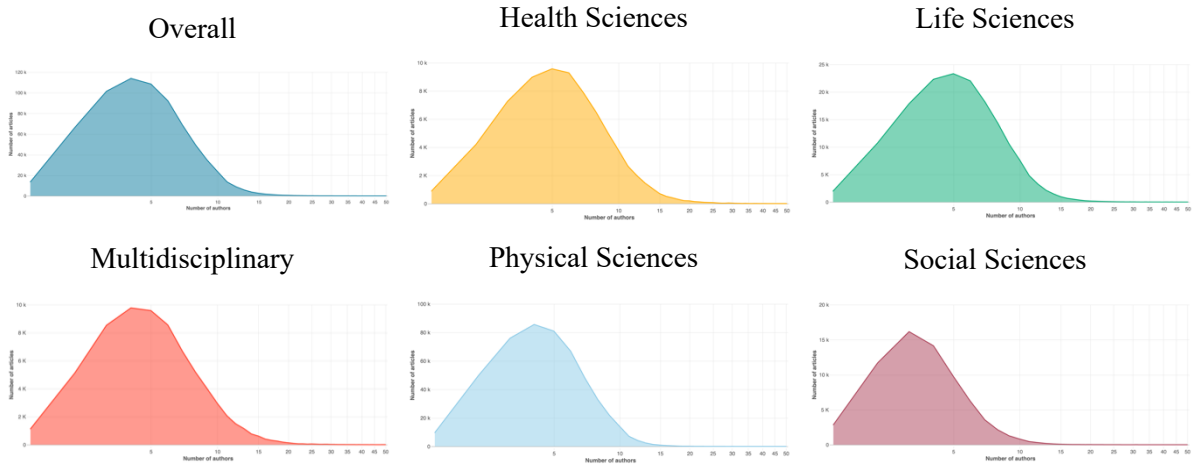

Figure 2. Distribution of number of authors (<= 50) by number of articles and field, log scale.

Similarly, in Figure 3 we examine the distribution on the number of contribution types per paper overall and by field. All fields except for Multidisciplinary follow a similar pattern, in which the number of papers decreases after reaching an average peak of 9 contribution types per paper. In Social Sciences the maximum number of contribution types per paper is 8, whereas Physical Sciences has 9, Health Sciences and Life Sciences have both 10, and Multidisciplinary 11 contribution types. Thus, Social Sciences is the field with the fewest contribution types, while Multidisciplinary presents almost 40% more contribution types.

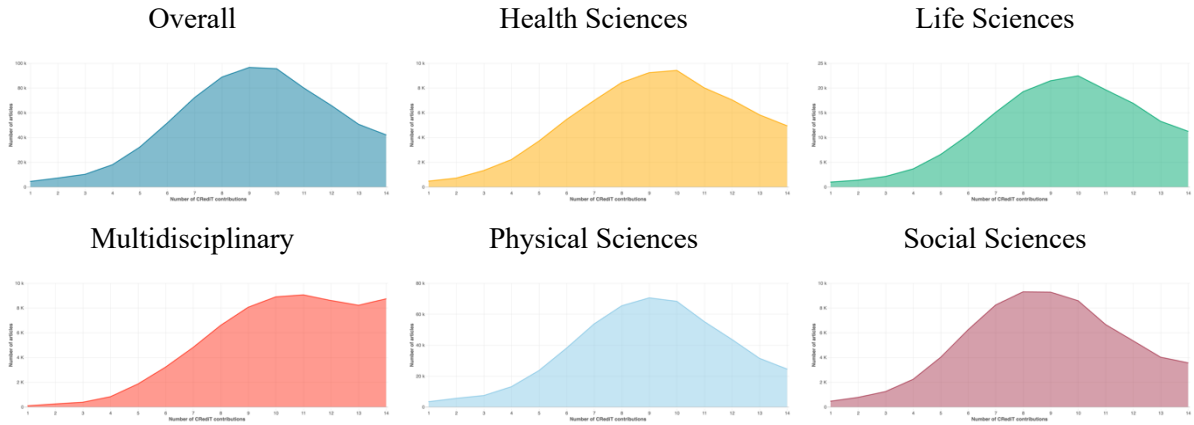

Figure 3. Distribution of number of CRediT contribution types per paper, overall and by field.

The average proportion of contribution types by discipline can be seen in Figure 4. Some particular statements, like Conceptualization, Methodology, Writing – original draft and Writing – review & editing, are clearly present in all disciplines at a high rate. Others, such as Resources, Project administration and Funding acquisition, are conversely seldom present. In between, we find a few mixed scenarios where the same contribution, for instance Software, Data curation and Formal analysis, has a high proportion of usage in some disciplines but very low in others.

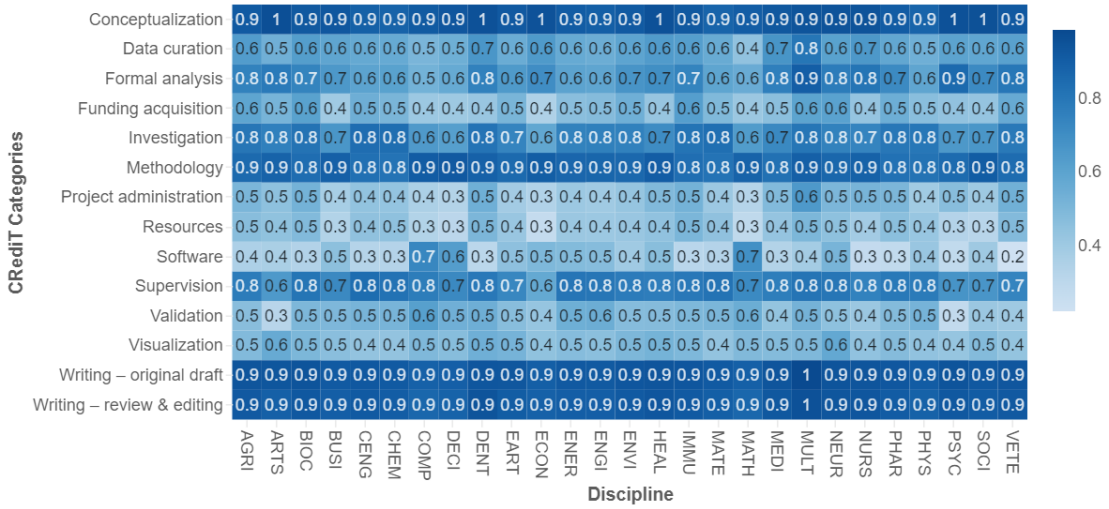

Figure 4. Average proportion of CRediT contribution types by discipline.

Author level analysis

Figure 5 shows the average number of contribution types each author contributes with on average per field. In this way, we show how distributed are tasks by field. For example, in Physical Sciences authors contribute on average with one task in 13.84% of the articles, whereas in 30.40% of them they conduct an average of 2 different tasks. In Health, Life and Physical Sciences the most common scenario is that of 2 contributions per author (26.52%, 27.80% and 30.40% respectively), while in Multidisciplinary and Social Sciences each author performs 3 tasks in 20.30% and 23.20% of the papers. If we look at higher numbers of contributions per author, we find that Health, Life and Physical Sciences are less represented in those groups (for instance, in less than 3% of their articles authors contribute with up to 7 different tasks), while Multidisciplinary and Social Sciences have a slightly more significant representation there (6.28% and 4.59%). Therefore, we observe a higher distribution of tasks in fields such as Health, Life and Physical Sciences. On the opposite side we find the Multidisciplinary field and Social Sciences, where authors are involved on a higher number of tasks on average.

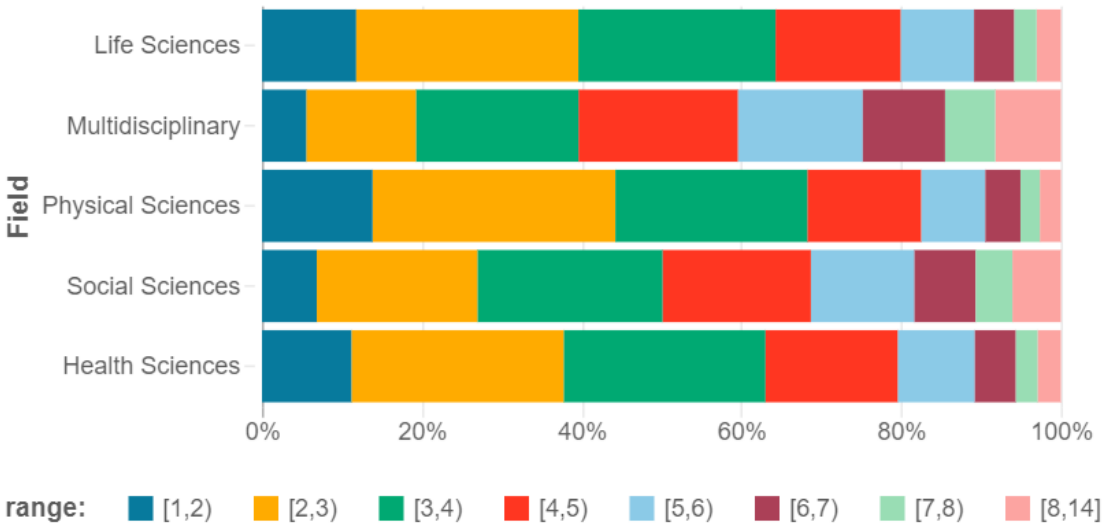

Figure 5. Average number of CRediT contribution types by author involved per article and field.

Co-occurrence of contributions by paper

Here we investigate the level of co-occurrence between contribution types, that is, how common it is for a contribution to be conducted by the same author in the same publication (Figure 6). The most common contribution co-occurrences take place between Conceptualization and Methodology (0.40), Writing - review & editing (0.35), and Writing - original draft (0.34). Furthermore, Writing - original draft, also tends to co-occur with Methodology (0.34) and Formal analysis (0.31). The contributions that appear together the least are mostly related to Software and other statements (e.g. Funding acquisition, Supervision and Project administration). The pairs formed by Funding acquisition and Data curation, Funding acquisition and Visualization, and Supervision and Data curation are also not usual contributions done by the same author within a publication (0.07 across all of them). There seems to be a higher co-occurrence of conceptual contributions whereby those who are involved in management related contributions (e.g. Resources and Project administration) do not typically perform technical tasks (e.g. Data curation and Software). When looking into specific fields, we observe that Methodology and Conceptualization is the most common combination in the Social (0.52), Health (0.41) and Physical Sciences (0.39), while Writing - review & editing together with Conceptualization co-occurs the most in Multidisciplinary (0.46) and Life Sciences (0.37). In general, there is a higher co-occurrence of contributions in the Social Sciences and Multidisciplinary fields than in the rest. This finding is consistent with the results shown in Figure 5, which indicate that authors in these fields are more likely to contribute to multiple roles simultaneously.

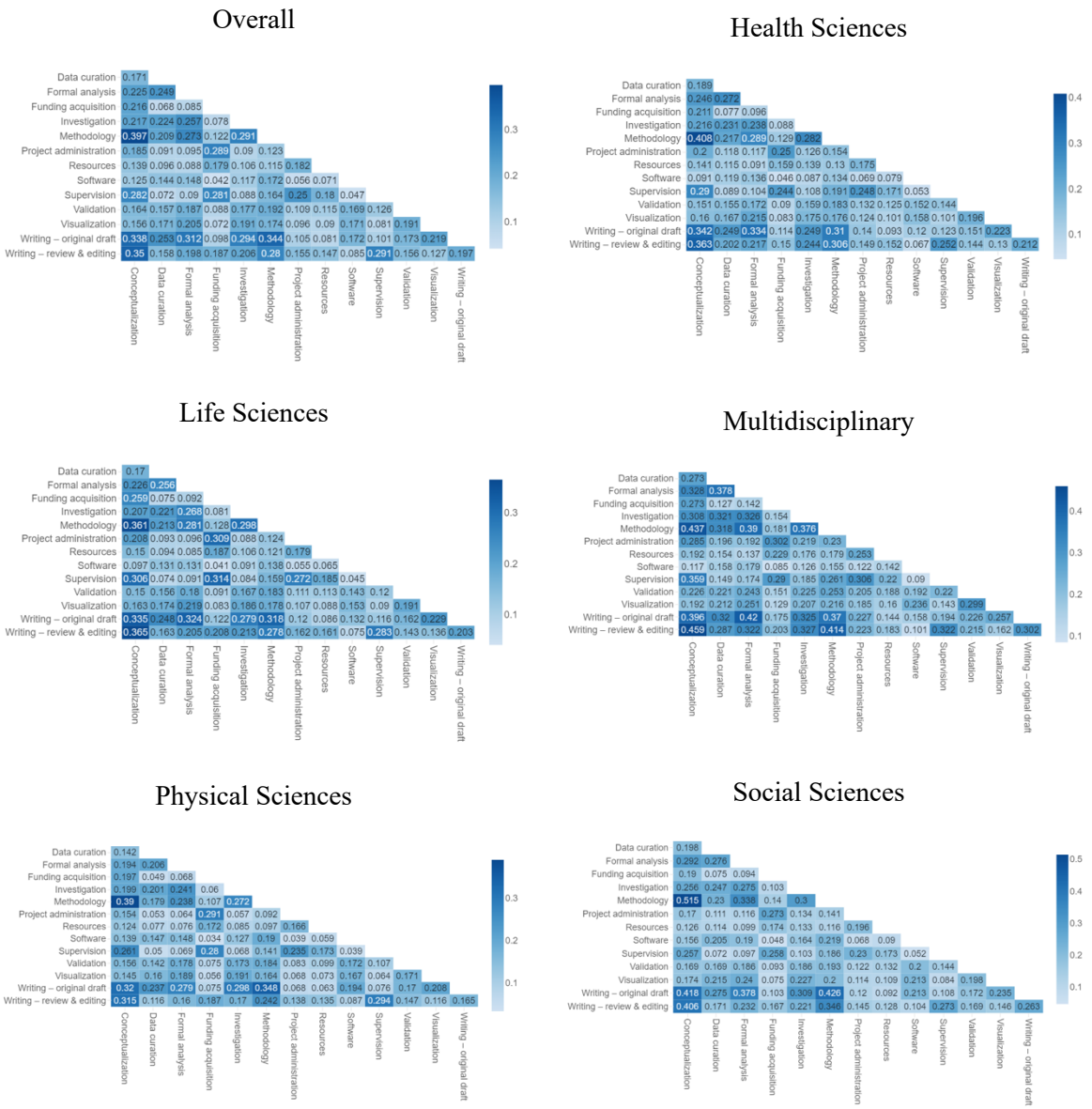

Figure 6. Co-occurrence of CRediT contribution types by field.

Relation between author order and contribution type

Finally, we analyze the relation between contributions and author positions. We do so by looking at the percentage of first, middle and last authors that perform each contribution overall and by field (Figure 7). First authors perform a higher number of contributions (37% on average) than last authors (27%) or middle authors (19%). The most common contribution among first authors is Writing - original draft (83.5%), followed by Methodology (67.2%), Conceptualization (63.3%) and Investigation (58.1%). Conversely, the contributions in which first authors participate the least are Resources (9%), Funding acquisition (10.8%), Supervision (10.9%) and Project administration (11.5%). In the case of last authors, their most common contributions are Writing - review & editing (66.3%) and Supervision (51%), while their least common contributions are Software (6.9%) and Visualization

(10.7%). As for middle authors, they participate the most in Writing - review & editing (45%) and Investigation (32.6%), but contribute very little to any of the rest (16% on average).

The results by field display a similar pattern where first authors always conduct a higher number of contributions. In all cases their main contribution is Writing - original draft, followed by Conceptualization (all fields but Multidisciplinary), Methodology (in Health, Social Sciences and Multidisciplinary) and Investigation (only in Life and Physical Sciences). The least common tasks among first authors are the most common ones among last authors, such as Supervision (across all fields) and Funding acquisition (in Health, Life Sciences and Multidisciplinary). Still, last authors particularly stand out in Writing - review & editing and Conceptualization. Lastly, middle authors' main contributions are Writing - review & editing (all fields but Multidisciplinary), Conceptualization (even with first authors in Health Sciences), Investigation and Methodology (especially in Multidisciplinary and Social Sciences), but they almost do not participate in Software, Visualization and Funding acquisition (with the exception of Physical Sciences in the latter).

Field		Overall			Health Sciences			Life Sciences		
Author position		First	Middle	Last	First	Middle	Last	First	Middle	Last
Author number		714671	2590052	701041	73723	364654	72821	164330	713629	162339
% of authors by CRediT	Conceptualization	63.3	21.8	48.0	68.3	23.4	62.5	61.9	60.5	42.7
	Data curation	38.9	19.8	12.2	43.8	24.5	17.1	41.5	19.7	13.1
	Formal analysis	49.0	21.0	16.5	57.9	21.0	21.9	56.1	21.8	18.6
	Funding acquisition	10.8	10.2	32.6	11.9	8.0	33.0	11.5	9.9	42.7
	Investigation	58.1	32.6	19.2	54.2	29.0	22.8	62.2	29.1	19.8
	Methodology	67.2	27.8	30.1	65.3	29.0	40.4	66.4	29.4	33.8
	Project administration	11.5	8.4	25.8	17.2	8.4	29.2	12.2	7.6	32.7
	Resources	9.0	14.7	20.9	10.9	15.9	21.3	8.7	15.8	23.8
	Software	25.4	9.2	6.9	18.2	7.2	5.7	19.8	8.1	12.1
	Supervision	10.9	19.2	51.0	12.0	16.0	59.1	9.4	16.4	61.2
	Validation	24.7	15.6	16.7	20.9	13.2	17.1	15.8	13.2	15.4
	Visualization	29.4	10.6	10.7	32.0	9.9	12.9	41.9	10.2	12.1
	Writing – original draft	83.5	14.1	18.8	86.4	15.6	27.0	85.1	13.4	24.2
	Writing – review & editing	39.9	45.0	66.3	47.9	54.5	75.0	41.9	43.6	72.0

Field		Multidisciplinary			Physical Sciences			Social Sciences		
Author position		First	Middle	Last	First	Middle	Last	First	Middle	Last
Author number		69564	305681	68416	503129	1654716	493423	69912	160627	67060
% of authors by CRediT	Conceptualization	72.0	30.5	70.7	61.2	40.6	40.6	77.5	46.4	46.4
	Data curation	61.9	31.1	26.1	34.7	17.0	9.3	38.0	20.4	15.0
	Formal analysis	74.5	27.6	33.1	42.7	19.3	12.7	54.1	21.7	19.4
	Funding acquisition	20.2	10.5	40.7	36.6	28.1	30.3	15.5	11.9	23.0
	Investigation	66.0	36.3	34.1	56.8	30.9	16.3	47.1	32.1	20.2
	Methodology	75.5	37.3	49.9	54.5	24.9	24.9	76.0	32.1	35.6
	Project administration	32.2	12.7	36.9	7.3	12.6	23.0	15.5	9.7	17.3
	Resources	18.0	16.0	26.8	36.8	14.0	19.7	10.7	12.2	15.7
	Software	9.0	8.0	15.7	36.8	10.2	21.9	31.0	11.1	9.3
	Supervision	20.7	21.7	71.0	14.2	20.0	54.1	14.1	22.0	34.2
	Validation	16.3	16.3	25.5	14.4	16.3	16.3	16.1	16.1	17.8
	Visualization	37.1	9.8	15.7	27.5	10.8	9.3	30.9	12.1	11.6
	Writing – original draft	91.5	18.2	37.4	81.6	12.2	13.6	83.6	22.7	23.6
	Writing – review & editing	64.6	0.1	82.8	34.6	34.6	61.7	48.9	56.5	67.9

Figure 7. Relation between CRediT contributions and author positions by field.

Discussion

This paper presents the first cross-disciplinary analysis of contribution statements performed. The novelty of this research lies in the unique dataset we analyze, which is the largest used so far to compare results at the field, discipline and author levels. Furthermore, we include a large sample of

journals that goes beyond PLOS, which have been the most thoroughly analyzed in previous studies (i.e. Ding et al., 2021; Larivière et al., 2021).

At paper level we examined the number of authors by article and across fields and found that less authors are usually involved in Social Sciences' publications. Although co-authorship has increased in the Social Sciences in recent decades (Henriksen, 2016), this field still presents fewer authors in line with what Fanelli and Glänzel (2013) as well as Thelwall and Maflahi (2022) detected. When looking at the number of contributions by paper, we observed that all fields but Multidisciplinary follow a similar pattern. At author level we identified considerable differences per field when delving into the average number of contributions by author. Social Sciences stands out the most here because, despite having fewer authors and contributions per publication, each author performs more of them. This suggests a lesser division of labor compared to the rest of the fields, which could be due to a lack of task specialization.

The findings related to contribution types' co-occurrence could indicate a concentration of intellectual tasks as well as a division between management responsibilities and technical activities. The higher values in the Social Sciences and Multidisciplinary contribution combinations could again point towards less specialization. Lastly, when delving into author order, we saw a further division of labor according to which first authors contribute with more tasks. The pattern that replicates across fields suggests a division of labor whereby the first authors' least performed contributions tend to be some of the most prominent among last authors. This strong complementarity between authors order and their contributions fits adequately with previous studies by Sauermann and Haeussler (2017), Larivière et al. (2021), and Escabias and Robinson-Garcia (2022) which suggest a relation between career stage and author order, with junior scholars in first position, while senior scholars occupying the last position in the author byline.

Throughout the entire analysis we have observed that both Social Sciences and Multidisciplinary differ from the other fields. The multidisciplinary field is a heterogeneous group, hence it is difficult to draw reliable conclusions from our findings. But in the Social Sciences we have seen some specific characteristics that are not new to scientometric research. Although usually studied in tandem with the Humanities, the Social Sciences have already shown some degree of single-authorships (Nederhof, 2006), as well as a faint trend towards larger work teams, probably due to more local-oriented research (Larivière et al., 2015). Moreover, its outputs are less common in mainstream databases like Scopus (Kulczycki et al., 2018) and social scientists still publish more monographs than other researchers (Giménez Toledo, 2020). All these nuances compel us to interpret the results carefully.

Conclusion

This research yields two main findings. First, looking at contributions and authorship order can enrich the way we understand science as a social endeavor. The future of research evaluation might be

partly in the nuanced application of the CRediT taxonomy (Allen et al., 2019). Why consider only authorship when contributorship may be more indicative of what is really happening in science? What is the future of authorship in an academia that increasingly works in teams? As research becomes more collaborative (Huang et al., 2023), CRediT contributions could embody a more comprehensive and fairer alternative for evaluating scientific outputs, one which ensures that recognition is appropriately distributed among authors. Second, delving deeper into contributorship differences by field is key. The Social Sciences particular dynamics highlight the need of evaluating each field applying ad hoc criteria that make varied uses of the contributions. In a research world where interdisciplinarity is increasingly common (Bolduc et al., 2023), the same approach could be taken for the Multidisciplinary field.

The analysis of this dataset led us to consider many potential areas that could be investigated further. For instance, in combination with data on gender, country of affiliation or academic age among others, contribution statements could provide a fuller understanding of the intersection between labor division and inequalities within science. Since it is not possible to gain access to every laboratory in order to study working patterns, the CRediT taxonomy could also be seen as an accessible and broad alternative to ethnographic approaches. CRediT statements are still recent, so for the time being it is not possible to conduct historical analysis. But thinking ahead, we find it very interesting to be able to witness the changes and trends that the use of this taxonomy is beginning to reveal.

Acknowledgments

We would like to thank the ICSR Lab from Elsevier and PLOS for providing access to ScienceDirect, Scopus and PLOS data on author contribution statements and bibliographic metadata.

Funding information

This paper is part of the COMPARE project (Ref: PID2020-117007RA-I00) funded by the Spanish Ministry of Science (Ref: MCIN/AEI/10.13039/501100011033 FSE invierte en tu futuro). Elvira González-Salmón is currently supported by an FPU grant from the Spanish Ministry of Science (Ref: FPU2021/02320). Victoria Di Césare is supported by a FPI grant from the Spanish Ministry of Science (Ref: PRE2021-097022). Aoxia Xiao is supported by a scholarship by from the China State Scholarship Fund. Nicolas Robinson-Garcia is supported by a Ramón y Cajal grant from the Spanish Ministry of Science (Ref: RYC2019-027886-I).

Authors contributions

Elvira González-Salmón: Data curation, Investigation, Formal analysis, Project administration, Writing – original draft.

Victoria Di Césare: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review & editing.

Aoxia Xiao: Data curation, Formal analysis, Methodology, Software, Visualization.

Nicolas Robinson-Garcia: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

References

Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, *32*(1), 71–74. https://doi.org/10.1002/leap.1210

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758–1770. https://doi.org/10.1002/asi.20380

Bolduc, S., Knox, J., & Ristroph, E. B. (2023). Evaluating team dynamics in interdisciplinary science teams. *Higher Education Evaluation and Development*, *17*(2), 70–81. https://doi.org/10.1108/HEED-10-2021-0069

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2), 151–155. https://doi.org/10.1087/20150211

Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, *54*(9), 855–871. https://doi.org/10.1002/asi.10278

Di Césare, V., & Xiao, A. (2024). *CRediT descriptive code* [Python]. https://github.com/vdicesare/CRediT-Descriptive

Ding, J., Liu, C., Zheng, Q., & Cai, W. (2021). A new method of co-author credit allocation based on contributor roles taxonomy: Proof of concept and evaluation using papers published in PLOS ONE. *Scientometrics*, *126*(9), 7561–7581. https://doi.org/10.1007/s11192-021-04075-x

Elsevier (2024). *CRediT author statement.* https://www.elsevier.com/researcher/author/policies-and-guidelines/credit-author-statement

Escabias, M., & Robinson-Garcia, N. (2022). *The relation between contribution statements and academic age*. 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain. https://doi.org/10.5281/ZENODO.6974732

Fanelli, D., & Glänzel, W. (2013). Bibliometric Evidence for a Hierarchy of the Sciences. *PLoS ONE*, *8*(6), e66938. https://doi.org/10.1371/journal.pone.0066938

Genova, Z. (May 16, 2023). *CRediT where credit's due.* Elsevier. https://www.elsevier.com/connect/credit-where-credits-due

Giménez Toledo, E. (2020). Why Books are Important in the Scholarly Communication System in Social Sciences and Humanities. *Scholarly Assessment Reports*, *2*(1), 6. https://doi.org/10.29024/sar.14

González-Salmón, E., Di Césare, V., Xiao, A., & Robinson-Garcia, N. (2024). Supplementary material of the manuscript "Beyond authorship: Analyzing disciplinary patterns of contribution statements using the CRediT taxonomy" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14131423

Haeussler, C., & Sauermann, H. (2013). Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. *Research Policy*, *42*(3), 688–703. https://doi.org/10.1016/j.respol.2012.09.009

Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, *107*(2), 455–476. https://doi.org/10.1007/s11192-016-1849-x

Huang, Y., Liu, X., Li, R., & Zhang, L. (2023). The science of team science (SciTS): An emerging and evolving field of interdisciplinary collaboration. *El Profesional de La Información*, e320204. https://doi.org/10.3145/epi.2023.mar.04

Jabbehdari, S., & Walsh, J. P. (2017). Authorship Norms and Project Structures in Science. *Science, Technology, & Human Values*, *42*(5), 872–900. https://doi.org/10.1177/0162243917697192

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A., & Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, *116*(1), 463–486. https://doi.org/10.1007/s11192-018-2711-0

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, *46*(3), 417–435. https://doi.org/10.1177/0306312716650046

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, *66*(7), 1323–1332. https://doi.org/10.1002/asi.23266

Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, *2*(1), 111–128. https://doi.org/10.1162/qss_a_00097

Macaluso, B., Larivière, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Academic Medicine*, *91*(8), 1136–1142. https://doi.org/10.1097/ACM.0000000000001261

Matheson, A. (2022). Attribution and credit bias in publication ethics. *Indian Journal of Medical Ethics*, 1–8. https://doi.org/10.20529/IJME.2022.023

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, *66*(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: a proposal to make contributors accountable. *Journal of the American Medical Association, 278*(7), 579-585. https://doi.org/10.1001/jama.278.7.579

Rennie, D., Flanagin, A., & Yank, V. (2000). The Contributions of Authors. *Journal of the American Medical Association*, *284*(1), 89–91.

Robinson-Garcia, N., Costas, R., Sugimoto, C. R., Larivière, V., & Nane, G. F. (2020). Task specialization across research careers. *eLife*, *9*, e60586. https://doi.org/10.7554/eLife.60586

Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances*, *3*(11), e1700404. https://doi.org/10.1126/sciadv.1700404

Sugimoto, C. R., & Larivière, V. (2023). *Equity for Women in Science*. Harvard University Press.

Thelwall, M., & Maflahi, N. (2022). Research coauthorship 1900–2020: Continuous, universal, and ongoing expansion. *Quantitative Science Studies*, *3*(2), 331–344. https://doi.org/10.1162/qss_a_00188

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Appendix

For the validation of the Scopus data, we manually checked 25 papers for 2017 (there were not enough publications to check 50) and 50 papers for each of the remaining years (2018-2022). We checked whether the contributions that our data stated matched those on the original publication and obtained the following results.

Table A1. Validation check of a random set of documents from Elsevier journals.

Year	% of correct CRediT statements	Nº of publications with correct statements
2017	100%	25/25
2018	72%	36/50
2019	90%	45/50
2020	94%	47/50
2021	92%	46/50
2022	88%	44/50
2023	92%	46/50
2024	54%	27/50

The most common mistakes we found on the contributions included the following:

- Authors missing in our data.
- Contributions from authors missing in our data.

- When data was missing on Scopus (that is, some author's contributions was "." or blank), our data had trouble identifying all contributions from that publication.
- When data was not normalized on Scopus (that is, it included contributions that are not part of the CRediT taxonomy such as "Problem designing" or "Wrote the manuscript"), our data had trouble identifying the contribution they belonged to.
- When more than one author shared a surname, our data sometimes mixed them up.
- Some contributions were sometimes duplicated in our data (for instance, an author having "Methodology" twice as their contributions).

We did the same for the validation of the PLOS data. That is, we checked 50 random papers for each year (2018-2023) to analyze the robustness of the data. As we can see in the following table, PLOS data had higher percentages of accuracy than that from Scopus.

Table A2. Validation check of a random set of documents from PLOS journals.

Year	% of correct CRediT statements	Nº of publications with correct statements
2018	100%	50/50
2019	94%	47/50
2020	98%	49/50
2021	100%	50/50
2022	98%	49/50
2023	100%	50/50

Table A3. Grouping of disciplines and fields.

Field	Discipline	Abbreviation
Health Sciences	Dentistry	DENT
	Health Professions	HEAL
	Medicine	MEDI
	Nursing	NURS
	Veterinary	VETE
Life Sciences	Agricultural and Biological Sciences	AGRI
	Biochemistry, Genetics and Molecular Biology	BIOC
	Immunology and Microbiology	IMMU
	Neuroscience	NEUR
	Pharmacology, Toxicology and Pharmaceutics	PHAR
Physical Sciences	Chemical Engineering	CENG
	Chemistry	CHEM
	Computer Sciences	COMP
	Earth and Planetary Sciences	EART
	Energy	ENER
	Engineering	ENGI
	Environmental Science	ENVI
	Materials Science	MATE
	Mathematics	MATH
	Physics and Astronomy	PHYS

Social Sciences	Arts and Humanities	ARTS
	Business Management and Accounting	BUSI
	Decision Sciences	DECI
	Economics, Econometrics and Finance	ECON
	Psychology	PSYC
	Social Sciences	SOCI
Multidisciplinary		MULT