Highly Visible Science: A Look at Three Decades of Research from Four Iberamerican Countries[1]

Jane M. Russell*, J. Antonio Del Río** and Héctor D. Cortés**

*jrussell@servidor.unam.mx
Centro Universitario de Investigaciones Bibliotecológicas, Torre II de Humanidades piso 11, Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, 04510 México, DF (Mexico)

**antonio@servidor.unam.mx, hdcg@cie.unam.mx
Centro de Investigación en Energía, Universidad Nacional Autónoma de México (UNAM), Apartado Postal 34, Temixco, Morelos (Mexico)

Abstract
The present research-in-progress illustrates how a simple analysis of high-ranking journals can give valuable information on the scientific visibility of countries, especially those from developing regions. Latin American research in the multidisciplinary journals *Nature* and *Science* was determined by analysing the papers published from 1973 to November 2004, by the region's three most productive countries, Argentina, Brazil and Mexico, and comparing these with Spain. A bibliometric and data mining software was developed to analyse corresponding records from the Web of Science with respect to the following parameters: total number of papers, publications by year and by author, collaborations by country and by institution, times cited, cited journals, and significant words in the abstracts of the papers. Preliminary results with respect to some of these parameters show the usefulness of the method developed.

Introduction
A special concern for developing countries is the extent to which their science is visible to the international scientific community and, especially, how much of their contribution to global science can be considered at the cutting edge of international research. Of all mainstream journals, the two multidisciplinary, weekly journals, *Nature* and *Science* are the best-known and most widely read. The specific criteria for publication of scientific papers in *Nature* are two-fold, that they be of *outstanding scientific performance* and *reach a conclusion of interest to an interdisciplinary readership.* Only about 10% of papers received are published hence selection criteria are rigorous (Getting published in *Nature,* 2004). In the case of *Science,* the guidelines for authors state that p*riority is given to papers that reveal novel concepts of broad interest* and that *competition for space in* Science *is keen* (Information and help for *Science* authors, 2004).

We can therefore consider that papers published in these two journals reach a far broader audience, both within and outside the international scientific community, than those published in specialist journals. Thus publication in these two widely-read journals can be taken as a measure of the international visibility of the research performed by countries such as those from Latin America, as well as identifying research work that is at the forefront of its field. In the present paper we develop a straightforward method for rating the scientific visibility of countries, using publication in *Science* and *Nature* as indicators.

Method
Taking three Spanish and one Portuguese-speaking countries, Spain, Argentina, Brazil and Mexico, we analysed all publications in the multidisciplinary journals, *Nature*[2] and *Science* from 1973 to November 2004. Full records including abstracts, references and citation data, were downloaded individually for each country from the *Science Citation Index Expanded* using the Web of Science, adopting the following strategy: source (Nature AND Science) AND address, either Spain, Argentina, Brazil or Mexico NOT New. Data were worked using software developed for bibliometric and data mining analysis at the UNAM's Centre for Energy Research (CIE). Groups of records for each country

[1] The work presented in this paper was supported by project FOMIX CONACyT-Morelos 9250.
[2] Only papers in the parent journal, *Nature*, were included in the analysis.

Proceedings of ISSI 2005, pages 256–259.
Edited by P. Ingwersen and B. Larsen.

can be analysed with respect to the following variables: Total production and its distribution over time, patterns of international collaboration, most productive authors and institutions, citation frequencies, and cited journals.

A text mining technique can be used (Ortuno *et al.*, 2002) to find the most relevant words occurring in the abstracts of papers of all four countries. A normalised standard deviation higher than 1 indicates that the distribution of the word within a particular abstract is not random allowing us to determine which words or string of words can be considered relevant for that particular text. The advantage of this technique is that it does not require a labour-intensive revision of individual words to extract the keywords from a text but rather provides a ready-made list of the most frequently occurring words and strings of words whose distribution within a text is not random and, therefore, likely to be significant.

Preliminary Results

The total numbers of papers in both journals were as follows: Spain, 625, Brazil, 322, Mexico, 202 and Argentina, 116. Both Spain and Brazil published over 70% of their total papers in *Nature*, while the production of Mexico and Argentina were more evenly distributed between the two titles. All four countries showed an increasing aggregate presence in the two journals over the 30 years analysed, as can be seen in Figure 1. Argentina, Brazil and Mexico first got into double figures after 1995. Spain showed a notable rise in publications from 1993 onwards even though a steady rise was apparent from the middle of the 80s coinciding with her entry into the European Union.

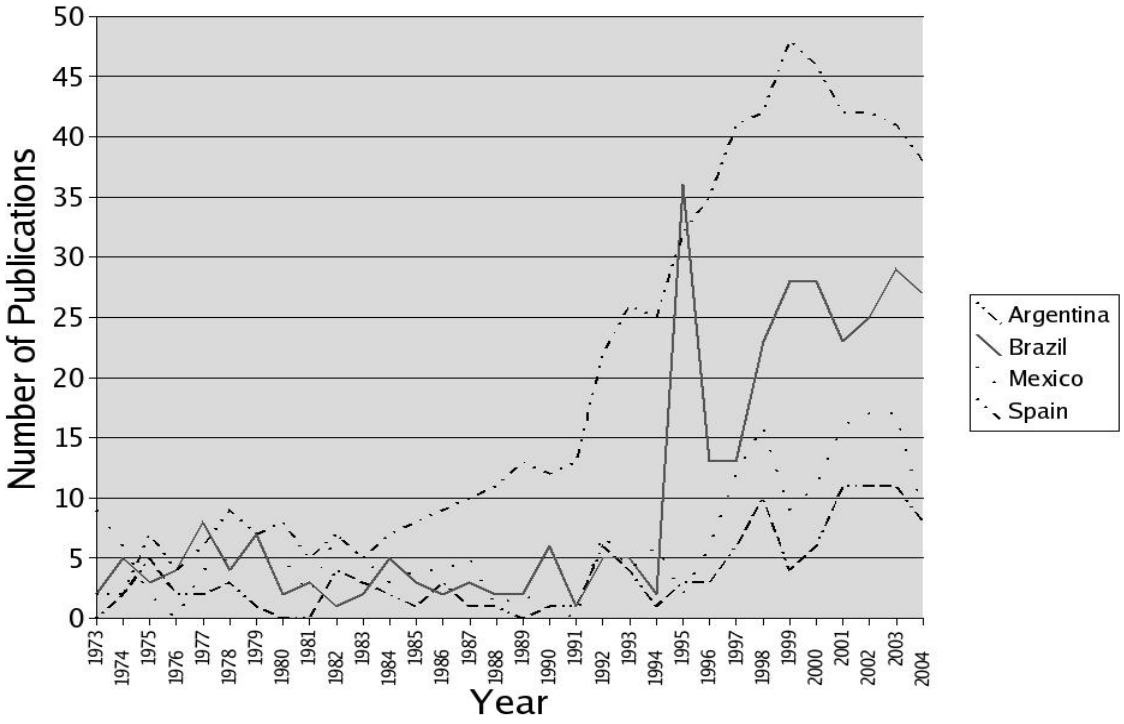

Figure 1: Publications by country per year

An analysis of the number of times the publications from each country were cited in subsequent years is given in Figure 2. Almost 28% of Mexican papers remained uncited, the corresponding figure for Argentina was 20%, for Brazil, 16% and for Spain, 15%. Fewer than 28% of publications from Argentina, 28% from Brazil, 25% from Mexico and 20% from Spain received between 1 and 9 citations. The most frequent citations levels greater than 9 were between 10 and 19 for Mexico, from 30-49 for Argentina, and from 50-99 for both Brazil and Spain. The highest level of citation, between 200 and 3000, was achieved by almost 9% of papers from Spain, around 5% from both Brazil and from Mexico and only 3% of Argentine publications.

The two most highly cited articles receiving more than 2000 citations, are on the subject of the genome, one of E. coli and the other the human genome, with Mexican and Spanish coauthors,

respectively. The most highly cited publications in the group of Argentine and Brazilian publications received less than 1000 citations. Two Spanish articles also in the field of genetics obtained more than 1000 citations.

An analysis of the journal titles most frequently cited in the reference lists of a group of papers gives us an indication of the main subject areas of these publications. In the present study, both *Nature* and *Science* are, perhaps not surprisingly, the most frequently cited journals, *Nature* more so than *Science* (Figure 3). Results show general agreement with the distribution patterns of the publications of the four countries between these two journals with the exception of Mexico which published more papers in *Science* than in *Nature* but showed a slightly greater inclination to cite papers from *Nature* than from *Science*. *Physical Review Letters* was the next most frequently referenced journal in the publications of Argentina, Brazil and Spain but not of Mexico. Other frequently cited titles were principally in the areas of astronomy and astrophysics (with the exception of Argentina), geophysics (predominantly Mexico but also Brazil) and, in the case of Spain only, in molecular biology. Papers from all four countries cited the multidisciplinary proceedings of the US National Academy of Sciences.

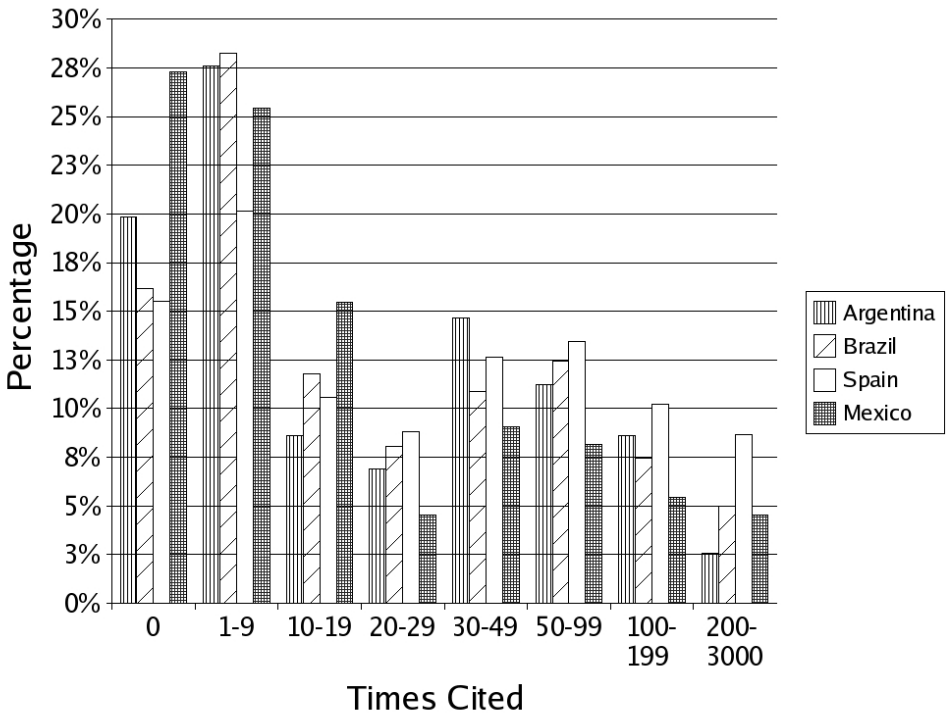

Figure 2: Citation distribution by country

The most significant single words found in the abstracts from Argentina were: *quantum, proteins, biodiversity, ice, jurassic,* and *larsen*, indicating the presence of papers in physics, biology, and in the geosciences. Analysis of strings of three words revealed important occurrences of the following phrases; *the larsen ice shelf, the southern hemisphere,* and *the australian continent,* revealing a concentration of papers published in geosciences of the southern hemisphere.

The most relevant words in the Brazilian papers were *fire, leprosy, stars, galaxies,* and *clusters. Gene expression, the amazon, forest fires, tree mortality* were common relevant occurrences. However, significant phrases were few and revealed no specific research topics. Our findings suggest that leprosy is a big concern for the Brazilians, as well could be the subject of forest fires. A general interest in astronomy is also apparent.

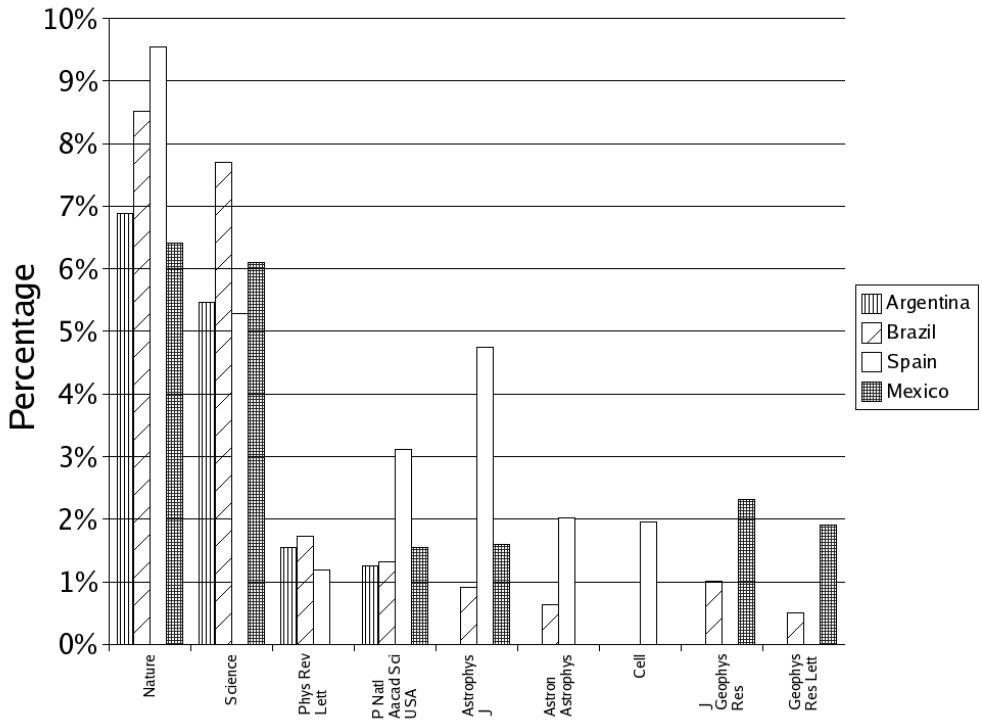

Figure 3: Most highly cited journals.

In the case of Mexico, *galaxies, nanotubes, chicxulub*[3]*, genomic,* and *planetary* showed standard deviations greater than 1 indicating three main research thrusts: astronomy, molecular biology, and geosciences. When looking at the phrases, the following themes became apparent: *cretaceous tertiary k t boundary, biological nitrogen fixation, and the chicxulub structure.*

Ozone, insulin, quantum, artic, and *mammals* were relevant words in the Spanish abstracts. *Ozone loss, stem cells, planetary nebulae, black hole, x ray emission* conformed relevant phrases corresponding to the general areas of environmental sciences, biological sciences, and astronomy. However, no particular research theme emerged from the present analysis suggesting that Spanish research in many branches of science is present in these mainstream journals.

Discussion

The present study has demonstrated the usefulness of analysing publications in the two influential, multidisciplinary journals, *Nature* and *Science*, to determine a country's high international scientific profile. Using this method we are able to identify trends in publications, collaborations and citations, as well as prominent institutions and actors, and main research thrusts. The validity of our analysis rests on the assumption that the scientific work published in *Nature* and *Science* effectively represents the advancement of scientific knowledge in the global scenario.

References

Getting published in *Nature* (2004). Retrieved December 10, 2004 from: http://www.nature.com/nature/submit/get_published/index.html

Information and help for Science *authors* (2004). Retrieved December 10, 2004 from: http://www.sciencemag.org/feature/contribinfo/home.shtml

Ortuno, M., Carpena, P., Bernaola-Galvan, P., Muñoz, E. & Somoza, A.M. (2002). Keyword detection in natural languages and DNA. *Europhysics Letters,* 57, 759-764.

[3] A crater created by the impact of a meteor in the Gulf of Mexico.