STI Conference 2016 · València Peripheries, frontiers and beyond
València, 14 · 16 September 2016
21st International Conference on Science and Technology Indicators | València (Spain) | September 14-16, 2016

# Indexed University presses: overlap and geographical distribution in five book assessment databases[1]

Jorge Mañana-Rodríguez*, Elea Giménez-Toledo**

*jorge.mannana@cchs.csic.es
**elea.gimenez@cchs.csic.es
Centro de Ciencias Humanas y Sociales, (ÍLIA Research Group), CSIC, Albasanz Street, 28037, Madrid (Spain)

**ABSTRACT:**
Scholarly books have been a periphery among the objects of study of bibliometrics until recent developments provided tools for assessment purposes. Among scholarly book publishers, University Presses (UPs hereinafter), subject to specific ends and constrains in their publishing activity, might also remain on a second-level periphery despite their relevance as scholarly book publishers. In this study the authors analyze the absolute and relative presence, overlap and uniquely-indexed cases of 503 UPs by country, among five assessment-oriented databases containing data on scholarly book publishers: Book Citation Index, Scopus, Scholarly Publishers Indicators (Spain), the lists of publishers from the Norwegian System (CRISTIN) and the lists of publishers from the Finnish System (JUFO). The comparison between commercial databases and public, national databases points towards a differential pattern: prestigious UPs in the English Speaking world represent larger shares and there is a higher overall percentage of UPs in the commercial databases, while the richness and diversity is higher in the case of national databases. Explicit or *de facto* biases towards production in English by commercial databases, as well as diverse indexation criteria might explain the differences observed. The analysis of the presence of UPs in different numbers of databases by country also provides a general picture of the average degree of diffusion of UPs among information systems. The analysis of 'endemic' UPs, those indexed only in one of the five databases points out to strongly different compositions of UPs in commercial and non-commercial databases. A combination of commercial and non commercial databases seems to be the optimal option for assessment purposes while the validity and desirability of the ongoing debate on the role of UPs can be also concluded.

**Key Words:** University Presses, Scholarly Books, scholarly book assessment, database coverage, research evaluation.

[1] This research has been carried out in the framework of the Project CSO2015-63693, 'Las prensas universitarias iberoamericanas y el libro científico en español: calidad, modelos de negocio y política institucional' funded by the Spanish Ministry of Economy and Competitiveness. The authors gratefully acknowledge Gunnar Sivertsen and Janne Pölönen for the information provided regarding book publishers' lists.

## INTRODUCTION

Books have remained on the peripheries of bibliometrics for a long time. Despite the emergence of bibliometric information systems in the 70's, it was not until 1996 that Eugene Garfield proposed the creation of a database for scholarly books (Garfield, 1996). Books were and are a key communication channel for Social Scientists and Humanists (Hicks, D., 2004; Engels, Ossenblok & Spruyt, 2012; Thompson, 2002; Giménez-Toledo *et al*., 2016). 53% of the output in SSH fields in Norway were published, between 2005 and 2009, in the form of monographs and book chapters (Sivertsen & Larsen, 2012) while the percentages in Finland were (for 2011-2012) 39 % and 47% in the case of Social Sciences and Humanities respectively (Puuska, 2014); 62% of the output of Spanish universities in Arts and Humanities were Books and Chapters (Michavila, 2012). Also, citation analysis considering books as a source of reference information shows that its relevance is far from residual (Gorraiz et al., 2013; Leydesdorff & Felt, 2012). Finally, in market terms, scholarly books represent a large percentage of total profits in the book market in Europe: 19.5%, being 'the second most important sales segment, after consumer (trade) books', in 2014 (Federation of European Publishers, 2014); in the case of Spain, 3.8% of the yearly turnover of the book industry corresponds to that from scientific-technical books and 10.8% to books in fields of the Social Sciences and the Humanities (FGEE, 2014).

University Presses, as part of the scholarly book publishing sector, are not necessarily a periphery in the scholarly book segment: 11,000 books a year are published by the ninety-two university presses belonging to the American Association of University Presses; Abel & Newlin, 2002, while 10% of all book publishing in Latin America was produced, in 2013, by University Presses (CERLALC, 2014, p. 30). In the case of Spain, University Presses published 6.5% of the total volume of books published in Spain in 2013 (FGEE, 2014) and the Presses Universitaires de France keep a catalogue with over 5,000 titles (PUF). Despite not being in the periphery in terms of publication volume and share, they can be considered a periphery since books are mostly produced in SSH disciplines and SSH is a small part of research in terms of funding and human resources Moreover university presses are not, in most cases, privately held companies but do compete with private companies while keeping a set of specificities (AEUP, 2015) that might set them closer to the periphery than to the core of scholarly book publishers: University Presses are often constrained by normative obligations from the entities they belong to. Often characterized by a local factor with regard to the works published (AAUP, s.d.), the languages used and the specific factors which regulate their publishing activity, it is the aim of this work to analyze the role of University Presses in five assessment oriented databases: Book Citation Index, Scopus, the lists from the CRISTIN system in Norway, the Finnish Lists and Scholarly Publishers Indicators. It is assumed that University Presses tend to be closely related to the activity of the university and, therefore the diffusion among different information systems shows recognition far from the closest institution or region. That international presence can be related to variables such as: a) recognition of the publisher by foreign specialists, b) improvements in the diffusion strategies of the publishers, c) professionalization and budget in marketing tasks, d) publication languages, e) business model and f) topics covered by the publisher (local topics would be less interesting for audiences abroad). The first step in the study of those conditioning variables is the analysis of the currently available data. Since the variables which explain the indexation of a given publisher differ, the comparisons should be done taking it into

consideration. There can be identified three systems for the inclusion of book publishers in the different information systems:

a) Book Citation index and Scopus apply different criteria including reputation and impact or content quality and, in the case of BCI: "English language full text is highly desirable, but books with full text in a language other than English are also considered for coverage in Book Citation Index"[2]
b) Finnish lists and Norwegian lists include scholarly publishers in which scholars from the respective countries have published research.
c) SPI includes book publishers which have been mentioned as relevant by a set of Spanish scholars through a survey methodology.

**OBJECTIVES**

The five databases studied here have in common the fact that they are recognition-based systems. The objectives of this work are the following ones:

a) Identify descriptive patterns in the geographical distribution of the UP's in the five databases.
b) Compare the coverage of the privately held databases to the coverage of the public databases
c) Identify the degree of overlap between the different databases.
d) Extract conclusions on the applicability of the different systems for assessment at the national or international level.
e) Identify variations in the visibility or recognition of UPs throughout the databases and which role they play in each one.

**METHODOLOGY**

Data origin:

The origin of data can be traced to the development Scholarly Publishers Indicators Expanded: http://ilia.cchs.csic.es/SPI/expanded_index.html), which included the five main databases on scholarly books. The lists of publishers were retrieved between December, 2015 and February, 2016 from the following sources:

-SPI: http://ilia.cchs.csic.es/SPI/
-Book Citation Index: http://wokinfo.com/mbl/publishers/
-Scopus:
http://www.elsevier.com/__data/assets/excel_doc/0005/154571/Scopus__books_29_4_15.xlsx
-Norwegian lists (CRISTIN): institutional exchange of files /Personal communication with Gunnar Sivertsen.
-Finnish lists: http://www.tsv.fi/julkaisufoorumi/lataa.php?id=

Although the selected sources allow multiple analysis, to be focused on UP let us know more about the role and the behaviour of one of the peripheries (UP) of the peripheries (evaluation of SSH)

*Data processing:* First, a master list containing the names of the publishers in all the five information systems was prepared; then, cleansing operations were performed: deletion of

[2] http://wokinfo.com/media/pdf/BKCI-SelectionEssay_web.pdf

non ASCII symbols, deletion of spaces, etc,. Then, an exact match search was performed with the master list and each individual list. The second phase involved the manual normalization of variants. The process yielded a total of 3948 distinct book publishers. The third phase of data processing involved the identification of university presses. For this purpose, the following word roots were searched in the names of the book publishers (Table 1):

Table 1: character chains used for the automatic identification of Ups.

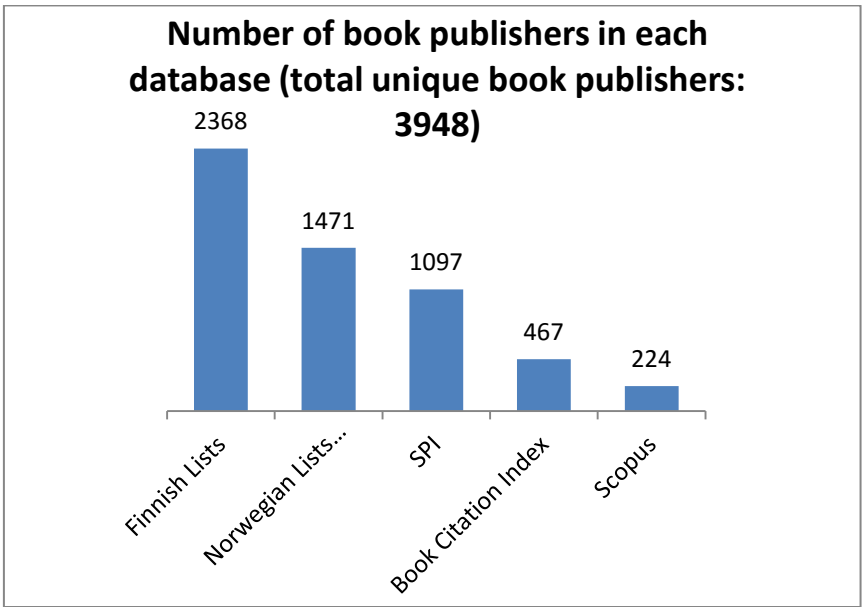
| | |
|---|---|
| Univers | Root of the term "university" in most romance languages. |
| Yliop* | Root of the term "Yliopisto", University in Finnish. |
| Korkea* | Root of the term "korkeakoulu", College and / or synonym of University in Finnish |
| Colleg | Root of the term "College" in most romance languages |
| Ülikoo | Root of the word 'Ülikool", University in Estonian |
| Egyet / össz | Roots of the word "Egyetem" in Hungarian |
| Uniwers | Root of the word Uniwersytet in Polish |

* Janne Pölönen, Coordinator of the Publication Forum, Federation of Finnish Learned Societies, was contacted by the researchers regarding the issue: he provided a wider list of Finnish UP's not necessary containing the roots detailed in this table as well as useful clarifications on the scheme for inclusion of UPs in the Finnish system.

Some other roots were also tested (I.E. for Pinyin Chinese and Hepburn Rōmaji Japanese "Daigak") with limited results.
Once the university presses were identified, a final manual depuration was carried out, excluding erroneously identified cases as well as normalizing variants. Also, a further review of the full list of publisher was manually carried out in order to identify university presses not previously identified.
Once the set of UPs was identified, the number of different databases in which each UP was included was counted. From this point, descriptive statistics were computed.

## RESULTS

**Figure 1.** Number of book publishers in each database (total n=3948)




The largest set of book publishers can be found in the Finnish System, followed by the Norwegian Lists and SPI (Fig. 1).

**Figure 2**. Distribution of UPs among databases

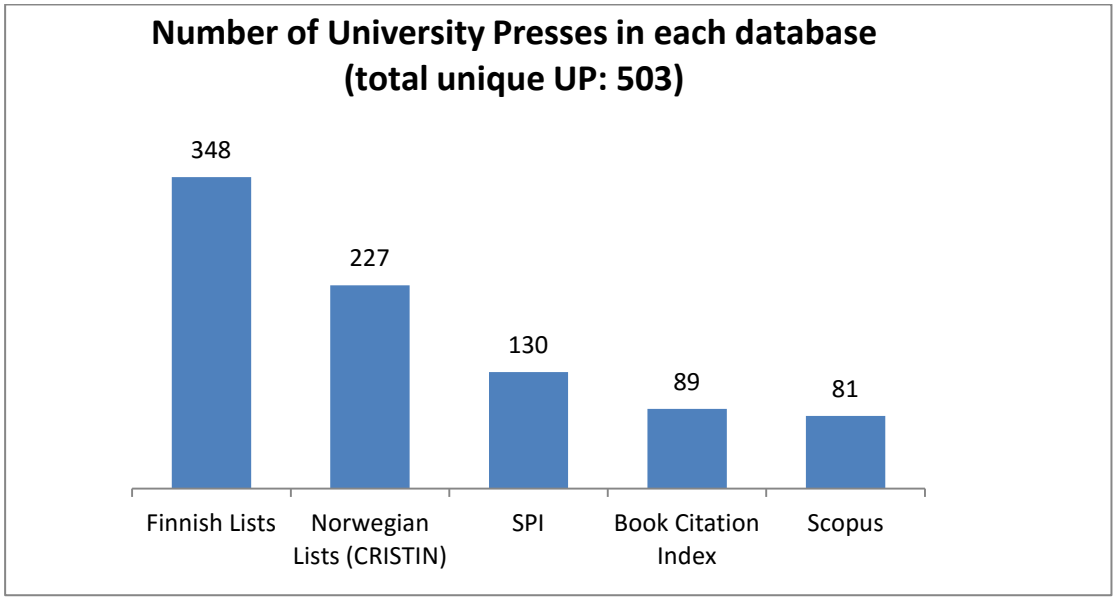
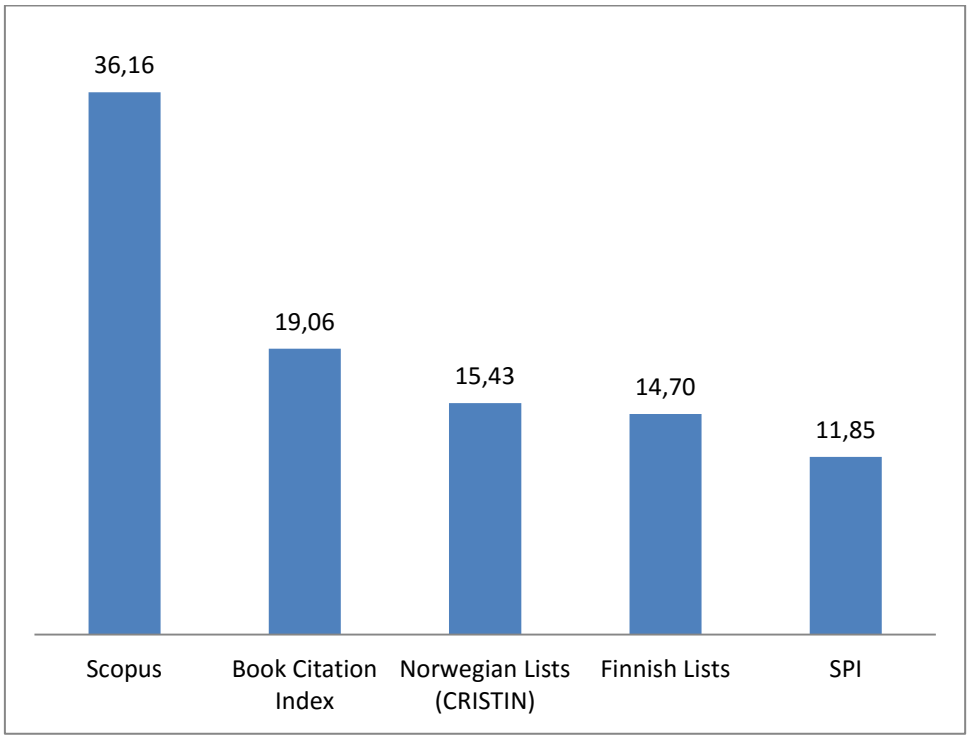
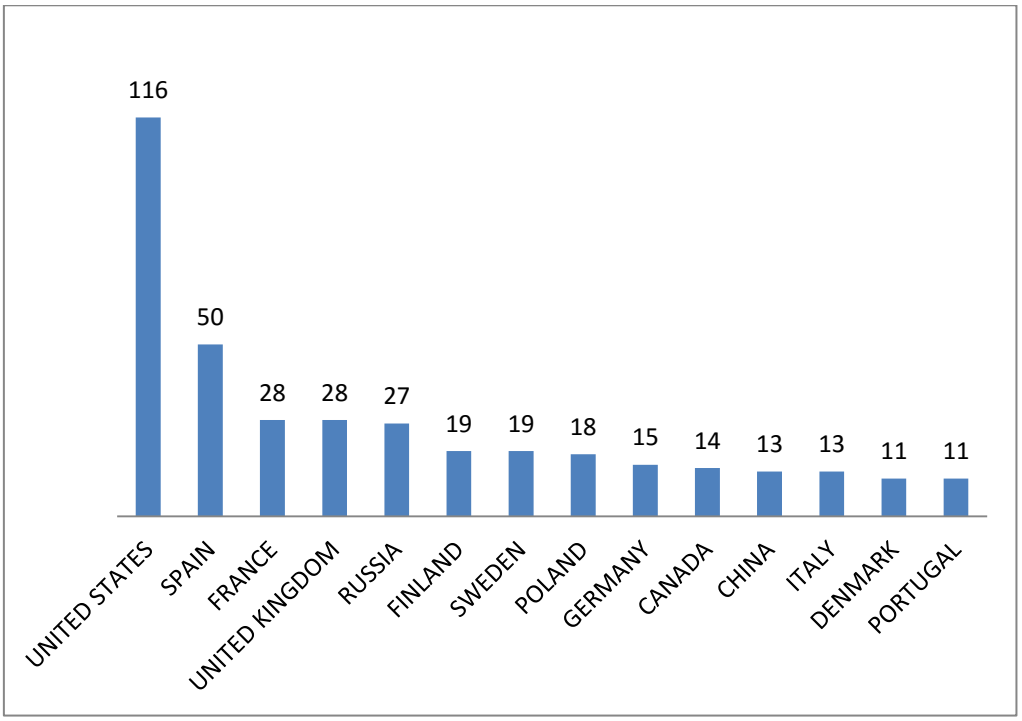


**Figure 3**. Percentage of UPs in each database



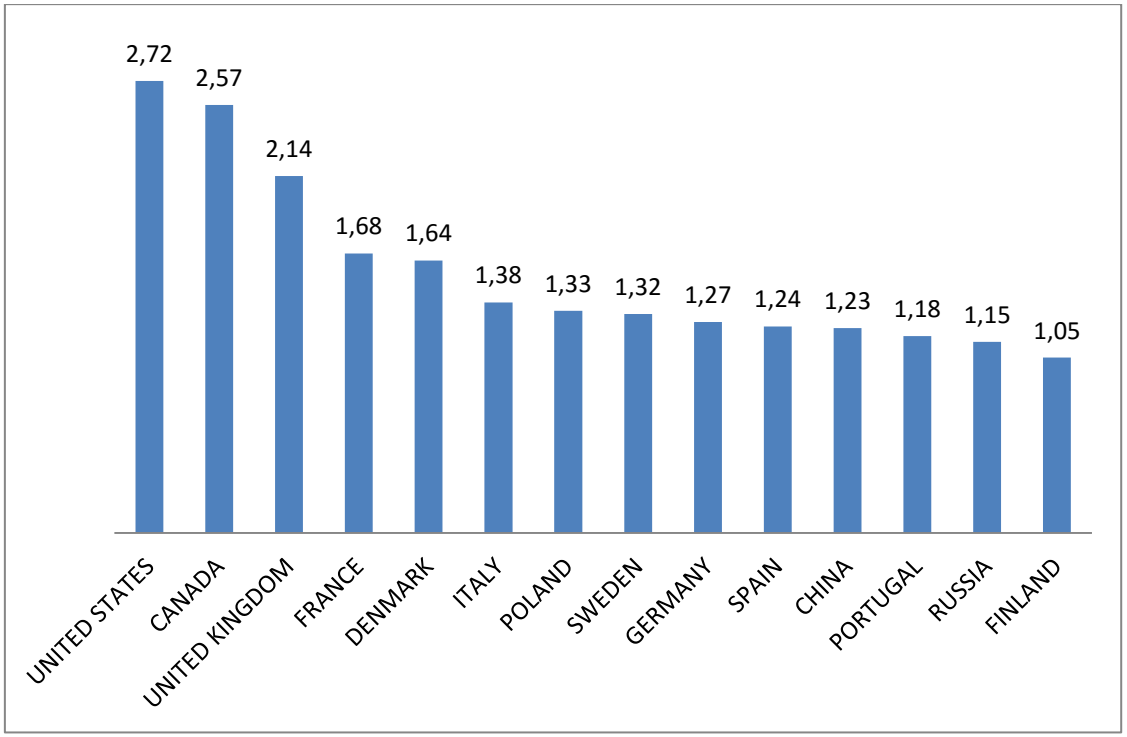

The percentage of UPs is significantly higher in the commercial databases, remarkably in Scopus, while SPI remains the database with the lowest percentage of University Presses (Fig. 3).

**Figure 4.** Countries with at least 10 UPs in any information system



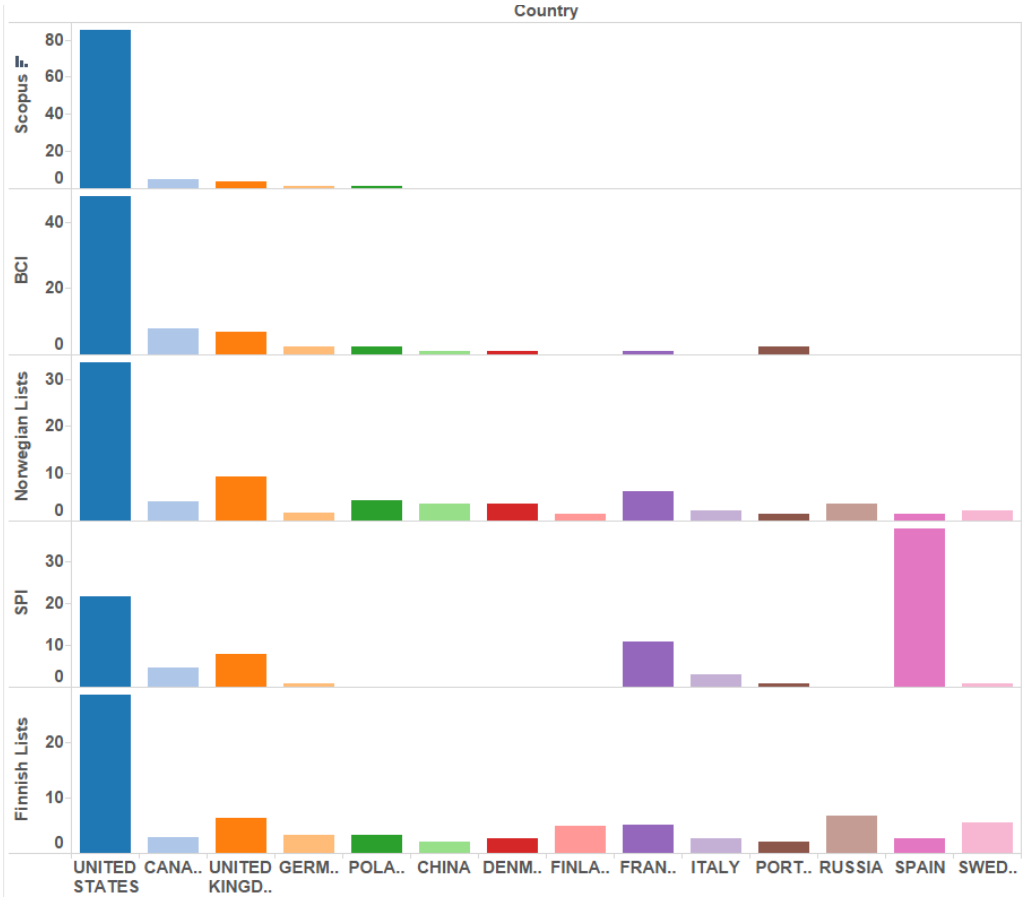
USA, Spain, France and Russia are the countries with a higher presence of UPs in any of the five systems (Fig. 4).

**Figure 5.** Average number of information systems for each country's UPs (if 10 or more book publishers in any information system).




The United States, Canada and the UK are the countries which UPs show the higher average presence in the five information systems (Fig. 5).

**Figure 6**. Distribution of each country UPs among databases




A thematic map with this information can be found at: https://public.tableau.com/shared/Z2ZGY8DXS?:display_count=yes

A large diversity of different countries in the public sources of Spain, Norway and Finland can be observed, while the concentration of countries (in most cases English-Speaking ones) in the case of the privately held products is observably larger.

**COMBINATIONS, OVERLAP & 'ENDEMIC' UPS**

The frequency of combinations in the five databases shows great variability in the overlap pattern, but also that a large number of UPs are indexed only in one database. Using a parallelism with the term used in biology we decided to term those UPs as 'endemic'. The full set of combinations showing the degree of overlap between the different databases can be found at: https://public.tableau.com/profile/publish/Frequencyofoverlapping/Sheet1#!/publish-confirm

**Table 2**. 'Endemic' University Presses

| Database | Number of 'Endemic UPs' | % respect total n of UPs | Composition (country of the UPs with higher frequency among 'endemic' Ups) | |
|---|---|---|---|---|
| | | | Country | Frequency |
| Finnish Lists | 146 | 42,2 | Russia | 19 |
| SPI | 65 | 50,0 | Spain | 39 |
| Norwegian Lists (CRISTIN) | 48 | 21,1 | China, Poland and United Kindgom | 5 |
| Book Citation Index | 24 | 27,0 | Czech Republic | 5 |
| Scopus | 5 | 6,2 | United States | 5 |

The large percentages of endemic UPs in the Finnish Lists and SPI strongly contrasts with the very low percentage of endemic UPs in the case of Scopus, while the lack of coincidence in the composition section of the table point towards a distinctive composition the 'endemisms' (Table 2).

**CONCLUSIONS**

Databases produced at the national level enable a clearer representation of the richness and diversity of book publishers relevant to the scholarly publishing activity of each country. National book publishers are better known, their catalogues are closer to the authors and those publishers might be more accessible for certain topics. Commercial databases might tend to be highly selective, thus possibly more prone to choose the least particular publishers of each country, region or language.

Also, some publishers with international scope and reach which might not be well known by experts in some countries. The direct association between presence in a given database and intrinsic quality should be avoided, being a combination of national and international classifications the best option for evaluation purposes.

The analysis of these data shows that there might be a close relation between the topics published and the readers of the works: local issues might play a significant role in national level book publishers, which does not necessarily imply a lower quality of the work. Nevertheless, the difference in composition and coverage and the analysis of the overlap of the various information systems shows that what the public and private databases cover is largely different.

The limited number of publishers covered by BCI and Scopus (both commercial databases) evidences their restrictiveness for assessment purposes. . Considering only the coverage towards UPs, commercial databases show a larger percentage, while in the case of the Spanish and Finnish systems the percentages are particularly low. Taking into account how the latter databases have been constructed, it can be concluded that scholars in these countries consider UPs less prestigious (Spain) or choose to publish less (Finland) in those publishers.

USA, UK and Canada UPs show high averages of presence among the different information systems (Figure 5): considering the bias towards English-publishing publishers in the commercial databases, UPs might occupy a similar position, in terms of recognition to publishers. Greater diversity can be observed among the three public databases concerning the number of countries with publishers in each database. Political and geographical influence

could be a factor contributing to this observation: Russian, Swedish, Polish and Danish UPs are covered in the Nordic countries' lists while the number of Spanish databases is low (3 in the Norwegian lists and 10 in the Finnish lists). Italy, Portugal and France are barely covered in BCI and not covered in Scopus and Finland, Sweden, Russia, Italy and Spain UPs are present in the public databases but are almost invisible for BCI or Scopus, the latter (being produced in The Netherlands) covering only two European UPs (the German Deutsche Universitäts Verlag and the Polish Deutsche Universitäts Verlag).
'Endemic' UPs could be indicative of the extent up to which local (in the nature of the topics published and / or in the languages of publication) research is relevant for each country.
Endemic' UPs in the case of Spain (SPI) are mainly Spanish while in the rest of the databases 'endemisms' are mainly from countries other than those where the headquarters of the database developers are.

**DISCUSSION**

The differences in the selection procedures among the five databases studied are considerable, as detailed in the introduction. Nevertheless, the composition and overlap of the databases shows a clear pattern when comparing the public databases with the private databases: in the countries where the public systems have been developed, including either publishers in which scholars have published their research or publishers which are considered the most prestigious, the diversity of publishers and the range of countries is wider than in the case of the private databases. The reasons for that observation are yet to be identified, since this study does not intend to provide a final set of explanations. Nevertheless, some ideas can be outlined. UPs indexed in all databases can be considered internationally recognized as relevant, since all the five databases are selective and imply recognition. Also, the larger diversity of publishers and countries in the public databases might correspond to the intrinsic features of research in SSH fields (often, of local interest, published in languages other than English; Hicks, 2004).
As main consequences of the analysis carried out, the sharp contrast between the coverage towards European UPs (with the exception of the UK and Germany) by the commercial and public databases can be useful in order to assess the suitability of both for assessment purposes: the latter might not be directly usable in national assessment processes. The linguistic bias of the commercial ones and the exclusion of large sets of European UPs is also a significant limitation for their use with assessment purposes for European outputs in SSH books..
The use of commercial databases which transparency is not equivalent to that of the public databases (although for understandable commercial reasons) should be the object of close scrutiny and analysis before taking a decision on its use. This is particularly relevant if the indicators provided by each database (citations in the case of the two commercial ones) are considered. Finally, since UPs are fewer than commercial databases in the public databases analyzed, the debate regarding the role of University Presses in each country, their publishing and business models might be still relevant.

**REFERENCES**

Abel, R. E. & Newlin, L. W. (2002). *Scholarly Publishing: Books, Journals, Publishers, and Libraries in the Twentieth Century*. US, John Wiley & Sons, Inc.

AAUP. *About university presses.* http://www.aaupnet.org/about-aaup/about-university-presses

AEUP (2015). http://www.aeup.eu/aeup/about/the-need-for-aeup/

Cerlalc/ UNESCO. 2014. *El espacio iberoamericano del libro 2014*. Available at http://cerlalc.org/publicacion/el-espacio-iberoamericano-del-libro-2014/

Federation of European Publishers (2014). *European Book Publishing Statistics 2014*. http://fep-fee.eu/European-Book-Publishing-741

Engels, T. C. E., Ossenblok, T. L. B., & Spruyt, E. H. J. (2012). Changing publication patterns in the Social Sciences and Humanities, 2000–2009. *Scientometrics*, 93, 373–390.

FGEE. 2014. *Comercio interior del libro*. Available at http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/libro/mc/observatoriolect/redirige/estudios-e-informes/elaborados-por-el-observatoriolect/sector-libro-abril2015/sector-libro-abril2015.pdf

Garfield, E. (1996, October). Citation indexes for retrieval and research evaluation. In: *Consensus Conference on the Theory and Practice of Research Assessment* (Vol. 7).

Giménez-Toledo, E.; Mañana-Rodríguez, J.;Engels, Tim C. E.; Ingwersen, P.; Siverten, G.; Verleysen, F. T. & Zuccala, A. A. (2016) Taking scholarly books into account: current developments in five european countries. *Scientometrics* (online first).

Gorraiz, J., Purnell, P.J., & Glänzel, W. (2013). Opportunities for and limitations of the book citation index. *JASIST*, 64(7), 1388–1398.

Hicks, D. (2004). The four literatures of social science. In: *Handbook of quantitative science and technology research* (pp. 473-496). Springer Netherlands.

Leydesdorff, L., & Felt, U. (2012). Edited volumes, monographs and book chapters in the Book Citation Index (BKCI) and Science Citation Index (SCI, SoSCI, A&HCI). *Journal of Scientometric Research*, 1(1), 28–34.

Michavila, Francisco (dir.). (2012). La Universidad española en cifras. Madrid: CRUE. http://www.crue.org/Publicaciones/Documents/UEC/LA_UNIVERSIDAD_ESPANOLA_EN_CIFRAS.pdf

Puuska, (2014). H.-M. Scholarly Publishing Patterns in Finland: A comparison of disciplinary groups. *Tampere University Press*.

PUF, 2015. Information available at: https://www.puf.com/Les_engagements_Puf#expertise-editeur

Sivertsen, G., and B. Larsen. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics*, 91(2), 567- 575.

Thompson, J. W. (2002). The Death of the Scholarly Monograph in the Humanities? Citation Patterns in Literary Scholarship. *Libri*, 52: 121–36.