# Outlining the Scientific Activity Profile of Researchers in the Social Sciences and Humanities in Spain: The Case of CSIC

Adrián A. Díaz-Faes[1], María Bordons[1] Thed van Leeuwen[2] and Mª Purificación Galindo[3]

[1] *adrian.arias@cchs.csic.es, maria.bordons@cchs.csic.es*
Quantitative Analysis in Science & Technology Group (ACUTE), IFS, Spanish National Research Council (CSIC), Albasanz 26-28, Madrid 28037 (Spain)

[2] *leeuwen@cwts.leidenuniv.nl*
CWTS-Centre for Science and Technology Studies, Leiden University, PO Box 905 2300 AX Leiden (the Netherlands)

[3] *pgalindo@usal.es*
Statistics Department, Salamanca University, Alfonso X El Sabio s/n, Salamanca 37007 (Spain)

## Abstract
Scientific activity of Social Sciences and Humanities researcher's comprises an assorted set of publication channels such as books, book chapters and national and international journal articles. Since knowledge dissemination in the field is characterised by a greater use of national journals and local languages, international bibliographic databases do not offer a suitable coverage. This work pursues to draw a comprehensive picture of the publication behaviour of CSIC researchers in the Social Sciences and Humanities from a micro-level perspective. For this purpose, Web of Science and an internal CSIC database called 'ConCiencia' were used along with a set of indicators describing the activity profile of researchers as well as the prestige of publication channels. Differences in the publication pattern of researchers in SSH were explored, and the relationship between their research performance and personal features such as age, gender and professional rank were analysed. In the Humanities, researchers with higher academic rank and age showed greater activity in books and non-WoS articles, whereas in the Social Sciences, higher rank was related to internationally-oriented scientific publications and a more collaborative activity. Considering only WoS articles would shrink meaningfully the visibility of CSIC researchers.

## Conference Topic
Science policy and research assessment

## Introduction
Outlining the scholarly work of researchers in the Social Sciences and Humanities (SSH) is often regarded as a challenge in bibliometrics, since the predominant publication types in these fields are not well covered by large bibliographic databases such as Web of Science or Scopus (Hicks, 2004). At this point, it is quite clear that dealing with journal publications, it is not enough for the SSH (Archambault et al., 2006; Sivertsen & Larsen, 2012) remaining books and books chapters as a major communication channel, chiefly in the Humanities. Moreover, due to the more local orientation of research in the SSH, knowledge dissemination in the field is characterized by a greater use of national journals and local languages (van Leeuwen, 2013). On the other hand, even though there has been a certain trend to consider SSH as a whole, different behavior between both communities can be expected (Mañana-Rodríguez & Giménez-Toledo, 2013).

The aforesaid factors hinder the potential capacity of the traditional bibliometric analyses to provide a reliable picture of the scientific activity of the SSH researchers and the development of national or regional databases to obtain full coverage of publications in the SSH has been suggested (Martin et al. 2010). This type of database has been developed in some countries such as Norway, Denmark, Finland and Belgium (Flanders), motivated by the need to monitor the performance of university scholars and in line with the development of performance-based

funding of university research (Sivertsen, 2010). Studying the activity of SSH researchers in Spain is difficult, because there is not such a full coverage national bibliographic database, but it can be addressed at the institutional level because many institutions collect the scientific output of their researchers, mainly with evaluative purposes.

This study focuses on the scientific activity of SSH researchers at the Spanish National Research Council (CSIC), the largest public institution dedicated to research in Spain which makes up more than 4,000 researchers and 125 institutes spread all over the country. This work pursues to draw a comprehensive picture of the publication behaviour of CSIC researchers in SSH from a micro-level perspective. An assorted set of publication channels such as books, books chapters, international and national journal articles are considered and specific indicators to assess the prestige of the different publication channels are introduced. Differences in the publication pattern of researchers in SSH are explored, and the relationship between their research performance and personal features such as age, gender and professional rank are analyzed.

## Methodology

This study analyses the scientific output of 268 active researchers in 2007 in the SSH area affiliated to the Spanish National Research Council (CSIC) and comprises both permanent researchers and postdoctoral research fellows. The time span under analysis is 2007-2011. Publications were collected from two different sources: Web of Science (WoS) (SSCI+AHCI+SCIE), which was used to download the more international articles; and an internal CSIC database called 'ConCiencia', to obtain other publication types not covered by WoS (books, books chapters and non-WoS journal articles). To cope with names inconsistencies and achieve a proper allocation of the publications to the researchers, different algorithms were used. A manual revision of the output collected, especially for the 'ConCiencia' database, was done. Based on the information retrieved, the following indicators were computed:

a) Activity profile of researchers

- % Books: proportion of books published by a researcher with regard to its total number of publications. In the same way, the next three indicators were calculated.
- % Book chapters.
- % WoS articles.
- % Non-WoS articles.
- Sum of publications: the total number of publications published by each researcher, including books, chapters in books and journal articles.
- Average number of authors/paper: this indicator measures the average number of authors per publication for the total output of a given researcher (WTI2, 2014).
- % International collaboration: share of the total output of each researcher co-authored with researchers affiliated with one or more foreign institutions.
- % English: proportion of a researcher´s output published in English.

b) Prestige of publication channels

- Top books and chapters (pptop10% Books & Chapters): proportion of books and chapters of a given researcher published by the top 10% publishers according to the Scholarly Publisher Indicators Project (SPI) (Giménez-Toledo, Tejada-Artigas & Mañana-Rodriguez, 2013). This project describes the Indicator of Quality of Publishers according to Experts (ICEE), which is based on a quality assessment of publishers rated by Spanish researchers in a national survey.

- Proportion of papers in first quartile journals (Q1): share of papers published in the top 25% journals of the impact factor journal ranking by subject category (source: Journal Citation Reports).
- Proportion of papers in top non-WoS journals (pptop10% non-WoS articles): % of non-WoS papers published in top journals according to the Integrated Scientific Journal Classification (CIRC) (Torres-Salinas et al. 2010). CIRC is a proposal for a categorization of journals in SSH developed by a group of experts in bibliometrics in Spain. It distinguishes four categories of journals (A, B, C and D) according to their visibility measured integrating the results of different journal classifications and assessments tools. For the purposes of this paper, "top journals" are those included in the categories "A" and "B".

**Table 1. Impact indicators for the different types of publication channels.**

| *Type of publication channel* | *Indicators of impact/prestige* |
|---|---|
| WoS articles | Impact factor (25% top journals by impact factor) |
| Non-WoS articles | CIRC (categories A and B) |
| Books/Book chapters | SPI (10% top publishers by expert opinion) |

c) Personal data: age, professional rank (P=postdoctoral research fellow, TS=tenured scientist, RS=research scientist and RP=research professor) and gender of researchers were provided by CSIC.

A preliminary inspection of the similarity between variables was explored by means of Multidimensional Scaling (MDS). Non-linear Principal Component Analysis (NLPCA) was used to explore the relationship between personal features of researchers and their performance. Statistical analyses were performed with SPSS (v.20).

## Findings

A total of 268 researchers had at least one publication in the period 2007-2011. In the whole SSH area, men represented 59% of all researchers, average age of researchers was 50 years old, and half of the researchers were in the lowest scientific category (tenured scientist). Postdoctoral research fellows accounted for only 7% of researchers in the area. Small differences between the Humanities and Social Sciences can be observed in Table 2.

**Table 2. Personal features and scientific rank of researchers in SSH.**

| | | *Humanities (N=192)* | | *Social Sciences (N=76)* | | *Total (N=268)* | |
|---|---|---|---|---|---|---|---|
| Gender | Men | 115 | 60% | 42 | 55% | 157 | 59% |
| | Women | 77 | 40% | 34 | 45% | 111 | 41% |
| Rank 2007 | Post-doc | 12 | 6% | 6 | 8% | 18 | 7% |
| | Tenured scientists | 98 | 51% | 42 | 55% | 140 | 52% |
| | Research scientists | 46 | 24% | 13 | 17% | 59 | 22% |
| | Research professors | 36 | 19% | 15 | 20% | 51 | 19% |
| Age | | 50 ± 9 (28-70) | | 49 ± 10 (32-70) | | 50 ± 9 (28-70) | |

Note: age expressed as average ± standard deviation (min-max).

A total of 3,004 documents were published by CSIC researchers in SSH during 2007-2011. Differences between Humanities and Social Sciences in the main publication types used are observed: WoS articles predominate in the Social Sciences while book chapters are the most frequent publication channel in the Humanities (Table 3).

**Table 3. Share of publication channels by area.**

| | *Books* | *Chapters* | *Non-WoS Articles* | *WoS Articles* | *Total* |
|---|---|---|---|---|---|
| Humanities | 14% (397) | 47 % (1,313) | 26% (717) | 13% (352) | 2,779 |
| Social Sciences | 8% (65) | 27% (214) | 29% (227) | 36% (289) | 795 |
| Total | 13% (462) | 43% (1,527) | 26% (944) | 18% (641) | 3,574 |

Note: the total is higher than 3,004, because the publication count is made at the individual level.

*Publication profile of researchers*

A MDS was applied to the set of variables which make up the activity profile of researchers to reveal their underlying structure. In terms of similarity, the plot gives away greater levels of international collaboration and English-written publications for WoS articles. The patterns for the remaining publications types (books, chapters and non- WoS articles) seems to be mainly related to higher levels of productivity and being written in national languages (Figure 1).

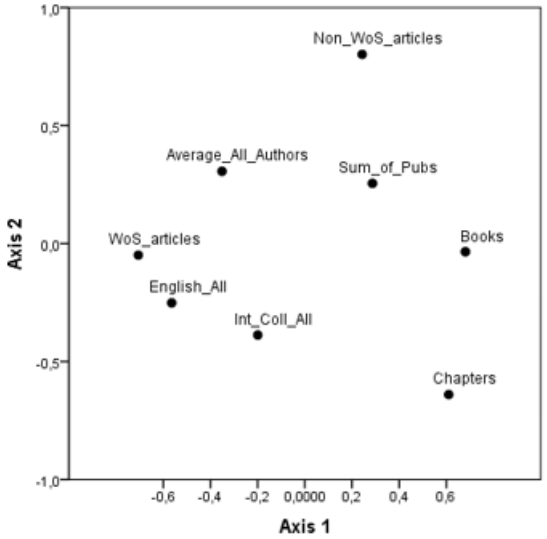


**Figure 1. MDS for the scientific activity profile.**

The diversity of publication channels in the output of researchers is the norm in SSH. Around 1/3 of the researchers presented output of the four different types considered: articles covered by WoS, non-WoS articles, books and book chapters. Three and two types of publication channels were observed in 40% and 17% of the researchers respectively, while only 12% of researchers had results of a single type. Several differences between Social Sciences and Humanities can be put forward: researchers who disseminate research among the four different types of publication channels considered are more frequent in Humanities (36% vs 24%), while using only WoS-covered journals is more common among Social Sciences researchers (16% vs 4). Finally, it is interesting to remark that around 22% of Social Sciences researchers and 41% of those in the Humanities may remain invisible in Web of Science-based studies since they do not show any publication covered by this database.

*Research performance of scientists*

Main statistics concerning research performance of scientists in SSH are shown in Table 4. A higher number of total publications is observed for researchers in the Humanities (15.1 vs 10.8), especially due to their high number of book chapters. Researchers in the Humanities exhibit a higher use of top publishers for books and chapters, while Social Sciences researchers present a greater share of articles in high impact factor journals.

**Table 4. Description of the research performance of researchers in SSH.**

| | *Humanities* | | *Social Sciences* | |
|---|---|---|---|---|
| | *Mean* | *SD* | *Mean* | *SD* |
| No. Books | 2.1 | 2.5 | 0.9 | 1.0 |
| No. Chapters | 7.1 | 5.7 | 2.9 | 3.3 |
| No. WoS Articles | 1.9 | 4.3 | 3.9 | 4.1 |
| No Non-WoS Articles | 3.9 | 4.7 | 3.1 | 3.7 |
| Sum of Publications | 15.1 | 12.2 | 10.8 | 7.5 |
| pptop10%_Books & Chapters | 35.9 | 26.5 | 23.7 | 28.9 |
| pptop10%_Non_WoS_Articles | 32.8 | 35.3 | 37.3 | 37.9 |
| % Q1 WoS Articles | 12.9 | 29.7 | 33.4 | 36.5 |
| Average number authors/publication | 1.7 | 1.4 | 2.6 | 1.1 |
| % International. collaboration | 16.9 | 23.0 | 24.1 | 29.8 |
| % English | 14.0 | 19.0 | 38.6 | 32.7 |

To explore the possible relations between personal features of researchers and their performance NLPCA was used, which allows reducing a large number of variables to a smaller number of uncorrelated non-linear combinations of these variables with miminum loss of information (principal components). Two different studies are conducted, since researchers in Social Sciences and Humanities are analysed separately. Preliminary results concerning the plots of component loadings (two-dimensional solution) are shown in Figure 2.

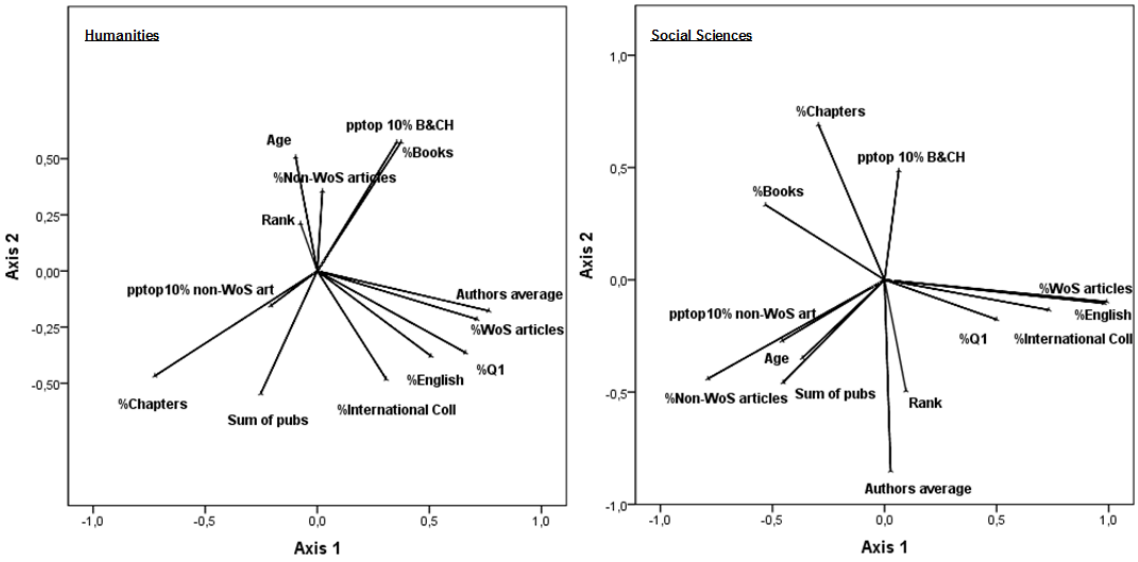


**Figure 2. Component loadings in: a) Humanities; b) Social Sciences.**

Note: only researchers with 2 or more publications considered

## Discussion and conclusions

At this point, some preliminary results can be pointed out in an attempt to provide a comprehensive picture of the activity of CSIC researchers in SSH from a micro-level perspective:

- Taking into account only WoS articles would shrink meaningfully the visibility of CSIC researchers in SSH, in particular in the Humanities.
- Different constraints of the 'ConCiencia' system are identified. More rigour in the input of data (carried out by researches themselves) as well as in the cleaning and validation processes (by the institution) would be advisable.

- In the Humanities, researchers who hold a higher rank and age present greater activity in books and non-WoS articles. However, a high number of total publications is apparently not associated to a higher rank.
- In the Social Sciences, a higher academic rank is associated to internationally-oriented scientific publications (high share of WoS articles) as well as a high productivity (high number of publications) and collaborative activity (high number of co-authors).
- Differences between the Social Sciences and Humanities are observed, but even within each of these fields different typologies of researchers according to their publication pattern, collaboration practices and international/national orientation may exist. These factors are being explored at present.
- Although our study focuses on four different types of academic output, it is still not comprehensive, since it does not consider the non-scholarly literature, which may have an important societal impact.

## Acknowledgments

Adrián A. Díaz-Faes is granted with a JAE predoctoral fellowship by the Spanish National Research Council (CSIC).

## References

Archambault, E., Vignola-Gagne, E., Cote, G., Lariviere, V. & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. Scientometrics, 68(3), 329–342. doi:10.1007/s11192-006-0115-z.

Giménez-Toledo, E., Tejada-Artigas, C. & Mañana-Rodríguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: results of a survey. *Research Evaluation*, 22(1), 64-77.doi:10.1093/reseval/rvs036

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems* (pp. 473–496). Dordrecht, The Nederlands: Kluwer Academic. doi:10.1007/1-4020-2755-9_22.

van Leeuwen, T.N. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie - Praxis und Forschung*, 2013, 1-18. Retrieved June 15, 2015 from: http://www.bibliometrie-pf.de/article/viewFile/173/215

Mañana-Rodriguez, J. & Giménez-Toledo, E. (2013). Scholarly publishing in social sciences and humanities, associated probabilities of belonging and its spectrum: a quantitative approach for the Spanish case. *Scientometrics*, 94, 893-910. doi:10.1007/s11192-012-0838-y

Martin, B., Tang, P., Morgan, M., Glanzel, W., Hornbostel, S., Lauer, G., et al. (2010). Towards a bibliometric database for the social sciences and humanities—A European scoping project. Research report produced for DFG, ESRC, AHRC, NWO, ANR and ESF.

Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter*, 6(1), 22–28.

Sivertsen, G. & Larsen, B. (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. doi:10.1007/s11192-011-0615-3.

Torres-Salinas, D, Bordons, M., Giménez, E., Delgado, E., Jiménez, E. & Sanz, E. (2010). Clasificación integrada de revistas científicas (CIRC): propuesta de categorización de las revistas en ciencias sociales y humanas. *El Profesional de la* Información, 19(6), 675-683.

WTI2 (2014). Scholarly publication patterns in the social sciences and humanities and their relationship with research assessment. *Science, Technology & Innovation Indicators 2014. Thematic paper 2*, 1-26 Retrieved June 15, 2015 from: http://dialogic.nl/documents/other/sti2_themepaper2.pdf