Articles on Data and Software Extraction From Scientific Literature:

[1] A. Allen, P. J. Teuben, and P. W. Ryan. Schroedinger's code: A preliminary study on research source code availability and link persistence in astrophysics. *The Astrophysical Journal Supplement Series*, 236(1):10, may 2018.

[2] A. A. Alsheikh-Ali, W. Qureshi, M. H. Al-Mallah, and J. P. A. Ioannidis. Public availability of published research data in high-impact journals. *PLoS ONE*, 6(9):e24357, sep 2011.

[3] C. W. Belter. Measuring the value of research data: A citation analysis of oceanographic data sets. *PLoS ONE*, 9(3):e92590, mar 2014.

[4] K. Boland, D. Ritze, K. Eckert, and B. Mathiak. Identifying references to datasets in publications. In *International Conference on Theory and Practice of Digital Libraries*, pages 150–161. Springer, 2012.

[5] H. Chrapary, W. Dalitz, W. Neun, and W. Sperber. Design, concepts, and state of the art of the swmath service. *Mathematics in Computer Science*, 11(3):469–481, Dec 2017.

[6] A. Coppin. Finding science and engineering specific data set usage or funding acknowledgements. 2013.

[7] G. Duck, A. Kovacevic, D. L. Robertson, R. Stevens, and G. Nenadic. Ambiguity and variability of database and software names in bioinformatics. *Journal of biomedical semantics*, 6(1):29, 2015.

[8] G. Duck, G. Nenadic, A. Brass, D. L. Robertson, and R. Stevens. bioNerDS: exploring bioinformatics' database and software use through literature mining. *BMC bioinformatics*, 14(1):194, 2013.

[9] G. Duck, G. Nenadic, M. Filannino, A. Brass, D. L. Robertson, and R. Stevens. A survey of bioinformatics database and software usage through mining the literature. *PloS one*, 11(6):e0157989, 2016.

[10] B. Ghavimi, P. Mayr, C. Lange, S. Vahdati, and S. Auer. A semi-automatic approach for detecting dataset references in social science texts. *Information Services & Use*, 36(3-4):171–187, 2016.

[11] B. Ghavimi, P. Mayr, S. Vahdati, and C. Lange. Identifying and improving dataset references in social sciences full texts. *arXiv preprint arXiv:1603.01774*, 2016.

[12] M. Grechkin, H. Poon, and B. Howe. Wide-open: Accelerating public data release by automating detection of overdue datasets. *PLoS biology*, 15(6):e2002477, 2017.

[13] G.-M. Greuel and W. Sperber. swMATH – an information service for mathematical software. In H. Hong and C. Yap, editors, *Mathematical Software – ICMS 2014*, pages 691–701, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[14] M. Haeussler, M. Gerner, and C. M. Bergman. Annotating genes and genomes with DNA sequences extracted from biomedical articles. *Bioinformatics*, 27(7):980–986, feb 2011.

[15] T. Henderson and D. Kotz. Data citation practices in the CRAWDAD wireless network data archive. *D-Lib Magazine*, 21(1/2), jan 2015.

[16] J. Howison and J. Bullard. Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. *Journal of the Association for Information Science and Technology*, 67(9):2137–2155, 2016.

[17] Y.-H. Huang, P. W. Rose, and C.-N. Hsu. Citing a data repository: A case study of the protein data bank. *PLOS ONE*, 10(8):e0136631, aug 2015.

[18] Ş. Kafkas, J.-H. Kim, and J. R. McEntyre. Database citation in full text biomedical articles. *PloS one*, 8(5):e63184, 2013.

[19] Ş. Kafkas, J.-H. Kim, X. Pi, and J. R. McEntyre. Database citation in supplementary data linked to europe pubmed central full text biomedical articles. *Journal of biomedical semantics*, 6(1):1, 2015.

[20] P. W. Kirlew. Life science data repositories in the publications of scientists and librarians. 2011.

[21] J. Li, S. Zheng, H. Kang, Z. Hou, and Q. Qian. Identifying scientific project-generated data citation from full-text articles: An investigation of TCGA data citation. *Journal of Data and Information Science*, 1(2):32–44, 2016.

[22] K. Li, X. Lin, and J. Greenberg. Software citation, reuse and metadata considerations: An exploratory study examining LAMMPS. *Proceedings of the Association for Information Science and Technology*, 53(1):1–10, 2016.

[23] K. Li and E. Yan. Co-mention network of R packages: Scientific impact and clustering structure. *Journal of Informetrics*, 12(1):87–100, 2018.

[24] K. Li, E. Yan, and Y. Feng. How is R cited in research outputs? structure, impacts, and citation standard. *Journal of Informetrics*, 11(4):989–1002, 2017.

[25] M. Lu, S. Bangalore, G. Cormode, M. Hadjieleftheriou, and D. Srivastava. A dataset search engine for the research document corpus. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 1237–1240. IEEE, 2012.

[26] N. Mahrholz, A. Reinhold, and M. Rittberger. Data citation quantity and quality in research output of a large-scale educational panel study. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, page 31. ACM, 2015.

[27] G. R. Major. Impact of NASA EOS instrument data on the scientific literature: 10 years of published research results from Terra, Aqua, and Aura. 2011.

[28] H. Mooney. Citing data sources in the social sciences: do authors do it? *Learned Publishing*, 24(2):99–108, apr 2011.

[29] H. Mooney and M. P. Newton. The anatomy of a data citation: Discovery, reuse, and credit. *Journal of Librarianship & Scholarly Communication*, 1(1), 2012.

[30] U. Nangia and D. S. Katz. Understanding software in research: Initial results from examining nature and a call for collaboration. *arXiv preprint arXiv:1706.06527*, 2017.

[31] A. Névéol, W. J. Wilbur, and Z. Lu. Extraction of data deposition statements from the literature: a method for automatically tracking research results. *Bioinformatics*, 27(23):3306–3312, 2011.

[32] S. A. Ochsner, D. L. Steffen, C. J. Stoeckert Jr, and N. J. McKenna. Much room for improvement in deposition rates of expression microarray datasets. *Nature Methods*, 5(12):991, 2008.

[33] X. Pan, E. Yan, M. Cui, and W. Hua. Examining the usage, citation, and diffusion patterns of bibliometric mapping software: A comparative study of three tools. *Journal of Informetrics*, 12(2):481–493, 2018.

[34] X. Pan, E. Yan, and W. Hua. Disciplinary differences of software use and impact in scientific literature. *Scientometrics*, 109(3):1593–1610, 2016.

[35] X. Pan, E. Yan, Q. Wang, and W. Hua. Assessing the impact of software on science: A bootstrapped learning of software entities in full-text papers. *Journal of Informetrics*, 9(4):860–871, 2015.

[36] H. Park and D. Wolfram. An examination of research data sharing and re-use: implications for data citation practice. *Scientometrics*, 111(1):443–461, 2017.

[37] A. Pepe, A. Goodman, A. Muench, M. Crosas, and C. Erdmann. How do astronomers share data? reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers. *PLoS ONE*, 9(8):e104798, aug 2014.

[38] H. A. Piwowar, J. D. Carlson, and T. J. Vision. Beginning to track 1000 datasets from public repositories into the published literature. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–4, 1 2011.

[39] A. Prlic, M. A. Martinez, D. Dimitropoulos, B. Beran, B. T. Yukich, P. W. Rose, P. E. Bourne, and J. L. Fink. Integration of open access literature into the rcsb protein data bank using biolit. *BMC bioinformatics*, 11:220, Apr 2010.

[40] P. H. Russell, R. L. Johnson, S. Ananthan, B. Harnke, and N. E. Carlson. A large-scale analysis of bioinformatics code on GitHub. *PLOS ONE*, 13(10):e0205898, oct 2018.

[41] M. Servilla, J. Brunt, D. Costa, J. McGann, and R. Waide. The contribution and reuse of LTER data in the provenance aware synthesis tracking architecture (PASTA) data repository. *Ecological informatics*, 36:247–258, 2016.

[42] J. E. Sieber and B. E. Trumbo. (Not) giving credit where credit is due: Citation of data sets. *Science and Engineering Ethics*, 1(1):11–20, mar 1995.

[43] A. Singhal and J. Srivastava. Data extract: Mining context from the web for dataset extraction. *International Journal of Machine Learning and Computing*, 3(2):219, 2013.

[44] A. Yan and N. Weber. Mining open government data used in scientific research. In *International Conference on Information*, pages 303–313. Springer, 2018.

[45] Q. Yu, Y. Ding, M. Song, S. Song, J. Liu, and B. Zhang. Tracing database usage: Detecting main paths in database link networks. *Journal of Informetrics*, 9(1):1–15, jan 2015.

[46] W. Zenk-Möltgen and G. Lepthien. Data sharing in sociology journals. *Online Information Review*, 38(6):709–722, sep 2014.

[47] Q. Zhang, Q. Cheng, Y. Huang, and W. Lu. A bootstrapping-based method to automatically identify data-usage statements in publications. *Journal of Data and Information Science*, 1(1):69–85, 2016.

[48] M. Zhao, E. Yan, and K. Li. Data set mentions and citations: A content analysis of full-text publications. *Journal of the Association for Information Science and Technology*, 2018.