

# Scholarly book publishers' ratings and lists in Finland and Spain. Comparison and assessment of the evaluative potential of merged lists

Jorge Mañana-Rodríguez & Janne Pölönen

# 1. Objectives

**First objective:** Compare the lists of scholarly publishes in SPI (Spain) and VIRTATA (Finland).

**Second objective:** Obtain information on the potential uses and / or interest of merged lists of publishers' ratings

## 2. Methodology (I)

1. – **Original datasets**
2. -- **A master list is created** with all the names of the publishers regardless of name variants and repetitions [Full lists of publishers' names on top of one another] Specific information is kept: ISBN Prefix if available in the original source and source name of the name variant as well as other relevant data (level of the publishers, etc).
3. -- **Cleansing in MS Excel:**
  - 3.1 -- *Automatic cleansing and preparation*

All name variants will be tracked in an adjacent column (or more): column A contains the original name and column B contains the correct name.

Duplicates / Additional / Special characters (non visible ASCII) are removed if present / All names are set in capital letters / Common abbreviations are removed when these do not alter the name of the publisher ( “.LTD” gets removed but ‘Verlag’ does not).

Publishers with ISBN prefixes are grouped together and VLOOKUP formula is used to identify the same ISBN for two names. These can be the same or different. If they are the same, the step ends. If they are different, both variants are kept in separate columns and the ISBN prefix queried in the Global Register of Publishers (GRP). The variant shown in GRP is kept as valid.

## 2. Methodology (II)

### 3.2 -- Manual cleansing:

The full dataset is alphabetically ordered according to the first column (master list).

A first visual inspection of the alphabetically ordered list a decision has to be made on the identified variants: only one will be kept. Several cases can occur:

- It is clear that the longest version of the name is the same as the other variants. The longest version, retaining the most information is kept.
- It is unclear that all the versions refer to the same publisher. In that case, for those with ISBN prefix, the name of the publisher is checked with the Global Register of Publisher (GRP). If it coincides, the longest variant is kept. If it does not, the variant in the GRP is kept.
- Several cases should be reviewed carefully: University presses (The University of Wisconsin Press / Wisconsin University Press / University of Wisconsin Press), acronyms (CSIC / Spanish National Research Council / Consejo Superior de Investigaciones Científicas) and variants starting with articles. The word publisher is a common source of error (Verlag, Wydawnictwa). These are manually reviewed and either the longest version (excluding the word 'publisher' and derivatives) or the name in the GRP is kept.

A final review involves selecting the most distinctive part of the name of a publisher (entropy measures could be used in an automated procedure) and searching all occurrences of that part in the rest of the list. If coincidence leads to match, then either GRPC or longest variant is kept. Else, the two denominations are kept.

## 2. Methodology (III)

### 4. – Search backwards

Once the two columns, the original name and the correct name are obtained, the rest of the information is processed (specific operation rules apply in each case) and a final list is obtained. It contains the original names for which no variant was identified and the repeated final names. Once deduplicated, the version is considered final.

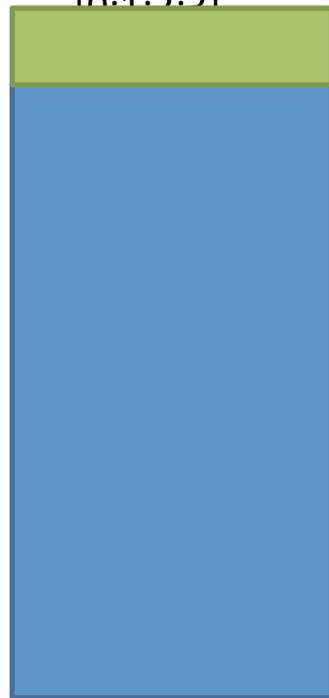
**Issues with ISBN prefixes:** not all ISBN prefixes are available in the original lists. The prefixes can be recovered, from the name, using GRP but for that purpose the correct name has to be identified. It makes the ISBN prefix search an 'educated guess'. Thus, despite an excellent tool when available, it is not always present.

**Issues with imprints and same institution publishers:** Imprints can have ISBNs different to those of the group they belong to (in example, a publisher is acquired by a publishing house in 2010: prior to 2010, the publisher would have its own ISBN and after that date it might or might not use the publishing house's ISBN prefixes). Imprints are kept as independent elements in the list except if they use ISBN's belonging to a publishing house, in which case they are included as part of the publishing house. The same applies to publishers from the same institution with the exception of departments: these are merged with the *mater* university.

## 2. Methodology (IV)

VIRTA: DISCRETE. 4 LEVELS

[0,1,2,3]



LEVEL 2 OR 3

=ELSE

SPI: CONTINUOUS. ICEE INDICATOR



=% LEVEL 2 OR 3 IN VIRTA

=% ELSE

### 3. Results (I). Core international publishers

Common publishers for the top 10 more prestigious publishers in SPI  
and top 10 publishers with more submissions to VIRTAsystem

<b>Publisher</b>	<b>Frequency of coincident fields (of publications in VIRTAsystem and prestige in SPI)</b>
Routledge (Francis & Taylor Group)	11
Oxford University Press	7
Springer	6
Cambridge University Press	4
De Gruyter	2
Ashgate Publishing (Ashgate Publishing Group)	1
Blackwell	1
John Benjamins	1

### 3. Results (II). Discrepancies in international publishers

Selection of publishers with the same and different levels in JUFO and SPI lists

<b>Publishers with level 2 in both lists</b>	<b>Publishers with level 2 in SPI and level 1 in JUFO lists</b>	<b>Publishers with level 2 in JUFO lists and level 1 in SPI lists</b>
Cambridge University Press	Peter Lang	Columbia University Press
Oxford University Press	McGraw Hill	Oxbow Books
Routledge (Francis & Taylor Group)	Wolters Kluwer	University Of California Press
Springer	Pearson	Honore Champion Editeur
Elsevier	Macmillan	Max Niemeyer
Blackwell	Giuffre	Multilingual Matters
De Gruyter	Cnrs	Suhrkamp
Sage	Armand Colin (Hachette Livre)	Peeters
Harvard University Press	L'Harmattan (Grupo L'Harmattan)	Mohr Siebeck

### 3. Results (IV). Divergence in national publishers

Cluster of more agreed-upon prestigious publishers in all fields and presence in VIRTAs lists.  
Spanish publishers in SPI lists.

<b>Publisher</b>	<b>SD</b>	<b>Number of fields in which it has been voted in SPI</b>	<b>Presence in VIRTAs lists</b>
Ariel (Grupo Planeta)	3.42	14	No
Síntesis	4.99	12	No
Biblioteca Nueva	5.41	13	No
Alianza (Grupo Anaya, Hachette Livre)	6.32	15	No
Tirant Lo Blanch	6.38	12	Yes
Marcial Pons	6.8	13	No
Siglo XXI De España (Akal)	7.62	15	No
Cátedra (Grupo Anaya, Hachette Livre)	7.94	14	No
Trotta	8.26	14	No
Akal (Akal)	9.1	15	No

### 3. Results (V). Divergence in national publishers

Publishers with activity in all fields in VIRTAs and presence in SPI lists.

<b>Publisher</b>	<b>Number of fields with output</b>	<b>Level</b>	<b>Presence in SPI lists</b>
Routledge	14	3	Yes
Gaudeamus	14	2	No
Springer	14	2	Yes
Palgrave Macmillan	14	3	Yes
Vastapaino	14	2	No
Tampere University Press	14	1	No
Peter Lang	14	1	Yes
Helsingin yliopisto	14	0	No
Ashgate	14	2	Yes
Lapland University Press	14	1	No
Cambridge University Press	14	3	Yes
Wiley-Blackwell	14	2	Yes
Sage publications	14	3	Yes

### 3. Conclusions

1. Overlap in the case of international publishers.
2. Limited extent of the overlap: relevant international publishers rated differently.
3. 'National publishers' show great disparity in terms of presence and rating.
4. Merged lists would provide potentially useful information for the rating of 'national' publishers in lists of other countries.
5. Merged lists would provide nuanced information on the rating in other evaluation systems of the international publishers.