

Historical bibliometrics using Google Scholar: the case of Roman law, 1500-2016

Janne Pölonen¹ and Björn Hammarfelt²

¹ janne.polonen@tsv.fi

Federation of Finnish Learned Societies, Snellmaninkatu 13, 00170 Helsinki (Finland)

² bjorn.hammarfelt@hb.se

Swedish School of Library and Information Science, University of Borås, Allégatan 1, Borås, 503 32 (Sweden)

Introduction

Bibliometrics can be a useful resource for social sciences and humanities (SSH) research beyond its role in research evaluation and funding-schemes (Scharnhorst & Garfield, 2010). Predominately research in the field of bibliometrics focus on contemporary developments using datasets that rarely provide historical perspectives. Still, historical approaches are not unheard of, with de Solla Price's seminal studies of the growth of science being one key example (de Solla Price, 1986). Yet, the field of bibliometrics has not fully explored the potential of what Hérubel (1999) calls "historical bibliometrics". There have been several attempts of going beyond established databases to study for example Catalan literature (Ardanuy, Urbano, & Quintana, 2009), Swedish literature (Hammarfelt, 2012) and Venetian histography (Colavizza, 2018). Yet, the approaches and methods used are often time-consuming and not easily transferred to other contexts and materials.

Considering the lack of coverage in established citation databases (Web of Science and Scopus), and the limitations of local and specific approaches, in this paper we investigate the potential that Google Scholar (GS) data has for studying the development of research fields from a historical perspective using Roman law as an example. Roman law (RL) has constituted an international research field within academia since the 12th century (Stein, 1999). After Latin ceased to be the *lingua franca*, there remain five international RL publishing languages: English, French, German, Italian and Spanish. Thus, Roman law literature provides a good case for probing the historical and linguistic coverage of GS.

Methods and materials

All publication records including in the title words denoting "Roman law" in English, French, German, Italian and Spanish, published between years 1500 and 2016, were retrieved from Google Scholar in August 2017, in blocs not exceeding 1000, using the Publish or Perish interface. The publication records were copied to Excel in RIS format, and

processed with BibExcel tool-box (Persson, Danell, & Schneider, 2009).

The dataset of Roman law publications is analyzed to establish the number of publications and authors, differentiating between the five language groups, from 1500 to 2016. The growth of the field is estimated on basis of the development of the absolute number and average yearly number of publications, as well as the number of authors involved in producing them, in different periods. Also bibliometric measurements are performed on the data to investigate its properties and consistency. These include the average number of publications per author (publication productivity), as well as the concentration of publications and citations.

Findings

The data retrieved from GS contains a total of 21300 publications published between years 1500 and 2016 and including the title words "Roman law" in the five languages (Table 1). The oldest publication year in French is 1727, in German 1730, in English 1772, in Spanish 1796 and in Italian 1833. Largest group of records consists of 9983 French publications that account for 47 % of all records. English language publications make up 18 %, Italian publication 13 %, Spanish publication 13 % and German publications 9 % of the records.

Table 1. GS records for publications 1725-2016

Period	En	Fr	De	It	Es	All
1725-1749	0	3	7	0	0	10
1750-1774	4	1	4	0	0	9
1775-1799	3	8	14	0	1	26
1800-1824	3	23	41	0	1	68
1825-1849	8	354	85	5	26	478
1850-1874	32	2540	116	22	32	2742
1875-1899	106	5592	213	148	56	6115
1900-1924	259	162	218	214	48	901
1925-1949	406	276	168	415	125	1390
1950-1974	616	414	359	629	301	2319
1975-1999	928	294	359	613	721	2915
2000-2016	1347	300	363	708	1352	4070
No date	71	16	53	49	68	257
Total	3783	9983	2000	2803	2731	21300

The number of RL publications has increased from around 10 publications in the earliest periods to 4000 publications in 2000-2016 (the latest timeframe is only 17 years). The early 19th century is a period when the number of publications begins to increase in all language groups (Table 1).

The largest number of RL publications is attested in the late 19th century, when there is a very large number of French publications (mostly thesis and dissertations). This can be related to requirements of French legal education. Following the introduction of Code Napoleon in 1804 and the reform of law schools, between 1808 and 1895 doctoral thesis in law consisted of two dissertations, one of which had to be based on Roman law (Imbert, 1984).

According to the GS data, the 21300 Roman law publications have a total of 11420 different authors. The largest number of authors is attested in 1875-1899, vast majority being related to the French publications. The average number of publications per author in the GS dataset has somewhat increased (Figure 1).

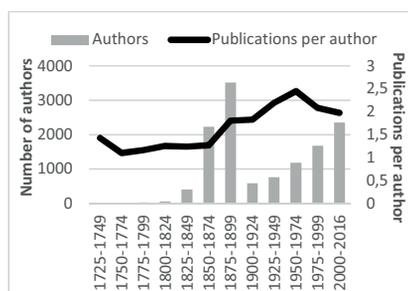


Figure 1. Publications per author 1725-2016

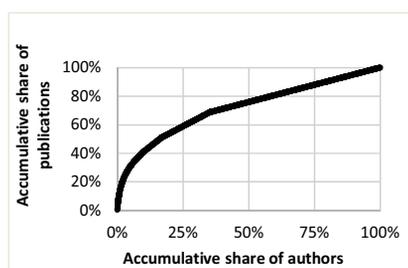


Figure 2. Concentration of publication to authors

Publication are unevenly distributed among the authors: one-half of all publications is produced by 16 % of the most prolific authors (Figure 2). Citations are even more unevenly distributed: 73 % of the publications have received no citations recorded in Google Scholar, and only 1 % of the

most highly cited publications account for one-half of all the citations.

Discussion and conclusions

We find Google Scholar to be a promising data source for historical bibliometrics: it is accessible, has broad coverage and has quite a historical depth. At the same time there are distinct disadvantages: the quality of data is low, and the database is continuously updating which renders it difficult to reproduce earlier searches and data collections. Still, the possibilities for historical bibliometrics will most likely increase as the digitisation of older materials progress. Hence, while the approach taken here is a probing one, with many difficulties to solve, we find that employing Google scholar data for historical studies of fields and disciplines is a promising path for the future, and it is likely that such a path might attract travellers among bibliometricians as well as historians and other digital humanists.

References

- Ardanuy, J., Urbano, C., & Quintana, L. (2009). A citation analysis of Catalan literary studies (1974–2003): Towards a bibliometrics of humanities studies in minority languages.
- Colavizza, G. (2018). Understanding the History of the Humanities from a Bibliometric Perspective: Expansion, Conjunctures, and Traditions in the Last Decades of Venetian Historiography (1950–2013). *History of Humanities*, 3(2), 377–406.
- De Solla Price, D. J. (1986). *Little science, big science... and beyond*. Columbia University Press New York.
- Hammarfelt, B. (2012). Harvesting footnotes in a rural field: citation patterns in Swedish literary studies. *Journal of Documentation*, 68(4), 536–558.
- Hérubel, J.-P. V. M. (1999). Historical Bibliometrics: Its Purpose and Significance to the History of Disciplines. *Libraries & Culture*, 34(4), 380–388.
- Imbert, J. (1984). Passé, présent et avenir du doctorat en droit en France, *Revue d'histoire des Facultés de droit*, 1, 11–35.
- Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström, R. Danell, B. Larsen & J. Schneider (eds.), *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday*. International Society for Scientometrics and Informetrics, 9–24.
- Scharnhorst, A., & Garfield, E. (2010). Tracing scientific influence. arXiv Preprint arXiv:1010.3525.
- Stein, P. (1999). *Roman Law in European History*. Cambridge: Cambridge University Press.