September 2009

Identification and dissemination of lessons learned by institutions participating in the Research Excellence Framework (REF) bibliometrics pilot

Results of the Round Two Consultation – report to HEFCE by Technopolis

Table of Contents

Table of Figures

Identification and dissemination of lessons learned by institutions participating in the REF bibliometrics pilot

Results of the Round Two Consultation

1. Summary

1.1 Introduction

The Round Two survey sought pilot-institutions' views on each of the following:

- Accuracy of the bibliometrics results
- Surprising results or outliers
- Notable issues regarding the coverage of the two reference databases
- Likely scale of future costs to develop/implement the preferred model
- Implications for institutions, research management or individual academics

This paper presents a synthesis of respondents' views as to the attractiveness and feasibility of each of the bibliometrics approaches tested, in light of their experience of the submission process and their reaction to the results. It offers a commentary on these views and draws out key messages and possible implications for Higher Education Funding Council for England (HEFCE).

1.2 Bibliometrics models

The consultation began by inviting pilot institutions to comment on each of the three bibliometrics models trialled through the pilot process:

- Model 1 (M1), based on **institutional address**. In this model the papers associated with each higher education institution (HEI) are taken directly from either the Web of Science (WoS) or Scopus databases, based on address data within those databases
- Model 2 (M2), based on **authors, all papers**. In this model, the pilot HEIs sought to list all papers published by all staff within each unit of assessment (UOA)
- Model 3 (M3), based on **authors, selected papers**. In this model, the pilot institutions' publication lists comprised the most highly cited papers (6) for all staff submitted to the 2008 Research Assessment Exercise (RAE)

1.3 Consultation results

1.3.1 Accuracy and robustness

Model 3 – author, selected papers – was the preferred model, under present conditions. Its results tended to fit best with the 2008 RAE results and pilot institutions believe it will be the simplest of the three approaches to implement.

On balance, pilot institutions see Model 3 as producing indicators of institutional research quality that are sufficiently robust to be a useful input to a REF peer-review process, subject to certain checks and balances, which are that: it should apply to a subset of UOAs; and HEIs should be allowed to select their papers for analysis.

Model 1 – institutional address – was widely regarded as a potential future solution, minimising the risk of inaccuracies caused by institutions building their bibliographic

databases themselves, bottom up. However, there was a universal view that its accuracy was unacceptable at present.

Model 2 – author, all papers – was seen as having the potential to produce the most accurate results of all three models, however in practice it was generally seen to be the worst solution. Its real-world accuracy was compromised by the variable coverage of the two key reference databases. Getting to a point where Model 2 would be acceptable in practice is beyond reach for the first REF.

1.3.2 Cost-effectiveness

At first appearance, Model 1 and Model 3 would require HEIs to do very little additional work: the former could in principle be carried out by HEFCE without direct input by HEIs and the data requirements for the latter are close to those of the 2008 RAE process. By contrast, Model 2 would require a majority of HEIs to upgrade their research information management (RIM) systems.

In practice, however, it seems likely that the differences would be much less pronounced, and that all three models would oblige HEIs to press forward with ongoing efforts to strengthen their central RIM systems and capture a majority of all research outputs.

On Model 1, pilot institutions were unanimous in the view that the entire higher education (HE) community would insist on being involved fully in the validation of the bibliographic databases for use in the REF citation analyses. The experience of the pilot suggests that for most universities and colleges, such a process would require substantial further development of their information systems.

On Model 3, the pilot institutions are certain the wider community will demand that individual universities and colleges be allowed to select their papers. Picking out the most appropriate portfolio for the institution is done best by running a simulation of the REF citation analyses against a *majority* of all papers. As with Model 1, the implementation of Model 3 is likely to oblige HEIs to upgrade RIM systems and to increase the proportion of all research outputs captured.

Model 3 is likely to be the most tolerable and cost-effective solution for the first implementation of the full Research Excellence Framework (REF), which is expected to happen in 2013. It leaves control over the publications lists with institutions and the implied scope of systems development feels manageable. Ultimately, Model 3 goes with the grain of current policy and systems development.

1.3.3 Behavioural incentives

There is an evident split in expectations between the two comprehensive models (M1 and M2) and the selective model (M3). The two comprehensive models were seen as positively reinforcing the notion that all research outputs should be as good as they can be. On the downside, people anticipate a possible reduction in the diversity of the types of research outputs, perhaps to the particular detriment of relationships with non-academic communities. In addition to narrowing the communication bandwidth, M1 and M2 might cause HEIs to discourage early-career researchers from publishing except as secondary or tertiary authors.

Model 3 by contrast appears less dramatic, being closer to the present RAE arrangements. The positive points are that it will reinforce current incentives for academics to seek to publish in higher impact journals and it seems unlikely to be a disincentive to publishing by early-career researchers. The negatives are a mix of old and new; M3 looks set to continue to encourage the poaching of key staff; and there is concern that it will encourage a narrow focus on journals indexed by Scopus and WoS. On balance, M3 seems to offer the best fit with the high-level policy objectives for the REF bibliometrics element, which is to say it offers robust indicators of research quality at the institution level while avoiding the creation of major new negative or perverse behavioural incentives.

1.3.4 Coverage of the WoS and Scopus databases

The pilot confirmed there are differences between the two databases, the coverage of which tends to vary by subject. This outcome suggests that the bibliometrics element of the REF should ideally make use of both databases in order to maximise HEFCE's ability to capture institutions' most highly cited work.

The coverage issue is likely to become less problematic in time as both databases index more publications and as researchers consciously target a proportion of their output on higher impact indexed titles.

1.3.5 Accuracy and coverage at the level of the UOAs

The consultation produced too few data points to carry out any systematic analysis of the accuracy of the results at the level of the UOAs.

The consultation did produce some useable statistics, overall. This aggregate analysis suggests that the bibliometrics results were judged to be 'good to reasonable,' in terms of their accuracy, in 80% of the 100 or so instances where ratings were given.

Where the aggregate statistics point to concerns with accuracy, it is evident that respondents see the bibliometrics as tending to understate their true standing with regard to research excellence.

As with accuracy, the consultation produced too few data points to carry out any systematic analysis of database coverage at the level of the UOAs. The aggregate statistics suggest the bibliometrics results were judged to be 'good to reasonable,' in terms of their coverage of HEIs' highest quality work, in around two thirds of the cases where ratings were given.

1.3.6 Anticipated future costs

A significant minority of respondents was able to provide a first approximation of the likely total development costs implied by the REF bibliometrics module. The majority of these provided a single estimate, independent of the model adopted.

The estimates covered a broad range from a low of £0K, for the two institutions that believe they have systems in place already to deal with the REF information requirements, to a high of £460K. The median estimate was around £100K.

The spread is a reminder that the actual cost will reflect the current state of development of an institution's research information management system and the institution's size.

In most cases, the bulk of the cost is associated with internal staff time with a split of around 60:40 between internal and external costs.

The majority of respondents providing financial estimates anticipate having to underwrite some additional costs, over and above their current investment plans, in order to comply with the REF bibliometrics requirements. Those estimates ranged from an increase in costs of 5-80%, with a median increase of 33%.

A significant number of respondents was able to provide a first approximation of likely running costs. In the majority of these cases, respondents provided similar estimates for all three models. The figures show a broad range, from no additional running costs, to a high of £160K a year, with a median estimated additional annual cost of £75K. The data also reveal that internal costs are expected to be much greater than external costs.

1.3.7 Implications for research information management systems

The pilot has produced an emerging consensus on the likely direction of the future development of RIM systems, and suggests that the REF will reinforce development

trajectories rather than signal the need for radical new departures. Taken together, the responses suggest there is a general ambition to

- Strengthen *corporate* research information management systems,
- Have central research and planning functions play a fuller role in collating and checking information recorded at departmental level,
- Have a central publications database that links to other university systems such as HR (staff) and finance (contracts, research income),
- Move towards a system where all research outputs are captured and catalogued,
- Develop internal bibliometrics expertise through additional training and possibly new appointments of specialist staff, and
- Take more value out of these investments, by developing systems to inform institutions' routine performance monitoring and periodical strategic planning.

1.3.8 Implications for research management

It is clear respondents believe the REF will have an impact on the strategic management of HEIs: the overall research undertaking will be more actively considered and directed by the centre. Taken together, the comments suggest there will be developments along each of the following dimensions:

- Increased use of bibliometrics and citation data to inform research management at the institutional level
- Increased use of bibliometrics and citation data at the faculty level, possibly effecting the full range of performance management from strategic planning to department reporting on the one hand, to individual academic appointments and annual reviews on the other
- Increased use of bibliometrics and citation data to inform dialogue between HEIs central management and individual faculties

Lastly, there were several comments that serve to remind all concerned that there is still some way to go to secure academic buy-in to these developments.

1.3.9 Other remarks and advice to HEFCE

Three points were picked up by several respondents, which are:

- Clear and precise guidelines are vital, and respondents would like this information as early as possible
- A recognition that whatever system is chosen, it will influence institutional and academic behaviour
- The use of bibliometrics should be focused on those areas where it is robust, and must only be used to inform a peer-review process

There were several separate points of note also, including:

- HEFCE should ensure institutions select the papers for citation analysis
- HEFCE should consider providing results from both Scopus and WoS
- HEFCE should improve the presentation and explanation of the citation results
- HEFCE needs to bear in mind that the costs of compliance will vary greatly across institutions, and even under Model 3 might prove quite demanding for some

There remains an undercurrent of concern about the use of citations in performance assessment, which while it reduced greatly across the two consultation rounds,

suggests it will take one or two iterations of the REF bibliometrics proper before the community is fully acclimatised.

2. Introduction

2.1 **This paper**

This paper presents the results of the second-round consultation with the 22 universities and colleges involved in the Research Excellence Framework (REF) bibliometrics pilot, conducted after the bibliometrics results had been shared with pilot institutions.[1]

It presents a synthesis of pilot institutions' views as to the attractiveness and feasibility of each of the bibliometrics approaches tested, in light of their experience of the submission process and their reaction to the results. The paper offers a commentary on these views and draws out key messages and possible implications for the Higher Education Funding Council for England (HEFCE).

2.2 **The round two consultation**

2.2.1 The questionnaire

The results of the various bibliometric analyses were shared with the 22 pilot institutions in April 2009.

The Round 2 questionnaire was circulated in mid-May for return in mid-June, by which time 21 of the 22 institutions had been able to provide a written response. In broad terms, the survey sought respondents' views on each of the following:

- Strengths and weaknesses of each of the three bibliometrics models trialled
- Accuracy of the bibliometrics results
- Number and explanations of any surprising results or outliers
- Notable issues regarding the coverage of the two reference databases
- Likely scale of future costs to develop/implement the preferred model
- Likely scale of running costs of the preferred model
- Implications for research information management systems
- Implications for institutions, research management or individual academics

2.2.2 Potential models

The consultation began by inviting pilot institutions to comment on the apparent strengths and weaknesses of each of the three different bibliometrics models trialled through the pilot process. As a result, much of the feedback refers to Model 1 or Model 2 or Model 3. To help readers decode these remarks, it is helpful to note that HEFCE sought to construct publication lists for input to its citation analyses in three different ways:

- Model 1, based on **institutional address**. In this model the papers associated with each higher education institution (HEI) are taken directly from either the Web of Science (WoS) or Scopus, based on address data within the databases.

[1] The paper on the first-round consultation is available to download from the HEFCE web site, at www.hefce.ac.uk/pubs/rdreports/2009/rd09_09/. Readers can find out more about the REF Bibliometrics Pilot itself on the same site, at www.hefce.ac.uk/pubs/rdreports/2009/rd13_09/

Papers are assigned to HEIs based on the addresses of their authors, and they are assigned to a subject category (or multiple subject categories) depending on the journal in which they are published

- Model 2, based on **authors, all papers**. In this model, an attempt is made to identify all the papers published by specific staff within each unit of assessment (UOA). For the purposes of the pilot, HEFCE included all staff that were selected for the 2008 Research Assessment Exercise (RAE), in relevant UOAs
- Model 3, based on **authors, selected papers**. In this model, the pilot institutions' publication lists comprised the most highly cited papers (6) for all staff submitted to the 2008 RAE

2.2.3 Bibliometric results

The bibliometric results are the starting point for much of the feedback gathered in this second-round consultation and they merit a brief introduction here. The HEFCE Interim report of the REF bibliometrics pilot exercise covers the subject in detail, should readers wish to know more.

HEFCE presented the results to pilot institutions using two performance indicators:

- **% above 2x world average (2x)**: proportion of the HEI's papers that are above twice the world average (normalised) citation score
- **% above 4x world average (4x)**: percentage of the HEI's papers that are above four times the world average (normalised) citation score

Results were compiled on both indicators for each of the 35 UOAs included in the pilot process. For each UOA indicator, the results were presented for each of the three main bibliometrics models, for the WoS data and again for the Scopus data.

There were two charts for each UOA; one showing each model using the lower threshold (2x world average) and one showing each model using the higher threshold (4x world average).

To aid interpretation of the bibliometrics data, HEFCE included an additional data series within each chart, taken from the 2008 RAE. For the 2x indicator, this 'fourth' model was computed using the proportion of outputs rated 3* or above in the 2008 RAE. For the 4x indicator, the data were computed using the proportion of outputs rated 4* in the 2008 RAE.

Lastly, in reporting the bibliometric results, HEFCE applied a size threshold to the summary graphs, wherein the results for a pilot institution are included in those UOA charts where they had submitted 50 or more papers in either WoS or Scopus in Model 2 (rather than the selective Model 3). Bibliometricians take the view that the citation analyses become less robust as the number of papers decreases, and 50 papers is a common cut-off.

2.2.4 Organisation of the report

The report is organised around the responses to the questionnaire, following a similar treatment in each case: starting with a synthesis of the main points arising, then illustrating these points through a comprehensive presentation of excerpts from the written responses and concluding with a short discussion of the implications of what has been said.

3. Results

3.1 **Accuracy of the three models**

3.1.1 Perceptions as to the accuracy of the three models

Question 1 invited pilot institutions to provide their views as to the strengths and weaknesses of each of the three models in turn, in terms of its adjudged accuracy.

Respondents based their opinions on an overall impression of their bibliometric results' fit with their 2008 RAE profiles. While the 2008 RAE is not a definitive benchmark of research quality, in practice all respondents took it as their reference point. Checks amounted to a visual comparison of citation performance for each UOA with their 2008 RAE profile and their rankings within the sub-set of pilot HEIs with sufficient outputs to include in the citation analyses. For Model 1, most pilot institutions examined the set of papers attributed to their institution for a given UOA, to check the degree of alignment with their own records, in terms of missing papers or additional papers that should not have been attributed to the institution or UOA.

The following sub-sections synthesise the balance of opinion on strengths and weaknesses, model by model.

3.1.2 Perceptions as to the accuracy of Model 1

Just four of the 19 pilots that answered this question considered the institutional-address model to have any evident strength in terms of its accuracy. These observations were as follows, in no particular order:

> *In terms of recognising % citations and above world averages for citations, Model 1 proved the most successful.*
>
> *Positive reflection of our interdisciplinary research, in some areas.*
>
> *It usefully captures relevant outputs associated with people who have since left the institution.*
>
> *This model would be a better reflection of work actually undertaken at an HEI, rather than work associated with a selection of members of staff as at a census date. Less reliance on academics to provide data, because of automatic data mining and therefore potentially more accurate data.*

The comments reflect two points that might be of more general interest. The first strength is that this top down approach has the potential to generate more accurate records through its automatic data mining, and through avoidance of a reliance on individual academics as the starting point for constructing individual bibliographic records that contain all the necessary meta data, accurately and in the correct format.

The second strength also reflects concerns about the ability of institutions to fully capture their research outputs, and in particular in regard to people that had left the institution.

The balance of opinion was strongly negative as regards the accuracy of Model 1. It is seen to be the least accurate of all three models, albeit with only two respondents stating specifically that it had produced substantially weaker results on the pilot's key metrics (e.g. proportion of an institution's total set of publications that achieved citation scores better than twice the world average) as compared with the other models and 2008 RAE profiles. However almost all respondents mentioned multiple shortcomings with their institution's data set and by implication the robustness and accuracy of any subsequent citation analyses.

There were three recurrent concerns, which together meant that, for most institutions, the automatically-generated lists had significant numbers of omissions, misallocations and false positives (papers not attributable to the HEI in question). Specifically, those limitations were:

- Limitations in the subject and journal coverage of the WoS and Scopus databases mean that any attempt to construct a complete data set must omit a proportion of an institution's total research output, which in turn must compromise the accuracy of the citation analyses. This might be work published in journals not indexed by either database or work that is disseminated through other channels, from online archives to conference proceedings to books
- Using institutional addresses produced an incorrect data set, with both missing papers and included papers that are not associated with the university or college. The model proved to be sensitive to the fact that many institutions operate from multiple sites and most do not require a uniform treatment of addresses across those different sites or research centres
- Differences between the two databases' subject classifications and the location of authors within particular schools and departments lead to the misallocation of papers to the *wrong* UOA. This means a proportion of papers might be excluded from the subsequent citation analyses (due to an institution having an insufficient number of papers in that area to safely run those analyses) or judged against the publication/citation behaviour more typical of a different discipline

The following selected excerpts illustrate the nature and extent of these concerns.

> *This is the least preferred model. The main weaknesses identified are: incomplete retrieval of publications; and assigning papers to UOAs using database-subject groupings that do not relate to how staff are assigned to UOAs.*
>
> *Journal classification and subsequent mapping back to UOAs could be misleading. Concern regarding how multidisciplinary papers are classified and how the model aligns with internal departments, e.g. we have no biology department, do not teach biology and would not submit to UOA 14 (biological sciences), and yet this model suggested we have papers linked to UOA 14.*
>
> *The issue of mapping of journals onto UOAs would need to be better defined to enable this model to work, particularly in areas of multidisciplinary research. We have identified some key journals in which our staff publish, which were not mapped onto either of the UOAs into which we submitted in the RAE. This occurred mainly in WoS. For example, the journal 'Physics in Medicine and Biology' which was not mapped to any UOA. Other journals in which we publish but which are missing from WoS UOAs 2 (cancer) and 14 (biological sciences) are 'Blood', 'British Journal of Haematology', and British Journal of Radiology'.*
>
> *The association of journals to UOAs is not always as expected, for example UOA 40 (social work and social policy) has items from Studies in Nonlinear Dynamics and Econometrics and the International Urology and Nephrology Journal. An article from IEEE Microwave Magazine, appears in UOA 2 (cancer studies), while UOA 14 (biological sciences) has been allocated outputs in textiles.*
>
> *Coverage is generally poor compared with Models 2 & 3. Visiting staff and medical staff affiliated to the institution are often not picked up in this model.*

This model includes outputs by individuals who are no longer associated with the university, but who have used, or are still using, the institutional address.

The institution's address needs to be very well defined for this model to work. We identified a significant number of publications in the Scopus model that do not belong to the institution.

More than 40% of papers found were not attributed to staff in the UOA. It could bring in authors who are not in that particular UOA. It could miss out the outputs published in an institute address or centre address and not institutional address. It could discard multidisciplinary publications as it pulls in a narrow-band of publications, e.g. a Health Economist might have a publication in a health journal but that will go missing. It would be very difficult to align the other metrics, for example, income and student data, which are recorded against a specific individual in a UOA. It is likely to pull in publications for lots of people whom we have not counted in other statistics.

Several respondents also noted the seemingly undifferentiated nature of the results from Model 1, across institutions and for most UOAs. This outcome raised concerns about the credibility of the model more generally, although it is conceivable that this apparent flattening of quality simply relates to the choice of scale. Had HEFCE presented the results using charts with gradations in single percentage points rather than tens of percentage points, there would have been marked differences evident across institutions. This is probably a matter of presentation, wherein finer gradations are necessary for comparing institutions on their *total* research output as opposed to their *best, selected* outputs.

The poor correlation with the RAE results appear to stem from two problems in particular, which are the degree to which the automatic data mining is able to capture a complete and accurate set of publications for any given institution; and the reliability of the mapping of database subject classifications to the RAE UOA classification.

The great majority of respondents object to the institutional-address model because it produced an incorrect listing of publications, which could compromise citation analyses and possibly distort future income. If the inaccuracies were uniform across subjects and institutions, it would be less problematic as it would be less likely to impact on quality judgements or future funding. However, it is likely that the degree of alignment would vary by institution, and pilot institutions foresee a risk that the citation results might overstate or understate actual performance and thereby produce an unfair outcome. Gaps are likely to reduce with time as the principal databases – Scopus and WoS – are continuously expanding their subject and journal coverage.

A second wrinkle relates to the allocation of a proportion of academic work to non-obvious subject areas, which dilute output numbers in core subjects and 'waste' citations by their allocation to subject areas where an institution has little or no presence and where bibliometricians might reasonably decline to report the citation counts due to the small numbers of publications. Under Models 1 and 2, institutions were only included in a UOA where they had more than 50 publications.

There is another issue, which might be driving people's perceptions of poor fit, which is the difference in the scope of the papers under review between Model 1 and the RAE. Model 1 is inclusive by design, and whether it captures 40% or 80% of an institution's total publications, it will tend to pick up papers from across the spectrum of research excellence. By contrast, the papers submitted to the RAE were provided on a selective basis, few in number and the best available. Placing the results from these two assessment modes on the better-than world average scale will almost always produce a big gap. Indeed, the anonymised charts show Model 1 returning results in the range 5-10%, while the RAE range is closer to 40-90% (of the proportion of all items submitted/identified that achieved citation scores more than twice the world average for that field).

3.1.3 Perceptions as to the accuracy of Model 2

Looking at strengths, a small number of respondents noted that Model 2 had produced results that were rather more closely correlated with the 2008 RAE results than had Model 1. Just one of the 19 respondents stated that Model 2 had proved to be the least accurate of the three models, when judged against its 2008 RAE performance.

Most respondents offered comments on the underlying strengths of Model 2. The majority noted the same one or two features:

- Citation analyses based on an institution's total research output ought to provide a more accurate and instructive view of quality than a selective view (Model 3), albeit this is presently not the case due to the limitations of institutions' internal databases
- With Model 2, the creation of the bibliographic record is the responsibility of the submitting institution and not some third party, as with Model 1. Other related positive features were the use of publication data that had been validated by the submitting institution and the assignment of outputs to individual authors

With respect to weaknesses, Model 2 attracted perhaps half the volume of criticisms of Model 1. That said, in four cases, respondents report significant discrepancies between their RAE profiles and their citation analyses. Moreover, the majority of other comments suggest widespread concern over the real-world accuracy of Model 2, which relates to:

- The completeness and accuracy of institutions' own bibliographic records, which a majority of pilot institutions stated was some way away from where things would need to be. A related concern included the feasibility and tolerability of undertaking the amount of work implied by the need to prepare a comprehensive submission of research outputs (for all research-active staff across all UOAs and across the 6-year review period). Other comments picked up on concerns around institutions' abilities to adequately track and record sufficient information on joiners and leavers
- Concerns over the ease with which any external validation process might work, and the extent to which missing records might be captured and attributed to the right author and institution
- The degree of alignment/coverage of an institution's overall submission with the absolute subject and journal coverage of Scopus or WOS

3.1.4 Perceptions as to the accuracy of Model 3

Model 3 is the preferred model, with five institutions stating that its results align more closely with RAE2008 results than either of the other two models. Another three respondents noted it was the best of the models in terms of its ability to accurately reflect the excellence of the research being undertaken by the institution concerned.

> *On balance this model comes closest to RAE outcomes and may therefore be regarded as more accurate than Models 1 and 2*
>
> *Best model; most representative of the quality of research we undertake*
>
> *Good results for AHP & Engineering*

A majority gave Model 3 their qualified support, viewing the author-selected arrangements as being better than the two other options albeit with various reservations and qualifications being expressed about its absolute accuracy. Two respondents commented that this model still produced very different results for the institution in question as compared with their RAE2008 profile.

There were far fewer comments regarding Model 3's weaknesses, as compared with either of the other two models. However, several respondents indicated that the concerns expressed about Model 2 also held for Model 3. Indeed, most of the express

concerns about Model 3 echoed those for Model 2. This is understandable as it is a variant of that model, just with a reduced scope.

> *Still the potential for errors as papers may not be submitted by the institution or found during the data verification process. Prior and post addresses may not be available or accurate. Reliant on academics to provide accurate and complete data. An automated ranking by citation count may not necessarily provide the most accurate ranking of impact and quality. Problems of top ranking papers appearing for two or more authors and therefore reducing the number of papers, which can be considered for an individual author.*

> *The wider coverage of the Scopus model is both a benefit and weakness. It has correctly included some publications that WoS has not, but also seems to have included a few papers where we could not find any link with our address or staff. The most significant of these is a paper in Immunity with a high impact factor, which has skewed the selected papers model. We have only checked some of the data but the consequence of the data errors and variances uncovered is such that we would wish to check all data were the exercise being run for real. The WoS omissions appear to reflect difficulties with an address or to being the output from very large multi-author studies. This leads to a suspicion that the databases and/or analysis tools treat the issues differently (does WoS restrict the number of authors in searchable fields? Do the analyses differ in how they treat self-citation?).*

A small number of respondents continue to have concerns about the fundamentals of citation analyses, which suggests that some pilot institutions are still tending to see the bibliometrics as a large-scale quality assessment of *individual* researchers rather than a means by which to test quality at the level of an *institution*.

> *The highest impact papers in terms of normalised citations are not necessarily the 'best' papers for an individual.*

> *Selection of papers by high citation rates: citations do not always correlate to high quality. It would be better if HEI selected the papers.*

This concern over how one might select the 'best' papers reveals something of a split in opinion, at least amongst the half-dozen instances where it was raised. On the one side, several respondents state that each institution should select its papers based on authors' judgements as to what is best. In contrast, there are a smaller number that saw fit to remark upon the potential for using citation analyses as the basis for selecting work that places the institution in the best light.

> *It selects the "best" results for the institution based on metrics, rather than an individual or institution's opinion. It only draws in academics who belong to that particular UOA, so we can select appropriate academics and align the other metrics. It ignores weaker/lower impact papers, which allows for career progression and for academics to "grow" their skills and contribution without penalty.*

A significant proportion of all of the comments – both on strengths and weaknesses – relate to the implementation of the model in the future and the issues that people feel HEFCE and its contractors will need to pay especial attention to.

> *Need equal systems for access to citation data across HEIs to ensure all have same information in order to understand the papers selected by bibliometric analysis or for HEIs to make judgements on which papers to select.*

> *HEIs can more readily check the data.*

> *Direct link to college staff enable comparisons with institutional internal data.*

There must be clear definition of eligible staff in this model. Given the inevitability of some degree of peer review, it would be preferable to leave selection of outputs to the submitting UOA, and not to automatic criteria.

How will the REF deal with large multi-authored papers and assign credit to the main author(s)? Early career researchers may be disadvantaged. It is unclear how joint submissions and research pools will be treated.

3.1.5 Discussion

Model 3 – author, selected papers – was the preferred model, under present conditions. The reason for this was that it tended to fit best with the RAE2008 results and pilot institutions believe it will be the simplest of the three approaches to implement for the HE community overall, as it is very close in its data requirements to the 2008 RAE process.

Respondents were somewhat ambivalent as to whether the pilot had demonstrated robust and rounded indicators of research quality. On balance, respondents conclude that Model 3 is sufficiently accurate for use as an input to a peer-review process, subject to institutions being given responsibility to select authors and papers. Prior modelling of the REF citation analyses is seen to be a necessary step to overcome the vagaries of journal coverage and subject mapping of Scopus and WoS.

Model 1 – institutional address – was widely regarded as a potential future solution, minimising the risk of inaccuracies caused by institutions building their bibliographic databases themselves, bottom up. It has the potential to be a low-cost solution too. However, there was a universal view that its accuracy was unacceptable at present, and that economies notwithstanding HEFCE should not move forward with Model 1.

Model 2 – author, all papers – was seen as having the potential to produce the most accurate results of all three Models, however in practice it was generally seen to be the worst solution. Its real-world accuracy was compromised by the variable coverage of the two databases, subject area by subject area. This holds for Model 3 also, but the effects would be amplified here by the uneven development of the pilot institutions' research information management systems. Getting to a point where Model 2 would be acceptable would entail very substantial additional investment by UK HEIs in order to arrive at a situation by 2012/2013 where there is essentially a level-playing field as regards research information. Several respondents noted that this community-wide development effort might be beyond our reach for the first REF, and that the community might prefer Model 1, with a comprehensive verification exercise in support, over Model 2.

Model 3 does pose some questions from the perspective of bibliometrics analysis, as it naturally constrains the numbers of publications in any given pool and this can cause accuracy problems of its own. In part for this reason, Model 2 is closest to the approach adopted in Australia; however the Australian approach was made feasible by virtue of the country having required its universities to record all research outputs as part of a national reporting framework almost 20 years ago.

3.2 **Operational qualities of the three models**

Question 2 invited pilot institutions to provide their views as to the strengths and weaknesses of each of the three models, in terms of their operational implications, behavioural incentives and cost effectiveness. Respondents were invited to offer comments on each model in turn and the following sub-sections synthesise the balance of opinion on strengths and weaknesses, model by model.

3.2.1 Perceptions as to the operational qualities of Model 1

A majority of respondents see the potential for the institutional-address model to emerge as the most cost-effective means by which to obtain bibliographic data with which to run citation analyses. However, no pilot institution went as far as to recommend this model for the first implementation of the full REF, expected to be run in 2013.

The top-down approach is clearly appealing because, in principle, it could be implemented in such a way as to require little or no additional work by individual HEIs. It could be the low-burden solution the community aspires to. However, in practical terms respondents recognise there is a need for substantial improvements in database coverage, data quality and data processing techniques before the citation analyses would be considered to be robust.

> *Automated process, potentially less burdensome for institutions.*
>
> *This model would require little information from HEIs in terms of an initial submission.*
>
> *This model should be the most cost-effective and successful in reducing the workload for institutions.*
>
> *If the automatic collection and correlation is shown to be robust, this has the potential to be a low burden model, however this is not evident from the pilot.*

Several respondents remarked on the potential for such an inclusive model to reduce incentives for game-playing behaviour, and to provide strong encouragement to HEIs to capture and publish information on all classes of research output and in a more consistent manner.

> *Possibly viewed as more objective, with less intervention from individuals, thereby minimising adverse behaviour.*
>
> *Potentially minimise poaching behaviour between institutions in order to obtain publishing track records.*
>
> *Might result in less perverse behaviour, because the subject mappings don't correlate well to internal units, so not much point in trying to second-guess where to publish! Alternatively, may result in restructuring to try to make things fit, which would be quite inappropriate.*

On weaknesses, respondents were unanimous in signalling their concerns regarding the adequacy of the coverage of the two databases and the reliability of the address-based approach to constructing an institutional record. As a result, respondents see Model 1 as being a high-burden process, and anything but a cost-effective solution to the production of an objective analysis of research excellence across institutions.

> *Whilst this model appears to imply there is no need for institutional data collection, we would want to check the data and this could prove to be a complex process as the data does not map to our internal units. So cost could be high.*
>
> *Contrary to the stated objectives of REF, this model seems to add to the administrative burden of the RAE.*
>
> *Substantial checking would be required to ensure that HEIs were content with the final data set.*
>
> *Given the defects in the operation of this model it does not in fact present a low-cost, low-burden model. Institutions would have to check the data before they could trust its use.*

Very high level of administrative burden for evaluation (checking for false positives as well as missing outputs).

HEIs would wish to check the accuracy of this data. However, the structure of the data does not align to internal systems and extensive work would be required to match outputs via co-author surnames to individual staff. This would duplicate work already done by institutions in collecting their own data for internal management purposes.

There were numerous remarks about the likely behavioural implications of Model 1, and several observations did stand out as being particularly noteworthy, whether that related to the confidence people had around securing academic buy-in to the issue of bibliometrics more generally or the tension that might be created between the promotion of the institution and the nurturing of early-career researchers.

Potential lack of trust in the data mining processes by academics.

Potential for academics not to see the value of maintaining internal records.

Potential for creating a transfer market in publications as well as academic staff; which HEI would be cited if work was done while at one HEI, but paper was published after the author had moved to another HEI?

Potential for creating a culture where non-cited publications (frequently with early-career researchers, junior research staff and research students) would be discouraged.

The inclusion of all staff, all outputs would discourage the nurturing of staff establishing their research careers.

Could encourage production of large quantities, rather than high quality publications.

It has the potential to encourage superfluous increase in citation rates.

Most researchers would put almost all effort in specific journals in UOA.

The publishing behaviour of researchers may be influenced by the journal and subject categorisation of the databases used. This may also have an adverse affect on interdisciplinary research.

If we have to stop people publishing case reports and CPD reviews this will have major resource implications to police and will have a negative impact on our scholarly activity/relationship with the profession.

3.2.2 Perceptions as to the operational qualities of Model 2

As with Model 1, Model 2 is seen as being a potentially cost effective solution. Indeed, several see Model 2 as being the most cost-effective approach as it would use an institution's own bibliographic data and there would be no need to run a separate verification process. This view appears to hold for that sub-set of pilot institutions that already have a comprehensive central record of research outputs.

Whilst in theory this would require the most extensive amount of data collection (bibliographic data for all outputs), in practice the university is doing this anyway.

Both Models 2 and 3 would be cost effective for us, although building staff selection into the process would make Model 3 more expensive.

No work to select papers or research active staff, just need to supply lists of all papers and academics employed.

As institutions are compiling the data, there would be less data-checking at the time of submission.

In practice, most pilot institutions do not have research information management systems that are sufficient to meet the data requirements for Model 2, which suggests that for the great majority of UK HEIs there would need to be substantial further investment in new systems and changed procedures. In that sense, Model 2 is not seen as being practicable in the medium term.

All respondents cited several weaknesses. The majority signalled a concern that the adoption of Model 2 would require very substantial additional effort/investment in the next several years, in order to capture bibliographic data for all research outputs. In addition to concerns about the burden through this transitional period, several respondents suggested that such an inclusive approach would add to the research overhead permanently in that it would require additional work to log a larger amount of bibliographic metadata for all outputs where this happens only selectively at present.

This model involves a large volume of data and has resource implications for HEIs, probably in terms of investing in new systems in order to be confident of having accurate publications data.

This would most likely be the least cost-effective as institutions compile the data and including all outputs would require a lot of work to upkeep databases in the interim.

Large burden of data collection and verification.

HEIs will need to validate enormous volumes of data before supplying. Will involve infrastructure cost and validation cost. HEIs would still want to validate outputs. Cost of checking and validating is enormous.

Extensive data collection and validation exercise would be required. Significant validation of data would be needed to ensure HEIs have submitted all research outputs.

Respondents also raised concerns about the potential negative impact on publishing behaviour. Respondents anticipate researchers focusing on higher impact journals and perhaps reducing the range of types of research output. Several foresee problems for early career researchers.

Most researchers would put almost all effort to publishing in specific journals relevant and important to the UOA.

This model would be likely to encourage publication of work in forms that are more able to be cited.

It has the potential to encourage a superfluous increase in citation rates.

Potential for creating a culture where non-cited publications (frequently with early career researchers, junior research staff and research students) would be discouraged

Might discourage students and early career researchers from publishing as the model will include everything.

The inclusion of all outputs for selected staff would disadvantage staff at an early stage in their careers.

3.2.3 Perceptions as to the operational qualities of Model 3

Model 3 is seen as being the most cost-effective solution under present circumstances, as the following excerpts reveal.

No need to collect additional bibliographic data – seen as lowest burden in this respect.

Most cost-effective model.

More manageable amount of data.

This would be cost-effective, and relatively easy to implement as it is the model most similar to the RAE. Also, it supports the most effective institutional management of submissions.

Likely to be less challenging to adopt than other models, due to similarities with RAE.

Processes already in place to collect data and put together submissions.

Model 3 is also seen to have some additional positive features, which include being an incentive to academics to concentrate their written output, producing fewer, but stronger papers published in higher profile publications. Moreover, Model 3's selection of papers avoids creating the situation where early-career researchers are encouraged to publish only as second authors, as their early work, which will tend to be less widely cited, does not pose any threat to an institution's excellence profile.

Will act as an incentive to publish excellent papers in high profile journals

This model only looks at the best papers so the results are not skewed by early work.

This model would encourage HEIs to continue to nurture staff establishing their research careers.

The model enables HEIs to make strategic decisions about the work that they wish to be assessed.

Incentive to maintain our research data in the best possible way.

On weaknesses, several respondents take the view that Model 3 will require new investment as well as the implementation of new processes in order to ensure a much greater proportion of their research output is indexed by the HEI's central bibliographic databases. In short, picking the best four papers for any author requires some overview of all of his or her publications.

This view appears to revolve around the assumption that the selection of papers would be made by the HEI using citation analyses to simulate the analyses that would be run by HEFCE. Using citation analyses means institutions will want to run these analytical exercises with much larger samples of publications.

These organisations take the view that bibliometrics modelling using one or both databases (Scopus, WoS) is likely to produce a different set of papers as compared with a portfolio built bottom-up by authors and colleagues. The experience of the pilot has led people to conclude that an author-constructed list might earn the institution fewer credits in a national assessment exercise for two reasons: it is very likely that a proportion of submitted papers will not be mapped to either database; and it is very likely that a proportion of those papers will be less well cited, diluting the overall institutional scores/credits.

May cause significant additional work in terms of selection of outputs, if we have to have regard to academic quality of outputs, citations profile (not necessarily the same as quality!) and evidence of impact. Could lead to academics chasing very high citation counts through inappropriate behaviours. Could influence choices about form of publication (i.e. encourage people to publish in "citable" form) and discourage applied or knowledge–transfer (KT) work, which might generate lower citation rates (but meet the impact agenda).

Highly cited papers will not necessarily produce the same corpus of publications that a strategic selection of outputs by the institution would choose.

Much more work would need to be done to understand the significant differences between the results from the two data sources. However a comparison between those selected by us for RAE submission and those selected by these means highlights concern that peer-review would take a very different view. For example a numerical-citations selection biases towards large multi-site clinical trials whereas our (peer-review) selection would bias towards those papers where the intellectual leadership came from our authors.

Six respondents stated that the selection of papers ought to be carried out by the HEIs. None of the respondents made any explicit remarks about HEFCE making the selection. One respondent noted that the operational implications depend on how the outputs are to be selected, and by whom. Several respondents suggested that the selection should involve both citation analyses and academic advice, as the REF submission will comprise both citation data and actual papers, which will inform assessment by a peer-review panel. While two other respondents take the view that the schools should continue to make the selection in much the same way as was done for the RAE.

Various behavioural issues are anticipated.

Could lead to (unwanted) changes in publication behaviour, with outputs directed away from outlets, which are not indexed by the database used for the REF, if bibliometric indicators were seen as having a strong influence on REF outcomes.

Most researchers would put almost all effort to publishing in specific journals relevant and important to the UOA.

It has the potential to encourage superfluous increase in citation rates.

Similar behaviour encountered in the RAE may still occur, for example institutions poaching staff before the relevant date.

Potentially encourages poaching behaviour between institutions.

3.2.4 Discussion

There appear to be several important differences between each of the three bibliometrics models, in terms of their operational implications.

At first appearance, Model 1 and Model 3 would require HEIs to do very little: the former could be carried out by HEFCE without their direct input and the latter is close to the RAE process. By contrast, Model 2 would require the majority of HEIs to upgrade their research information management systems.

In practice, however, it seems very likely that the actual differences would be much less pronounced, and that all three will oblige HEIs to press forward with their efforts to bring their central research information management systems up to scratch.

On Model 1, pilot institutions were unanimous in the view that the entire HE community would insist on being involved fully in the validation of the bibliographic databases for use in the REF citation analyses. In order to participate in such a process, however, HEIs would first need to have comprehensive bibliographic information on the majority of their research outputs, for most of their staff. The experience of the pilot suggests that for most universities and colleges, such a process would require substantial further development of their information systems.

There seems little prospect that the database coverage and mapping issues, which so compromised the acceptability of Model 1, will be improved sufficiently by 2012 for HEFCE to even consider moving forward with Model 1. The database situation is

improving steadily and it is possible that things will have improved sufficiently by the time we come to implement the second full REF, perhaps in 2018, for the community to invite the funding bodies to switch to a more top down procedure.

On Model 3, the pilot institutions are certain the wider community will demand that individual HEIs be allowed to select the authors and papers for the bibliometrics work. This remark masks an expectation that choosing papers would be done by mimicking the REF citation analyses, in order to be confident of the optimal bibliometrics results being fed into the relevant peer-review panels. The pilot confirmed that in a good proportion of cases, a bibliometrics-driven process would produce a slightly different basket of 'best' papers as compared with a peer-review panel within an HEI department. Picking the most widely cited and highest impact papers, as determined by WoS or Scopus, cannot be done easily through a bottom-up process.

Model 3 is still likely to be the most tolerable, as it leaves control over the publications lists with institutions rather than the funding body and the scope of the development implied feels manageable. Ultimately, Model 3 goes with the grain of current policy and systems development.

Any new performance measurement system is likely to change behaviour, however the fact that these new metrics will inform a peer-review judgement of research quality has clearly lessened concerns that it will produce obvious and strongly negative changes. However, respondents foresee both positive and negative impacts on behaviour with each of the three models, and therefore signalled the need for HEFCE to keep the situation under review.

There is an evident split in expectations between the two comprehensive models (Model 1 and Model 2) and the selective model (Model 3). The comprehensive models were seen as positively reinforcing the notion that all research outputs should be as good as they can be. On the downside, people anticipate a possible reduction in the diversity of the types of research outputs, perhaps to the detriment of relationships with non-academic communities. In addition to narrowing the communication bandwidth, M1 and M2 might cause HEIs to discourage early-career researchers from publishing except as secondary or tertiary authors. Prohibiting the direct exposure of early work and ideas to external debate and third-party peer review could inhibit, and even stunt, the personal and professional development of tomorrow's leading researchers.

Model 3 by contrast appears less dramatic, being closer to the present RAE arrangements. The positive points are that it will reinforce current incentives for academics to seek to publish in higher impact journals and it seems unlikely to be a disincentive to publishing by early-career researchers. The negatives are a mix of old and new; Model 3 looks set to continue to encourage the poaching of key staff, although it is not clear how extensive or problematic this is; and there is concern that it will encourage people to focus on that subset of journals indexed by Scopus and WoS.

On balance, Model 3 seems to offer the best fit with the high-level policy objectives for the REF bibliometrics element, which is to say it offers robust indicators of research quality at the institution level while avoiding the creation of major new negative or perverse behavioural incentives.

3.3 Notable differences between the two databases

Question 3 invited pilot institutions to provide their views on any notable differences between the two databases, when looked at in the round.

The great majority of respondents stated that Scopus produced a wider coverage overall, as compared with WoS, which produced higher citation levels. This assessment is based on each institution's own submissions, and several respondents noted that the differences in coverage were better or worse depending upon the UOA in question.

The two databases give rise to different numbers of outputs. In some cases Scopus appears to have a wider coverage (up to UOA 14, very broadly speaking). Thereafter WoS appears to have wider coverage.

Generally there was greater coverage in Scopus for all models, but the differences were lesser or greater depending on the subject area, e.g. UOA 4 – 130 outputs found on Scopus and 3 on WoS when looking at papers written by staff while employed at HEI.

Different results depending on which UOA you look at, no 'one-size-suits-all' approach possible, although WoS had slightly better disambiguation in a few UOAs than Scopus.

Scopus: significantly higher degree of matching (approximately 75%) to submitted outputs achieved in most subjects, particularly across UOAs 1-14 in which the college made REF pilot submissions. Slightly higher degree of differentiation between pilot outcomes.

WoS: Lower degree of matching (approximately 65%) to submitted outputs in most subjects. Higher degree of matching in UOA 23 – computer science and informatics (60% compared to 40% in Scopus). Comparatively low matching in UOA 6 – epidemiology and public health (50% compared to 75% in Scopus).

There was a fuller coverage in Scopus for some of our social science subjects, such as sports related studies and social work and social policy & administration. However, in general, both databases gave more coverage to the sciences.

A number of respondents kindly shared their internal analyses, which help to illustrate the situation with respect to the two databases. Figure 1 shows an analysis of the variance between the internal and external bibliographic databases, for one pilot institution and a sub-set of units of assessment. It serves to illustrate the variability in coverage across subjects, ranging from a low of 43% to a high of 96% and a median of around 85%. It also suggests that in these natural sciences and medical research areas, Scopus consistently mapped a larger share of the institution's total research output in the period. The view for this single institution appears to fit the view from the pilot institutions more generally.

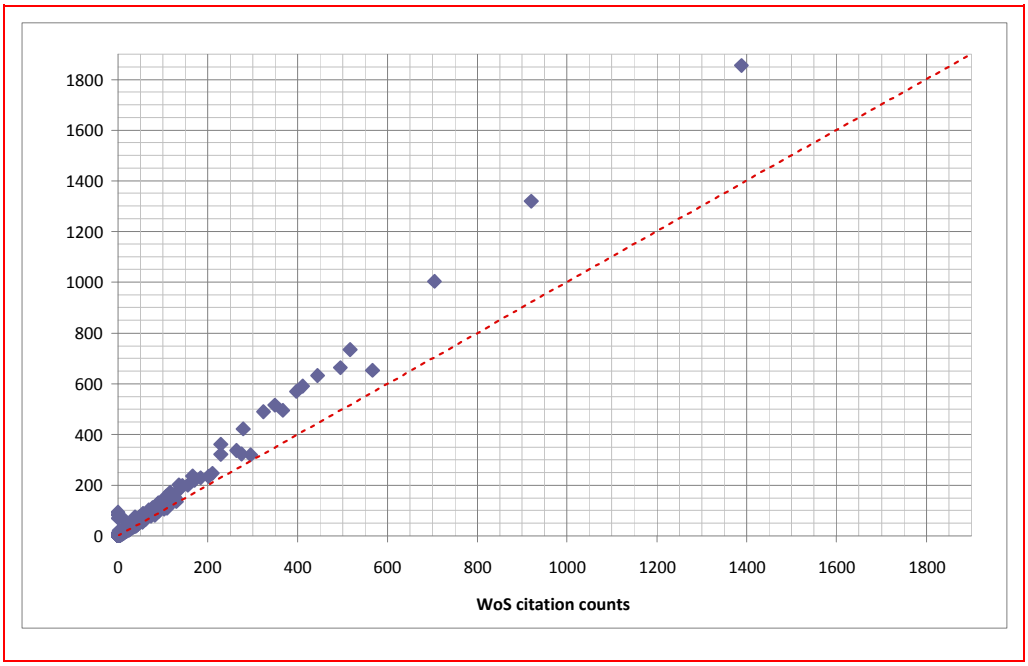
Figure 1 Analysis of HEI total research outputs captured by internal and external databases, for a selection of units of assessment (2001-2006)

UOA	UOA name	SYMPLECTIC	Scopus		WoS		VARIANCE
		total no. of unique outputs per UOA	outputs all staff	%	outputs all staff	%	between Scopus & WoS
1	Cardiovascular Medicine	1570	1271	81.0%	1066	67.9%	205
2	Cancer Studies	913	785	86.0%	492	53.9%	293
3	Infection & Immunology	899	802	89.2%	717	79.8%	85
4	Other Hospital Based Clinical Subjects	846	754	89.1%	579	68.4%	175
6	Epidemiology & Public Health	619	490	79.2%	387	62.5%	103
9	Psychiatry & Clinical Psychology	225	170	75.6%	128	56.9%	42
10	Dentistry	288	238	82.6%	220	76.4%	18
12	Allied Health Professions (Nursing)	58	32	55.2%	25	43.1%	7
14	Biological Sciences	1573	1508	95.9%	1361	86.5%	147
16	Agriculture, Veterinary / Food Science	1219	1044	85.6%	905	74.2%	139

Source: Pilot institution's internal analyses

Figure 2 shows a comparison of citation counts for the same pilot institution for cardiovascular medicine, with Scopus counts shown on the 'x' axis and WoS counts on the 'y' axis. The table suggests Scopus is indexing a larger number of journals in this field and is capturing 30-40% more citations for the same paper, as compared with WoS, at least for the most cited papers.

Figure 2 Comparison of citation counts across Scopus and WoS, for UOA 1 (cardiovascular research) for a single pilot institution (papers published 2001-2006)



Source: Pilot institution's internal analyses

Several respondents noted other concerns, such as the poor or unacceptable performance of both databases when using the institutional address model or the relatively greater importance of correctly mapping journals to subject categories or the more defined citation profile produced by WoS.

> *Both databases performed poorly on the address-based model. Scopus appeared to perform better for engineering disciplines, as expected. The address-based model yielded more results in WoS than Scopus, the quality of the results indicating either the subject categorisation or address mapping was inadequate or poor.*
>
> *The correct assignation of journals to subject categories seems to be critical to the assessment process. We feel that this issue is as important as the coverage of each database. We did not find evidence of one database being consistently "better" than the other one. Their coverage also seems to vary across disciplines.*
>
> *WoS tends to provide a more defined citation profile*

3.3.1 Discussion

The pilot has confirmed there are differences in coverage between the two databases, subject by subject. This outcome suggests that the bibliometrics element of the REF should make use of both databases ideally in order to maximise HEFCE's ability to capture institutions' most highly cited work.

In practical terms, coverage and subject mapping are more or less problematic for different UOAs, and HEFCE is likely to err on the side of caution by insisting on a bibliometrics element only for those subject areas where the databases already align reasonably well with the core of a discipline's high quality research outputs. In other areas, it seems appropriate to allow the community – and the REF peer-review panels

– to signal whether or not it believes bibliometrics amounts to a useful additional source of intelligence regarding international performance and impacts.

The coverage issue is likely to become less problematic in time as both databases index more and more publications and as researchers consciously target a proportion of their output on higher impact indexed titles.

3.4 **Accuracy at the level of Units of Assessment (UOAs)**

3.4.1 Rating of each unit of assessment

Question 4 returned to the issue of accuracy, this time from the perspective of the individual UOAs. Analysis of these data was intended to reveal any specific subject areas where pilot institutions had collectively expressed reservations.

For each UOA where they had made a return, institutions were invited to indicate the extent to which the result was aligned with their view of the HEI's standing. Respondents were invited to rate the degree of alignment using a 1-5 scale, where 1 indicates a result that significantly understates performance (to the HEI's detriment) and a 5 is significantly overstated).

Respondents found this question difficult to answer and many elected not to provide a rating, although in some cases pilots offered explanatory remarks without scores (and vice versa). Several respondents ascribed their decision to not answer the question on the limited explanatory data that accompanied the results.

> *Given that graphs compiled showing best and worst performing institutions based on data collected were anonymised, it is difficult to say whether or not we performed as expected in the pilot project. If we had known where we were placed within the group of institutions taking part, we may have been able to compare this to RAE2008 results for example, to check if we ranked similarly by UOA.*

Notwithstanding this challenge, respondents returned a total of 103 votes across the 35 UOAs, and Figure 3 shows the distribution across the five possible ratings. A score of 3 indicates a good match between the bibliometrics and the RAE2008 results, and a score of 2 or 4 indicates some over or understatement of performance. A score of 1 or 5 is an indication that the accuracy of the result is unacceptable, and on this analysis that holds in 20% of cases. On the other hand, the analysis suggests that the results are good to reasonable in the other 80% of cases.

Lastly, it is evident from the distribution that respondents see the bibliometrics as tending to understate their true standing with regard to research excellence. This is possibly the result of natural bias; as is often noted, a majority of us believe we are better than average drivers. It might also reflect the inclusion of Models 1 and 2, which are more comprehensive than was Model 3 and as such tend to produce lower 'scores' than Model 3 or the RAE on the chosen citation indicators (e.g. proportion of outputs assessed achieving better than x2 world average citation score for the field).

Figure 3 Distribution of scores across all UOAs

Score	Number of scores	Share of all scores
Total 1s	15	15%
Total 2s	41	40%
Total 3s	23	22%
Total 4s	19	18%
Total 5s	5	5%
Total all votes	103	100%

It did not prove possible to carry out any systematic analysis of opinions at the level of the UOAs, as we secured five or more ratings in just eight of the 35 UOAs. Looking

briefly at these eight UOAs, UOAs 14 (biological sciences) and 21 (applied mathematics) stand out as the two where the largest proportion of people, still small numbers, rated the results as unacceptable in terms of their accuracy (all understated).

3.4.2 Comments on outliers

In addition to inviting respondents to rate their results for each unit of assessment, Question 4 invited people to provide a comment on any UOAs where there were significant differences (outliers) between the bibliometrics results for any of the three models and other evidence of performance in the field in question.

In a majority of cases, respondents did provide a comment on those UOAs they believed to be problematic but typically referring back to the results for Model 3, their preferred model. The reasons were essentially threefold: poor coverage of the field by one or both databases; undue influence of individual academics in areas where the institution had a small number of papers; and heterogeneous publishing behaviour within a single UOA.

> *Looking at the graph showing WoS and Scopus results and RAE outputs 3* and 4*, it is clear that a lot of outputs in nursing & midwifery are not accounted for in these databases. This might hold for much of the health policy work carried out in this area.*
>
> *Poor match because in REF pilot, only included 20% of RAE submission.*
>
> *No correlation at all. Possibly due to low extent of journal coverage and type of outputs.*
>
> *Output types not suitable for citation analysis.*
>
> *No correlation. Discipline publishes heavily in conference proceedings.*
>
> *Discipline publishes heavily in conference proceedings. There will be large errors associated with the low citation rates in this discipline. Scopus appears to correlate more closely with RAE scores than WoS.*
>
> *Smallish numbers, and one or two individuals had a significant impact on the average score.*
>
> *Top 6 papers were significantly better than expected across the board. This UOA included some prominent academics with papers that have achieved a greater proportion of citations compared with the field.*
>
> *There are obvious outliers, which may reflect the diversity of the discipline; from physical to human geography and social sciences.*
>
> *Applied mathematics has a large number of sub-disciplines, which display very different citation behaviours and approaches. It is likely that the normalisation is not at a suitable level to reflect this, and that the overview graphs show 'noise' from the sub-disciplines.*
>
> *Differences might be due to the sub-disciplines in UOA 44 psychology, which covers different journals with different citation behaviour.*
>
> *Outliers might be explained by highly diverse UOA, but can't tell due to anonymised institutions (agriculture, food and vet science which all publish diversely).*
>
> *Looking at our own anonymised data there appears to be a poor correlation between RAE expert peer review and any of the measures derived from citations data. This might be explained in part by the greater proportion of interdisciplinary research in computer science departments.*

In the very great majority of cases, the comments related to explanations as to why the bibliometrics results were unreasonably lower than the RAE profile. In just two cases,

respondents noted that the bibliometrics – Model 3 – had returned a better performance for the institution than had the RAE, and both were content with this and the relative position of the organisation within the sub-set of pilots in that UOA.

> *Higher relative position than RAE results but comfortable with this as more weighting in REF given to citation (and thus the application of our research which we know to be high) than in RAE.*

3.4.3 Discussion

The consultation produced too few data points to carry out any systematic analysis at the level of the UOAs, however it did produce some useable statistics overall.

This analysis suggests that the bibliometrics results were judged to be 'good to reasonable,' in terms of their accuracy, in 80% of 100+ cases where ratings were given.

Where the statistics show concerns about accuracy, it is evident that respondents see the bibliometrics as tending to understate their true standing with regard to research excellence. Comments suggest that such negative assessments vary across subjects and between institutions within subjects, at which point people foresee the possibility of winners and losers.

The pilot institutions' explanations of the outliers were essentially threefold: poor coverage of the field by one or both databases; undue influence of individual academics in areas where an institution had registered a small number of papers; and heterogeneous publishing behaviour within a single UOA.

3.5 **Database coverage at the level of Units of Assessment (UOAs)**

3.5.1 Rating of coverage of each Unit of Assessment

Question 5 returned to the issue of database coverage, this time from the perspective of the individual UOAs. Analysis of these data was intended to reveal any specific subject areas where pilot institutions had collectively expressed reservations.

For each UOA where they had made a return, institutions were invited to indicate the extent to which the bibliometrics analyses (Models 1, 2 and 3) had included the institution's best quality research. Respondents were invited to rate the coverage using a 1-5 scale, where 1 indicates very poor coverage and 5 is excellent coverage.

Respondents struggled with this question to an even greater extent than with the analysis of accuracy by UOA and again in relation to limited information. In this instance, just 88 scores were recorded. Figure 4 shows the distribution across the range of possible scores, with around two-thirds of the scores suggesting the coverage of quality work had been reasonable to good (3s, 4s and 5s) and a third suggesting it was poor to unacceptable (1s, 2s). There were too few scores to analyse the feedback at the level of individual UOAs, however in two cases, computer science and social policy, the scores were entirely 1s and 2s.

Figure 4 Distribution of 'coverage' scores across all UOAs

Score	**Number**	**Share**
total 1s	11	13%
total 2s	16	18%
total 3s	31	35%
total 4s	27	31%
total 5s	3	3%
total all votes	88	100%

3.5.2 Comments on 'coverage' outliers

The majority of comments provided related to those UOAs where people deemed the coverage to have been unacceptable, in as much as this could be determined. Most of those offering comments attributed the poor coverage to an intrinsic problem with

bibliometrics inasmuch as the discipline concerned tends to publish more often through media and channels other than journal articles.

One respondent usefully clarified the assumptions underpinning their scores, raising a point of difference with the majority. This respondent judged the databases on their coverage of the institution's journal articles and reviews, where most others score and comment on the bibliometrics' coverage of a discipline's total research output, from papers to conference proceedings to books.

> *Please note that the information entered for coverage denotes the degree of matching to journal articles and reviews only. There is no measure of coverage of other types of high quality research outputs. The values have been entered according to the following criteria: 1 is ≤ 55%; 2 is > 55%; 3 is > 65%; 4 is > 75%; and 5 is > 85% coverage. The average coverage of the two databases has been used. No explanation of poor coverage has been entered as the reasons for the low frequency of mapping/inclusion in the commercial databases is not clear.*

The following excerpts provide further insight into pilot institutions' thinking on poor coverage of their quality research, which mostly revolves around differences in publishing behaviour across disciplines, where for example, computer science makes extensive use of conference proceedings and much less use of journal articles and reviews. There are various other types of publications and communication channels referred to, from monographs and book chapters to online databases and archives. There is also a sense expressed by several respondents that strongly applied and inter-disciplinary areas are poorly covered.

> *Neither of the databases picks up special publications or books, which are significant outlets for this UOA.*
>
> *Sports studies can publish in sports science areas or more social science, policy areas, grey literature (commissioned reports, etc.), and historical journals and monographs. This may account for the very poor coverage found in this discipline.*
>
> *Moderate coverage. Mathematics traditionally served by A&I databases MathSciNet (Maths Reviews from the AMS) and Zentralblatt MATH, with foreign language publications (Russian, French, Japanese) not served by WoS. Scopus covers these disciplines better. Also, citations in these disciplines expected to take longer due to nature of the subject*
>
> *Likely to be limited coverage of human geography and social science area of discipline.*
>
> *Computer science was badly affected across the board as conference proceedings are not captured, and these are a very common output form for this discipline.*
>
> *Economics does not seem to be widely covered as a discipline by either Scopus or WoS*
>
> *There is a strong culture of repository online archives, which are not covered by the databases. Furthermore, it is citations from these sources (such as arXiv) rather than published papers that generate the essential impact. Astronomers have found that coverage of outputs is better in the database ADS (Astrophysics Data System) than WoS. Some sub-areas publish work as technical reports, design studies and conference proceedings; these will not be included in the analysis.*
>
> *Within our units of assessment, WoS and Scopus produce different results for some UOAs both in terms of inclusion of highest quality research outputs and coverage. Without a comprehensive analysis of why that is, we suggest that panels have access to both.*

Very poor coverage. Hugely interdisciplinary area, only two pilot submissions and very small sample of publications. Poor performances even by material classed as reviews. Not unexpected due to multi-disciplinary nature of material, and diverse publishing sources, namely reports.

3.5.3 Discussion

As with accuracy, the consultation produced too few data points to carry out any systematic analysis of database coverage at the level of the UOAs; however it did produce some useable statistics overall.

This analysis suggests that the bibliometrics results were judged to be 'good to reasonable,' in terms of their coverage of HEIs' highest quality work, in around two thirds of the cases where ratings were given.

The majority of comments provided related to those UOAs where people deemed the coverage to have been unacceptable, in as much as this could be determined. Most of those offering comments attributed the poor coverage to an intrinsic problem with bibliometrics in as much as the discipline concerned tends to publish more often through media and channels other than journal articles.

3.6 **Estimated costs to develop the different models**

3.6.1 First-approximation estimates of development costs

In Question 6 of the survey, respondents were asked to provide an approximate estimate of the likely total development costs/effort the institution would incur in order to implement each of the three models. The estimates included both external purchases (e.g. capital investment, licences and bought-in development expertise) and internal development effort (e.g. staff costs).

This is a challenging question given that the final specification for the REF bibliometrics is yet to be finalised, however eight pilot institutions did offer an estimate. Of those that did not provide an estimate, most remarked that they felt unable to offer a figure without further details on the REF bibliometrics requirements. Figure 5 shows the estimates for the eight institutions for each of the three models, internal and external costs, as well as a total cost for Model 3.

The first point of note is that the majority of respondents that provided an estimate take the view that the combined costs are likely to be very similar no matter which model is followed.

The next point of note is that the estimates covered a broad range from a low of £0, for two institutions that believe they have systems in place to deal with the likely information requirements, to a high of £460K. The median was around £100K. Differences in estimates will no doubt reflect the current state of development of an institution's research information management systems and institutional size, however there are too few data for any meaningful segmentation on these grounds.

In most cases, the bulk of the cost is associated with internal staff time. Estimates of staff days have been converted into pounds sterling using an average annual cost for an administrator of £40K for a 220 person day year.

Figure 5 First approximation estimates of costs to implement each of the three models (£K)

	M1 External	M1 Internal	M2 External	M2 Internal	M3 External	M3 Internal	M3 Total
Institution 1	60	400	60	400	60	400	460
Institution 2	30	140	30	140	30	140	170
Institution 3	2.5	120	2.5	120	2.5	120	122.5
Institution 4	50	50	50	50	50	50	100
Institution 5	65	40	65	40	65	40	105
Institution 6	50	30	50	30	TBC	30	80
Institution 7	0	0	0	0	0	0	0
Institution 8	0	0	0	0	0	0	0

Source: Pilot institutions' internal estimates.

3.6.2 Comments on estimates of development costs

In light of the fact that these estimates can only be a first approximation, respondents were invited to explain any assumptions they had made or otherwise offer explanatory remarks regarding estimates.

Comments were varied with few clear patterns, beyond a sense that the costs would be broadly similar across models and that the full implication for costs cannot be determined until the final REF bibliometrics data requirements have been specified. As this will no doubt be a topic of general interest to a majority of readers, the explanatory remarks are represented here in full.

> *Although Model 1 is theoretically lower cost than Models 2 and 3, we would want to check any data derived using an address field search in bibliographic databases and therefore would want to collect on an institutional basis the same level and kind of data as for Models 2 and 3. There is likely to be little difference in cost between Models 2 and 3 as the institution may want to see an author's total research outputs before selecting those for the REF. The data may also be required for other research management purposes. Enabling WoS to work with EPrints to provide real-time citation data – £2,500. Development of internal systems to make them fit for purpose – c. £100,000. Retrospective population of publications database, 6 months Grade 6 cataloguer – £17,765.*
>
> *Possibly licence costs to monitor citations performance (e.g. In-Cite at $40,000 pa), dependent upon whether HEFCE negotiates sector wide arrangements. 80% of grade 8 between Jan 08 to present (approx £45,767), already incurred, plus possible 30% to work on citations data, Spring 2010 – Autumn 2010 (approx £9,059). Might wish to invest in an additional post as "back-up" to current senior analyst/systems developer (grade 8, approx £50k pa).*
>
> *Institutional repository costs and subscriptions to ISI and Scopus are already incurred. Systems are already functioning and form part of budgeted costs.*

Licensing costs and other purchases can only be estimated once REF parameters are known, so that it is possible to assess whether the costs and efforts of implementing a system are justified.

It is impossible to assess this without details of REF requirements. We would wish to run a REF model with citation experts overseeing processes to make sure that we maximise our return and REF will cost more in terms of this continuing effort.

Our estimates are calculated on the basis that the 'real' REF has very similar data requirements to the pilot REF.

I am not sure I can answer this for each model. We did not have a publications database and are now releasing the first version of one which would serve our purposes whatever the model used I hope. This has been developed in house with our IT department. I estimate it has taken 1 person two months to develop (£9–10K in staff time).

Model 1 requires the lowest expenditure working on the assumption that database trawling will be carried out by the funding council. However, the data submitted will be checked by the institution and this would require the highest number of staff days as data-checking in the pilot took considerable effort, with most of the additional publications assigned to our institution being investigated and subsequently rejected. Model 2 would require the highest investment and also a considerable number of staff days as we would identify the staff and the publications. At present we do not have a Scopus subscription, for example, and areas like this would need to be addressed to ensure we could locate all appropriate publications – this alone would cost us around £17,000 per annum.

Model 3 would require a lower expenditure than Model 2, but higher than Model 1 as this would involve internal location of information for submission, but as this covers selected papers, less expenditure would be required. This would also require the least number of staff days of all three models, as it is something quite similar to the RAE, so a familiar process to what has taken place in the past would be needed. It is difficult to put a precise figure on each of these questions. The pilot process does not offer much of a clue as we only returned data for selected UOAs. We stated 80 working days for the pilot project, so maybe this should be taken as a minimum starting figure. 80 days would cost around £32,000 and we would anticipate extra staff would be required as per the RAE. If we were to cost this work for 18 UOAs (as submitted to the 2008 RAE) it would cost around £72,000.

This is impossible to answer given that any development would be intrinsically linked with other ongoing research information management systems developments, the total cost of which may be anything up to £5M. However, we can be confident that internal staff effort would be likely to increase from Models 1 through 3.

These figures are derived from costings for similar projects. It is felt that the initial development will require this approximate level of resources, irrespective of the model.

One respondent provided a detailed explanation of likely costs, which readers might find useful in gauging the range of expenditure items and staff inputs, although the final tally is not necessarily generalisable.

Approximate cost of staff time for preparing submission in current form and with existing university systems = e.g. £600K admin, library and management, plus approximately £250K individual effort by academics (e.g. 10 hours each for 500 research active academics). Total: £850K

Additional costs of validating citation data:

- Assume bibliometrics/bibliographic data management software acquired from eprints and Symplectic = e.g. £13K
- Assume purchase of 2x IIS servers required to run Symplectic system = e.g. £2K
- Assume citation data and license purchased from WoS = e.g. £15K
- Assume web interface developed in house – 1 full-time equivalent (FTE) web developer for 6 months and 0.5 FTE technical support for 12 months = e.g. £90K
- Assume 1 hour additional work per research active academic and 1 hour extra work per academic for both library and administration = e.g. £44K
- Assume 3 hours additional academic management and 1 hour additional senior management per submission (x20) = e.g. £5K
- Approximate total additional costs = e.g. £169K (19% increase)

No differentiation in costs between the three models is possible with the information available. The cost would depend on what action is required by the REF process and the role of the data in the exercise.

3.7 **Estimated share of additional development costs**

Several of the pilot institutions had noted they were already embarked on substantial development of their research information systems. In light of this fact, Question 7 invited all pilot institutions to estimate the likely additional costs of any system development, which might be attributable to compliance with a future REF bibliometrics exercise. This was another challenging question, and nine respondents felt able to offer an estimate.

Figure 6 shows the range of responses to this question, anonymised and sorted by estimated share of additional costs. It shows a broad range, from no additional cost to fully additional, with a median estimated additional development cost of 33%.

Figure 6 Estimates of additional share of future systems development costs needed to implement Model 3 (%)

Institution	% development costs
Institution 1	100
Institution 2	100
Institution 3	80
Institution 4	50
Institution 5	33
Institution 6	10
Institution 7	10
Institution 8	5
Institution 9	0

3.7.1 Discussion

A significant minority of respondents were able to provide a first approximation of the likely total development costs the institution would incur in order to implement each of the three models.

Given the provisional nature of the information on likely final requirements, the great majority of respondents provided a single estimate of development costs, independent of the final bibliometrics model adopted.

The estimates covered a broad range from a low of £0, for the two institutions that believe they have systems in place already to deal with the REF information requirements, to a high of £460K. The median estimate was around £100K, but this might be an unduly conservative working estimate of the likely development cost/HEI, given the range of estimates and the fact that the majority felt unable to offer a realistic estimate.

This spread is an important reminder that the actual cost will reflect the current state of development of an institution's research information management system and institutional size.

In most cases, the bulk of the cost is associated with internal staff time with a split of around 60:40 between internal and external costs.

We did invite people to estimate the additional cost of complying with the REF bibliometrics requirements, over and above current investment plans. Here again, there was a full spread of responses, with the great majority stating that costs would go beyond current or anticipated investment plans. The range of the estimated share of additional costs was 5-80%, with a median of 33%.

3.8 **Estimated cost to run research information systems**

Question 8 invited pilot institutions to estimate the likely additional running costs that might be attributable to an upgraded research information management (RIM) system. As with the first two financial estimates, this proved to be another challenging question, and just seven respondents felt able to offer an estimate.

Figure 7 shows the range of responses to this question, anonymised and sorted by total additional running costs. In most cases, respondents provided similar estimates for the running costs for all three models, and therefore Model 3 figures are presented here for reasons of brevity.

The table shows a broad range, from no additional running costs to a high of £160K a year, with a median estimated additional running cost of £75K a year. The table also reveals that in most cases internal costs are expected to be greater than external costs (for licences and such like). As with the estimates for development costs, the figures presented here will reflect differences in the current state of development of individual RIM systems and the size of the institution.

Figure 7 Estimates of annual cost to run/maintain research information systems needed to implement Model 3 (£K)

Institution	Model 3 External	Model 3 Internal	Model 3 total
Institution 1	65	95	160
Institution 2	-	100	100
Institution 3	30	50	80
Institution 4	5	68	73
Institution 5	15	60	75
Institution 6	10	10	20
Institution 7	0	0	0

As before, respondents took some care to explain their reasoning and in several cases to detail the manner in which they had built up their estimates. These responses are provided below in full for readers' information.

> *The university already has subscriptions to both Scopus and WOS, and to e-Prints, and will continue with these regardless of the REF and the model chosen. The internal costs would be part of existing staff costs,*

and constitute additional reporting to Senior Management on the bibliometric performance of academics.

It is unlikely that we will invest in new information systems as we are satisfied that the system we already have is fit for purpose although some refinements would enhance the effectiveness of the system.

Institutional subscription to Scopus: c. £38,500 per annum. Enhanced subscription to WoS (to allow us to model citation data): £26,240 per annum. Ongoing cataloguing of publications data, 1 cataloguer at 50% Grade 4 and 1 cataloguer at 50% Grade 3 – £23,490 per annum. 40% of Grade 5/6 for Outreach/Faculty follow-up and day-to-day management of publications and staff £12,400 per annum. 1 day per week for 9 Faculties at Grade 5/6 c. £50,000 per annum. Ongoing maintenance of systems £12,000 per annum [0.25 FTE of Grade 8 analyst/developer].

Recurrent costs will more than likely be the same across the board because we would still want to rely on our own systems whichever model is chosen.

Given the uncertainties in the data, we would need to keep parallel data and also have the ability to link citations to authors and to work out rebased citation impact. This means buying data and analysis tools. The data is not static so would need frequent updating.

This is very difficult to gauge so these are ballpark suggestions. It is clear however, that regardless of which model is implemented, a comprehensive publications database will be necessary to follow good practice in the institution.

£5,000 per year external costs for maintaining and upgrading hardware/software systems. The internal costs increase from model 1 to 3, reflecting the increasing demands on academic and administrative time required to check and select authors, papers and data accuracy.

Companies such as Thomsons and Elsevier offer citation analysis services but there are potentially significant costs to the sector if each institution is purchasing enhanced access to WoS, taking out subscriptions to Scopus etc. HEFCE might want to consider how these costs might be defrayed – e.g. national licences for these resources.

3.8.1 Discussion

A significant minority of respondents was able to provide a first approximation of the likely running costs the institution would incur in order to implement each of the three models. In a majority of these cases, respondents provided similar estimates for all three models.

The figures show a broad range, from no additional running costs, to a high of £160K a year, with a median estimated additional annual cost of £75K. The data also reveal that in most cases internal costs are expected to be much greater than external costs.

As with the estimates for development costs, the figures reflect differences in the current state of development of individual RIM systems and the size of the institution.

3.9 **Implications for research information management systems**

Question 9 invited pilot institutions to comment on the implications of a REF bibliometrics element for their RIM systems.

Several respondents stated that participation in the pilot had provided a useful boost to institutional thinking with respect to central systems, and that it had already lead to improvements in RIM. Further improvements are anticipated, which respondents believe will improve their institution's capacity for strategic management.

We believe that the development work we have undertaken over the last 18 months has already improved our Research Information Management system, by making our publications database more robust and more easily populated, and providing significant additional management information on which we can now draw. We would view likely future activity to be around refining these improvements, and possibly building in the capacity for more detailed citations work to be undertaken locally, but the latter will depend upon the future formulation of the REF.

As a result of the pilot, it was decided that there was a need to overhaul our systems in order to develop a flexible approach to the management of research information. The goal of this initiative is to improve the way we store, manage and report our research data. It will also allow us to enhance the visibility of our research.

Several respondents stated that decisions are on hold until the final requirements of the REF bibliometrics exercise are announced. In turn, one respondent signalled a note of caution, arguing that HEFCE will need to finalise its data requirements in light of the external reporting required by other funders. HEFCE should avoid becoming so prescriptive as to make the REF requirements wholly additional to the research information required more generally in corporate planning or indeed for other major research funders.

Awaiting final REF guidelines to be published so that internal data management and reporting mechanisms can be established and final REF format requirements so that e-Prints can develop a reporting tool for our institutional repository.

The majority of respondents do foresee the need to continue with their efforts to develop stronger, central RIM systems with better cross-linking to other databases in finance and human resources (HR). The many and various comments provided by pilot institutions are reproduced below.

Increased need for a 'cradle-to-grave' research management system, which interfaces with other institutional systems. Possibility of HEFCE negotiating sector-wide subscriptions to WoS and Scopus at a reduced price. Internal research systems need to be made more efficient and useful to academics, i.e. information is only provided by the academics once and is then used to populate all necessary systems.

The university is reviewing its Research Information Management systems in order to align data capture across a range of functions to improve the quality of general research management information for both planning and monitoring of activity and resource. Resources, systems and staff are located across a number of established and separate functions (Research Office, Library, ITS, Finance, Academic Departments etc) and these will all be required to support an enhanced and coherent Research Information Management system. Requirements for the REF will be a key input into development.

Our internal database, which was used to gather data for the RAE submission, Content Management System (CMS), a web database, could present a viable system for an improved Research Information Management system to include REF data collection requirements. System development would be needed to achieve improvements needed in front end forms and reporting capability and this will require dedicated resource across a range of units.

The pilot has underlined the need for the university to have a central publications database which links with other university systems such as HR and the central University Research System. Such a system is under

development. It is likely that the university will want to model citation data in advance of the REF, and HEFCE might want consider how institutions can be helped to do this.

Research Information Management will have to be further developed in light of the pilot project. We are to make recommendations to departments and HR in order to ensure better RI management in the future. The Research & Enterprise Office also expects to take a more 'hands-on' role in collating and checking the information recorded at departmental level (e.g. this information may be submitted to a central database on an annual basis from now on). We will record publication details for all staff in the university to ensure that we can collate research assessment data in the future with minimal effort.

The changes we are planning to implement in the near future have been informed by the pilot, but will not be driven by REF criteria. In the near future, we expect to be able to have an accurate picture of our research capital at any point in time.

The pilot has demonstrated to us the need to take forward discussions on a university-wide publication repository. It has also brought home the need to move forward with an integrated information database although work on this will be dependent on HEFCE information on the extent to which they will use Higher Education Statistics Agency data.

Use of REF as a driver to improve our corporate research information systems for on-going use rather than simply providing data for one-off exercises. Need for mechanisms to more consistently objectively assess bibliometrics. Need for coordination and standardisation of data between the requirements of different bodies including for example the research councils, the higher education funding bodies, the Higher Education Statistics Agency and other funders.

The platform that the university has adopted for Research Information Management appears to be effective, but some refinements could improve the system further. Consequently the further development of the university's publication database and Institutional Repository will need to be priorities. In particular the integration of the two systems would be highly desirable.

For us, given our size and monotechnic nature the implications are not so far reaching. It has been helpful to take part in the REF pilot as I think we will be better prepared for the real thing when it happens.

Increased training in bibliometrics will be required for Library, Research & Enterprise and Academic Staff and some institutions may be considering the appointment of a "bibliometrics expert".

The adoption of metrics such as these for research assessment will mean that all HEIs will wish to review their arrangements for maintaining a repository and will want access to the ability to undertake a citations analysis.

Citations data and its analysis are likely to feature more strongly than in the past. We will require access to citations data and tools to analyse it. Training in the interpretation of the data will also be essential.

Citation data are not used by the university as a proxy for research quality or performance. If citations are used in REF the university will have to develop systems to establish how citation patterns will affect REF placings (and other performance measures such as league tables) and to establish mechanisms for managing the data and optimising

performance in that context. This work is unlikely to be of relevance to other aspects of the university's activity or performance.

3.9.1 Discussion

The pilot has produced an emerging consensus on the likely direction of the future development of RIM systems, and suggests that the REF will reinforce development trajectories rather than signal the need for radical new departures. Taken together, the responses suggest there is a general ambition to

- Strengthen the *corporate* RIM systems;
- Develop central or corporate systems for ongoing use to inform institutions' internal strategy and planning exercises, rather than just one-off assessment exercises driven by HEFCE or other funding bodies;
- Have central research and planning functions play a fuller role in collating and checking information recorded at departmental level;
- Have a central publications database (institutional repository) and bibliographic database that links to other university systems such as HR (staff) and finance (contracts, research income);
- Move towards a system where all research outputs are captured and catalogued; and
- Develop internal bibliometrics expertise through additional training and possibly new appointments of specialist staff.

3.10 **Strategic implications for research management**

Question 10 invited pilot institutions to comment on the strategic implications of a REF bibliometrics element for the university management of research more generally.

It is clear respondents believe the REF will have an impact on the strategic management of HEIs: the overall research undertaking will be more actively considered and directed by the centre.

In around half of all cases, respondents' remarks suggest that a national assessment of institutional research quality will almost certainly impact on institutional research management procedures, from strategic planning to research partnerships to human resources management. The following excerpts are included to provide the reader with rather more texture than is achieved within the preceding synthetic remarks.

> *Research Information Management is strategically important for the university, irrespective of REF. It helps the university to analyse and identify where the university is currently with respect to outputs, income, student numbers and studentships. In the REF context, the university will make strategic decisions to capture information on economic and social impact as well.*

> *We anticipate making more use of the range of bibliographic data that we now have as part of our normal research management processes (e.g. analysis of institutions with which we collaborate, analysis of linkages with research grants); precisely how we will build in bibliometrics data will depend on the future formulation of the REF. We envisage that should citations data become of significance for particular disciplines in the REF, this would impact on the full range of performance management including appointments, annual research reviews and promotions.*

> *There needs to be a culture change so that academics opt into REF procedures, understand what is being assessed and utilise internal systems such as institutional repositories. The university needs to*

determine whether to invest in a bibliometrics expert who can analyse bibliometric data and academic performance. There is a potential tension between academics and senior managers, and how academic performance is assessed and managed.

The university has already established faculty-led reviews of all disciplines with a clear focus on maximising quality of output. Management of research will become more intensive in terms of planning and monitoring of research outputs, progress and quality. Research Information Management, for planning and monitoring of research activity and resources, must be developed and improved.

We anticipate that the introduction of bibliometrics to the assessment process will have the following effects: Increased focus on the citation performance of academic staff in staff appraisal processes; Increased focus on citation performance when recruiting new staff; A greater level of monitoring and analysis of citation performance at central and departmental level; The provision of institutional guidance and training to staff to staff advising them on where and how to publish. These developments would have both positive and negative effects: positive in terms of sharpening up performance management and negative in terms of the potential impact on publications behaviour.

The publication strategy of research staff will have to be considered to ensure that there is a focus on quality, high impact publications that are likely to gain greater citations. In addition, it will be important to review and enhance the impact of the research undertaken by the university.

We are assuming that the pilot has demonstrated that a system that requires papers to be selected by the institution will still be required for the REF. Our strategy is to continue to place strong emphasis on quality publications as an important key performance indicator in research. We are relieved that we are unlikely to have to spend time deterring academics from publishing papers that are likely to be lowly cited and agree that metrics on citations should be used to inform peer review rather than replace it. As papers of 3 years or less in age were the least robust in terms of the citation metric, it is tempting to suggest that this age of paper might be over-represented in the future REF exercise.

Whatever research assessment system is chosen, it should ideally not be unduly disruptive to ongoing research management. Nor should it artificially constrain research activity. Models 1 and 2 would appear to require much more careful management of research output to eliminate all weaker outputs from the total body of work. That could have significant negative implications for more developmental or speculative outputs or could damage the research process. Since Model 3 is closest to the RAE output methodology, it would be least disruptive to research management. We will, of course, have to reflect carefully on our research management once the full details of REF are clear.

3.10.1 Discussion

Taken together, the comments suggest there will be developments along each of the following dimensions:

- Increased use of bibliometrics to inform institution-level research management processes, from strategic collaborations to the targeting of research income
- A presumption that management of research at faculty level will become more intensive, and more exacting in terms of planning, monitoring and reporting of progress and achievements

- A presumption that if citations data come into more general use, it will impact on the full range of performance management from appointments to annual reviews to promotions. There is recognition that this might affect some disciplines more than others
- A presumption that central management will encourage departments to focus on the quality of output, possibly targeting higher impact publications that are more likely to gain greater citations. This might include institutional guidelines and training for academics on where and how to publish

HEIs will need to develop a more corporate approach to the management of research performance. This will include technological approaches (e.g. central repository), human systems (e.g. bringing faculty heads into a performance relationship with the corporate centre) and organisational processes (e.g. for strategy development and planning).

HEFCE might wish to give further thought to how best to respond to the suggestion that a national assessment exercise might lead institutions to begin to manage academic appointments and performance reviews against citations. Bibliometricians go to some lengths to argue that these metrics work well at scale – institution, field, country, etc – and that they are unlikely to be appropriate to judgements on performance of very small units, and especially individuals. Individual appraisals that review publication behaviour would seem reasonable, however some caution might need to be exercised if institutions hope to begin to use these metrics in a more instrumental fashion to determine promotions and other awards for individuals.

Several concerns were flagged by respondents, which ranged from a worry that academic behaviour might be changed for the worse, for example, narrowing points of engagement with business and other communities. Other concerns included the potential for citation metrics to impact negatively on the relationship between senior academics and early-career researchers.

One respondent signalled a note of caution regarding the extent of the implementation of the REF bibliometrics. They foresee a need for HEFCE to work with the community to manage the risk, evident from the pilot, that institutions will be required to develop comprehensive RIM systems, across all disciplines, that will go far beyond the requirements of REF and other management information needs, and that this will amount to a substantial additional burden on institutions and academics at a time when institutions are taking all possible steps to increase efficiency.

Lastly, there were several comments that serve to remind all concerned that there is still a long way to go in many institutions and fields to secure academic buy-in to these developments.

3.11 Other observations

Question 11 invited respondents to offer any additional remarks they might wish to make in light of the feedback of results. This last open question secured a wide range of rather fulsome observations, advice and requests for clarification.

As elsewhere in this paper, the following excerpts present respondents' further observations in some detail, to provide texture and insight for readers.

> *Involvement in the REF pilot, although demanding on staff time over last summer, has provided an excellent insight into the challenges of the proposed system and the requirements for institutions to provide the relevant information. It has been a very useful project to be involved in, and we were delighted to be able to make our contribution.*

> *All models flatten out the discrimination in quality achieved in the RAE 2008 assessment. The selection of pilot institutions has presented quite a flat range of RAE results in some UOAs, which has made assessment of the pilot outcomes difficult.*

The assessment of research performance driving the allocation of QR funding has enormous potential to influence researcher behaviour and decisions taken at the highest level by institutions. Institutions will inevitably seek to do well by adapting their behaviour in response to the requirements of the exercise. Therefore, in establishing a framework for assessment HEFCE is explicitly identifying the behaviours it wishes to reward. If the REF explicitly rewards institutions for achieving high levels of citations it will be understood that the HEFCE wishes institutions to focus resources on improving their performance in this regard.

It is clear that the use of a bibliometric algorithm will drive changes in behaviour. Institutional behaviours will adapt to optimise the outcome and this will impact adversely on both scientific and collegial approaches.

Using international impact as a driver will drive "artificial" authoring choices.

In terms of recruitment, appointments will be influenced by the citation records of academic researchers.

The use of bibliometric indicators in the REF will increase the commercialisation of citation databases, and will vest even more power with journals and publishing houses. An industry and market will grow up around creating citation analyses, marketing outputs and promoting citations. In addition, there is likely to be an impact on many research sponsors' open access policies and discipline-specific repositories. This would stifle innovation, information sharing and be counter to the interests of the academic communities concerned.

Data definitions need to be consistent across HEFCE and HESA to reduce costs of REF

We would like to emphasise that bibliometric data should be used to inform and support expert assessment of research quality, and not be used formulaically to arrive at quality scores.

Citations are no substitute for peer-reviewed judgment, and should at most be used as subsidiary evidence in expert review. Normalisation of citations to journals tends to favour publication in journals of lesser standing, where competition is weaker, thus creating a perverse incentive for academics to publish in weaker journals. Given the inherent weaknesses of citations, and the consequent need to use them at most as subsidiary to peer review, we believe that the HEI should select the staff and select the papers

We aim to implement processes to record more information regarding publications at the university. However, until the funding bodies make clear decisions regarding the future of research assessment, our exact plans for recording and managing data cannot be finalised.

Any response by the HEI in adjusting Research Information Management systems and research management strategy obviously depends on the release of guidelines for the next REF exercise.

We would ask HEFCE to issue clear statements well in advance of the formal REF exercise as to how bibliometrics are to be used to support institutions in their REF preparations.

The specification must provide clear, unambiguous and user-friendly information regarding the data required. To set this in context, the specification for publications data in the pilot was very unclear. For example, with regard to volume number, the bibliometrics contractors

apparently did not want parentheses and just wanted the volume number (not issue). Similarly with page numbers, we gave the range rather than just the first page itself. The data specification was a single table, without real examples and this led to interpretation by individual pilot sites. We labelled all the fields correctly and in the correct order. If we had been told, for example, don't use ( ) for volume and don't include the issue number, or that only the first page number was required, it would have been very easy to correct the data to comply with those requirements.

In our response to the initial REF Consultation in February 2008, we supported a position whereby there would be no selection of research active staff, all outputs should be assessed and credit should remain with the institution identified by the author's address on the output. At that time, however, it was proposed that research assessment for science-based disciplines would be driven by quantitative indicators (to include bibliometrics). It has since become clear that bibliometrics are not fit for purpose as a standalone measure of research performance and all evidence argues for the continuation of peer review in a REF which is becoming increasingly like the RAE. It would not be appropriate for peer review to consider all outputs for an institution and there must therefore be a return to selection of outputs. Given this, the operational simplicity we had anticipated for submitting all academic staff would no longer arise.

The different costs of each model and any assessment of costs needs to consider the different level of preparedness of HEIs and costs/development that has gone on to date. The cost moving forward is very different to the total cost as it depends on the starting point. Any true cost assessment should take into account prior investment. For example, an address-based model would be more expensive to an HEI with advanced information systems since it would require checking of data from another source when the HEI may feel that its own systems are adequate.

For Model 3, who would select the papers on which the bibliometric analysis is based? We suggest that this should be the HEI. It would need to be made clear if additional outputs (including published papers) could be used by HEIs as evidence supporting the impact Statements being proposed within REF. Any bibliometric model needs to factor in the presence of Early Career Researchers otherwise this will be a significant (and unwanted) impact on publishing behaviour where new researchers are not encouraged to publish or be co-authors. We suggest that panels be given bibliometric analysis based on both WoS and Scopus.

3.11.1 Discussion

Three points were picked up by several respondents, as follows:

- Clear and precise guidelines are vital, and respondents would like this information as early as possible, to inform systems development and to minimise the risk of wasted effort in any future REF
- A recognition that whatever system is chosen, it will influence institutional and academic behaviour: HEFCE will get what it measures. For example, respondents anticipate the use of international impact will affect authoring choices
- The use of bibliometrics should be focused on those areas where it is robust, and must only be used to inform a peer-review process

There were several separate points of note also, including:

- HEFCE should do all it can to ensure institutions are able to select the authors and papers that will be submitted for citation analysis
- HEFCE should consider the feasibility of providing peer review panels with citation results on any set of papers, from both Scopus and WoS
- HEFCE should consider ways in which it might improve the presentation, packaging and explanation of the citation results, as the pilot required institutions to devote substantial intellectual effort to assessing the implications
- HEFCE needs to bear in mind that the actual costs of compliance will vary greatly across institutions, and even under Model 3 might prove quite demanding for some. Ultimately, costs will depend upon the final specification and the interplay between the state of development of institutions' current systems and the nature and extent of their research endeavour

There remains an undercurrent of concern about the use of citations in performance assessment, which while it reduced greatly across the two consultation rounds, suggests it will take one or two iterations of the REF bibliometrics proper before the community is fully acclimatised.

With the benefit of time, pilots have examined their returns more closely and the relative performance of other institutions, as best they can, and one part of the group of 22 seems to have returned to the rather gloomier outlook of the early days of the pilot. Confidence is shaken by evident problems in database coverage, subject mapping and a range of detail issues with the completeness and accuracy of the individual records on which the citation analyses were based. It seems to matter less that the results across pilot institutions amount to a reasonable approximation to the 2008 RAE profiles, when faced with demonstrable errors and inaccuracies in items within the underlying data set.

This opinion is not going to go away, however HEFCE needs to document the nature and extent of these limitations and reflect upon the sensitivity of the final results to the level of quality problems evident in each broad area. Ultimately, several pilot institutions appear to have been left with the impression that this is not a positive step forward in performance assessment and that the results are likely to be driven by the somewhat unpredictable but not random interplay between one's particular research focus and the completeness of the bibliometrics record related to that area.

A smaller but still significant number of pilot institutions has been much more matter of fact about the exercise throughout, providing feedback on what works best and what sorts of things HEFCE will need to do to make a success of the REF. This difference in outlook is possibly based in a difference in philosophy, or at least pragmatism as they are more openly supportive of the move towards a more strategic approach to research management.

Post pilot, this continuing split in opinion suggests that the autumn consultation is almost certain to attract responses from a large and vocal group of HEIs and academics that will continue to be ill-disposed towards bibliometrics under any circumstances, and see the pilot as having confirmed this position.

No doubt the consultation will ask whether people prefer one option over another, under a set of given circumstances, and will not seek a mandate from the community to proceed with a REF bibliometrics element. However, there might be value in HEFCE including a comprehensive simulation, or worked example of how the citation analyses might be used within the context of a national peer-review exercise.

The trust in the infallibility of peer review is striking, and feels rather contrived, game playing perhaps, especially in light of the numerous studies carried out over the years that note it has limitations as well as great strengths (subjective, orthodox, non-replicable, non-transparent, etc). Those limitations become more problematic as the level of aggregation increases, and while argument and judgement might be the critical last step in even a national-level assessment of research performance it seems

incontrovertible that such judgements will be more robust for having considered multiple streams of data and intelligence, both subjective and objective.

The last observation is that the gradual movement across the life of the pilot towards a decision to implement a limited bibliometrics model has improved pilot institutions' comfort levels. However, this reaction suggests the community more generally might begin to downgrade its rating of the strategic importance of this work. Certainly, the open discussions with the community, rather than just pilot institutions, in the latter stages of the pilot had begun to take much greater interest in the black box that is research impact. In light of developments elsewhere, it would be unfortunate if early momentum was lost and the HE community missed this opportunity to upgrade its access to strategic intelligence on many billions of pounds of public investment.

4. List of abbreviations

FEC	Full economic cost
FTE	Full-time equivalent
HE	Higher education
HEFCE	Higher Education Funding Council for England
HEI	Higher education institution
HESA	Higher Education Statistics Agency
HR	Human resources
RAE	Research Assessment Exercise
REF	Research Excellence Framework
RIM	Research Information Management
UOA	Unit of assessment
WoS	Web of Science

Technopolis Ltd
3 Pavilion Buildings
Brighton BN1 1EE
UK
T +44 1273 204320
F +44 1273 747299
E info@technopolis-group.com
www.technopolis-group.com