

Consultation on the second Research Excellence Framework

Page 1: Respondent details

Q1. Please indicate who you are responding on behalf of

Subject association or learned society

Please provide the name of your organisation

This is a joint submission on behalf of: The UK Computing Research Committee (UKCRC): an expert committee of the British Computing Society (BCS) and the Institution of Engineering and Technology (IET): <http://www.ukcrc.org.uk/> and The Council of Professors and Heads of Computing (CPHC): <https://cphc.ac.uk/> and The BCS Academy of Computing (academy.bcs.org) Several members of UKCRC who served on sub-panel 11 in REF 2014 contributed significantly to the authorship of this response.

Page 2: Overall approach

Q2. 1. Do you have any comments on the proposal to maintain an overall continuity of approach with REF 2014, as outlined in paragraphs 10 and 23?

We do not consider there is any need for fundamental change in the approach that should be on assessment of outputs, environment and impact as in 2014 REF. There should not be changes that increase the burden on academic institutions or assessors; indeed a reduction would be welcomed. Keeping the assessment process broadly similar would obviously assist in longitudinal comparisons.

The process is already overly-complex and every chance should be made to simplify the current process so that academic time spent on the process itself is limited. In particular, implementing the recommendations that arise out of this survey, could easily lead to a big increase in complexity; thus the big picture must be kept in mind, and the goals of the REF kept to the forefront, rather than trying to micro-manage every aspect of a submission. Thus we welcome *simplifying* process change in particular rather than major changes to the *assessment* process.

Page 3: Unit of assessment structure

Q3. 2. What comments do you have about the unit of assessment structure in REF 2021?

There are no changes required for main Panel B or the Computer Science and Informatics UoA (Sub-panel 11).

Reporting sub-profiles makes sense and would be logical if multiple submissions to the same UoA ceased to be allowed. This would allow more direct comparison across institutions. Multiple submissions to UoA11 in REF2014 was not an issue and we do have a view as to whether or not this is desirable for other UoAs.

Page 4: Expert panels**Q4. 3a. Do you agree that the submissions guidance and panel criteria should be developed simultaneously?**

Yes

Comments:

As there are likely to be comparatively minor changes to submissions guidance and panel criteria compared to REF 2014, they could be developed simultaneously. Indeed, we believe that submission guidance and panel criteria and working methods should be developed in parallel, otherwise information critical to assessment may not be requested of submissions. Unambiguous guidance is essential for panels and HEIs.

Q5. 3b. Do you support the later appointment of sub-panel members, near to the start of the assessment year?

No

Comments:

Only involving main panel members in the development of criteria (and working methods) and appointing sub-panel members later in the process would be disastrous – for much the reasons outlined in the consultation document. Our computing discipline is a wide-ranging one (for example the behaviour of subatomic particles through software to general public end-user interfaces. The early appointment of experts on sub-sections of our discipline is essential. Sub-panel members should thus be appointed at an early stage and their combined expertise should be used to develop criteria and working methods. A sub-panel chair could not adequately foresee the issues that would need addressing across a broad field. Equally, the buy-in of sub-panel members to the assessment process they have to deliver is essential. Sub-panel members need a great deal of time to get their activities and their minds in order as preparation for REF; the individual and their host department need time to plan and to be able to reduce the workload of the panel member – this is much harder if it is close to the academic year in which the assessment happens (particularly for small departments). We also note that, potential panel members may be asked by universities to participate in internal assessments. This is not acceptable if the person is to be a REF sub-panel member and so this needs to be signalled early. It was difficult to find suitable people who were willing to serve on SP 11 due to the work involved and the amount of time required for the assessment; thus, in exceptional circumstances, later appointment for some panel members in large sub-panels may be acceptable, providing a sufficient number of panel members are appointed to ensure adequate sub-area coverage before the criteria are set. Late appointment should be the exception rather than the rule however.

Q6. 4. Do you agree with the proposed measures outlined at paragraph 35 for improving representativeness on the panels?

No

Comments:

Expertise, competence and coverage of the subject area should be the main criteria in selecting sub-panel members. However, Equality and Diversity are also important criteria since they may lead to more accurate assessments, and as such, reflecting the population from which the panel is drawn from may be regarded as desirable. However, since panels contain relatively small numbers of people, any conclusions should be drawn with care, though we note the proposal is only for monitoring, not to directly influence selection/recruitment. Also note our comment in (3b) concerning the difficulty of finding people to serve which inevitably creates difficulties in ensuring diversity. We have marked “no” above since we do not believe it should be mandatory to require individuals to disclose demographic information, although it should be strongly encouraged.

Q7. 5a. Based on the options described at paragraphs 36 to 38 what approach do you think should be taken to nominating panel members?

Self-nomination could lead to a more open process, but may result in a large number of inappropriate nominations which would have to be filtered, causing a good deal of work for the sub-panel chair. Also, self-nomination is likely to impact on diversity, at least in terms of gender, as it is widely recognised that women tend not to put themselves forward to the same degree as men.

It is also not clear how self-nominations would demonstrate sufficient community support; at a minimum, seconders from senior colleagues in two other Universities should be required. We support nominations coming from subject communities and professional bodies, though these should be encouraged to allow their members to self-nominate to their own nomination committees (as happened within UKCRC in REF 2014). Such associations and bodies should be reminded of their responsibility as regards ensuring fair and proportionate representation; given their status, one might assume that this is (increasingly) foremost in their minds in any case.

Q8. 5b. Do you agree with the proposal to require nominating bodies to provide equality and diversity information?

No

Comments:

Nominations are generally for people known, respected and trusted by the community and so it is not difficult to take into consideration equality and diversity when making final selection of members. However, this may be perceived to be a good thing by the community at large, so worthwhile, though it would pose an additional administrative burden to the process. Learned societies and such are certainly aware of the challenges of equality and diversity and should be entrusted to ensure that they put forward appropriate individuals for sub-panels. They could be encouraged to provide additional E&D information, though this should not be onerous nor mandatory, and not so as to inhibit nominating bodies. There is an argument that individuals being nominated should not be *compelled* to provide demographic information (e.g. when applying for a job one is rarely *required* to do this) although this could be encouraged for monitoring purposes.

Q9. 6. Please comment on any additions or amendments to the list of nominating bodies, provided alongside the consultation document.

We believe the institutions listed for REF2014 were sufficient and appropriate.

Q10. 7. Do you have any comments on the proposal to use HESA cost centres to map research-active staff to UOAs and are there any alternative approaches that should be considered?

We disagree for these reasons.

1) HESA data and REF returns are for different purposes, on different timescales and using different definitions. Both research and teaching considerations come into play in the HESA return whereas REF is concerned exclusively with research.

2) HESA cost centres do not map directly on to the REF UoAs and there will always be individuals whose research is not easily defined using either set (REF or HESA) and who do not form a critical mass on which a separate submission might be based. HESA data was not designed to be used in this manner and adding another requirement is likely to create problems now and in the future.

3) It artificially limits the ability of HEIs, as independent organisations, to manage their own research environment to engender excellent research. Institutions are the best judge of how staff are best allocated to UoAs.

4) HESA requires classification by job-role and seeks to determine the allocation of staff time to those different roles. Difficulties are posed by, e.g. staff who are fulfilling Faculty roles, and research staff that are Directors of Interdisciplinary Research Institutes who do not feature in what would be their home UoA by virtue of that role.

5) UoAs (such as UoA-11) which include cross-disciplinary staff, may be compelled to return them to a different UoA to the one to which the majority of staff in their department will be returned -- HESA cost centres do not embrace interdisciplinarity. E.g. Digital Humanities scholars may be classed as academics in English, Computing, Psychology, Theology, Business Studies, Music, Modern Languages etc. However, the HESA data would not make this interdisciplinary affiliation clear; thus there is the potential for outputs in the field to be misunderstood, or for there to be a perverse incentive not to hire interdisciplinary researchers in future.

6) The HESA returns based on academic year 2019-20 will not reflect accurately the size of each cost centre nor define the UoA at the point of REF submission due to staff movements and the HESA-data time-lag.

7) If there is no selection of which staff to submit, then institutions are likely to try and move some academics into different employment categories. The divisiveness relating to selection may be moved to contractual changes in employment. SP11 did see a number of 20% employment contracts, but no real evidence of involvement with the institution. Possibly the percentage should be increased (cf Q14).

8) A number of SP11 institutions submit only their small pockets of research activity; if they have to submit all academics this will substantially increase the assessment burden with no real benefit in terms of QR distribution.

Notwithstanding our arguments above, if HESA data is used for staff selection, then UoAs must be allowed to make a case for moving a limited number of staff between HESA categories on the grounds of fair assessment of interdisciplinary research. However, our ****overwhelming**** recommendation is that HESA returns should not be used for staff selection/assignment.

Q11. 8. What comments do you have on the proposed definition of 'research-active' staff described in paragraph 43?

There is a need to have very clear specification of research active staff that would allow the submission of senior independent Research Assistants and those on research fellowships but prevent the submission of junior research assistants. One definition would be to have secured funding covering (part of, e.g. >50%) their own salary in open competition, but this could be viewed as too restrictive (eliminating "University Academic Fellows", funded by the University who have not yet secured their own external funding but who have a brief to develop their own research programme, and who will have been appointed competitively).

Q12. 9a. The proposal to require an average of two outputs per full-time equivalent staff returned?

Our view is that there should be selective submission of staff with a norm of 4 outputs per person as in the 2014 REF. This is the best compromise on assessment burden and accuracy of research assessment.

If, notwithstanding our strong preferences for no change in the number of outputs required per staff number, a decision is made to force the submission of all staff, this will substantially increase the assessment burden for SP11 as a substantial number of institutions submitted a comparatively small proportion of staff, reflecting pockets of good research; the output reviewing load was already at the maximum and individual could bear, so any increase in the number of outputs overall would require further panel members in an already very large panel.

On the other hand if the number of outputs required per staff member is reduced, e.g. to an average of 2 or 3 outputs, this will reduce the accuracy of output assessment. If the number of outputs per staff member is reduced, then we much prefer 3 over 2. We recommend some modelling of the expected number of outputs and panel size to allow all papers to be reviewed by three panel members with a total load of no more than 1100 papers each.

For those UoAs with roughly the same number of FTEs returned as in 2014, it will lead to assessment of a smaller sample and thus likely higher profiles; on the other hand for UoAs who return a substantially greater percentage of their staff compared to 2014, it is likely that their profiles will be lower.

If the proposed range in number of outputs per person is imposed, this will also result in higher profile ratings. There may also be internal and staff issues for UoAs regarding the selection of how many outputs for any given member of staff. Additionally, this might encourage "gaming" between research groups within SP11, with (probably) greater focus on core computer science, and less on interdisciplinary research (that is regarded as more risky to include) and also less on applied research where it is more difficult to demonstrate methodological excellence; this would not be a benefit for the health of the field in the long term.

A sampling approach is not appropriate as it would undermine confidence in the results, with the potential that poor performance will be excused as unlucky sampling. In many UoAs there will not be sufficient outputs submitted to make sampling sufficiently accurate.

Q13. 9b. The maximum number of outputs for each staff member?

Many in our community would prefer the maximum to equal the minimum. However if this practice is ended, then removing the fixed relationship of 3 or 4 outputs for each staff member and allowing a maximum of e.g. 6 has a number of disadvantages. It would lead to a biased assessment of only a small part of a UoA's research, as it may result in 6 outputs from a few staff with only 1 from the rest; there is likely to be even more 'poaching' of star academics for very high salaries with detrimental effect on junior appointments; the divisiveness for those academics who have only 1 output selected compared to those with 6 would probably be greater than that for non-selection for REF 2014 as the numbers of academics with only 1 output selected is likely to be much higher than those not selected in 2014, particularly if the everyone has to be submitted. A maximum of less than 6 would be better particularly for small units where otherwise a disproportionate benefit may be gained from the unit's "star researchers". We suggest that the use of modelling to help determine whether 4 or 5 would be better.

One key point, whatever maximum is chosen, we hope that this is only the maximum associated with any individual; i.e. it should be possible for a researcher to be a co-author on papers associated with other individuals submitted in the same UoA. Thus the outputs section of a return should be structured as a set of individual staff submissions, as has been the case previously.

Q14. 9c. Setting a minimum requirement of one for each staff member?

If the maximum is not to be equal to the minimum, then a minimum of 1 is appropriate with the possible exception of very recent ECRs. We suggest that the spread between minimum and maximum is small; that is, we would prefer a fixed number of outputs with reductions for special cases as in 2014.

Q15. 10a. Is acceptance for publication a suitable marker to identify outputs that an institution can submit and how would this apply across different output types?

Outside mainstream publications (e.g. journal articles) the date of acceptance is difficult to audit, and as such it would not be a suitable approach for determining which outputs an institution is eligible to submit. The only readily auditable data is the date of publication of an output, or, in the case of certain outputs, the date of their public disclosure. However, there are reasons why the date of publication is not an optimal marker – the sometimes-lengthy time between production of the research and its eventual publication loosens the correspondence between the resource that an HEI has invested in generating the research underpinning an output and the credit that it would receive from the REF submission of that output.

However, imperfect though it is, the use of an output's publication date would make the process uncomplicated to implement using current systems, as well as auditable. In most cases the date of publication could equate to the availability of a digital object identifier (DOI) – thus “online –first” publications' date would equate to their online appearance rather than their print appearance.

We also suggest that the rules concerning the eligibility of journal articles based on earlier conference papers be clarified and improved. The rules in 2014 (paragraphs 43 and 44 on page 23 of the guidance) were problematic in SP11. It is common in Computer Science for research to be reported first in published workshops and conferences and then, for the best work, to write it up for a journal, often combining several earlier papers. If all these publications occur within a REF period there is no problem. But for journal articles occurring early in the period they will typically be based partly on research first published in the previous period; those articles may well have been considered “not good enough” for submission to the previous REF, whilst the journal article, combining several papers would be an excellent potential output. Rule 44 gives discretion to the panel as to how to treat such papers which “cross” a REF boundary in this way. In 2014 some HEIs did not want to risk submitting such papers, however excellent; in fact SP11 decided that it would not penalise such papers (assuming that the earlier paper(s) had not been submitted to the previous REF). Other HEIs guessed that the panel would be lenient and were lucky. This element of chance of how the panel views such outputs should be removed. Since it seems unfair to penalise journal papers published early in the period (that might be based on conference papers in the previous period), we suggest that all such papers should be allowed, only penalising those parts of papers based on conference/workshop paper(s) submitted to a previous REF.

Q16. 10b. What challenges would your institution face in verifying the eligibility of outputs?

Institutions have already a considerable increase in verification burden of Open Access for eligibility verification. Portability would probably not make a big difference to this, though any rules regarding (non)-portability must be clear and unambiguous. Certainly the situation should be avoided where institutions have to argue about "whose paper it is". Since any other definition than the institution(s) named on a paper as defining where the work is done would be very hard to audit, it is still not clear that non-portability would really prevent porting of research done at one institution to the REF submission of another (if the author moves during the research, as is very likely as most research takes an extended period).

We note that whereas affiliation at the date of acceptance might be viewed as fairer, since this date is often un-auditable, so the affiliation on the paper may be better, though if there is an extended time to publication, during which an author moves, there is an opportunity for authors to change the affiliation on a paper. Whatever the rules, academics and institutions will seek to maximise their individual and collective benefits. In summary, a rule of non-portability is not as simple as it might first appear.

Q17. 10c. Would non-portability have a negative impact on certain groups and how might this be mitigated?

It may have the effect of making it more likely that ECRs will emigrate, since they may be less competitive in recruitment exercises as a result; this would be detrimental to the UK, having trained them. One mitigating action would be to exclude from non-portability the outputs of individuals who were themselves too junior to be returned to REF2014. The same rule could be applied for individuals recruited from overseas who were not returned to the previous REF.

We consider that preventing portability of outputs is addressing a problem that does not currently exist in our community. There is some movement of staff between institutions due to REF, but on the whole we consider this to of benefit overall. Movement of people between institutions is good for research as it brings in fresh ideas and new expertise, so we can see no reason to make it less attractive. However if the maximum and minimum number of outputs is set to 6 and 1 then poaching of star staff for publications will become a problem which would have to be addressed by preventing portability, i.e. changes to the number of outputs submitted to REF will create the problem. It may be necessary to limit the number of publications that can be ported, possibly pro-rata to the period in the institution. Setting a cutoff date (mid period) for porting is another possibility to mitigate institutions poaching "last minute".

However the view above is not universal in our community for non-portability: there is also some support for non-portability for all except those in their first REF. This has the advantage of simplicity

Q18. 10d. What comments do you have on sharing outputs proportionally across institutions?

This has been done in the past and was dropped as it was an administrative burden on institutions. This is trying to address an issue for the comparatively few staff who switch institutions. It would be better to leave criteria as they were in REF 2014, where institutions cannot submit outputs from staff (or students) who are no longer at the institution and allow portability of outputs.

If this were to be done, then an issue would arise for those staff previously at a UK HEI, and those recruited from overseas, or not eligible at their previous HEI (e.g. PhD students).

Q19. 11. Do you support the introduction of a mandatory requirement for the Open Researcher and Contributor ID to be used as the staff identifier, in the event that information about individual staff members continues to be collected in REF 2021?

No

Comments:

We are of the view that the wider adoption of ORCID as a unique and persistent digital identifier for researchers should be encouraged but that the timescale for its implementation is too short for this REF cycle -- we note that whilst some publications now require authors to specify their ORCID, this is by no means universal yet (though it would be helpful to use it where available). We would further point out that there will be either a financial or workload penalty on any institution implementing it. This might however be a good time to signal to the community that it is intended to use ORCID's mandatorily for REF 2028. We note that Open Researcher and Contributor IDs are necessary for automated collection and analysis of metrics such as citation counts for staff submitted; however, as there are no plans to do away with human assessors in sub-panels, who can easily determine validity of authors or institutions in the outputs, there is no need for this mandatory requirement.

Q20. 12. What comments do you have on the proposal to remove Category C as a category of eligible staff?

We have no objections.

Q21. 13. What comments do you have on the definition of research assistants?

In the 2008 RAE there were many problems with inappropriate junior RAs being returned, but this did not occur in 2014 REF as the definitions were clearer. There will always be some ambiguity in the definition of independent researcher, particularly for university funded research fellowships. There were no issues about independence of externally funded research fellows e.g. funding from RCUK or learned societies.

Q22. 14. What comments do you have on the proposal for staff on fractional contracts and is a minimum of 0.2 FTE appropriate?

We understand that SP11 did identify some problems with 0.2 FTEs whose main research was outside the submitting university e.g. visiting academics from industry or overseas, academics who had left for employment elsewhere, ex PhD students now employed elsewhere. There is a difference between semi-retired staff, and staff who have a substantive employment elsewhere. A statement outlining the connection of the staff member might help (though this would require difficult decisions by the SP).

Increasing the fractional contract to 0.4 would also remove many of the problem cases (though this would mean that many partly retired staff could no longer be returned – an exception might be made for those staff who were already in the HEIs employment at the start of the REF period to have the currently lower FTE threshold of 0.2).

Incorporating a maximum on the number of outputs a fractional staff member could contribute would also mitigate issues associated with such staff (e.g 1 for 0.2, 2 for 0.4, 3 for 0.6). We also suggest that small fractional appointments shortly before the submission date be disallowed (unless already employed by the institution), with a minimum period of, say, 1 year of appointment before the cutoff. A longer period would be yet more robust against game playing. Care should be taken to stop institutions suddenly increasing a smaller fractional appointment to one just large enough for the REF threshold just for a short period around the census date.

Page 6: Collaboration

Q23. 15. What are your comments in relation to better supporting collaboration between academia and organisations beyond higher education in REF 2021?

We support reducing the number of outputs required for staff who have moved from outside academia to academia similar to ECRs. An issue may be in policing this effectively.

Although industrial collaboration is identified in the environment, it usually takes the form of a list of industries with which there is some form of collaboration. Instead, the environment submission should explain using a few case studies, what benefits arose from collaborations and give some information on papers co-authored with industry etc.

Outputs co-authored with industry might be specially marked in the outputs section, though these are likely to only form a very small percentage of a UoA's submission.

Page 7: Outputs

Q24. 16. Do you agree with the proposal to allow the submission of a reserve output in cases where the publication of the preferred output will post-date the submission deadline?

Yes

Comments:

There is little need for this in SP 11. In the case of conference publications, the date is fixed and easily validated. However, there will always be a small number of papers that appear after the REF deadline but before the end of the year. As such we have no objection to allowing a reserve publication for such cases, as was done last time for double weighted outputs and indeed some in our community recommend this strongly.

Q25. 17. What are your comments in relation to the assessment of interdisciplinary research in REF 2021?

Sub-panel 11 assessed a very broad remit of inter-disciplinary work including outputs submitted to venues that essentially related to other engineering disciplines, mathematics, various science subjects, medicine and even psychology. Very few outputs were referred to other sub-panels. A number of sub-panel members did have experience of inter-disciplinary work as much of Computing research is now inter-disciplinary. The sub-panel as a whole did not think inter-disciplinary work was unfairly assessed. Of the suggested further steps that could be taken to support interdisciplinary research, we favour only proposal (c): the environment submission can capture the institutional and local approaches that promote and reward excellent interdisciplinary research. We are not sure that champions at the main panel level would be of any benefit.

Very few outputs that were interdisciplinary had set the indicator for the submission so it was not really possible to do any post-REF analysis of assessment of inter-disciplinary work. Submitted outputs in UoA 11 had a topic identifier in their 100 words, based on ACM classification system. Analysis showed that on the whole fundamental topics were rated slightly higher than more applied topics which may have been more interdisciplinary but life and medical science was one of the more highly rated topics.

One feature of REF2 014 that was helpful in encouraging interdisciplinary research was the ability to submit the same output to two or more different sub-panels. We strongly recommend that this feature be retained in REF 2021.

Q26. 18. Do you agree with the proposal for using quantitative data to inform the assessment of outputs, where considered appropriate for the discipline? If you agree, have you any suggestions for data that could be provided to the panels at output and aggregate level?

Yes

Comments:

Although we have answered “yes”, we do have concerns regarding the robustness/completeness of available data and the use of measures such as citation metrics to assess research quality across the various units of assessment. We note the outcomes and recommendations of “The Metric Tide”: Report of the Independent Review of the Role of Metrics in Research Assessment and Management (2015) and concur that it is not currently feasible to assess research outputs or impacts in the REF using quantitative indicators alone. We note also that there is evidence of gender bias in citations, which suggests that women may be disadvantaged by using such a measure (see DOI: 10.1073/pnas.1510649112, <https://arxiv.org/abs/1610.08984>, <https://arxiv.org/abs/1607.00376>) In REF 2014 SP 11 did use Scopus citation data but about 55% of outputs had 0 to 5 citations and less than 25% had a meaningful number of citations. SP11 had approached Google for access to Google Scholar citation data but they refused to provide access. After completion of the REF output assessment SP11 managed to unofficially get Google scholar citation data from the web, as the coverage of SP11 outputs is much better than Scopus and showed there was some rather limited correlation between output scores and citation counts. It would be worth trying again to get Google cooperation on citation data for SP11 for 2021. Download counts have been suggested as another possible metric for individual outputs, but there are difficulties of aggregating the many different download sources, and are subject to manipulation, so we do not recommend their use. We do not consider any of the ‘alt-metrics’ such as tweets or facebook-likes to be relevant to UoA 11. Aggregate citation information for a UoA could be a possible metric for Environment assessment as an indication of overall influence of past research by the UoA. Care will need to be taken by the panel to interpret any citation data appropriately, for example even with the topics covered by SP11, average citation rates vary widely (theory low, computer vision and bioinformatics much higher for example). Certainly citation data is not a substitute for evaluation of individual outputs by panel members. We endorse the REF 2014 guidance: “No sub-panel will make any use of journal impact factors, rankings, lists or the perceived standing of publishers in assessing the quality of research outputs”. Citation data that could be useful includes recording the field-weighted citation impact (FWCI) for each UoA, as well as the % of submitted outputs that fall into the top 1%, 5%, and 10% most cited in the world.

Page 8: Impact

Q27. 19. Do you agree with the proposal to maintain consistency where possible with the REF 2014 impact assessment process?

Yes

Comments:

The overall weighting of impact should remain unchanged (though we support moving the impact template to the environment section, which would effectively increase the weight of the ICS). One noticeable feature of REF-14 was the number of institutions submitting 14.9, 24.9, 34.9... FTE, presumably because they could not "find" another ICS. One way of mitigating this unfortunate pattern, would be to weight the lowest scored ICS for each submission proportionately to the FTE over the threshold. Assuming the same thresholds as last time for the number of ICS required, the worst graded ICS in each UoA should be graded with a weight of $\text{MOD}(y+5,10)/10$, where y is the FTE submitted. Thus there would be much reduced incentive for a UoA not to submit an extra person or two, which would take them over the threshold. (Of course, if Universities were required to submit all research active staff, then this suggestion becomes essentially irrelevant.) Though we approve of maintaining consistency, this should not be at the expense of making changes which are definite improvements following lessons learned.

Q28. 20. What comments do you have on the recommendation to broaden and deepen the definition of impact?

It would be beneficial to broaden the specification of what can be submitted as impact case studies. The time limit on when the research was done could be increased, provided the impact is still within the assessment period. This would allow for more cases where the impact takes a long time to come to fruition. There is some support in our community for the possibility to submit specific case studies related to substantial academic impact but this view is not universally shared. In particular, there should be some scope to allow impact of one discipline on another. For example Mathematics research impacts massively on (and generally underpins) other disciplines (which go on to impact in a more "conventional" way) but this impact of Mathematics is not counted.

We also note that there is some fundamental research (e.g. the initial work on Model Checking) that substantially influences more applied research in other institutions which in turn may create considerable end impact. However, the originating institution may not have the evidence to submit the impact as a case study as they may not have had any interaction with the exploiting organization.

One area where it may be useful to clarify eligibility is open source software. Often this is primarily used by other academics. It seems to us that this kind of impact from research can be extremely valuable with high impact and is not directly measured already within any accompanying outputs.

It may also be helpful to note that wider impact can also mean impact in teaching and learning, at all levels. Indeed, "wider impact" can be generated in multiple dimensions and a definition that brought out these dimensions would be welcomed. Indeed, established literature on Knowledge Exchange includes "direct instrumental impacts" of a particular piece of work in a particular application, the "conceptual impacts" that can reshape a whole field, "capacity building" impacts through education and training, "attitude or cultural change" and "enduring connectivity" of long term relationships with research users. See Meagher, L. and Martin, U. (2017) Slight dirty maths: The richly textured mechanisms of impact", research evaluation, <https://doi.org/10.1093/reseval/rw024>

Q29. 21. Do you agree with the proposal for the funding bodies and Research Councils UK to align their definition of academic and wider impact?

Yes

If yes, what comments do you have on the proposed definitions?

This alignment of definitions is very welcome. In terms of specific recommendations, the definition of 'academic impact' could explicitly mention the impact on teaching. We also note that the more theoretical aspects of our subject area (akin to mathematics) are likely to have impact on other disciplines, or sub areas of computing, rather than directly outside academia.

Q30. 22. What comments do you have on the criteria of reach and significance?

We believe these criteria were understood by HEIs and sub-panels last time and worked well.

One point that might be clarified further how reach and significance are weighted against each other, i.e. must a top-rated impact activity have both large reach and significance, or does excellence in one measure offset a less extensive performance in the second (i.e. would a local activity with very deep impact score 4*?).

Q31. 23. What do you think about having further guidance for public engagement impacts and what do you think would be helpful?

Further clarity is certainly needed here. Although the results of the last REF have been instructive in distinguishing between mere dissemination and impact, it would be helpful for the guidance to emphasise that public-engagement-based impact should be subject to the same level of rigour as other types of impact with respect to the articulation of the evidence of change. Any such ICS should clearly describe the purpose of the engagement.

We also believe that the assessment of public engagement impact shouldn't be solely in terms of counting numbers directly impacted (e.g. attending an event) but should include influence on thinking or driving demand for uptake (visits to associated web sites, though this is subject to manipulation).

Q32. 24. Do you agree with the proposal that impacts should remain eligible for submission by the institution or institutions in which the underpinning research has been conducted?

Yes

Comments:

We did not identify any issues caused by non-portability of impact, apart for one new group which did not have any ICS to submit. The non-portability of impact does not in general affect hiring of ECRs, given the relatively low number of ICS required compared to the number of staff and the relatively low number of ECRs. The consistency of strategy for supporting the impact submitted by the institution outweighs the potential disadvantages of locating data once a staff member leaves an HEI. In any case, the evidence-capture systems adopted by a growing number of HEIs as part of good record-keeping should simplify the process of locating the evidence underpinning a case study that was developed (at least in part) at an HEI. We do note though that sometimes research at one university may be further developed at another which then realises the external impact – arguably both institutions should be eligible to claim impact (though the former institution may have difficulty in acquiring corroborating and detailed evidence).

Q33. 25. Do you agree that the approach to supporting and enabling impact should be captured as an explicit section of the environment element of the assessment?

Yes

Comments:

This would be much easier to assess as part of the environment where it belongs. The profile of the impact strategy could be retained by ensuring, as proposed, that impact is covered in a separate section of the environment template.

Q34. 26. What comments do you have on the suggested approaches to determining the required number of case studies? Are there alternative approaches that merit consideration?

The minimum number of case studies should be reduced from two to one. If all staff are submitted, the number of staff members per case study should be increased so that the overall number of case studies is roughly the same. If all staff have to be submitted, some SP11 UoAs, which submit a small proportion of their staff may decide not to submit as they will not be able to find enough impact case studies.

One issue that arises is that departments vary in how applied their research is; small, "theory-heavy" departments find it much more difficult to find the requisite number of case studies; the proposal to reduce the minimum number to one may alleviate this issue.

Q35. 27. Do you agree with the proposal to include a number of mandatory fields in the impact case study template to support the assessment and audit process better (paragraph 96)?

Yes

Comments:

We agree that a structured template would be very beneficial both for helping to ensure that a case is described accurately and consistently, and for ease of reporting data to government. Associated research outputs should be submitted (as pdfs) as finding these proved quite time consuming for assessors.

Q36. 28. What comments do you have on the inclusion of further optional fields in the impact case study template?

As per the answer to Q28, we are very supportive of approaches that allow impact cases to be described to their full potential and that also facilitate the analysis and reporting of the impact that has been achieved.

Additional fields should not impinge on the limited word counts though.

Q37. 29. What comments do you have in relation to the inclusion of examples of impact arising from research activity and bodies of work, as well as from specific research outputs?

We agree that case studies with good impact could arise from bodies of work which do not have 2* outputs, and there is support in our community for reducing the minimum requirement to 1* (see also the answer to question 31 below). However, we believe that there should still be a link between impact and a research output.

Although we appreciate and support the attempt to capture as much impact deriving from our research as possible, completely removing the need for research output risks losing the link between the impact and the discipline. There will inevitably be some casualties of this approach – such as preventing the submission of impact deriving from committee membership (in the absence of a direct link to an output). However, we believe that this approach is justified given that the aim of the exercise should be to measure the impact of the research, and not of the researcher. Moreover, measuring the quality of “research activity and bodies of work” seems non-trivial, and complicated. At the least it should be trialled before any implementation.

Q38. 30. Do you agree with the proposed timeframe for the underpinning research activity (1 January 2000 - 31 December 2020)?

No

Comments:

While much computing research matures quickly and can have an impact in just a few years, other research may take more than 20 years to have an impact or may continue to have new impact after 20 years. Thus there is a good case that the time should be increased to at least 25 years; indeed there is a question as to whether there should be any limit, providing there is impact in the current REF period. As an example, Type polymorphism was introduced by Girard in 1972 and independently by Reynolds in 1974, and a variant system (underlying the type systems of ML and Haskell) was introduced by Hindley in 1969 and independently by Milner in 1978. These four systems strongly influenced the design of generic types for Java, which was not released until Java 5 in 2004. Or, for an even more long range example, Boole, Peano, Peirce, and Frege, among others, made seminal contributions to logic in the 19th century, but these did not have an impact on computing until the mid-20th century, when computers were first built.

Q39. 31. What are your views on the suggestion that the threshold criterion for underpinning research, research activity or a body of work should be based on standards of rigour? Do you have suggestions for how rigour could be assessed?

The requirement for the original research to be 2* placed an additional burden on the assessors as they had to trace the original papers and assess more papers. Some good impact was considered ineligible or not submitted as the underlying research was not 2*. Some impact may come from reports not published in academic venues and hence difficult to rate as 2* research outputs. All that should be needed is proof that the impact was based on research within the institution, e.g. by relating to grants, reports written etc. However case studies should not be based purely on consultancy work or "general eminence" of the staff member.

Allowing the submission of impact that derives from expert ise, research activity or body of work would be counterproductive to ongoing efforts to incentivise and embed impact, as it would detract from the role of impact as an activity that sits alongside and strengthens recognised research activities.

Q40. 32a. The suggestion to provide audit evidence to the panels?

All audit evidence (e.g. letters from companies) should be provided to sub-panels as part of a UOA's submission. SP11 found that requesting evidence for audit during the assessment process was complicated and time consuming to request, and so often assessors just did not bother, with a resulting unease about the resulting score. Web references that are easy to access would be acceptable way of submitting such evidence. We understand that the provision of such evidence impinges on commercial confidentiality, but there should be a mechanism to ensure that only the assessors involved see it, if requested by the HEI.

Q41. 32b. The development of guidelines for the use and standard of quantitative data as evidence for impact?

Very little numeric data was provided as evidence in REF 2014. Many organisations refuse to provide generated income related to the impact. For start-ups, the number of people currently employed as well as external funding would be useful.

However, clarity would be welcome as to which quantitative data should be cited, where available, and how, whenever it is available and where it is relevant to the type of impact that is being described. This is consistent with our support for a more structured template for reporting impact (Q27).

It is also the case that ICS vary very widely and a uniform approach to metrics to assess ICS is neither possible nor desirable; assessors must be free to make judgements based on their interpretation (as is the case in the assessment of outputs and citation counts, which only have a very loose correlation with star ratings according to the SP11 panel).

Q42. 32c. Do you have any other comments on evidencing impacts in REF 2021?

Narratives connecting bodies of work and sets of activities by a researcher/research group should be allowed in evidencing impact.

Otherwise, we have no other comment other than the importance of providing clear guidance, which should be aimed in particular at reducing game playing.

Q43. 33. What are your views on the issues and rules around submitting examples of impact in REF 2021 that were returned in REF 2014?

Continuation of impact should be acceptable, i.e. assuming there is a link to research still within the relevant time period, and there is new impact within the current assessment period, then the ICS is valid.

There should be no requirement to make REF2021 panels aware that the impact has been submitted previously, and indeed to mandate this disclosure would cause unjustified biases: the impact should stand on its merits. In accordance with this principle, there should be no requirement for the REF2021 impact case to show additionality in the form of additional outputs or activities. It is sufficient to show the additional impact in the new assessment period.

Page 9: Environment

Q44. 34a. Do you agree with the proposal to change the structure of the environment template by introducing more quantitative data into this aspect of the assessment?

Yes

Comments:

Yes, such a development would simplify the process of providing information that is also readily comparable between UoAs. However, we underscore the importance of retaining an accompanying narrative to provide context for the data and a description of future plans. We also note that although additional metrics could be useful, it is not easy to identify what metrics really correspond to a good research environment. In SP11, analysis showed there was a correspondence between Research Income/FTE and overall GPA, but no correspondence with respect to PhDs/FTE. Some metrics that could be used in environment assessment include numbers of funded research fellowships, overall citation metrics, numbers of fellows of learned societies etc. We understand from SP11 that they did notice last time that there was a good correlation between environment and outputs in particular, suggesting that the existing methodology is reasonably robust. Finally, although not strictly answering the question here, we take this opportunity to note that the scoring system imposed last time on the different parts of the environment template was too granular, having the effect that the minimum quanta of any *rating that could be awarded was 10% (0.5 of 20%). Being able to award a full profile to an environment subsection rather than just a 50%/50% split between two adjacent *ratings would be better in our view.

Q45. 34b. Do you have suggestions of data already held by institutions that would provide panels with a valuable insight into the research environment?

Relevant data already held by institutions could include:

Trajectory of PGR/FTE.

Trajectory of income across different sources, both academic and industrial.

Benchmarked publication metrics, % of papers in the top 1%, 5%, 10% most cited.

Narratives and data that are provided as part of regular reporting to funders e.g. relating to support for research integrity, research data management, equality and diversity data (e.g. status of Athena Swan awards or applications), or the compliance with UK Concordats relating to e.g. researcher development.

Survey data, e.g. PRES.

Number of ECRs holding an independent fellowship.

Number of patents and licensing agreements.

Numbers of papers co-authored by members of another organisation, (particularly international) indicating collaboration might be one such additional piece of data, but each such data requirement may increase submission burden.

Q46. 35. Do you have any comment on the ways in which the environment element can give more recognition to universities' collaboration beyond higher education?

It is appropriate for this activity to be described in the narrative of the environment statement. It may be appropriate to explicitly flag this in the template.

Q47. 36. Do you agree with the proposals for providing additional credit to units for open access?

No

Comments:

The open access policy for outputs has already put considerable burden on institutions to manage this. In some quarters, there seems to be a view that REF should be forcing institutions to behave in a particular way rather than an assessment of research. However we recognise that the REF can have some benefits in helping reinforce good practice. We would note though, that while the principle seems sound, we would caution against using crude metrics to award such extra credit. For instance, resource-rich institutions will most likely be able to evidence a larger spend on open access publications for their researchers (particularly if they are recipients of the RCUK Open Access Block Grant). Taken in absolute terms, such metrics tend to measure existing wealth and the ability to spend it rather than a propensity to openness. We also note that the current requirement to act within 3 months of acceptance is awkward, inconvenient and cumbersome, in particular since the actual date of acceptance in some cases is non auditable. Acting within 3 months of publication (including "on-line first") would be better in our view. This should of course not discourage researchers from acting earlier!

Q48. 37. What comments do you have on ways to incentivise units to share and manage their research data more effectively ?

Environment submissions could indicate what support is provided for open access to the data underpinning research. Some conferences and journals now require this to be made available for submitted papers to allow reproducibility of experiments or evaluation.

Research data generated with the support of taxpayers' money should be publicly available. We suggest that institutions should be rewarded for embedding research data sharing within their culture, for example by providing training on research data sharing to both staff and students – particularly postgraduate research students - so that data sharing becomes part of standard research practice.

Page 10: Institutional level assessment

Q49. 38. What are your views on the introduction of institutional level assessment of impact and environment?

Most institutions would find difficulty in identifying many impact case studies that truly involve multiple departments. Often interdisciplinary research is between different institutions within EU or other research projects rather than between departments in the same institution. If institutions have to submit institution level case studies, it is likely that in many cases these would have to be taken from departments that would then have to find other case studies to submit. It is likely that institution level case studies would increase both the overall assessment and submission burden. It is also unclear then how QR income associated with such ICS should be distributed within the institution (this is a particular issue for those HEIs that distribute QR to departments according to the national formula).

There is perhaps an argument for all ICS to be at UoA level, or all at HEI level, but not a mixed approach. It is also worth noting that HEIs are probably more likely to submit STEM ICS rather than Arts/Humanities, and this is likely to cause friction in an HEI when it comes to allocating QR arising from ICS if all done at institutional level. On the other hand, it is the case that departments falling in the remit of SP11 that are heavily theory based are disadvantaged by having to find sufficient ICS within their UoA (as are relatively new departments, which do not have a long history of research-leading-to-impact).

There are some aspects of environment which are defined at institutional level and hence may be repeated in UoA environment submissions e.g. Institutional research strategy, approach to multi-disciplinarity, staff development etc. Trying to move this into an institutional level submission will just result in additional institutional level submission burden and assessment burden. It is unlikely to make any substantial difference to the UOA assessment burden. Within the same institution there is considerable differences between UoAs as to how their research strategies fits with the institutional one: e.g. what the staffing strategy is, even though there may be a common staff development program; how they manage PhD students etc.

It is not clear how the scoring of institutional environment and impact would fit in with REF profiles which are UoA based. Would some part of the profile come from institutional level scores and would that depend on the size of the UoA?

We also note that in past REFs the quality of environment and impact has varied across UoAs within a HEI, so it is not clear to what extent institutional level assessment is appropriate in assessing an individual UoA.

Q50. 39. Do you have any comments on the factors that should be considered when piloting an institutional level assessment?

We cannot see sufficient benefits in institutional level assessment and so do not think it should be trialled. The only benefits might be a small reduction in assessment of institutional level environment indicators and the ability to assess multi UoA ICS more effectively; there is also a possibility that such a kind of assessment might promote more cross disciplinary activity but we do not regard this as likely.

Moreover, the existing challenges of developing and assessing cross-disciplinary work makes the push for institutional-level impact seem counterproductive and somewhat contrived. It would also be managerially difficult to achieve; for example, it could have divisive consequences when trying to establish the boundary between the UoA-level submission of a multi-disciplinary impact case, versus a truly institutional-level cross-disciplinary case. The proposed expectation that 10–20% of impact submissions would capture impact that occurs at institutional scale is therefore unrealistic.

Excellent cross-disciplinary cases are usually achieved at a smaller scale, often at the level of a Main Panel and consequently expertise within HEIs and REF Main Panels would be better placed to assess such cross-cutting work. REF could provide an opportunity for HEIs to showcase their cross-UoA impact, without expecting a fixed proportion of contributions.

Thus an improved approach might be to review submissions at Main Panel level: this is the level at which there is less redundancy with UoAs and more relevant shared activity, thus ensuring better complementarity between UoA- and higher level strategies. However, there will still be activity which cuts across Main Panels (e.g. health technologies or digital humanities), and we suggest that the overwhelming amount of activity is best assessed at Sub-panel level. Thus it is essential that the ability of the REF process to highlight excellence delivered by individual UoAs within an HEI's overall submission is not lost. Such beacons act as examples of best practice within HEIs and hence are drivers of future quality.

See Q42 re the need to balance UoA-level vs above-UoA-level assessment.

Finally, we recommend that if there is overriding opinion across the country that some assessment take place above the level of a Sub-panel, then a trial would be a good idea.

Page 11: Outcomes and weighting

Q51. 40. What comments do you have on the proposed approach to creating the overall quality profile for each submission?

It does not seem to make much sense for all UoAs to get 5% of their impact scores from case studies they may have had nothing to do with. If impact within a UoA is reduced to 15%, does that imply fewer case studies are needed for the UoA submission? Currently if impact arises from collaboration between a small number of departments, then all the departments can submit the case study written up from their own viewpoint. We also note that it would be undesirable to discourage collaboration on impact-producing research across departments if departments did not benefit subsequently from QR income associated with the ICS.

The suggestion that half the Environment score i.e. 7.5 % comes from institutional assessment seems exceptionally high and implies that the half the research environment is the same for all UoAs in an institution which we do not agree with. There may be some similarity between UoAs within a faculty but there are often very big differences in research culture between faculties such as Arts, Medicine and Engineering.

Different departments have different Athena Swan ratings, CDTs are often within one or a few UoAs so assessing this type of information at Institutional level seems to make little sense

We think introducing institutional level assessment creates many problems with little benefit.

Q52. 41. Given the proposal that the weighting for outputs remain at 65 per cent, do you agree that the overall weighting for impact should remain at 20 per cent?

Yes

Comments:

Increasing Impact weighting to more than 20% as suggested by some would have a detrimental effect on fundamental research. The Research Councils already place a strong emphasis on impact in their pathways to impact requirement and funding of special topics often requiring collaboration with users or industry. Increasing impact rating would mean institutions would put more emphasis on appointing researchers who focus on shorter term impact-producing research rather than people who focus on risky, blue sky fundamental research which really underpins our research culture. Moreover, moving the Impact Template to Environment effectively increases the weighting on ICS. We also believe that the assessment of impact is less reliable than the assessment of outputs, and thus the weighting should not be increased. The environment section is probably the most subjective part of the assessment process (and also the hardest part to apply calibration measures to -- SP11 used automated methods to assist with the calibration between different SP members in both outputs and impact, but there is not enough data (only one environment section per UoA) for any automated methods to be applicable, leaving the results more vulnerable to individual SP member mis-calibration.

Q53. 42. Do you agree with the proposed split of the weightings between the institutional and submission level elements of impact and environment?

No

Comments:

See comments above.

Page 12: Proposed timetable for REF 2021

Q54. 43. What comments do you have on the proposed timetable for REF 2021?

We agree with the timetable. We note that there is some risk associated with the potential disruption caused by the implementation of the HE & Research Bill.

Page 13: Other

Q55. 44. Are there proposals not referred to above, or captured in your response so far, that you feel should be considered? If so, what are they and what is the rationale for their inclusion?

In REF 2014, sub panels could not attribute more than two different star ratings to any subsection of the Environment section (e.g. one section could be 100% 3*, or 50%* and 50% 2*, but no other combinations were possible). We understand that SP11 felt that this was over constraining as sometimes they felt that a sub-component had aspects of, e.g. 2*, 3* and 4* activity. Moreover, the granularity of attribution of star ratings to an Environment section was felt to be too coarse -- the minimum amount of any star rating possible for the whole aggregated Environment section was 10% (if a 20% subsection scored a mixed rating). It would be desirable to allow smaller scores, at least at the 5% level to be possible.

Page 14: Contact details

Q56. If you would be happy to be contacted in the event of any follow-up questions, please provide a contact email address.

a.g.cohn@leeds.ac.uk