International Comparative Performance of the UK Research Base - 2011

A report prepared for the Department of Business, Innovation and Skills.

APPENDICES C-E

ELSEVIER

CONTENTS

APPENDIX

Appendix C: Data Sources

Data conventions
UK standard usage for the term 'billion', which is defined as one thousand million (10^9), is adhered to throughout this report.

Association of University Technology Managers
(AUTM; http://www.autm.net/Home.htm)

AUTM provides annual surveys of university licensing activity in the US and Canada. In addition to statistics on IP-generated income, the surveys contain data on patent applications, granted patents, spin-offs formed and spin-off survival.

Higher Education Statistics Agency
(HESA; http://www.hesa.ac.uk/)

HESA provides accurate statistics on all UK HEIs including, for example, post-graduate statistics. HESA's list of HEIs per constituent country is particularly useful in addressing article performance indicators for each constituent country separately.

HEFCE's Higher Education-Business and Community Interaction
(HE-BCI; http://www.hefce.ac.uk/pubs/hefce/2011/11_25/)

HE-BCI survey database contains statistics on patent applications, IP income, spin-offs formed and spin-off survival for all UK HEIs – data from 1999-2000 to latest 2008-2009, annually.

The Organisation for Economic Co-operation and Development
(OECD; http://www.oecd.org/home/)

The OECD collects internationally comparable data on research and development. Two of its databases, the Main Science and Technology Indicators (MSTI 2010-2) and Education and Training Database (ED 2008), were used as the source of data to generate many of the indicators presented in this report. Extensive notes on the collection and usage of these OECD data are provided in the report footnotes, and the reader is referred to these resources for details. A useful history of the development of the OECD's R&D statistics is available in an article by Benoît Godin[1].

Data are presented for the most recent five years for which data are available, though some countries may lack data for certain years. In such cases, missing values have been filled by interpolation or forecasted by extrapolation using established methods. Financial data are given in constant US$ at 2000 prices and corrected for Purchasing Power Parity (PPP), allowing comparability over time and between countries. Full-Time Equivalent (FTE) counts are used for all human capital data in this report.

OECD Main Science and Technology Indicators
(OECD-MSTI; www.oecd.org/sti/msti)

This biannual publication provides a set of indicators that reflect the level and structure of the efforts undertaken by OECD Member countries and nine non-member economies in the field of science and technology. The indicators cover the resources devoted to research and development, technology and international trade in R&D-intensive industries.

ProTon Europe
(www.oecd.org/sti/msti)

Since 2003 ProTon Europe has run a comparative meta-survey of the activities of Knowledge Transfer Offices (KTOs) in European universities and other Public Research Organisations (PROs). The latest available survey was for fiscal year 2008.

ScienceDirect
(http://www.sciencedirect.com/)

ScienceDirect is an Elsevier-owned full-text scientific platform covering over 2,000 journals across various disciplines. With a large customer base (12,000 institutions worldwide, more than 11 million active users and over 600 million full-text article downloads in 2010), ScienceDirect.com usage analytics data have been aggregated by country, region and discipline. The usage statistics from ScienceDirect are COUNTER[2] compliant and also adhere to new initiatives such as SUSHI[3]. Because of its large customer base, ScienceDirect.com usage analysis provides a different look at performance measurement.

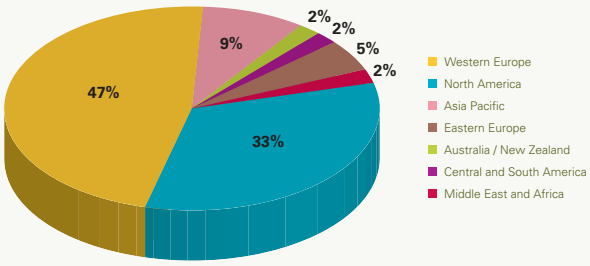
[1] Godin, B. (2008) "The Culture of Numbers: Origins and Development of Statistics on Science, Technology and Innovation" *Project on the History and Sociology of S&T Statistics*, Working Paper No. 40, Canadian Science and Innovation Indicators Consortium.
[2] For more information about COUNTER (Counting Online Usage of Networked Electronic Resources) see http://www.projectcounter.org.
[3] For more information about SUSHI (Standardized Usage Statistics Harvesting Initiative) see http://www.niso.org/workrooms/sushi.

Scopus

(http://www.scopus.com/home.url)

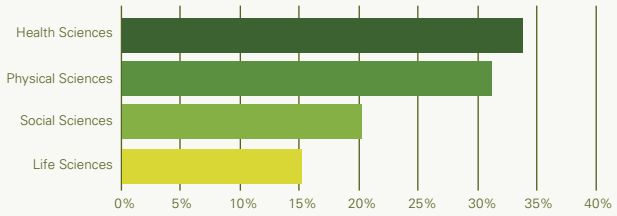
Scopus is the largest abstract and citation database of research literature available, with abstracts and citation information from more than 45 million scientific research articles in 18,000 peer-reviewed journals published by over 5,000 publishers spanning all science sectors. Scopus covers more than 3,000 publications in the fields of arts & humanities and approximately 5,900 titles from North America, 8,400 from Europe and 2,800 from Asia-Pacific and 800 from Latin-America and Africa. Geographical regions covered (including non-English titles where an English abstract is available) are shown in *Figure 1.*

Figure 1 Percentage of journals in Scopus based on geographical regions (July 2010).



Approximately 21% of titles in Scopus are published in languages other than English (or published in both English and another language). Subject coverage is extensive across the sciences, technology, and medicine as well as social sciences and arts & humanities (*Figure 2*). Titles which are covered are predominantly serial publications (journals, trade journals, book series and conference material), but considerable numbers of conference papers are also covered from stand-alone proceedings volumes (a major dissemination mechanism, particularly in the computer sciences). While a great deal of important literature in all fields (but especially in the social sciences and arts & humanities) is published in books, there are many challenges to the inclusion of such materials in abstracting and indexing databases. For this reason, Scopus does not currently index books.

Figure 2 Percentage of titles in Scopus by subject area (July 2010). The total percentages add up to over 100% because titles may be attributed to more than one subject category.



For this report, a static version of the Scopus database covering the period 1996-2010 inclusive was extracted in May 2011 and aggregated by country, region, and subject. Subjects were defined by BIS for comparative purposes as follows: clinical sciences; health & medical sciences; biological sciences; environmental sciences; mathematics; physical sciences; engineering; social sciences; business; humanities. To account for indexing delays, 2010 article counts (and all indicators derived from this data-point) were corrected by multiplying by 1.05. When aggregating article and citation counts, an integer counting method was employed where, for example, a paper with two authors from a UK address and one from a French address would be counted as one article for each country (i.e. 1 UK and 1 France). This method was favoured over fractional counting, in which the above paper would count as 0.67 for the UK and 0.33 for France, to maintain continuity with previous reports in this series.

A body of literature is available on the limitations and caveats in the use of such 'bibliometric' data, such as the accumulation of citations over time, the skewed distribution of citations across articles, and differences in publication and citation practices between fields of research, different languages, and applicability to social sciences and humanities research. In social sciences and humanities, the bibliometric indicators presented in this report for these fields must be interpreted with caution because a reasonable proportion of research outputs in such fields take the form of books, monographs and non-textual media. As such, analyses of journal articles, their usage and citation, provides a less comprehensive view than in other fields, where journal articles comprise the vast majority of research outputs.

SciVal Spotlight
(http://info.spotlight.scival.com/)

SciVal Spotlight is an analytical tool that has been implemented on the basis of Scopus data. Through a co-citation algorithm, the tool shows articles clustered by citation behaviour rather than by pre-set categories. See Appendix E for a more detailed discussion of this approach.

World Intellectual Property Organisation data and definitions
(WIPO; http://www.wipo.int/freepublications/en/intproperty/941/wipo_pub_941_2010.pdf)

WIPO collects internationally comparable data on patenting activity. Its recent report entitled "World Intellectual Property Indicators 2010[4]" provides access to datasets on patent applications and was used as the source of data to generate the patent indicators presented in this report. Extensive notes on the collection and usage of these data are provided elsewhere[5], and the interested reader is referred to these resources for details.

[4]Economics and Statistics Division, WIPO 2010; Publication n°: 941. Updated January 2011.
[5]WIPO Patent methodological information. Wipo Patent Information available at http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/patent_stats_methodology.pdf

Appendix D: Countries Included in Data Sources

Country	ISO 3-character code	G7	G8	EU27	OECD[6]
Australia	AUS				☑
Austria	AUT			☑	☑
Belgium	BEL			☑	☑
Bulgaria	BGR			☑	
Brazil	BRA				☑
Canada	CAN	☑	☑		☑
Chile	CHL				☑
China	CHN				☑
Cyprus	CYP			☑	
Czech Republic	CZE			☑	☑
Denmark	DNK			☑	☑
Estonia	EST			☑	☑

Country	ISO 3-character code	G7	G8	EU27	OECD[6]
Finland	FIN			☑	☑
France	FRA	☑	☑	☑	☑
Germany	DEU	☑	☑	☑	☑
Greece	GRC			☑	☑
Hungary	HUN			☑	☑
Iceland	ISL				☑
Ireland	IRE			☑	☑
India	IND				☑
Israel	ISR				☑
Italy	ITA	☑	☑	☑	☑
Japan	JPN	☑	☑		☑
South Korea	KOR				☑

[6]OECD+ includes OECD members and others such as China, India, etc.

Country	ISO 3-character code	G7	G8	EU27	OECD[6]
Latvia	LVA			☑	
Lithuania	LTU			☑	
Luxembourg	LUX			☑	☑
Malta	MLT			☑	
Mexico	MEX				☑
The Netherlands	NLD			☑	☑
New Zealand	NZL				☑
Norway	NOR				☑
Poland	POL			☑	☑
Portugal	PRT			☑	☑
Romania	ROM			☑	
Russia	RUS		☑		

Country	ISO 3-character code	G7	G8	EU27	OECD[6]
Slovakia	SVK			☑	☑
Slovenia	SVN			☑	☑
Spain	ESP			☑	☑
Sweden	SWE			☑	☑
Switzerland	CHE				☑
Turkey	TUR				☑
United Kingdom	GBR (UK used throughout this report)	☑	☑	☑	☑
United States	USA	☑	☑		☑

Appendix E: Elsevier Methodology

For all bibliometric analysis, only the following document types are considered: Article (ar), Review (re) and Conference Proceeding (cp).

For all bibliometric analysis, a "citation rooftile" approach has been applied. This employs a sliding 5-years publication and citation window. For example: the citation rooftile 2006-10 considers citations received in the period 2006-10 inclusive to all articles published in the same period, 2006-10. The same concept applies to the rooftiles 2005-2009, 2004-2008, etc.

Indicators

Research Quantity and Quality Indicators

Performance Indicator	Description	Data sources
Publication output	Number of publications per country where at least one author from that country figures among the authors listed	Scopus 1996-2010
Publication output per researcher	Number of publications from a country relative to the number of researchers in that country	Scopus 1996-2010 OECD MSTI 2010
Share of publication output	Global, G8, OECD+ and EU27 share of publications of a country	Scopus 1996-2010
Single authorship output	Number of publications per country with one single author in the publication's author list	Scopus 1996-2010
Domestic co-authorship output	Number of publications per country with multiple authors from different institutions within that country	Scopus 1996-2010
International co-authorship output	Number of publications per country with multiple authors from multiple countries, including the country in question	Scopus 1996-2010
Patent output	Number of patents per country assigned to an inventor from that country	WIPO
Citation output	Citations received for the publications of a particular country	Scopus 1996-2010
Citation output per researcher	Citations received for the publications of a country relative to the number of researchers in that country	Scopus 1996-2010 OECD MSTI 2010
Share of citation output	Global, G8, OECD+ and EU27 share of citations received by a country's publications	Scopus 1996-2010
Citation distribution	A geographic distribution noting the country of origin of the citing paper (e.g. Which countries are citing UK publications)	Scopus 1996-2010
Cited vs. non-cited output	• Number of publications from a country receiving citations • Number publications from a country that do not receive citations • Proportion of cited and non-cited publications from a country	Scopus 1996-2010
CPP	Average citation per publication from a country	Scopus 1996-2010

Research Quantity and Quality Indicators

Performance Indicator	Description	Data sources
Field weighted citation impact	Discipline-weighted citations-per-publication of a country relative to global discipline-weighted citations-per-publication	Scopus 1996-2010
Citation-based Centiles	Number of publications a country has in the top 1%, 5% and 10% most-cited publications in World, G8, OECD+ and EU27	Scopus 1996-2010
Single authorship citation output	Number of citations received by single-authored publications from a country	Scopus 1996-2010
Domestic co-authorship citation output	Number of citations received by domestic co-authorship publication output	Scopus 1996-2010
International co-authorship citation output	Number of citations received by internationally co-authored publications	Scopus 1996-2010
Downloads output	Number of downloads of publications from a country	Scopus 1996-2010 ScienceDirect.com usage statistics 2002-2010
Share of downloads	Global, G8, OECD+ and EU27 share of downloads of publications from a country or set of countries	Scopus 1996-2010 ScienceDirect.com usage statistics 2002-2010
Downloads vs. citations	Number of downloads of publications from a country relative to number of citations received by publications of that country	Scopus 1996-2010 ScienceDirect.com usage statistics 2002-2010

Human Capital Indicators

Performance Indicator	Description	Data sources
Students and PhDs	• Number of students per country • Number of PhDs (enrolled and graduates) per country	OECD, UNESCO
Students and PhD vs. researchers	• Number of students in a country relative to the number of researchers in that country • Number of PhDs (enrolled and graduates) in a country relative to the number of researchers in that country	OECD, UNESCO
Researchers	Number of researchers per country	OECD, UNESCO
Researchers vs. workforce	Number of researchers in a country relative to the total workforce in that country	OECD, UNESCO
R&D workers	Number of R&D workers per country	OECD, UNESCO
R&D workers vs. researchers	Number of R&D workers in a country relative to the number of researchers in that country	OECD, UNESCO
Researchers inflow	Number of new[7] researchers moving to an organisation in a specified country	Scopus 1996-2010
Researchers outflow	Number of researchers previously affiliated to an organisation in a specified country, who have since emmigrated	Scopus 1996-2010
Researchers repatriation	Number of researchers returning[8] to a country where they have previously published	Scopus 1996-2010

[7] Researchers who have not worked before in the country in question.
[8] Researchers who have started their career in a country, left that country and then returned to it.

Research Productivity Indicators

Performance Indicator	Description	Data sources
Publications vs. GDP, GERD, HERD, BERD and GOVERD	Number of publications from a country relative to the GDP, GERD, HERD, BERD and GOVERD of that country	Scopus 1996-2010 OECD MSTI 2010
Publications vs. researchers, PhDs	Number of publications from a country relative to the number of researchers or PhDs in that country	Scopus 1996-2010 OECD MSTI 2010 UNESCO
Citations vs. GDP, GERD, HERD, BERD and GOVERD	Number of citations received by publications from a country relative to the GDP, GERD, HERD, BERD and GOVERD of that country	Scopus 1996-2010 OECD MSTI 2010
Citations vs. researchers, PhDs	Number of citations received by publications from a country relative to the number of researchers or PhDs in that country	Scopus 1996-2010 OECD MSTI 2010 UNESCO
Researchers vs. GDP, GERD, HERD, BERD and GOVERD	Number of researchers in a country relative to the GDP, GERD, HERD, BERD and GOVERD of that country	OECD MSTI 2010 UNESCO
Students vs. GDP, GERD, HERD, BERD and GOVERD	Number of students in a country relative to the GDP, GERD, HERD, BERD and GOVERD of that country	OECD MSTI 2010 UNESCO

Knowledge Transfer Indicators

Performance Indicator	Description	Data sources
Academia to Corporations	Number of researchers moving from an academic to an industrial organisation(e.g. corporation) per country; both domestically and internationally	Scopus 1996-2010
Corporations to Academia	Number of researchers moving from an industrial (e.g. corporation) to an academic organisation per country; both domestically and internationally	Scopus 1996-2010
Academia/ Corporations publication output	Total number of publications of a country co-authored[9] by academic and corporate researchers.	Scopus 1996-2010
Academia/ Corporations citation output	Citations-per-publication from a country co-authored by academic and corporation researchers	Scopus 1996-2010
GERD by funding source	GERD by HERD, BERD and GOVERD per country	OECD MSTI 2010

[9] Academic and corporate authors are both in the same country in question or one of them is from the country in question.

Human Capital Methodology

Methodologically, studies of international researcher mobility have relied on census or migration data[10], interviews[11], surveys of researchers, or more recently, CV-based analysis[12]. The use of address data from publication databases has only recently been employed, often in conjunction with other data[13]. The measurement of international researcher mobility by co-authorship in the published literature is complicated by the difficulties involved in teasing out long-term migration from short-term mobility (such as doctoral research visits, sabbaticals, secondments, etc.), which might be deemed a form of collaboration. The approach presented here uses Scopus author profile data to derive a history of an author's affiliations recorded in their publications and to assign them to mobility classes defined by the type and duration of observed moves. Author nationality is not captured with article or author profiling data, and there are serious methodological difficulties in using cultural indicators (such as family names) as a proxy for nationality of birth[14]. Therefore, in this study, authors have been assumed to be from the first country from which they have published (for migratory mobility), or from the country where they published the majority of their articles (for transitory mobility). These criteria may in individual cases result in authors being assigned migratory patterns that may not accurately reflect the real situation, but such errors may be assumed to be evenly distributed across the groups and so the overall pattern remains valid.

It is a limitation of this analysis that mobility is determined solely from publication profiles and only for relatively productive researchers (more than 1 paper between 1996 and 2010), so that researchers without sufficient publications are not included. Work published in recent years has demonstrated that net rates of researcher migration may mask trends in fields or specialties with inherently lower publishing activity[15]. Moreover, it is clear that (at least amongst elite researchers), that it is not the current elite that migrates but those who will go on to become the next generation elite (who are attracted to move and co-locate with the current elite). Other work has shown that it is typically junior researchers that move internationally[16] which suggests that a proportion of UK researcher mobility is not reflected in this analysis.

A longstanding problem in studies of researcher productivity and mobility has been the unambiguous identification of the individual[17]. With common family names in every language and country, such as Smith, Wang, and Lee, and multiple variants of a given person's name in the published literature (e.g. Lewis, M; Lewis, M.J; Lewis, Michael) this is significant problem and one that has the potential to make research assessment very difficult.

[10] Johnson, J.M. & Regets, M.C. (1998) "International Mobility of Scientists and Engineers to the United States–Brain Drain or Brain Circulation?" *NSF Division of Science Resources Studies Issue Brief no. 98–316.*

[11] *Debackere, K. & Rappa, M.A. (1995)* "Scientists at major and minor universities: mobility along the prestige continuum" *Research Policy 24(1) pp. 137–150.*

[12] Dietz J.S., Chompalov I., Bozeman B., Lane E.O., Park J. (2000) "Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment" Scientometrics 49(3) pp. 419–442; Cañibano, C., Otamendi, J., Andújar, I. (2008) "Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal programme in Spain" *Research Evaluation* 17(1) pp. 17–31.

[13] Pierson, A.S. & Cotgreave, P. (2000) "Citation figures suggest that the UK brain drain is a genuine problem" Nature 407(6800) p. 13; Laudel, G. (2003) "Studying the brain drain: Can bibliometric methods help?" *Scientometrics* 57(2) pp. 215–237.

[14] Jonkers, K. (2009) "Emerging ties: Factors underlying China's co-publication patterns with Western European and North American research systems in three molecular life science subfields" *Scientometrics* 80(3) pp. 775–795.

[15] Laudel, G. (2005) "Migration currents among the scientific elite" *Minerva* 43(4) pp. 377–395.

[16] Bekhradnia, B. & Sastry, T. (2005) "Brain Drain: Migration of Academic Staff to and from the UK". Available at http://www.hepi.ac.uk/466-1181/Brain-Drain–Migration-of-Academic-Staff-to-and-from-the-UK.html.

[17] Qiu, J. (2008) "Scientific publishing: Identity crisis" *Nature* 451 pp. 766-767.

In order to overcome these problems, Scopus has improved its author-profiling algorithm in order to identify individual researchers precisely. The Scopus Author Identifier[18] distinguishes between author names; it gives each author a separate ID and groups together all the documents written by that author. This sophisticated algorithm recognises authors based on various data elements associated with the article, including affiliation, publication history, subject area and co-authors. The algorithm behind the author profiling function matches alternate spellings and variations of the author's last name, international name variations, author names with and without initials and middle names, and all possible combinations of first and last names.

Despite international variations, for example, an author will be matched based on various data elements associated with the article, including affiliation, publication history, subject area and co-authors.

Looking ahead, an industry-wide initiative called ORCID (Open Researcher & Contributor ID) has been gathering pace[19]. This initiative represents "a community effort to establish an open, independent registry that is adopted and embraced as the industry's de facto standard. Our goal is to resolve the systemic name ambiguity, by means of assigning unique identifiers linkable to an individual's research output, to enhance the research discovery process and improve the efficiency of funding and collaboration[20]".

The method at the core of SciVal Spotlight

Co-citation clustering is a key method[21] at the core of SciVal Spotlight. Clusters of 4-100 articles are assigned via a deconstruction-assignment approach where a paper can be assigned to more than one category. A co-citation analysis instead of bibliographic coupling or co-word analysis is then used to develop the reference relationships.

The threshold SciVal Spotlight uses for selecting references has three benefits: it resolves the issue of disciplinary bias (all disciplines are well represented); it increases the number of current papers that can be unambiguously assigned to clusters; and it makes the linking of clusters over time more accurate (clusters are linked via the references they have in common).

The unique differentiators of how SciVal Spotlight uses co-citation methodology include:

- A large number of reference papers in the model
- A clear measure of paper-paper relatedness
- An intuitive clustering algorithm for reference papers
- A large number of current papers assigned to paradigms

The methodology has been published in peer-reviewed literature, and relies on open source algorithms[22].

[18] For more information about Scopus Author Identifier see http://www.info.sciverse.com/scopus/scopus-in-detail/tools/authoridentifier/.
[19] For more information about ORCID (Open Researcher & Contributor ID) see http://www.orcid.org.
[20] Klavans, R. & Boyack, K.W. (2008) "Identifying distinctive competencies in science" SciTech Strategies Inc. Working Paper. Available at http://mapofscience.com/images/pdf/STS08-01.pdf; Klavans, R. & Boyack, K.W "Measuring Multidisciplinarity Using the Circle of Science" SciTech Strategies Inc. Working Paper. Available at http://mapofscience.com/images/pdf/KWB_RK_IDR_wkshp.pdf; SciVal Spotlight "User Guide for SciVal Spotlight 2011". Available at http://www.info.scival.com/UserFiles/3384_SciVal_Spotlight_User_Guide_LO_singles_0.pdf.
[21] Boyack, K. W., Klavans, R., Börner, K. (2005) "Mapping the backbone of science" Scientometrics 64(3), pp. 351-374; Klavans, R. & Boyack, K. W. (2006) "Identifying a better measure of relatedness for mapping science" *Journal of the American Society for Information Science and Technology* 57(2) pp. 251-263.
[22] Klavans, R. & Boyack, K. W. (2006) "Quantitative evaluation of large maps of science" Scientometrics 68(3) pp. 475-499; Small, H. (1973) "Co-citation in the scientific literature: A new measure of the relationship between two documents" *Journal of the American Society for Information Science* 24(4) pp. 265-269.

For more information on this report:
Reed Elsevier Group Offices
1-3 Strand London WC2N 5JR UK

Email: reportenquiries@elsevier.com