

ARC ERA and E&I Consultation: Submission by Professor Cameron Neylon, Curtin Open Knowledge Initiative

Background

The Curtin Open Knowledge Initiative is a strategic research initiative supported by Curtin University, based in the Centre for Culture and Technology in the School of Media, Creative Arts and Social Enquiry. As part of a broader research project engaging the future of universities we have developed a large (12 trillion+ data points) data resource focussed on research outputs and university performance, which we believe to be one of the largest available globally.

We are one of the world's leading groups focused on how evaluation is related to innovation in practice and culture, and how research practice relates to wider impacts and engagement. We take a critical and data-led approach to questions of how evaluation leads, or does not lead, to change and whether that change corresponds to substantive performance or merely improves metrics. One of our particular concerns is how evaluation and indicator design can provide robust and reliable information on performance. In most cases we find that existing, widely accepted, indicators are not fit for purposes.

With the global pandemic focusing attention on the openness and transparency of research, and the effectiveness of communications and engagement strategies for research, evaluation and assessment processes offer an effective way of signalling a desired direction of change, and for encouraging that change. Situating ERA and E&I Assessment in the context of an agenda for shaping an Australian research capacity that engages research end-users in driving impact offers significant potential.

At the same time, the effectiveness of evaluation processes in driving change can be reduced if they create significant burden. Diversion of resources from supporting improved practice into enhancing evaluation performance is a significant risk. Attention must be given to a constant process of reducing reporting burden, reducing opportunities for gaming and supporting a diversity of approaches to achieving national and local goals. In this sense this submission is focused on supporting greater delivery of the indicator principles for ERA, those of *flexibility* that underpins and supports a diversity of approaches and responses, *robustness*, that prevents issues of gaming and instrumental approaches domination, and a methodological *reliability* that means that proxies deliver information that is of value and can be applied. Alongside this we focus on means of reducing burden.

We recommend a move towards automated data collection approaches for quantitative proxies, underpinned by a National Data Asset that provides shared capacity. This would enable a reduction in reporting burden, an increase in data quality through stakeholder curation, and provide for the re-use of data in other contexts including strategy and policy development.

Summary and recommendations

Detailed responses to the submission questions are given below. Here I summarise our core recommendations and observations:

- The current exercise involves a high reporting burden, a substantial stop-start cycle of preparation for higher education institutions, and uses quantitative indicators with limited reliability and methodological robustness.
- The data used is non-transparent, proprietary and limited in its scope in ways that bias the evaluation processes against specific disciplines and cross-disciplinary and applied work.
- The proxy indicators used in ERA and E&I are in many cases adequate (although there are significant issues with the ERA citation benchmarks and indicators) but should be supplemented with a much wider range of indicators to support a diversity of activities and approaches.
- Reporting burden and the stop-start cycle could be addressed through the creation of a shared national data asset that continuously collects relevant data in real time and allows for curation of that underpinning data. Such a system is technically feasible today.
- Such a collection process has a number of substantial benefits:
 - It supports transparency and openness of the evaluation system, reducing opportunities for gaming and structural bias against less rewarded institutions.
 - It substantially reduces reporting burden through automation.
 - It creates flexibility and enables experimentation with improved evaluation proxies that can better capture a dynamic and diverse research environment.
- Such a system could not be readily implemented and fully validated in time for the next ERA/E&I Assessment cycle, but pilot data collection could be conducted and real-time comparison of alternative indicators and benchmarks explored. This would enable a full investigation of the feasibility of implementing such a system for future ERA/E&I exercises.
- E&I Assessment should be conducted alongside the ERA with many common elements of data being collected. A three-year cycle seems appropriate, provided reporting burden can be released. A regular formal process including qualitative reporting could be supported by continuous reporting of key proxy indicators by a real time system.

Detailed Responses

ERA policy

Value of ERA

Q3.1 To what extent is ERA meeting its objectives to:

a. Continue to develop and maintain an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australian higher education institutions. A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

A small amount: The consultation document notes that the sector and other stakeholders respond to ERA outputs. This suggests engagement with the results of the ERA process but this does not mean that the actual underpinning framework provides robust results.

b. Provide a national stocktake of discipline level areas of research strength and areas where there is opportunity for development in Australian higher education institutions. A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

A moderate amount: ERA provides a valuable sense of the volume of activity in the categories (FoRs) to which outputs are assigned. Due to limitations in the methodology applied in provides only a narrow view of the qualities and strengths of works in disciplines.

c. Identify excellence across the full spectrum of research performance. A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

Not at all: What constitutes “excellence” does not appear to be defined, interrogated or discussed in any of the supporting documentation. Rather a shared conception is assumed throughout. Further, the use of narrow quantitative metrics is then assumed to capture that (undefined) shared conception. The connection between the metrics applied and an agreed conception of what “excellence” means is not articulated, or understood, by the communities whose research performance is being evaluated.

d. Identify emerging research areas and opportunities for further development. A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

Not at all: The overall approach limits the visibility of new strands and approaches to research by forcing them into the existing FoR system which, as we argue below, is not fit for this purpose. As an example our own research has no obvious place within the FoR system. More generally, innovative data-led work with social relevance is frequently assigned to FoRs relating to their methodology, where it will generally be masked by the higher average citation counts captured for computer science. If there is an FoR code appropriate for the object of study it may either be highly cited in comparison or lower cited in comparison. However, the design of the ERA categorisation and FoR system means new disciplines will be reported across multiple FoRs and therefore any signal will be lost.

e. Allow for comparisons of research in Australia, nationally and internationally, for all discipline areas. A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

A moderate amount: ERA results allow for comparisons to be made. We would argue that those comparisons are not robust to small changes in methodology and should not be used as the sole basis for policy or intervention design.

Q3.2 The ERA objectives are appropriate for meeting the future needs of its stakeholders. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: The Objectives are generally good ones, with the caveat that the concept of “excellence” needs to be better defined for the purposes of the assessment. Objective 3 is unclear and possibly internally contradictory without clarity of definition. The definition of excellence used by the UK REF is one example of this. As a point of principle we would argue that it would be more productive to replace “excellence” with “qualities”.

Q3.4 How do you use ERA outcomes? Please describe.

Not applicable. We do not use ERA outcomes as they are not useful data for our purposes.

Q3.5 ERA outcomes are beneficial to you/your organisation. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Not applicable

Q3.6 Do you have any suggestions for enhancing ERA's value to you/your organisation? Please explain your answer.

Not applicable

ERA methodology

ERA methodology at a glance

Q3.7 The current methodology meets the objectives of ERA. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly disagree: Current methodology is limited both from a technical and robustness perspective. More details below.

Q3.8 What are the strengths of the overall methodology? Please describe.

Not applicable.

Q3.9 What are the weaknesses of the overall methodology? Please describe.

The quantitative measures of quality utilised have substantial weaknesses both as metrics in their own right and with respect to the methodology applied. The data used is non-transparent and relies on closed metadata (the Journal List).

The collection and submission of data creates a massive burden on the sector; and the dataset created for each ERA exercise cannot be built upon or re-used by the sector in a meaningful way. Alternative approaches could transform ERA from a burdensome data collection and submission exercise with limited wider benefit for the institutions being measured; into an opportunity for institutions to contribute to the creation of a national data asset that captures key data relating to higher education and research, and which supports strategic decision making for stakeholders at all levels of a dynamic sector.

Citation analysis methodology

Q3.10 The citation analysis methodology for evaluating the quality of research is appropriate. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strong disagree: The citation analysis methodology has substantial issues across the entire framework. Here we divide those issues up by data issues, methodological issues, potential for gaming, and comparative issues, and then consider the implications of those issues for the goals articulated for ERA.

Data issues:

- The input data used is proprietary and non-transparent.
- This creates an equity issue, particularly given the opportunities the methodology provides for gaming.
- It also privileges a specific provider of data, with known limitations in coverage for both the outputs themselves, and for the input metrics (see also Huang et al, (2020) https://doi.org/10.1162/qss_a_00031).
- **Alternative Approaches** that address these issues would be to use a publicly available and open dataset or to create a common national data asset that was shared amongst stakeholders. This would create a level playing field and allow for innovation in continually enhancing the set of indicators applied.
- Such an approach could enable institutions to define their own goals and select appropriate indicators for them to be judged against, allowing greater diversity of approaches and enabling evaluated institutions to feature the qualities of innovative work, supporting ERA to *identify emerging research areas and opportunities*

Methodological issues:

- Normalization as a whole is methodologically questionable. A robust normalisation strategy requires clarity on the characteristics that are being normalised for and those that are not. The category of “outputs assigned by some university to an FoR code” is not well defined (see also issues with the FoR-based approach).
- As an example with respect to the goals of ERA (eg “stocktake of discipline level areas of research strength”, “opportunity for development”) the current normalisation strategy will mask strengths

with a national scope, and strengths and opportunities that are tracking with international developments as these signals will be normalised out.

- The use of an arithmetic mean is generally regarded as inappropriate for any normalisation against highly skewed distributions that citations present. A geometric mean would be more appropriate in this case.
- The use of a Journal List to provide the global-normalization for an output-level analysis is inappropriate as it applies two different sampling strategies (and therefore samples with differing variance) to numerator and denominator.
- Self-assignment of outputs to FoR codes creates similar issues, as does the small number threshold (see also gaming below) as the sampling strategy is not uniform.
- Grouping by year for recent years means that there will be a differential across month of publication for normalised metrics as some outputs have had a longer time to accrue citations.
 - As an example in an analysis on COKI data run in October 2020 for citation counts of articles in Medicine published in 2019, articles published in January had on average over twice the citations of articles published in December. That is, outputs from January would on average receive a normalised RCI twice that of outputs from December.
 - For 2018 the difference for medicine was still 40%. The difference for Philosophy for 2018 was two-fold. See also Donner (2018) <https://doi.org/10.1016/j.joi.2018.01.012>
 - Note also that some journals and publishers manipulate the formal publication date for their own reasons and journals vary in their publication schedule. This in turn has a disciplinary dependency.
- **Alternative approaches** include a different approaches to identifying normalisation benchmark sets such as describing the co-citation set for outputs (the set of all outputs cited by those outputs that cite the target output), the use of a smaller set of categories, automated assignment of categories to reduce gaming.
- The year issue could be addressed by applying a higher resolution of date buckets to the normalisation approach (Donner et al recommend the month level, and our small scale analysis suggests this is viable) or to using a citation rate since publication indicator (although both of these may be subject to small number effects).

Gaming Issues:

- The combination of the FoR categorisation, self assignment and access to data makes a range of strategies available for gaming. All of these will reduce the value of the information provided and limit the ability to deliver on ERA goals.
- Self selection of FoRs is the primary area of gaming as well as being the greatest burden for submission. This creates a perverse incentive whereby the opportunity to enhance scoring increases incentive to adopt further burdens. This also creates an equity issue and dilutes the signal from new areas of innovation (which will either be dumped into low number or poorly performing FoRs, or alternatively used to bolster existing strengths).
- As a secondary effect the opportunities for gaming may lead to perverse incentives for publication practices and venues, which have the potential to reduce both academic and wider impacts. Journals might be sought based on their supposed FoR rather than on the optimal target audience for the research. Again this will dilute signals of innovative new fields and potentially lead to noise amongst high performing fields. This mode of gaming could potentially be exacerbated by an automated categorisation process.

Issues for Comparisons:

- The goal of the normalisation process is to provide indicators that are comparable across disciplines and FoR categories. Discipline variance and methodological issues mean that they do not achieve this goal and do not deliver on the ERA goals or guiding principles (“allow for comparisons of research...for all discipline areas”, “robust”)
- Due to the lack of definitional clarity there is no underlying model for what a comparison across disciplines would seek to achieve. Addressing this would require more targeted indicators that report on specific aspects of research activity and practice. For example, a discipline might

collaborate more with industry (or policy implementation) than another and co-authorships with industry (or citation from policy documents) might be a proxy for this.

Implications for delivery on the Goals and Guiding Principles of ERA:

...continue to develop and maintain an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australian higher education institutions

The consultation document notes current confidence in the ERA exercise. We suggest that the methodological issues raised above are serious enough that the presumption that the current methodology will *continue* to deliver confidence is not *robust* or *reliable*.

...provide a national stocktake of discipline level areas of research strength and areas where there is opportunity for development in Australian higher education institutions...identify emerging research areas and opportunities for further development

ERA in its current form does provide a stocktake of research volume across broad disciplines (a *reliable* and *robust* approach if raising some issues of limited *flexibility*) and some narrow (if methodologically problematic) indicators that relate to some specific qualities of that research. As noted above we do not see it as *robust* or *reliable* in identifying opportunities and areas for development.

In particular the methodology is very weak on identifying new strands of research and practice. This relates both to the narrowness of the measure (citations, which are known to be biased against novel practice, under represented researchers and emerging areas, making it not *flexible*) and the methodology (specifically that a normalisation approach will mask emerging areas and dilute the signal from them, which is not *robust* for the intended purpose).

...identify excellence across the full spectrum of research performance

The citation methodology and approach is wholly inadequate to address this goal, which as noted above is not clearly defined enough to be actionable.

...allow for comparisons of research in Australia, nationally and internationally, for all discipline areas

The methodology provides the appearance of allowing for comparison but such comparisons could not be regarded as *robust* or *reliable*.

Q3.11 Does the discipline-specific approach for evaluating research quality (citation analysis or peer review for specific disciplines) continue to enable robust and comparable evaluation across all disciplines?

A discipline specific approach is crucial and is more important than comparability across disciplines. Despite the misgivings detailed above we would note that some citation-based indicators are likely to remain important across many disciplines, while being totally inappropriate in others. Therefore to address the need for *flexibility* as well as *robust* evaluation it is crucial to support discipline specific approaches.

Noting the binary emphasis in the question there might be some value in considering how a mixed approach could be applied across all disciplines where a “basket” of consistent evidence is provided across all disciplines but selected from and applied in ways that are specific to each discipline.

Q3.12 What are the strengths of the citation analysis methodology? Please describe.

See above

Q3.13 What are the weaknesses of the citation analysis methodology? Please describe.

See above

Q3.14 Can the citation analysis methodology be modified to improve the evaluation process while still adhering to the ERA Indicator Principles? Yes/No.

The citation analysis methodology must be changed if it is to adhere to the ERA Indicator Principles and address the ERA objectives as discussed above.

Contextual indicators

We note with concern the statement that the “...presence or absence [of contextual indicators] has virtually no effect on the rating given to a unit of evaluation”. This is directly opposed to the goals articulated for ERA with respect to identify *excellence across the full spectrum of research performance* and *emerging research areas and opportunities for further development*. Both of these goals imply a desire to support a diversity of approaches as does the principle of *flexibility*.

As a trivial example, a research unit conducting world-leading research in an area of social relevance may choose to prioritise communication to affected communities over formal scholarly publishing. This might mean choosing to not play the wasteful game of submitting to multiple journals in the prestige hierarchy but going directly to a good journal, or it might mean focussing communications on the affected communities rather than conference attendance to boost citations. This is a valid strategy for excellent research involving engagement that drives impact.

The contextual indicators provide the most important opportunity to address these objectives and unlike citation data have the potential to provide *leading* rather than *lagging* indicators of activity. If they do not have a substantive effect on the ratings given, particularly in those disciplines driven by citation analysis, then this should be urgently addressed.

Q3.19 The volume and activity indicators are still relevant to ERA. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly agree: Volume and activity indicators provide a valuable longitudinal indicator that is *robust* over time and will allow for further analysis. While imperfect and not sufficiently *flexible* to allow for comparison across disciplines it is a useful indicator within disciplines for observing trends and opportunities, allowing for the limitations of the FoR system.

Q3.20 The publishing profile indicator is still relevant to ERA. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree. No strong view except to note that diversity of publishing approaches within a research unit may be a strength and naive measures of “depth” and “breadth” may have limited value in identifying the qualities of excellent units.

Q3.21 The research income indicators are still relevant to ERA. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: Research income indicators at least have the value of being leading indicators of research activity and outputs. Diversity of income is a valid indicator, alongside evidence of engagement. As we argue for integration of ERA and E&I we would see this as not necessarily raising a tension but an opportunity to combine the concerns of E&I into assessment of excellence.

Q3.22 The applied measures are still relevant to ERA:

- a. Patents. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*
- b. Research commercialisation income. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*
- c. Registered designs. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*
- d. Plant breeder's rights. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*
- e. NHMRC endorsed guidelines. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*

This set of applications are all valuable but are limited in scope. An improved approach would enable a range of evidence to be collated and presented on the application of research that could be expanded over time. Substantial areas of application including use in environmental assessments, policy design and implementation, adoption by user communities are not addressed here. Application is a different phase of the research cycle than Impact and care is needed to ensure that good practice in the application of research is not encouraged, even where downstream Impact may not (yet) be realised.

Expanding the evidence base for the applied measures would create challenges of comparability. However as part of the contextual indicators a more qualitative approach may be appropriate.

ERA rating scale

Q3.23 The five-band ERA rating scale is suitable for assessing research excellence. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: The five band scale provides a reasonable level of granularity without creating an explicit ranking which reduces but does not remove the risk of perverse incentives the pure rankings create. Alternatives are generally no better and change will achieve little so retaining the status quo makes most sense.

Q3.24 Noting that 90% of units of evaluation assessed in ERA 2018 are now at or above world standard, does the rating scale need to be modified to identify excellence? Yes/No.

- a. If you answered, 'Yes', please explain how the rating scale can be modified to identify excellence.*

This is a predictable product of the evaluation process. Recommended best practice is to group UoE into appropriate bands (quartiles or quintiles) and to use indicators that are not top-limited (i.e. analytic variables rather than rankings). Reporting on a range of relevant indicators for top performing UoE's still allows them to improve in future assessment while also allowing other institutions to use comparative performance as a guide to strategy development.

ERA low-volume threshold

Q3.25 The ERA low-volume threshold is appropriate. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: A low volume threshold is necessary to address statistical issues. This would largely disappear if the Units of Evaluation had a higher level of granularity.

Q3.26 Are there ways in which the low-volume threshold could be modified to improve the evaluation process? Please describe.

As above we would recommend that evaluation occur at a higher level of granularity (eg two-digit FoR level). This would allow interesting trends to surface through categorisation at higher levels of granularity without requiring that the statistical behaviour be monitored so thoroughly. Through removing the perverse incentives that arise from the current evaluation at the 4-digit level.

ERA staff census date

Q3.27 What is the more appropriate method for universities to claim research outputs—staff census date or by-line? Please explain your answer.

This issue was extensively explored in the consultations leading to the 2020/21 UK REF process. There are clear arguments for and against both approaches. All other things being equal we would argue for retaining the status quo census date so as to minimise unintended side effects of a shift in approach. Major concerns were raised in the UK consultation that focused on issues, particularly for ECRs of a shift in approach that would be caused by a change.

Q3.28 What are the limitations of a census date approach? Please describe.

The primary risk with a census based approach is of “ERA-shopping” of academics which creates cliff-edges effects with particularly damaging consequences for ECRs who happen to be at the wrong stage of their career at the critical point of the ERA cycle.

A limitation of census-based approach is that it necessarily requires a manual data collection process, whereas the byline approach can be more easily automated. This could be addressed through more systematic use of ORCID (or encouraged by allowing ORCID to be used as the means of providing census-date status) but this would require substantial background work to achieve sufficient data quality.

Q3.29 Would a by-line approach address these limitations? Yes/No. Please explain your answer.

A byline approach allows for more automated data collection and also reduces some of the risks of cliff-edge census date effects. There are substantial issues of data quality to be addressed, with affiliation completeness and correctness an issue across all providers of bibliographic data. A shared national data asset that collected and curated such data could address these issues.

Byline approaches (including those adopted by international rankings) encourage “affiliation-washing” where high profile academics are given honorary or even paid appointments on condition of adding affiliations to their research outputs. This would be a substantial risk in a byline-based approach and would be challenging to address in an automated pipeline. For these reasons we recommend retaining that status-quo census approach.

Q3.30 What are the limitations of a by-line approach? Please describe.

See above.

ERA interdisciplinary research and new topics

Q3.31 ERA adequately captures and evaluates interdisciplinary research. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strong disagree

- a. If you disagreed with the previous statement, how could interdisciplinary research best be accommodated? Please describe.*

ERA along with many evaluation exercises focuses on disciplinary analysis. The unique aspects of cross- and interdisciplinary work are rarely effectively captured. The application of disciplinary assumptions of quality, and how those are expressed, rarely allows interdisciplinary work to shine as it generally combines qualities and approaches from across the input disciplines. Allowing for a wider diversity of indicators and supporting case study approaches can help, but a global analysis of output by discipline, especially at the level of granularity that ERA applies will never be well suited to identifying outstanding interdisciplinary research.

Short of a radical overhaul in approach the most effective way of identifying high quality interdisciplinary research is to incorporate the engagement and impact assessments as reflecting valued qualities of research. This enables research that is important, but does not perform well on traditional measures, to be nonetheless identified and valued.

ERA process

Collection of ERA data

Q3.35 ERA should move to an annual collection of data from universities. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree. If the collection of ERA data is streamlined and contributes to/draws from a continuously updateable national data asset (see section 5 below) then the burden of the ERA process on research institutions will become more manageable. We recommend the creation of a national data asset that would enable continuous real-time collection of data alongside regular formal assessment points to ensure a focus on quality assurance and curation of that data by HEPs.

An annual collection of data in the current form will create substantial additional burden and is therefore not recommended.

Q3.36 What would be the advantages and/or disadvantages of an annual data collection? Please describe.

N/A - see above

Publication of ERA data

Q3.37 In future ERA rounds, should the volume of outputs submitted for each unit of evaluation be included in the National Report?

Yes

a. *Please explain your answer.*

Making the volume of outputs submitted publicly available is an important step towards increasing the transparency, and therefore the credibility, of the ERA process. It provides *robust* longitudinal measures that are of value in tracking the development of Australian research.

Q3.38 In future ERA rounds, research outputs should be published with their assignment to specific disciplines following completion of the round. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Agree.

a. *What would be the advantages? Please explain your answer.*

Publishing this information will increase the transparency and trustworthiness of ERA, allowing universities, research managers, and researchers, to gain a deeper insight into how discipline categories are understood and determined within the ERA process. It will also drive a greater scrutiny of choices in discipline categorisation. Finally it can support the development, improvement and validation of systems for automated assignment of discipline that could reduce the burden in future exercises.

b. *What would be the disadvantages? Please explain your answer.*

N/A

Q3.39 What other data do you think the ARC should publish following an ERA round? Please describe.

In the interest of transparency ARC should publish all input data into the assessment process. This will allow for critique and future improvement. It also allows for a level playing field and limits inequities due to uneven data access or capacities amongst institutions. Data would include details of the implementation of benchmarks ideally in the form of running code. Transparent approaches will allow for sensitivity analysis, improving *robustness* and *reliability* as well as enabling stakeholder to identify issues that arise in the evaluation improving future *flexibility*.

An issue with all data-led evaluations is that they embody existing biases in data collection and analysis and therefore existing power relationships. This is a particular issue in addressing the serious problems of inclusion and bias within our research institutions. Critique of representation and bias within datasets is a crucial means of improving equity and diversity in Australian research.

Section 4—Engagement and Impact Assessment

EI Overview

Q4.1 Considering that EI is a new assessment, to what extent is it meeting its objectives to:

In answer to the question as a whole the pilot implementation of E&I was too narrow to address the objectives at scale and it is too early to make a full assessment. The following is therefore based on experience of the REF Impact assessment and its evolution rather than a detailed assessment of the impact of the E&I pilot.

a. encourage greater collaboration between universities and research end-users, such as industry, by assessing engagement and impact? A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

At this point only a small impact. Evidence from the UK REF Impact assessment is that case studies and the inclusion in the REF raises awareness and has, after several years, raised the standard of discussion and awareness of the qualities of collaboration.

b. provide clarity to the Government and the Australian public about how their investments in university research translate into tangible benefits beyond academia? A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

In the UK case the REF case studies have provided powerful narratives that have supported government and public awareness of the impact of research investments. This is probably the most positive outcome of the whole process.

c. identify institutional processes and infrastructure that enable research engagement? A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

At this point a limited effect. The UK case demonstrates a risk that investment will be in the production of high quality case studies rather than support for systems and infrastructures that enable research engagement. A far more sophisticated approach is called for in delivering this objective that is fully aware of the complexities of investment, the long term nature of its benefits, and the way in which evaluation processes drive instrumental rather than substantive effects.

d. promote greater support for the translation of research impact within institutions for the benefit of Australia beyond academia? A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

Here the E&I exercise has had an effect on the awareness within institutions of these issues. The challenge lies in shifting that awareness from wariness and antagonism towards positive engagement. This is a highly complex issue that will only occur over an extended period.

e. identify the ways in which institutions currently translate research into impact? A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.

In the UK there has been limited synthetic analysis of case studies that enable a broad understanding of modes of engagement, impact and the details of institutional support. It is deeply problematic to expect such a complex system to be effectively revealed through an evaluation exercise. Case studies and their deep analysis offer one potential route, but these are necessarily advocacy documents. Actually understanding Australian research impact pathways would be better supported by a separate policy evidence development approach that would be disconnected from evaluation. There are questions to be asked about whether institution-level modes of impact optimisation are appropriate in the Australian context vs State or National approaches.

Q4.2 The EI objectives are appropriate for the future needs of its stakeholders. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: The E&I objectives cover important issues with respect to these issues and for the future development of Australian research and its underpinning support systems. However some of those objectives are not a good fit for an evaluation-driven policy implementation, particularly as noted the goal of identifying and understanding pathways to impact.

EI definitions

Q4.7 The current Engagement definition is appropriate. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree.

Agree: The definition is broadly appropriate. We would argue that two modifications would strengthen the definition, and also reduce its overlap with that of impact: “Research engagement is the *productive* interaction between researchers and research end-users outside of academia, with *the goal of mutually beneficial* transfer of knowledge, technologies, methods or resources.”

“Productive” emphasises that this should be of real value (and not trivial or instrumental in nature) and emphasising goals rather than outcomes reduces a risk that the definition only focuses on successful engagement, rather than on a continual process of improving engagement, which will necessarily involve some failures.

Q4.8 The current Impact definition is appropriate. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree.

Agree: The definition is adequate. I have long argued that the distinction between “academic” and “non-academic” impact and that we should understand impact as all effects of research, with the differences lying in which stakeholders are involved. That is, we can learn as much about impact from looking at uptake of research results in research as we can from looking at uptake in industry or society. Further that distinguishing between research that is excellent in its uptake in research vs research that is excellent in its uptake beyond research clouds the important issues.

However, this is a substantial shift in approach and definition and outside the scope of the current consultation.

Q4.9 The current end-user definition is appropriate. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree.

Neither agree nor disagree: The definition is broadly adequate but limited (see below). Given the focus on E&I it would seem useful to add “engagement” to “output, outcome or result” as end-users may benefit from engagement separate to the expected or unexpected results of research. Indeed this specifically addresses objectives of the E&I assessment *vis* “provide clarity to the...Australian public...”. As noted, the definition focuses on the drawing of a line between academia and “external to academia”.

a. If you don't agree, what are your suggested amendments to the end-user definition? Please describe.

See above.

b. Are there any end-user categories excluded in the current definition of research end-user that you think should be included? Please explain your answer.

As a group focussed on research on research it is not in our interest that research organisations, HEPs and others are excluded. This exclusion of research on research or metaresearch extends beyond the E&I program to the FoR and SEO classifications which also fail to capture this kind of work. Given the structure of the assessments and the division between “excellence” and “engagement and impact” there is a need to make that distinction.

However, given the HEPs are one of Australia’s largest industries and a national strategic asset in difficult times, research into improvement in practice and delivery needs to be recognised as valid and considered within the evaluation landscape.

EI methodology

Unit of assessment

Q4.11 Are the two-digit Field of Research codes the most appropriate method to define units of assessment for Engagement and Impact? Yes/No. Please explain your answer.

As we argue for a less granular approach to ERA and for the combination of the two assessment processes this is an appropriate categorisation. It broadly maps to the UK REF Impact Case Studies which are generally regarded as best current practice.

Q4.12 Are there other ways to classify units of assessment in EI, for example, SEO codes? Yes/No. Please explain your answer.

There are many possible classifications, including a potential to reduce burden through more automated classification. In the absence of a clearly better alternative we would argue to retain the status quo.

Selectiveness of EI

Q4.13 Should there be more or fewer units of assessment per university? More units of assessment; The same number as in EI 2018; Fewer units of assessment.

More.

a. How many and why? Please explain your answer.

There would be a benefit to the sector and to institutions in increasing the number of case studies. The collection and further analysis of case studies from the UK REF has been one of its most positive outcomes. However, this clearly needs to be balanced against burden. Case studies are intensive and challenging and there is a risk of creating further work which distracts from the actual goals of increasing engagement and impact.

A modest increase, with an option to consider the number and adjust in future would be of value. The UK REF level of one per ten staff members returned would likely be too high. The perverse effect on staff returns for the UK REF where there were cliff edges at staff numbers just below those that would trigger a requirement for a further case study is clear evidence that this burden was too high. An increase to two or three per area of evaluation, allowing for a low-volume threshold might be an appropriate step.

EI low-volume threshold

Q4.14 The EI low-volume threshold should continue to be based on the number of research outputs submitted for ERA. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree.

Agree: The biggest risk of the E&I Assessment is that it creates a level of burden that has a negative effect. A low-volume threshold is therefore necessary. The outputs volume is probably the simplest criterion to be used and is already collected. An alternative would be HERDC Category 1 income assigned to the two-digit FOR level but it is unclear what this would add.

Q4.15 The low volume threshold is set at the appropriate level. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: No strong view.

Engagement indicators

Q4.16 Overall, the engagement indicator suite for the assessment of research engagement is suitable. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strong disagree: The indicators are all useful but are wholly inadequate to address the broader objectives of the E&I Assessment. While defining indicators of community engagement, stakeholder involvement and other more social than industrial forms of contact is challenging there are opportunities to develop a wider range of proxies. The current set of indicators focus almost exclusively on direct economic impacts, whereas broader impacts through society may in fact be much larger, including in economic terms.

Q4.17 The cash support from research end-users indicator using HERDC data is appropriate for the assessment of research engagement? Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree. Please explain your answer.

Neither agree nor disagree: The question is ambiguous. Such cash support is an appropriate indicator amongst others.

Q4.18 The research commercialisation income is appropriate for the assessment of research engagement. Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree. Please explain your answer

Neither agree nor disagree: As above. Such income is an appropriate indicator amongst others.

Q4.19 Are there additional metrics that would be appropriate across many or all disciplines? Yes/No. If you answered 'Yes', please outline the metrics. If you answered 'No', please explain your answer.

Yes:

A non-exhaustive list of additional proxies that might be of value are:

1. Outputs in specialist and end-user media (both digital and print), including evidence of sustained engagement with such venues (rather than one-off events)
2. Evidence of public attendance at on campus events, potentially including attendance numbers logged through ticketing systems but also potentially proportion of all users of campus parking or campus public transport that are non-university members.
3. Value of investment in the support of stakeholder fora (potentially including industrial translation and tech transfer, indigenous community platforms, and programs for public engagement)
4. Count of participants in research-led continuing professional development.
5. Count of citations from patents, policy documents, legislation, regulation and other relevant stakeholder media.
6. Count of submissions to government and community consultations. Count of invitations to give evidence as experts.
7. Count of participation, or logged work-load contributions to expert panels for government, community or industrial stakeholders.
8. Awareness of institutions and their activities through a market-survey like approach run on behalf of the E&I assessment.

Many of these are feasible in part currently, and most would require further development to represent the full diversity of engagement activities.

Q4.20 Are there alternative metrics that would be appropriate across many or all disciplines? Yes/No. Please specify the metrics.

All proxies, including those that are current applied, will have some discipline bias. For instance the current set of income measures are heavily biased towards representation of heavy industry and expensive research disciplines such as engineering and will substantially under-represent engagement that is either un-funded or in research areas that are less expensive. One example of important engagement that is under-counted by current proxies is the work of colleagues at Curtin on advocacy for regulatory requirements on closed-captioning on Australian broadcast TV.

Q4.21 Should any of the current Engagement metrics be redesigned? Yes/No. If you answered 'Yes', which ones and how?

Yes: The current indicators are adequate as part of a broader and more inclusive set of metrics. Their use in the pilot was based on existing reporting and minimising burden and this continues to be sensible. They are useful measures within a broader set of proxies.

Q4.22 The co-supervision of HDR students should be made an engagement indicator in future rounds of EI. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: Co-supervision is one useful proxy in a broader basket of indicators. This is also biased towards technical industries and away from research of social relevance where collaborators may not be qualified

to act as formal supervisors for HDR students. It may also be effectively double counting certain classes of research income, in those some technical industry areas.

Q4.23 In your opinion, are any of the ERA applied measures appropriate indicators of research engagement in EI?

All of the ERA applied measures are potentially useful as part of a broader basket of proxies. They each are biased towards specific fields of research. If used there should also be a focus on seeking not merely counting outputs but also seeking to identify influence across these areas (i.e. patents from citations, references from registered designs and NHMRC guidelines to research) and to ensure that the scope is not merely national (eg clinical guidelines in other countries should be included where feasible)

Engagement narrative

Q4.24 The narrative approach is suitable for describing and assessing research engagement with end-users. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: The narrative approach is valuable as one part of a larger set of evaluation information. Given the diversity of engagement approaches some qualitative evidence is crucial and there are a limited set of alternatives to a narrative or case study approach.

Q4.25 One engagement submission per broad discipline is sufficient for capturing the research engagement within that discipline. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: As with the UoE we have no strong opinion on the optimal number of submissions, and would argue that some increase over time, with a deliberate and careful monitoring of burden would be appropriate.

Q4.26 The engagement narrative needs to be longer. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: No strong opinion.

Q4.27 Additional evidence is needed within the narrative. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: Drawing on the experience of the REF Impact case studies it would appear that emphasising evidence within the narrative is valuable. This should not be circumscribed, limited or standardised but rather be encouraged. The largest value created in the REF Impact case studies was the strength of narrative and qualitative evidence of impact. Here, encouraging an approach that is both *flexible* but also encourages the use of *robust* and *reliable* evidence is of value.

Impact narrative

Q4.28 The narrative approach is suitable for describing and assessing impact. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly agree: Best practice globally is to take a case study narrative approach to allow flexibility in describing the routes towards achieving impact while encouraging high quality evidence. See also above for engagement as many of the same issues apply.

Q4.29 One impact study per broad discipline is sufficient for capturing the research impact within that discipline. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: As above, some increase is appropriate provided it is done gradually and burden is monitored.

Q4.30 The impact narrative needs to be longer. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: No strong opinion.

Q4.31 There is a need for additional evidence to be provided within the narrative. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: See above under engagement as the same issues and arguments apply.

Q4.32 In your opinion, are there quantitative indicators that could be used to measure the impact of research outside of academia? Yes/No. Please explain your answer.

No: This is an issue of definition and quantitative indicators are not possible in principle, let alone in practice. The definition of impact is open ended intentionally. "Impact" as defined is not something that can be identified except through knowledge of its causes. That is it is defined not by its inherent qualities but by its origins in academic research.

There may be specific proxies of impact in specific fields, but these will be indirect and cannot be generalised across disciplines. The appropriate proxies for judging the importance of impacts will be tied to class of effects (eg increase in GDP, jobs created, children not dead i.e. "the body uncount", increase in democratic participation, reduced recidivism, greater artistic output) not to the fact that the causative factor is academic research. To suggest otherwise is a category error.

Approach to impact Narrative

Q4.33 The narrative approach is suitable for describing and assessing approach to impact. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: See above

Q4.34 One approach to impact narrative per broad discipline is sufficient for capturing the activities within that discipline. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: No strong opinion.

Q4.35 The approach to impact narrative needs to be longer. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Neither agree nor disagree: No strong opinion.

Q4.36 There is a need for additional evidence to be provided. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: See above.

Q4.37 Would there be benefit in combining engagement and approach to impact? Yes/No. Please explain your answer.

Yes: Broadly speaking as impact is defined, approaches to impact can all be tackled under the rubric of engagement. Combining qualitative assessments will reduce burden and encourage more connected thinking. This is equally an argument for seeing engagement as a component of research quality and impacts as a signal of research excellence, demonstrating the common aspects of academic and non-academic impacts.

EI rating scales

Q4.38 The engagement rating scale is suitable for assessing research engagement. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: There is little value in seeking more granular assessment in a necessarily qualitative evaluation for which community standards need to be developed over time.

Q4.39 The descriptors for the engagement rating scale are suitable. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Disagree: Distinctions between High and Medium are vague and might be better focussed by addressing a quartile or quintile description of what "highly" is meant to mean in modifying such vague terms as "effective" and "contributions".

Q4.40 The impact rating scale is suitable for assessing impact. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: As above.

Q4.41 The descriptors for the impact rating scale are suitable. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Disagree: As above.

Q4.42 The approach to impact rating scale is suitable for assessing approach to impact. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Agree: As above.

Q4.43 The descriptions for the approach to impact rating scale are suitable. Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.

Disagree: As above.

EI interdisciplinary research

Q4.44 Should EI continue to include an interdisciplinary impact study in addition to the two-digit Fields of Research impact studies? Yes/No. Please explain your answer.

Yes: This is a valuable opportunity to focus on interdisciplinary research, and to signal its importance in the evaluation process.

Section 5—Overarching Issues Common to both ERA and EI

Frequency of ERA and EI

Q5.1 How often should ERA occur? Every three years; Every five years; Other, please specify. Please explain your answer.

Given the current burden the frequency should not be increased unless that can be reduced. If the exercise is to be run in broadly its current form then approximately once every three years on a predictable schedule is to be preferred.

If burden can be substantially reduced through automation and shared processes then a continuous process for automated data collection with a two or three year cycle for formal assessment including qualitative submissions could be developed. This would need to pay particular attention to the risks of non-representation of some disciplines in continuous data and would likely need to consider the introduction of a wider range of quantitative proxies to address this.

Q5.2 What impact would a longer assessment cycle (i.e. greater than three years) have on the value of ERA results, particularly in the intervening years? Please explain your answer.

We currently have little better than anecdotal evidence of how policy interventions and changes in the landscape are affecting Australian research. This is particularly the case as we look to monitor the response to and recovery from the pandemic. Longer cycles will reduce our capacity for strategic planning. A low-burden continuous monitoring process based on the development of a national data asset would mean that lower frequency (e.g. five year) could be managed with less impact provided the caveats noted above were taken into consideration.

Q5.3 How often should the EI assessment occur? Every three years; Every five years; Other, please specify. Please explain your answer.

Three years/the same as ERA: We argue for a combination of the ERA and EI assessments (see below). This focuses attention on E&I as part of the high prestige ERA process and sends a message of its importance to government and society. Assuming that a goal of the E&I Assessment is to increase a focus on engagement amongst HEPs then sufficient regularity is required to prompt ongoing improvement in practice.

Q5.4 What impact would a longer assessment cycle (i.e. greater than three years) have on the value of EI results, particularly in the intervening years? Please explain your answer.

See above (5.2). In particular a longer cycle would lead to slower changes in practice and improvement.

Streamlining and simplifying ERA and EI

Q5.5 ERA and EI should be combined into the one assessment. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly agree: Provided that a combination can be implemented without substantial increases in burden, particularly through automation and sharing of data resources (see below) then a combined approach will offer substantial benefits. In particular it strengthens the case that excellent research is that that achieves real impact through productive and well planned engagement. Properly managed the combination could substantially reduce burden. The major risk is that combining the two creates an increased burden and significant stop-start cycles over the period of assessment. For this reason we argue that two to three year cycle combined with continuous data collection is the most productive route forward over the medium term.

Q5.6 Are there other ways to streamline the processes to reduce the cost to universities of participating in ERA and EI? Yes/No. Please explain your answer.

Yes: We argue in particular of the creation of a national data asset that provides a framework for continuous and automated collection of relevant data from HEP and third-party systems. This reduces stop-start workload issues, particularly when the evaluation schedule is changed, and also allows for real time monitoring (we note that an effort with some aspects of this is being developed in the “real time REF” project in the UK).

Q5.7 In your view, what data sources could ERA utilise? Please explain your answer.

ERA should be based on transparent and open data where feasible, that is available across the sector and ideally in real-time. It should be able to be corrected and critiqued by Australian HEPs and to guide strategic decision making. We argue for the creation of a national data asset on research performance to support this.

Q5.8 In your view, what are the most time-consuming elements of an ERA submission? Please describe.

No direct experience. We believe that output collection, validation and submission are amongst the most burdensome of the current processes.

a. Are there efficiencies that could be introduced? Yes/No. Please describe.

A shared national data asset building on a national capacity for collecting relevant outputs could substantially reduce reporting burden.

Q5.9 In your view what are the most time-consuming elements of an EI submission? Please describe.

No direct experience

a. Are there efficiencies that could be introduced? Yes/No. Please describe.

A shared national data asset building on a national capacity for collecting relevant outputs could substantially reduce reporting burden.

Utilising technological advances and pre-existing data sources

Q5.10 ORCID iDs should be mandatory for ERA. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly agree.

a. What are the advantages and/or disadvantages? Please explain your answer.

ORCID is a highly respected community initiative and provides a powerful mechanism for linking research outputs to individual researchers. By making ORCID IDs a mandatory requirement for ERA the ARC will create an incentive for ORCID uptake that will create benefits for research communities; as for research institutions and policy making communities that need access to high quality data relating to research activities. High levels of ORCID uptake across Australia's research community will transform possibilities for linking publication related data sets and capturing researcher activity efficiently and at scale. This will reduce reporting burden and substantially ease the creation of a national data asset.

Q5.11 The automatic harvesting of output data using ORCID iDs would streamline a university's submission process. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Strongly agree.

a. What are the advantages and/or disadvantages? Please explain your answer

Automatically harvesting output data using ORCID iDs would drastically improve the flexibility, as well as the efficiency, of ERA data collection. It would also allow the work carried out by individual institutions in identifying and verifying researcher outputs to contribute to the improvement of a publicly available data set (ORCID records), with long-term benefits for individual researchers, as well as institutions.

Q5.12 DOIs should be mandatory for ERA. Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

Disagree.

a. What are the advantages or disadvantages? Please explain your answer.

While there should be a strong push towards encouraging publishing platforms and providers to adopt DOIs to identify scholarly works and while we would argue that institutions should be required to submit DOIs where they are available, making them a requirement for submission will render key outputs in the Humanities, in particular, invisible and further narrow the definition of "research" further than is necessary or desirable. Although monograph publishers are increasingly assigning DOIs to books, ISBNs remain the key identifier for monograph publishers and DOIs are not yet always available. A DOI requirement also risks masking the significance of creative works, performances, exhibitions and other types of output, to the detriment of the dynamism and diversity of Australia's higher education and research sector. Creating a parallel pathway for DOI assignment to works that are otherwise without them is also not appropriate.

Any national data collection system should collect (and validate) DOIs where they are available but not require them. A flexible approach to persistent identifiers can be adopted which is inclusive of a wider range of work.

Q5.13 Are there new ways to collect data to reduce the cost and burden to universities of participating in ERA and EI whilst maintaining the robustness of the ERA and EI process? Yes/No. Please explain your answer.

Yes. Rather than asking universities to capture, verify and submit data on research outputs every 3 years, universities could be invited to contribute to the development and continuous improvement of a National Data Asset focussed on data relating to Australian research performance and outputs. Such an asset could build on existing open data sources (Crossref, Microsoft Academic), publicly available data and systems (eg Lens) and universities' own records to create a comprehensive set of outputs.

While further curation would be necessary to disambiguate and qualify metadata this could be handled through a shared infrastructure and system that would substantially reduce burden. This could further be

expanded to gather other relevant data that could support evaluation of engagement and impact as well as aiding institutions to identify potential E&I narratives.

In addition, such a system would allow alternative evaluation proxies to be trialled, create greater transparency in evaluation and support more *flexible* and *robust* evaluation exercises in the future.

a. What are the advantages and/or disadvantages? Please explain your answer.

An advantage of this approach is that the effort invested by universities in building, cleaning, verifying and auditing data already being made to support ERA submissions could, instead, be spent contributing to the creation of a high quality national data asset with multiple uses. By working with libraries; building on the work of groups like the Australian Access Federation and the Australian ORCID consortium; the ARC has an opportunity to help support the creation of national infrastructure that will provide genuine, long-term value for Australia's national innovation system.