Discussion Paper prepared as part of Work Package 2 'Thematic Collaboration Roadmaps' in the project entitled 'FEAST Enhancement, Extension and Demonstration' (FEED). FEED is jointly funded by the Australian Government's *International Science Linkages (ISL)* Programme (Grant number SP07-0017) and by the European Commission's *Seventh Framework Programme* (Grant number FP7-222-747). See: *www.feast.org*.

A Bibliometric Analysis of Australia's International Research Collaboration in Science and Technology: Analytical Methods and Initial Findings

FEAST Discussion Paper 1/09

Dr Mark Matthews[1]
Mrs Bev Biglia
Dr Kumara Henadeera
Mr Jean-François Desvignes-Hicks
Dr Rado Faletic
Mrs Olivia Wenholz

Introduction

This paper presents the initial findings from an exploratory bibliometric analysis of Australia's international collaboration in science and technology. This paper is focusses on:

(a) Assessing the methodological challenges faced in comprehensively mapping Australia's science and technology research activity from an international engagement perspective;
(b) Suggesting solutions to these challenges;
(c) Providing some policy-relevant findings of potential use to the Australian Government and the European Commission.

Seeking to generate useful results in order to meet objective (c) provides a useful means of addressing objectives (a) and (b) in a clear manner.

It is important to recognise that fields of research vary in the extent to which journal publications can be regarded as the primary mode of research dissemination (in some areas books are the primary dissemination channel). In addition, the Web of Science (WoS)[i] data used in this paper does not comprehensively cover the output of journals in:

- Most fields in the humanities and social sciences;
- Engineering, information sciences;
- Some other fields of research in the applied sciences; and
- The applied end of the research spectrum even for fields generally well covered in WoS.

However, although the coverage of WoS is not universal, for subfields where a large proportion of their research has a basic orientation the coverage of these fields is extensive. Therefore, the main focus in this paper is upon science and technology fields.

Most of the analysis in this paper relates to counts of publications in different research fields and in a limited number of different combinations of national authorship, together with citations per paper (CPP) associated with these publications.

[1] Corresponding author, email: *mark.matthews@anu.edu.au*. Mark Matthews is Executive Director of FEAST and Jean-François Desvignes-Hicks, Rado Faletic and Olivia Wenholz are members of the FEAST Secretariat. Bev Biglia and Kumara Henadeera are with the Research Evaluation and Policy Project (REPP). Both FEAST and REPP are units within the new *Centre for Policy Innovation* at The Australian National University, see *cpi.anu.edu.au*.

CPP is a useful indicator of publication quality. In this paper it is expressed as relative CPP, that is to say CPP performance normalised for world CPP rates in a given research field (see methodology section for a detailed definition). In this analysis there is a particular emphasis on examining how different clusters of international collaborative activity affect relative CPP performance. The clusters chosen for analysis involve Australian scientific research papers and co-authorship clusters between the USA, Europe and the European Union.

Policy issues

There is a growing emphasis on supporting international engagement in research. Effective international engagement can provide the following benefits:

- Reduced unnecessary duplication of research efforts;
- Enhanced economies of scale and scope in research teams;
- An improved ability to exploit synergies between different capabilities, types of instrumentation and natural circumstances;
- Improved knowledge transfer;
- Enhanced skills development and recruitment;
- More effective work addressing global challenges;
- Contributing to constructive international relations;
- Stimulating foreign investment flows; and
- Facilitating access to research infrastructure (see Borthwick 2008).

Adams *et al.* (2007:3) states that international research activity is a rapidly growing component of core research activity for all countries. It is encouraged at a policy level because it provides access to a wider range of facilities and resources. International collaboration is particularly important in small to middle range countries like Australia because it allows participation in and access to activities from which scientific and technological innovation largely derive (Bourke and Butler 1995) - especially where the cost of major research facilities and associated research projects would be prohibitive for a smaller economy.

Given the benefits, it is natural for governments seeking to maximise productivity in academic research to focus on the measured benefits arising from international collaboration. If, from a public policy perspective, the 'yield' from funding academic research is partly reflected in the impact of the resulting publications on peers and other knowledge users then citation rates provide a useful indicator of this aspect of quality.

In general terms, the best type of case to make for increasing the level of funding available to support international engagement in research will be the case most likely to influence the Treasury (or Central Economic Ministry to use the generic term). This will inevitably tend to take the form of a productivity equation and it is therefore important to recognise that citation impact will, in such circles, tend to be approached as one element in productivity assessments. Given this, the scope exists to assess the extent to which increased support for international research collaboration can be expected to generate a productivity dividend – as expressed in the impact on citation rates.

Collaboration Patterns and Fields of Research

Fields of research vary in the extent to which journal publications reflect the total publication output. As previously stated, bibliometric analyses using journal based datasets are problematic in some research fields, particularly disciplines in the applied sciences, social sciences, and humanities. Much of the research in these fields is not published in journal articles, and even when output is journal-based the capture rate in WoS indices can be incomplete. A study by Butler and Visser (2006) found that the proportion of total output that finds its way into the public domain in source journals indexed by ISI [WoS] ranges from 90% in chemistry down to 6% in law. A summary of these results is provided in Table 1, for full details refer to Butler and Visser (2006:329).

Table 1. WoS Coverage of Australian University Publications, 1999-2001

	WoS Coverage of Publications (%)
Chemistry	90
Physics	85
Biology	78
Medicine	73
Mathematics	68
Agric., Vet., Envir. Sciences	64
Earth Sciences	64
Psychology	54
Engineering	40
Philosophy	32
Economics	27
Human Society*	23
Computing	21
Politics and Policy	20
History	18
Management	16
Education	14
Language	14
The Arts	13
Communication	9
Architecture	9
Law	6

*Sociology, Social Work, Anthropology, Human Geography, Demography
Source: Butler and Visser (2006).

Consequently, bibliometric data based on journal outputs should be used cautiously in the humanities and the social sciences because the proportion of publications captured may be relatively small. In addition, a large number of papers may only be single authored papers, so the sample size when trying to analyse international collaborations will be even smaller. The number of co-authored papers in the Social Sciences Citation Index has increased over the 1991-2005 time period. However, the total number of internationally authored papers in these fields still represents only 21% of the total publications.

In contrast, in science and technology fields that are well represented in the journal citation indices, the analysis of co-authorship patterns are more robust. It is for this reason this paper has focussed attention on the major fields of research in science and technology.

Methodology

One potentially useful measure is relative CPP performance (see box below for details). Relative CPP tells us the extent to which a group of publications deviate, upwards or downwards, from the median or the average CPP for that field of research (depending on which summary measure is used).

Relative Citation Impact defined

Relative Citation Impact compares the citation rate with the relevant world average or median (as appropriate). It is calculated by dividing the average or median number of citations per publication in any given subfield by the average or median number of citations for all publications in that subfield i.e. the world citation rate for that subfield. A relative citation impact of more than 1.0 indicates a higher position relative to the world norm, whilst a relative citation impact less than 1.0 would indicate a relatively low performance.

In the analysis that follows, relative citation rates are calculated on the basis of median rather than average citations because this is the most appropriate metric to use for skewed distributions, a characteristic of citation patterns (see Bornmann *et al.* 2008).

The paper explores the co-authorship clusters between Australian authors and some of their traditional partners in international scientific research. The analysis focuses specifically on USA, Europe and the European Union.

Relative productivity in this context can be defined as the relationship between relative CPP performance and different co-authorship clusters. Note that sole authored Australian publications can be treated as a particularly simple cluster from this more global perspective. It is therefore interesting to consider the extent to which relative CPP performance is influenced by cluster *composition* – treating co-authorship clusters as the primary unit of analysis rather than nations *per se*. This cluster-based approach, when operationalised in appropriate data structures, provides a useful means of mapping collaboration patterns and of analysing how relative CPP performance relates to cluster composition.

Results

A Snapshot of International Collaborations in Australian Publications

Most of the growth in Australia's research publications is associated with international collaboration rather than purely domestic efforts.

The proportion of Australian publications in the Science Citation Index with international coauthorship has increased from almost 21% in 1991 to over 44% of total publications in 2005. The output of internationally collaborative papers is growing at almost double the rate of purely domestic papers. This trend is illustrated in Figure 1.

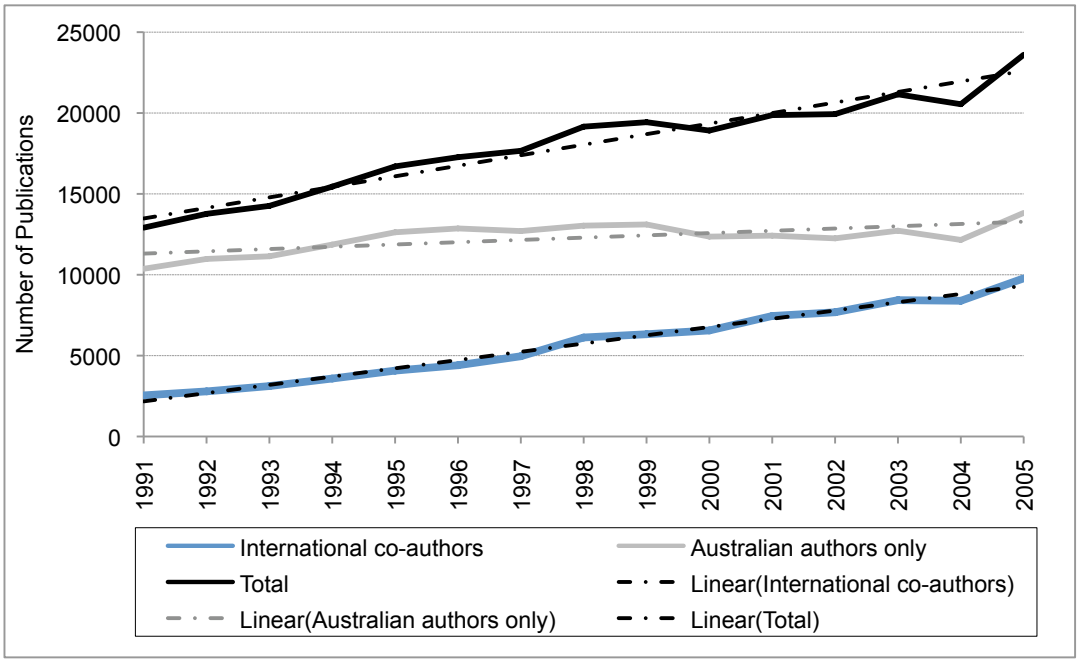

Figure 1. International Collaborations in Australian Publications, Science Citation Index, 1991-2005

This trend is investigated further in Figure 2, which compares the increase in the number of Australian collaborative papers with USA, Europe and with European Union based researchers. It is evident from the graph that the volume of international collaboration increased significantly after the 1990's. A report by Adams (2007) on patterns of international collaboration for the UK also revealed a similar increase in the volume of international collaboration for UK researchers (see also Bourke and Butler 1995:20 for UK University international collaboration by regions of the world).

The changing pattern of international collaboration between the early 1980's and after the 1990's is also influenced to some extent by the increased coverage of the journals captured in the science citation indices and by technological improvements in communication.

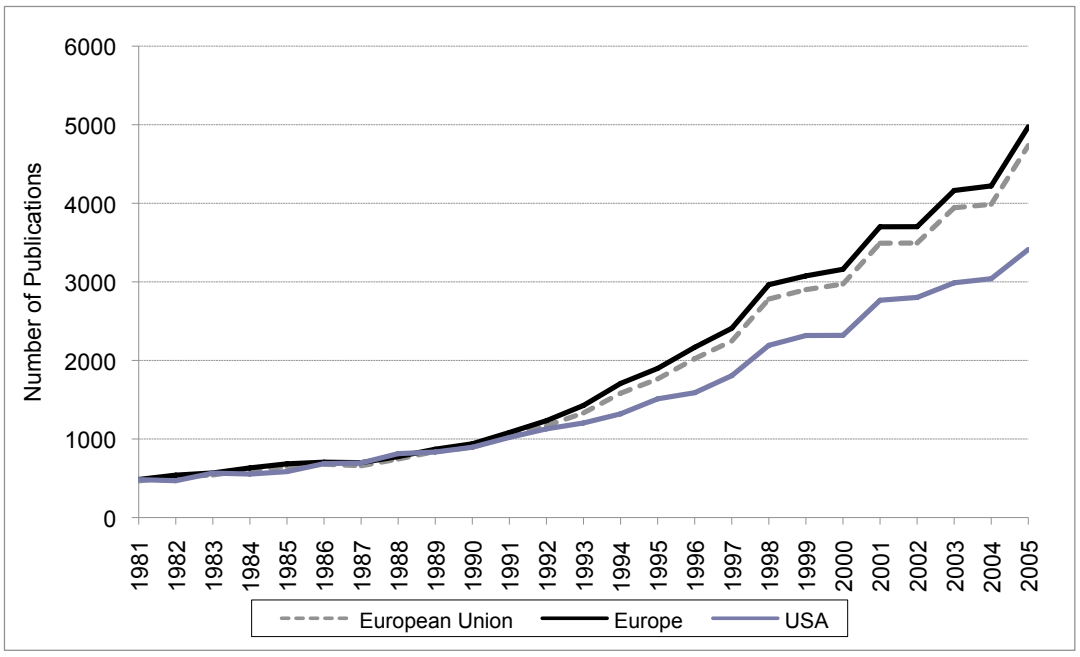

Figure 2. Australian Collaborations with the European Union, Europe and USA, Total publications, 1981-2006

Collaborative Patterns in Science and Technology

The following section analyses major fields of the Science Citation Index in terms of various co-authorship patterns and relative citation measures.

Figure 3 graphs the relative citation rates for collaboration between Australia-based researchers and a selection of collaborative groupings: Australia-US collaboration; Australia-Europe collaboration and Australia-USA-Europe collaboration. The higher relative citation performance of the latter, multilateral group, is noteworthy.

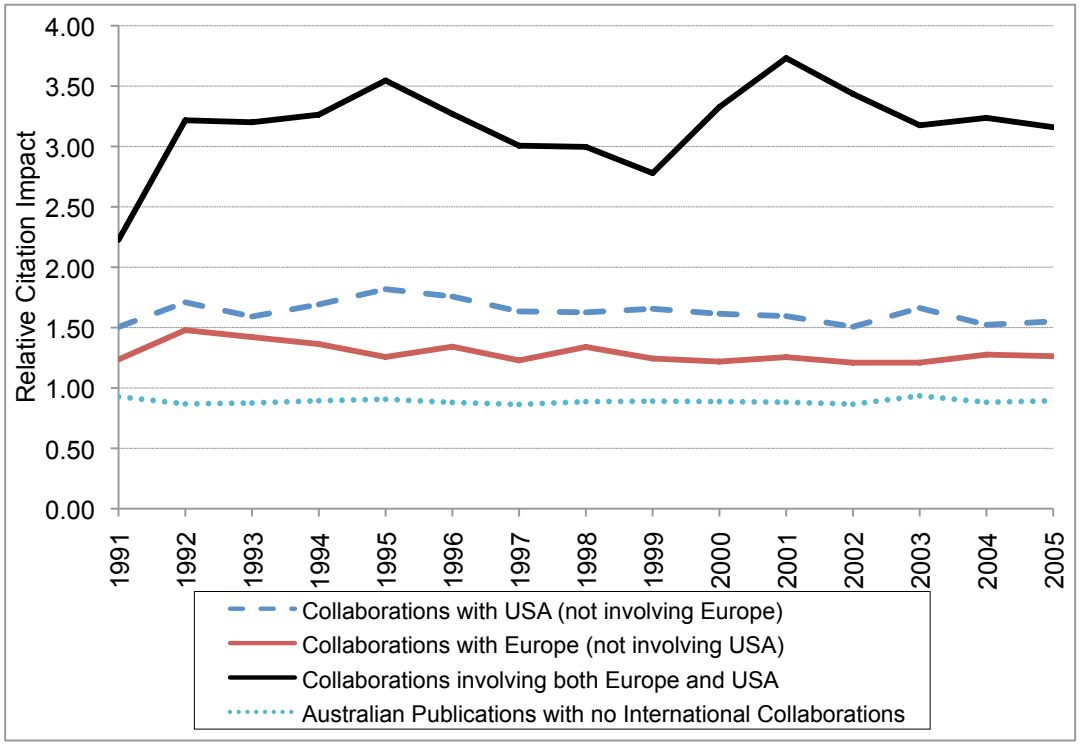

Figure 3. Relative Citation Impact of Science Citation Index Publications. 1991-2005

In order to investigate this issue further, data on relative citation impact by major research fields of science were examined. The results can be found in Figure 4 and Table 3 (below) for the period 1991-2005. Table 2 provides basic statistics by major science research fields.

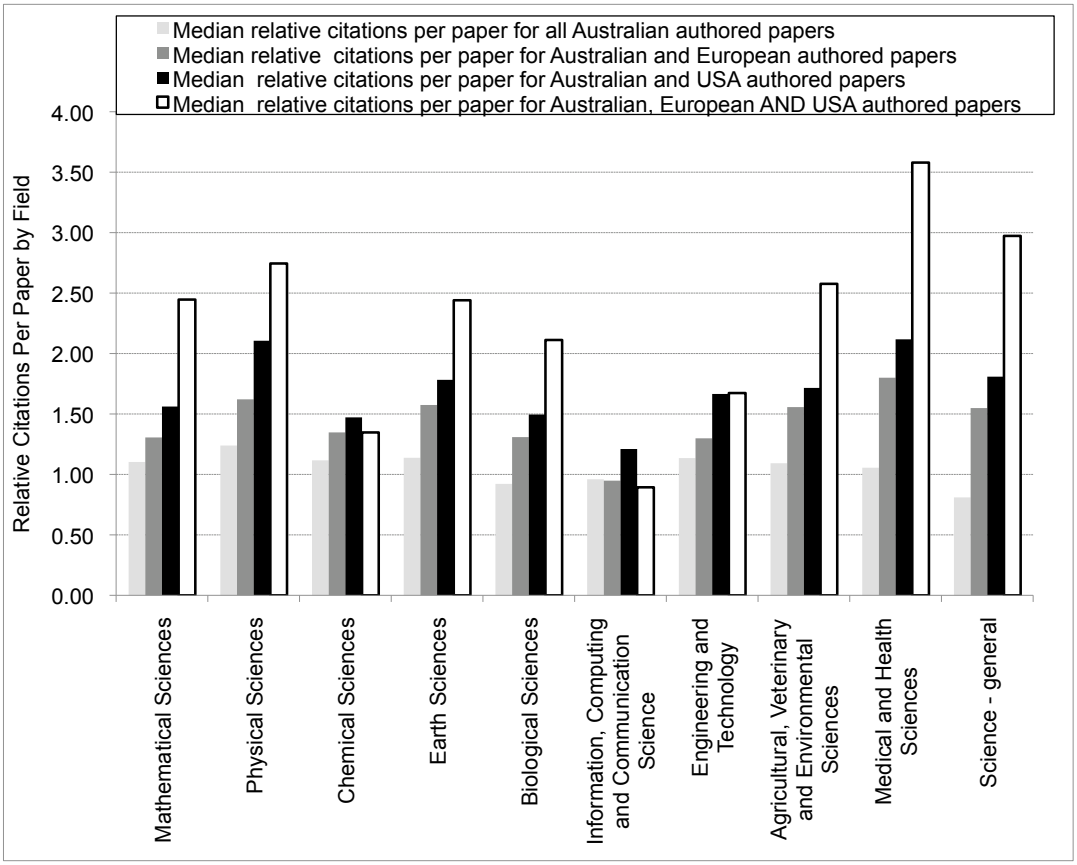

Figure 4. Median Relative Citations per Publication in Science Citation Index Publications by Major Research Fields, 1991-2005

These findings confirm that, at this level of aggregation, bilateral collaboration between Australia and Europe is associated with higher citation rates than are achieved by publications with no international co-authorship. Furthermore, in most major research fields multilateral collaboration involving Australian, European, *and* USA-based co-authors is associated with higher citation rates than bilateral cooperation alone.

The increase in relative citation performance is particularly pronounced for medical and health science and general science publications (the latter comprise broad-scoped high impact factor journals such as Nature). The only exceptions are chemical sciences and information, computing and communication science. Significantly, these are two research fields with relatively small numbers of papers with Australian, European and USA located authors, suggesting that this may simply be a consequence of small numbers and poor ISI coverage.

The case of information, computing and communication science is particularly noteworthy in this regard. There are only 3,229 publications captured in this dataset, mainly due to the fact that this research field tends to validate and disseminate results through channels other than journal publications. Therefore the results for that research field should be used cautiously.

The case of chemical sciences is different. With 23,177 publications tracked over this time period and 297 papers with Australian, European and USA located authors, the reasons for this type of multilateral collaboration to be associated with lower relative citation rates warrants further investigation.

In general however, these results suggest that there is a clear incentive to pursue bilateral collaboration with European researchers and an even stronger incentive for Australia-located researchers to collaborate multilaterally with European and USA-based authors. This is one logical explanation for the important role that international collaboration plays in driving increases in publications output (as summarised in Figure 1).

This finding lends support to the argument that an increased share of nationally available funding made available for international research collaboration would lead to increased relative citation impacts. In Central Economic Ministry terms, an increased proportion of national research funding made available for international research collaboration would increase the productivity of the academic research sector.

Table 2. Summary Statistics on Publications Output by Major Science Research Field,1991-2005

Publication numbers	Mathematical Sciences	Physical Sciences	Chemical Sciences	Earth Sciences	Biological Sciences	Information, Computing and Communication Science	Engineering and Technology	Agricultural, Veterinary and Environmental Sciences	Medical and Health Sciences	Science - general
Number of Australian authored papers	7,890	30,217	23,177	17,174	54,380	3,229	29,193	28,913	91,860	17,068
Number of Australian authored papers with international collaborators	3,710	14,395	7,696	7,122	18,568	1,297	9,644	7,440	9,644	5,865
Number of Australian authored papers with no international collaborators	4,180	15,822	15,481	10,052	35,812	1,932	19,549	21,473	82,216	11,203
Number of Australian and European authored papers	1,844	8,130	3,806	3,091	9,023	521	3,829	2,930	12,951	2,697
Number of Australian and USA authored papers	1,102	5,829	1,728	2,765	7,088	425	2,634	2,185	9,939	2,378
Number of Australian, European and USA authored papers	206	2,189	297	638	1,540	72	385	360	2,697	601
Number of world papers (science and technology fields)	356,875	1,809,067	1,523,503	373,961	1,787,226	1,787,226	1,510,057	647,854	3,501,262	519,630

Table 3. Median Relative Citations per Paper by Major Science Research Field, 1991-2005

Citations per paper medians (1991-2005)	Mathematical Sciences	Physical Sciences	Chemical Sciences	Earth Sciences	Biological Sciences	Information, Computing and Communication Science	Engineering and Technology	Agricultural, Veterinary and Environmental Sciences	Medical and Health Sciences	Science - general
Median relative citations per paper for all Australian authored papers	1.10	1.24	1.12	1.14	0.92	0.96	1.14	1.09	1.06	0.81
Median relative citations per paper for Australian and European authored papers	1.31	1.62	1.35	1.57	1.31	0.95	1.30	1.56	1.80	1.55
Median relative citations per paper for Australian and USA authored papers	1.56	2.11	1.47	1.78	1.50	1.21	1.67	1.72	2.12	1.81
Median relative citations per paper for Australian, European and USA authored papers	2.45	2.74	1.35	2.44	2.11	0.89	1.67	2.58	3.58	2.97

Discussion

These findings indicate that it is preferable to map and analyse Australia's competitiveness in research with reference to participation in *global collaboration clusters* (GCCs) rather than simply as sets of discrete bilateral collaborations. In the natural sciences at least, Australia's competitiveness needs to be framed in relation to participation in these *global collaboration clusters* rather than simply a nationally based definition of competitiveness in research.

To explain this point in financial terms the 'asset' may be the global collaboration cluster *itself* (which benefits from the economies of scale and scope in research that are associated with larger international projects) – not simply the sum of nationally based assets in that cluster. Consequently, determining the extent to which the improved citation performance associated with participation in GCCs is the result of benefiting from the higher citation rate norms found in the various partner countries (an 'extrinsic' effect) versus the beneficial citation impact generated by the cluster *itself* (an 'intrinsic' effect) requires further investigation.

This is a different perspective than that often adopted in bibliometric work – which tends to use a particular nation as the primary unit of analysis. Nation-based analyses are extracted from the wider dataset by selecting all records with at least one author with an address in that country. This generates a partial view of the more complex reality of publishing activity, which consists of a range of different types of publication clusters.

Arguably, the more narrowly nation-based approach has evolved partly because of the severe technical challenges involved in mapping and characterising more complex bibliometric datasets. The very large number of data queries required to carry out clustering work means that this process must be automated, and requires more sophisticated software algorithms.

If the performance of different nations is to be compared, starting only with nation-based datasets, then double counting problems arise because a proportion of each nation's publications are in fact joint publications. As this paper has demonstrated, in some cases such joint publications are multilateral and not simply bilateral. Furthermore, the more highly cited the paper the more likely it is to involve authors from a number of nations.

This is not to say that all Australia's competitiveness in research results mainly from participation in these *global collaboration clusters*, but it does imply that there are two major drivers of research competitiveness at work: a *national effect* (linked to both natural location advantages and advantages generated via accumulated research experience) and a *global collaboration cluster effect*.

Conclusions and next steps

As a result of these initial findings, work is now underway in FEAST aimed at automating bibliometric data queries in order to extract the data in a wide range of different permutations. This will allow us to create a large secondary dataset that can then be used to map Australia's key collaborative research relationships – as expressed in publications. The result will be a two-stage process that will allow for far more sophisticated analyses than can be carried out at present. The first stage requires access to fast and sophisticated computer processing power, but once completed this will allow for far easier and more flexible bibliometric analysis than can be achieved using case-by-case data queries.

This two stage process will allow a more thorough analysis of the *global collaboration clusters* in which Australian researchers operate and of the relative importance of the national effect and the global collaboration cluster effect on Australia's publications performance.

The next Discussion Paper in this series will build upon this analysis, focusing on a more detailed analysis of Australia's science and technology collaborations with the major EU member states (UK, Germany, and France) by major field of research and with regard to both bilateral and multilateral collaboration.

References

Adams, J, K. Gurney and S. Marshall. 2007. *Patterns of International Collaboration for the UK and Leading Partners (Summary Report).* A report commissioned by the UK Office of Science and Innovation, Evidence Ltd, June.

Bornmann, L., R. Mutz, C. Neuhaus and H.D. Daniel. 2008. 'Citation Counts for Research Evaluation: Standards of Good Practice for Analyzing Bibliometric Data and Presenting and Interpreting Results.' *Ethics in Science and Environmental Politics* 8: 93-102.

Borthwick, J. 2008. 'International Research.' *Australian International Education Conference*. URL: *http://www.aiec.idp.com/pdf/Borthwick_Fri_1140_M4.pdf*

Bourke, P. and L. Butler. 1995. *International Links in Higher Education Research*. Commissioned Report No. 37, National Board of Employment, Education and Training, AGPS, Canberra.

Butler, L. and M.S. Visser. 2006. Extending Citation Analysis to Non-Source Items, Scientometrics 66(2): 327-343.
