

# Improving the evidence for research institute assessment

## A report of a discussion forum

Ian Viney, Andrew Knowles, Matthew Thakur, Emily Ryen Gloinson,  
Susan Guthrie

WR-A965-1

February 2020

Prepared for The Medical Research Council, Cancer Research UK, and the Wellcome Trust

RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Europe but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission from the author, provided the source is clearly referred to as a working paper.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.



For more information on this publication, visit [www.rand.org/pubs/working\\_papers/WRA965-1.html](http://www.rand.org/pubs/working_papers/WRA965-1.html)

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2021 RAND Corporation

**RAND**® is a registered trademark

**Limited Print and Electronic Distribution Rights**

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions.html](http://www.rand.org/pubs/permissions.html).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**Support RAND**

Make a tax-deductible charitable contribution at

[www.rand.org/giving/contribute](http://www.rand.org/giving/contribute)

[www.rand.org](http://www.rand.org)

# Preface

---

In 2020, Medical Research Council (MRC), Cancer Research UK (CRUK) and Wellcome, supported by RAND Europe, held some discussions to examine the methods and data used to evidence progress in their research institutes. These discussions began with online meetings with a variety of funders and research institutes and concluded with a discussion forum to which research institute researchers, funding agency representatives, members of assessment panels and expert evaluators were invited to participate.<sup>1</sup> The aim of these discussions was to help improve approaches, primarily used to make the case for the value of investment in these institutes. This working paper compiles the learning and recommendations that emerged from these discussions.

The work was led by an organising committee with members from the Medical Research Council, CRUK, the Wellcome Trust, King's College London, the Francis Crick Institute, and RAND Europe who organised the discussion forum and reviewed drafts of this working paper.<sup>2</sup> The lead authors for the chapters that summarised the content and key messages from each section of the online discussion forum included: CRUK for section one, RAND Europe for section two, the Wellcome Trust for section three and the MRC for sections four and five. Lead author for the summary was MRC.

---

<sup>1</sup> A list of those invited to the forum is at Annex B.

<sup>2</sup> The organising committee members were Ian Viney (MRC, Chair), Andrew Knowles (CRUK), Matt Thakur (Wellcome), Susan Guthrie (RAND Europe), Cat Mora (King's College London), Rachael Stewart (Francis Crick Institute), Emily Ryen Gloinson (RAND Europe), Mark Pitman (MRC), Emily Gale (MRC), Kevin Dolby (MRC) and James Carter (MRC) provided additional moderator support for the discussion forum.

## Summary

---

Research funders want nuanced, holistic and high-quality expert judgement of progress for research institutes, and this expert judgement should be supported by evidence. In the discussion forum, we probed how we might contract with institutes to set goals, and then organise, compile and assess evidence of progress. This is particularly important as capturing robust evidence of progress can be burdensome, and there is usually a wide choice of evidence that could be collected. The discussion forum was introduced with a statement from Professor Ottoline Leyser, CEO of UK Research and Innovation (UKRI) on the topic.<sup>3</sup>

We found that institutes would benefit from a bespoke approach that acknowledges their unique missions and goals. These goals were likely to have been set by funders as part of the rationale for establishing an institute. Contributors to our discussions highlighted that comparisons between institutes, given the variety of missions, structures and funding models, needed to be made with care. Examples were identified of institute evaluations utilising key performance indicators to drive strategic change, and of more nuanced systems of rubrics for evaluation. Usually progress was independently assessed, but at least one example of a system of self-assessment was identified. Most forum participants suggested that it was good practice for institute and funder to work together to develop a theory of change or simple logic model that could help guide the evaluation. Examples of very complex frameworks were identified, and it was agreed that, while these can be useful in some contexts, it was preferable to develop a framework that was understood by the relevant parties, regularly tested and adapted to guide evaluation.

The merits of carefully selected quantitative evidence to support qualitative statements was discussed in the discussion forum. It was suggested by participants that assessments of progress should ‘demand data’, with an agreed evaluation framework providing clear guidance on the indicators that would be most useful. In contrast to reviewers being presented with everything that can easily be measured, without evaluative statements that set this data in context. Publications are clearly a major output for discovery science institutes and analysis of these can be useful to evidence collaboration, identify emerging/growing areas of work, the contribution of trainees and the impact of external funding. However, the analysis of citations to papers is often over-emphasised, and poorly explained as part of evaluations. Contributors also highlighted the importance of using indicators in evaluation responsibly, to avoid introducing incentives to game the system or engage in nugatory work. Some important aspects of research institute activity are generally under-emphasised in institute evaluations. This work highlighted the crucial importance of securing good evidence of a positive and diverse research culture, and of recognising activities outside of the production of journal

---

<sup>3</sup> The statement from Professor Leyser can be read in full in Annex A.

articles (such as peer reviewing, mentoring trainees, sharing datasets, etc.). Institutes and funders are less used to capturing evidence in these areas.

We found that some institutes had commissioned their own impact studies to look in more detail at their pathways to impact, and to secure independent evidence of the wider societal and economic benefits of their work. These studies were helpful to both funders and institutes in explaining the benefits of research, and there were opportunities for learning from institutes in other disciplines and nations in this area. We found that some institutes participated in networks set up with the aim of sharing good practice, benchmarking, and even independently reviewing their work (such as the public engagement/communication assessment run by the InterAction collaboration). Although many institutes are connected via scientific advisory groups, these networks of good practice seemed a helpful development.

The material identified through these discussions should be helpful to continuous improvement of the way that institutes capture and present evidence of their progress and the approaches that funders use to assess this progress.



# Table of contents

---

Preface .....	iii
Summary .....	iv
Figures .....	viii
Tables .....	ix
Abbreviations .....	x
Acknowledgements.....	xi
<b>1. Introduction.....</b>	<b>1</b>
<b>2. Planning and scoping institute evaluations.....</b>	<b>5</b>
<b>3. Specific components of institute activities.....</b>	<b>9</b>
<b>4. Stakeholders.....</b>	<b>12</b>
<b>5. Evaluation methods.....</b>	<b>14</b>
References .....	17
<b>Annex A. Summary of the discussion forum .....</b>	<b>23</b>
<b>Annex B. List of invitees to the discussion forum.....</b>	<b>54</b>

# Figures

---

Figure 1. A Research Institute on a page.....	32
Figure 2. Comparison of expenditure and Relative Citation Ratio of publication output for comparator institutes (2012-2016).....	40
Figure 3. Radar Plot of Activity Index .....	50
Figure 4. Econometric findings from EBI's economic impact study [52] .....	52

# Tables

---

Table 1. The 6 As of evaluation..... 25

Table 2. Experimental career stage/track framework for recording staffing data..... 38

## Abbreviations

---

CDW	Crick Data World
CERN	Conseil européen pour la recherche nucléaire/European Organization for Nuclear Research
CRUK	Cancer Research UK
DESY	Deutsches Elektronen-Synchrotron/German Research Centre for High Energy Physics (DESY)
ECRs	Early career researchers
ERIC	The European Research Infrastructure Consortium
E-RIHS	European Research Infrastructure for Heritage Science
EU	European Union
IDRC	International Development Research Centre
KNAW	Koninklijke Nederlandse Akademie van Wetenschappen/Royal Netherlands Academy of Arts and Sciences
MRC	Medical Research Council (part of UKRI)
NWO	Netherlands Organisation for Scientific Research
OECD	Organisation for Economic Co-operation and Development
REF	Research Excellence Framework
RI-PATHS	EU Research Infrastructure imPact Assessment paTHwayS
SEP	Strategy Evaluation Protocol
SNSF	Swiss National Science Foundation
UK PRP	UK Prevention Research Partnership
UKRI	UK Research and Innovation
VSNU	Vereniging van Universiteiten

## Acknowledgements

---

The organising committee wish to acknowledge the time taken by those contacted in advance of the discussion forum, and the participants in the discussion forum, who provided such helpful input.



# 1. Introduction

---

Medical Research Council (MRC),<sup>4</sup> Wellcome<sup>5</sup> and Cancer Research UK (CRUK)<sup>6</sup> support several of UK's leading biomedical research institutes. In 2018, these three funders provided core and/or external funding to at least 17 UK institutes, and this support totalled £260 million. This represents 15 per cent of the total investment in biomedical research from these three funders.<sup>7</sup> RAND Europe is a not-for-profit organisation that helps to improve policy and decision making through research and analysis.<sup>8</sup> RAND Europe has regularly worked with each of the funders on various aspects of research evaluation. Research institutes have complex and varied missions and tracking progress against sponsor and institute strategies, evidencing successes and capturing wider economic and societal impact in a meaningful way can be challenging.

Each funder has developed an assessment process for its institutes, and joint approaches are agreed for institutes with core support from more than one funder.<sup>9</sup> In common with most funding agencies worldwide [54], informed expert opinion is central to the review processes, usually involving a visit to the institute by a specially convened and independent subcommittee. This subcommittee then passes its recommendations to the funding agencies to make an allocation of resources. In acknowledgement that the data and methods available to evaluators are rapidly developing,<sup>10</sup> the funders recognised that there was merit in reviewing the state-of-the-art in the field. The three funders therefore agreed to conduct some work to generate ideas on improving the capture, presentation and assessment of evidence to support the review of institutes. The funders were interested in identifying any improvements to approaches used to evidence progress that might help to make the case for the value of investment in research institutes. An objective was to involve research institutes directly in discussion of the supporting methods and evidence used.

---

<sup>4</sup> <https://mrc.ukri.org/>

<sup>5</sup> <https://wellcome.ac.uk/>

<sup>6</sup> <https://www.cancerresearchuk.org/>

<sup>7</sup> Based on the analysis of UK health research expenditure in 2018 published at [www.hrcsonline.net](http://www.hrcsonline.net)

<sup>8</sup> <https://www.rand.org/randeurope.html>

<sup>9</sup> For example, the [Francis Crick Institute](https://www.frcr.ac.uk/) is core funded by MRC, CRUK and Wellcome.

<sup>10</sup> Examples of developments that make information about research more accessible and advance the methods available to evaluators include: i) the move to open access (e.g. [UKRI](https://www.ukri.org/), [CRUK](https://www.cancerresearchuk.org/), [Wellcome](https://wellcome.ac.uk/) policies), ii) the introduction of Researchfish<sup>®</sup>[44] (e.g. by [UKRI](https://www.ukri.org/), [CRUK](https://www.cancerresearchuk.org/), [Wellcome](https://wellcome.ac.uk/)), iii) new commercial entrants to the market such as [Digital Science](https://www.digital-science.com/) and iv) an increased interest in 'research on research' (e.g. Wellcome established a new [Research on Research Institute](https://www.researchonresearch.org/) in 2019).

Originally, the work was designed to include a workshop meeting to bring together funder and institute staff, research assessors and experts in evaluation – to involve as many different perspectives on the review process as possible. The funders wanted to start a conversation about what aspects of an institute’s work are important to capture, and how the information already collected can offer insight into this. We aimed to identify what works well (and not so well), and to showcase examples from varied organisations worldwide. Early in 2020 we contacted the institutes core funded by the three funders, colleagues in the other UKRI councils and contacts identified in a diverse set of international research organisations.

The COVID-19 pandemic meant that the plans for a traditional workshop were modified and we instead designed an online discussion forum.<sup>11</sup> The forum was a simple bulletin board divided into four sections, pre-populated with a few example questions to stimulate comment. The four sections were:

- **Scope and purpose of evaluation:** the reasons for evaluating research institutes, considerations for how to effectively and responsibly carry out an evaluation, and the intended audiences for the results.
- **Specific components of institute activities:** how the evidence for evaluations should be prioritised, gathered and organised.
- **Stakeholders:** involvement of different stakeholders in the evaluation and how they can help evidence outcomes and impact, and ensure credibility and rigour in evaluation.
- **Evaluation methods:** details of the latest and most robust approaches to evidence progress.

The material for the forum was compiled by an organising committee.<sup>12</sup> The forum was open for participants to view and, if they wanted, to add comments in response to the material for a week at the end of June 2020. The discussion was moderated by a team that consisted of members of the organising committee plus additional colleagues from MRC. Moderators read the responses, added some replies and further questions to stimulate the discussion, and registered new participants who wanted to join the forum. We invited 150 colleagues to participate in the forum, representing 20 research institutes and experts in the UK, USA, Canada, Netherlands, Germany, Denmark and Austria.<sup>13</sup> Exactly half of the invitees logged into the forum during the week (75) and a quarter (38) left comments. In total, 300 comments were added by invitees and moderators, providing a wide collection of ideas, links to further examples and research, as well as recommendations and suggestions for improvement. The organising committee agreed that this had been an effective way to compile current thinking and practice in the specific area of research institute evaluation. Material from the workshop discussions was combined with that of interviews with institute staff, evaluators and funders, plus contributions from the organising committee to produce the following report. The report sets out the findings of the discussion forum covering each of the four sections of the forum listed above in

---

<sup>11</sup> The [VisionsLive](#) platform was used to create the discussion forum.

<sup>12</sup> The organising committee members were Ian Viney (MRC, Chair), Andrew Knowles (CRUK), Matt Thakur (Wellcome), Susan Guthrie (RAND Europe), Cat Mora (Kings College London), Rachael Stewart (Francis Crick Institute), Emily Ryen Gloinson (RAND Europe), Mark Pitman (MRC), Emily Gale (MRC). Kevin Dolby (MRC) and James Carter (MRC) provided additional moderator support for the discussion forum.

<sup>13</sup> A list of invitees is at Annex B.

turn. More details on some of the materials and content provided in the discussion forum are supplied in Annex A.



## 2. Planning and scoping institute evaluations

---

The organising committee asked participants for their views on preparing to review research institutes.<sup>14</sup> A foundation for evaluation in the public sector in the UK is the HM Treasury ‘Green Book’ [2], the UK central government guidance on appraisal and evaluation. The Green Book states that appraisal and evaluation are essential activities to support evidence-based decision making at all stages of the policy development cycle. In line with this guidance, respondents in the discussion forum emphasised that careful planning for evaluation is essential, and that this should include an explicit plan to evaluate responsibly and be clear over the intended audiences for the results. The ‘6As’<sup>15</sup> is a list of potential reasons for evaluating, and respondents agreed they were a useful starting point. Evaluations may have multiple purposes (e.g. Advocacy and Allocation), but clarity over the intended reasons for evaluation will help select the most appropriate approaches to use and the best channels to reach the intended audiences. The forum discussed the recent work by the International Network of Research Management Societies (INORMS) [4] to develop a five-step process (‘SCOPE’) for evaluating responsibly, which emphasised starting with what you value and considering carefully the context, with some respondents noting that this could be a useful framework.

**Conducting an evaluation has costs as well as benefits, and care is needed to reduce the burden and maximise the return.** Participants in the discussion forum said that evaluations are resource intensive for evaluators and the evaluated, so the benefits need to be balanced against the effort required, and the burden on researchers reduced where possible. It is important to aim for high-quality evaluation, choosing reviewers and other parties with care to ensure they have the right level of experience and expertise. Participants argued that existing scientific advisory boards can provide a useful input into evaluations as long as conflicts of interest are managed and sufficient independence achieved (advisory board members may be invested in the setting up or significant transformation of an institute). The costs of evaluation are often overlooked when budgeting for the running and operation of a research institute. It might be useful for institutes to have a specific facility/role(s) in the institute responsible for compiling data and providing a point of contact for the evaluation. This will require additional investment but can free researchers from less effective approaches to capturing and organising information about progress. Ensuring that information gathered for internal management purposes is fit for external evaluation and can be reused in other reporting and communication activities clearly helps to maximise the benefits.

---

<sup>14</sup> See Annex, Section 2 for more information on the specific topics covered and responses received on this section of the discussion forum.

<sup>15</sup> The ‘6As of evaluation’ (Analysis, Advocacy, Allocation, Accountability, Acclaim and Adaptation) were adapted from [3].

**The purpose of an evaluation should be agreed at the outset.** For a successful evaluation, the purpose needs to be made explicit and agreed by all parties – institute, funder, evaluation professionals and any other stakeholders. Institutes have distinct goals<sup>16</sup>; equally different funders have different requirements and the results of a similar evaluation may be used quite differently by a government research council and a charity. There is no one-size-fits-all model, and an open and constructive dialogue is necessary to clarifying the goals of the evaluation and how the results are to be used.

**Sponsors may be interested in the performance of similar institutes in specific dimensions, but participants cautioned that broad comparisons could be misleading.** Institutes suggested that they should be evaluated against their own specific vision, mission or goals. Research institutes are unique environments, usually established by funders with specific strategic expectations, and evaluations should capture the added value of the institute model. This might include, for example, work to enhance connections between different sectors of society/industry, or research disciplines, training the next generation of researchers and accelerating translation towards patients and the public. Given that research institutes cover a broad range of organisations with different goals, funding models and scale, participants advised that broad comparisons between institutes risked potentially misleading results. However, it is important to understand ways to better deliver research outcomes and impacts by studying evaluation in different contexts. Participants in the forum highlighted that institutes are typically long-term investments that should (and do) evolve over time in terms of their work and objectives. Evaluations need to be flexible enough to recognise the experimental nature of science and to adapt to changes to research priorities and the system investigating them. Goals of the institutes should be re-examined at the start of each evaluation process.

**Evaluate the institute overall, as well as individual research groups.** Institutes are expected to deliver outputs beyond those that could be achieved from a group of programme grants, in part because of the opportunity they provide for intra- and extra-mural collaboration. Evaluations built up solely from the assessment of individual groups can miss this collaborative dynamic, underplay the value an institute adds and place too much emphasis on individual performance at the expense of teamwork. Evaluations need to consider the interactions between groups, and the overall benefit of the institute – the ‘added value’ of supporting an institute.

**Institute self-evaluation can complement and support external evaluation by funders or others.** Regular evaluation by institute management is an important governance function, allowing institutes to assess progress against their objectives and to assess – and where necessary, adapt – culture, processes and behaviour. To avoid unnecessary work, self-evaluation should be consistent/aligned with the approach used by funders.

**Funders should carefully consider the audiences for the evaluation.** Funders may evaluate an institute to ensure that it is delivering good quality science and training, being appropriately run, spending money effectively and efficiently and that it is contributing to the funder’s wider strategic goals. Evaluations can provide institute senior management with important learning about progress against strategic goals. Communicating success back to institute staff can boost morale and contribute to transparency – or provide

---

<sup>16</sup> Institutes may include elements of pure research, innovation activities, providing facilities in support of research and innovation, and activities designed to relate to wider society (e.g. public engagement).

valuable feedback where there are areas for improvement. Other important audiences are policymakers, and (where relevant) patients and the public. Where possible, evaluations should be designed such that the results can be used to address multiple stakeholders, using a framework that allows the format and language of results to be adapted to suit different audiences. For example, the outputs of an expert review can be used to develop case studies that can be used by policymakers. The evaluation community is also an audience for evaluations and would benefit from the method and results of evaluations being published. Some research institute evaluations are not published in full and details of the underlying data used to produce statistics for institute annual reports are often not transparent. Evaluators may be interested in how evaluations draw on best practice, elements of innovation and adaptation of approaches, as well as the long-term impact of recommendations made.

**Evaluation of research institutes should be used as a tool to support learning and improvement.** Evaluations work best when there is a positive relationship and high levels of trust between stakeholders. Ensuring evaluations highlight positive aspects, as well as any areas to improve, builds a positive feedback loop and supports a culture of enquiry and learning, and ultimately better research outcomes.

**Independent expert advice may benefit the design and delivery of evaluations.** Funders can help by providing advice and clarity to institutes over process and objectives for evaluations. All evaluations benefit from independent expert advice (beyond the expertise used to assess the quality of research), and while funders can be a critical friend to institute evaluation, both stakeholders can be assisted by occasionally seeking views from a third-party consultant or external academic teams to help design evaluations. Those carrying out an evaluation should consider an assessment of the impact of that evaluation at the outset, aiming both to remove any discrimination and to understand (and therefore minimise) the degree of disruption to both the institute and the review party. The independence of evaluations is an active area of discussion,<sup>17</sup> and funders are increasingly ensuring that they bring outside views to the design of their evaluations, to avoid an accusation that they are ‘marking their own homework’.

---

<sup>17</sup> For example, the UN Development Programme has provided a report on evaluation and independence [5].



### 3. Specific components of institute activities

---

We asked participants for their views on gathering evidence about various aspects of research institute work and how this can contribute to assessment. We started with questions about frameworks (such as logic models) that are useful to organise this work.

**Clear logic models or strategic narratives can be useful to support thinking and articulate a shared vision.** Frameworks such as logic models are tools to structure thinking, articulate strategy and to communicate the complex ways in which impact is generated, potentially to diverse audiences. They can provide guidance, clarify underlying assumptions, and support stakeholder buy-in to the process of evaluation. They can help internal communication, promoting shared values and aims. Although these are mostly diagrammatic, a clear strategy narrative can serve the same function.

**To be effective, logic models should be developed collaboratively at the outset and reviewed regularly.** Although a logic model can be developed at any time, this work is best tackled early in the process of setting up a programme to help inform the prospective collection of evidence about progress. They can include a baseline to remind stakeholders of the starting position and document the longer-term focus of the research institute. Logic models should be adapted to specific needs and contexts, and should be able to accommodate change. Participants in the forum agreed that a commitment to review and change plans to account for new evidence and a changing external context was needed. Views amongst respondents regarding the utility of logic models were mixed. Some felt logic models might not be applicable to the evaluation of a whole research institute, and that pathways to impact in a research institute are complex and difficult to fit into one framework. It might require a lot of resources to put a helpful logic model together, and simplified models focusing on just a few elements may not be robust or fit for purpose. An inclusive design process and consultation with experts familiar with logic models may mitigate against this.

The impact framework developed for the UK Prevention Research Partnership (UK PRP) [6] was highlighted as a helpful model developed at the outset of the programme, used to shape the monitoring and evaluation approach, and with a commitment to continue to develop as new evidence is captured. Although the UKPRP is a strategic funding partnership, not an institute, the way that this framework has helped guide evaluation of the partnership was thought to be a good example.

**Is research impact (or societal benefit) an appropriate lens for institute evaluation?** Funders often establish institutes to address longer-term strategic challenges, which may take more than a single (typically five-year) review period to progress. However, the discussion noted that in practice, most evaluations (whether for accountability and allocation, or for advocacy and acclaim purposes) focus on changes that occur in a shorter time-frame. Evaluations can be framed around understanding the shorter-term goals of the institute that will contribute to long-term impact, and the systems and processes in place to support the delivery of those.

**Governance structures play an important role in culture and operation of institutes.** The importance of governance structures was recognised in the discussions. These structures include hierarchies, transparent and effective procedures, functional supporting infrastructures and standards for data or protocol sharing. Governance structures can be complex, and in cases where organisations merge or implement transformation programmes it takes time for arrangements to change. However, because it is difficult to measure governance objectively, rubrics and audits are often used to track progress. Accreditation systems exist that can support this type of assessment. However, Professor Leyser had in her introductory remarks stated:

There are strong parallels to wider questions of accountability and trust, as eloquently set out by Onora O’Neil in her Reith Lectures. To quote her: ‘Intelligent accountability, I suspect, requires more attention to good governance and fewer fantasies about total control. Good governance is possible only if institutions are allowed some margin for self-governance of a form appropriate to their particular tasks, within a framework of financial and other reporting’.

**Taking in a broad range of views is needed to understand research culture.** Funders are increasingly realising the importance of a creative, supportive and inclusive research culture [7]. To measure positive research culture for an inclusive research economy, there should be a recognition of collaborations and ‘voices on the ground’. Evaluations should seek to capture the perspectives of PhD students to professors, technical and professional staff, support personnel, alumni, and internal and external consumers. The perspectives of these stakeholders can be included through surveys asking about research culture and their perceptions of management, having informal meetings, interviews, peer review and feedback, audits, exit interviews, and carrying out ethnographies that combine interviews and participation in daily practice. Productive collaboration should also be measured using a combination of qualitative and quantitative approaches, including through annual internal reports from group leaders, annual visits to the headquarters team, Researchfish® [44] reports, and using collaborative network maps.

Participants in the forum also suggested that some of the more difficult to measure aspects of progress could be captured via funder involvement in institute advisory and governance boards (so that discussion of progress could be independently recorded and reused), and institute staff surveys (to act as a barometer of research culture). Respondents also highlighted the crucial importance of the visiting expert subcommittee itself meeting, and the direct interaction this provided between institute staff and assessors. The Max Planck Society emphasised that it had learnt the importance of a good channel of communication from all levels in its institutes to the Society’s executive<sup>18</sup> and has set specific time aside in all its evaluations to gather views from institute staff. The Max Planck Society has active networks of students, early career researchers and group leaders across all its institutes that regularly feed views into the society’s executive.

**Longitudinal changes are important.** Measuring changes over time was thought to be important, and regularly overlooked, as ‘snapshot’ or annual output information is often used without reference to a previous baseline. If a reliable time series can be established, then the direction of travel will clearly be of interest to reviewers and may help support the argument that specific actions taken by the institute have had a positive impact.

---

<sup>18</sup> The Max Planck Society supports an intramural PhD student network across all its 84 institutes which provides an independent voice for 5000 students [8]

**Staff development is an important aspect of an institute's output.** The resources dedicated to employing and developing institute researchers are usually the major component of an institute budget. However, approaches to track the careers of staff while working at the institute, and over the long term as institute alumni, have not featured strongly in evaluations. Evidence of the contribution that institutes make to training, developing and exporting skilled researchers around the economy, was agreed by respondents to the forum to be an area for significant improvement. Evidence should cover both academic and non-academic staff, and career pathways within and outside of academic research. It could also include the impact that research institutes have on external researchers through training and by providing expertise or information through collaboration. A simple framework for capturing staff numbers against scientific track, or scientific support track, career pathways was proposed to help standardise statistics over time and across institute staff populations (this draft framework is included in Annex A).

**Different funding models should be considered in the evaluation approach.** Evaluation measures should capture how core and uplift/external funding to institutes support the institute's strategy and contribute to sustainability.

## 4. Stakeholders

---

We asked participants for their views on how to involve different stakeholders in institute evaluation.

**Involve diverse stakeholders at all stages of the evaluation.** Respondents emphasised that, in addition to providing evidence for the impact of an institute on different sectors and providing credibility and independent valuation of an evaluation, stakeholders can also contribute to developing the evaluation approach. For example, through consultation they can add valuable external perspectives to evaluation design, such as helping evaluators understand what are considered the most credible evidence of outcomes and impacts. This consultation could be gathered via an advisory board, or other novel routes such as a conference or online discussion forum. During evaluation, external stakeholders can inform the evaluator's interpretation of evidence. In the context of research institutes, this could take the form of including a diverse range of peer reviewers in the evaluation process (of particular value for institute research that is closer to being applied [9]). Once evaluation is complete, external stakeholders can inform communication strategies and support putting evaluation findings into practice [10].

**Evaluation should be a collaboration between funders and other stakeholders.** As the stakeholders most likely to perform evaluation of institutes, respondents thought it was important that funders proactively reflect on, share and refine their evaluation methods and approaches with other stakeholders, particularly those being evaluated, but also partners such as host institutions, key collaborators, etc. Respondents felt that co-producing the most appropriate and proportionate approaches for evaluating institutes would promote engagement, best practice and progressive thinking in this area.

Participative methods such as Evaluative Inquiry, developed by CWTS Leiden [11], can be used by funders to develop a shared understanding between the institute and sponsor of what an institute is aiming to achieve, and when. It was suggested that this may lead to better quality evaluation, appropriate benchmarking, and more engagement by institutes in monitoring performance against their strategies. Ways to increase the involvement of research institutes in their own evaluation were discussed at several points in the forum. Peer networks of institutes (such as EU-LIFE<sup>19</sup> for European life science institutes, or InterAction a communications resource for particle physics institutes<sup>20</sup>) that help share good practice were highlighted. Although few of these formal networks exist, it was noted that many institutes had strong informal connections via shared membership on their scientific advisory boards, governing

---

<sup>19</sup> EU-LIFE is an alliance of independent European research institutes in life sciences [12].

<sup>20</sup> InterAction has an initiative to conduct peer review of communication at laboratories that are members and other scientific institutions worldwide [13].

boards and via collaborative projects. In addition, MRC/CRUK/Wellcome already build in consultation and agreement over terms of reference and the arrangements for their scientific reviews.

**Local stakeholders for evidencing local impact.** Respondents identified an additional external stakeholder group - the local and regional connections of an institute. For many public funders, research funding, including funding for institutes, is part of an industrial strategy aiming to deliver economic growth. For example, in June 2020 the UK government announced a £400m investment in regional R&D projects expected to drive local economic growth as part of the UK R&D Roadmap [14]. As government justification for research funding becomes more explicitly linked to economic growth, evaluation is likely to require more focus on evidencing economic impact, including the impact of institutes on local and regional stakeholders. An institute's own strategy may also reflect its intent to make the best use of local resources, such as translational networks and partners, and evaluation should reflect this. Recent European Union (EU) [15] and Organisation for Economic Co-operation and Development (OECD) [16] projects have developed frameworks for evidencing the impact of research infrastructure on local and regional stakeholders. There are many places where these conceptual models and indicators are likely to be relevant to evaluation of institutes.

**Internal and external academic stakeholders are an important and diverse group.** Respondents cautioned against an over-simplified view of how institutes are expected to impact academic stakeholders, given the diversity of this stakeholder group. Again, evidence of academic impact is often overly focused on publication outputs. Where an institute's strategic objectives include goals such as improvements to research integrity, equity, diversity and inclusion and skills capacity building, respondents emphasised the importance of adequately monitoring impact on its internal and affiliated academic stakeholders, such as current and former staff and students, as well as external stakeholders. Other institutes were identified as another important academic stakeholder group, including similar and contrasting institutes internationally.

## 5. Evaluation methods

---

We asked participants for their advice on quantitative, qualitative and implementation of mixed method approaches for research evaluation. For the public sector in the UK, the HM Treasury ‘Magenta Book’ [17] provides extensive guidance on methods for process, impact and value for money evaluations, and so this section focused on aspects suggested as particularly pertinent to the evaluation of research institutes.

**Quantitative evidence should be used responsibly to support expert judgement.** Professor Leyser’s introductory remarks for the discussion forum highlighted that, while we are all keen to find ways to measure success quantitatively to ensure that we can track progress and assess the impact of strategic decisions, we should be careful in the use of proxy measures. The usefulness of metrics has been highlighted in the context of evidencing narrative statements [18], [19], and as facilitators to further enquiry [20], but they cannot reliably be used as stand-alone performance measures [21]. Many papers have been written on potential measures that could be chosen to provide insight into specific activities,<sup>21</sup> and some approaches to measuring and benchmarking progress quantitatively have gained traction.<sup>22,23,24</sup> These approaches emphasise that a ‘basket’ of information is better than a single indicator, helping to use independent facts to triangulate expert judgement [24].

Where performance measures/indicators/metrics are used to support, not supplant, expert peer review, evaluators should ask whether these indicators correlate sufficiently well with the desired performance criteria (e.g. quality). Many easy to collect metrics measure quantity not quality, and it is not clear what the best quantity is for each measure. It is also important to consider the likely consequences on researcher behaviours of using proposed metrics. While some metrics used at institutional level are reasonable, it is difficult to separate institutional from individual level incentives. This is at least as important in considering

---

<sup>21</sup> For example RAND published in 2016 ‘[100 Metrics to Assess and Communicate the Value of Biomedical Research](#)’, a collaboration between research-intensive UK universities working with Elsevier who published the first [Snowball Metrics Recipe Book](#) in 2012, and has followed up with further reports. The [LSE impact blog](#) is also a source of opinion articles on the merits of different metrics.

<sup>22</sup> In 2019 the European Strategy Forum on Research Infrastructures (ESFRI) concluded a project to recommend a common approach across its infrastructures to monitor their performance based on KPIs [22]. The proposed KPIs ranged from input to outcome indicators and were tested against the RACER criteria (whether they were ‘relevant’, ‘accepted’, ‘credible (for non-experts)’, ‘easy to monitor’, and ‘robust’) [23].

<sup>23</sup> The [Snowball Metrics Recipe Book](#) provides a set of metric definitions agreed between research intensive universities for benchmarking.

<sup>24</sup> The [Researchfish® system](#) common question set developed between 2009 and the present day, and used by UKRI, CRUK and Wellcome to collect annual feedback from all their investments, represents a consensus definition of a wide range of research outputs.

the things not measured as it is in considering the measured list. For example, if the papers that an institution publishes in specific journals are counted, this might devalue time invested in making negative results and datasets openly accessible. Although funders insist that they value these other aspects, if they are excluded from the assessment, the perception will be that they are not important. The same applies to important activities such as participating in the peer review process [25] and other services to the research community. Respondents highlighted that they wanted their ‘evaluations to demand data’. Too often, assessments began by cataloguing all the data available, rather than thinking about what the critical information to support the review might be.

Respondents emphasised that a rigorous approach can still be taken with qualitative research. While the core principles of rigour in qualitative research are debated, investigators aim for reliable and valid results by reducing sources of bias [26], increasing the diversity of views sought [27] and triangulating the results using multiple/independent sources of evidence.

**Bibliometrics can be a useful way of evidencing collaboration and the diffusion of knowledge.**

Publication data is the most widely analysed research output, and the analysis of citations is the most widely used quantitative approach. For example, the citations of publications submitted to the Research Excellence Framework 2021 (REF2021) exercise will be analysed and considered alongside other evidence by 11 of the 34 expert subpanels [28]. The principles of citation analysis and the responsible use of this data have been extensively reviewed elsewhere [29].

Analysis of the bibliographic information contained in journal articles allows the connection of information about funding, people, locations and subject matter. This may be used to study collaboration, changes in effort devoted to different research fields or the involvement of different funding sectors or research organisations (private, charity, hospital, etc.). The analysis of the citations these papers receive allows investigation of the attention that they secure in other papers, patents or policy documents. This may be used to track the diffusion of knowledge or its translation into intellectual property/new practice. However, evaluations often place undue attention on the citation counts or index itself, which is arguably less meaningful than evidencing knowledge transfer.

Respondents again emphasised the triangulating power of independent sources of information suggesting that bibliometric analysis can help by providing a starting point for review panels to understand priorities, research strategies and goals, and support directors to reflect on whether they are moving in the right direction, or to spot new opportunities.

**Self-evaluation.** The Strategy Evaluation Protocol (SEP – a joint effort of Royal Netherlands Academy of Arts and Sciences (KNAW), Vereniging van Universiteiten (VSNU) and Netherlands Organisation for Scientific Research (NOW) in the Netherlands) [30] stimulated some discussion on the forum. The SEP is used to assess groups against their own strategic goals, not against a fixed set of external criteria. It may usefully diminish the role of quantitative metrics and discourage comparisons between institutes. It does still rely on an independent expert assessment committee to consider progress and future proposals, but the guidance on the process has received praise for its simplicity [31]. Respondents questioned whether self-evaluation could provide enough accountability or credible assessment of value for money. In SEP, the role of providing challenge to the process is with the board of the university. For institutes, it would

depend on the units evaluated, the governance structure of the institute and the part that the sponsors took in agreeing the criteria for the review.

**The use of rubrics.** Rubrics usually consist of agreed criteria and standards, which allow ratings to be assigned to complex qualitative activities. An example of a rigorously evaluated framework that used rubrics was the 'RQ+ Assessment Framework'[32, 33], launched by the International Development Research Centre (IDRC) in Canada. This example was raised in discussion with stakeholders ahead of the forum. The RQ+ Framework has three main elements: i) key influences – contextual factors that may impact on research quality, which may include research capacity, political environment, etc., ii) research quality – dimensions include research design (integrity), importance and positioning for use, iii) evaluative rubrics – a set of criteria for each dimension in the first two elements is set out, so that evaluation subjects can be rated. Once ratings have been produced, these are aggregated in a way appropriate to the specific evaluation. The National Research Council in Canada reported that it had piloted the use of rubrics to an evaluation of its research centres [34] and found the development of rubrics more difficult than expected.

**Researcher CVs and biosketches.** Respondents highlighted the usefulness of CVs to list researcher achievements, capturing key input/output/outcome/impact information and explain their track record. CVs would benefit from i) emphasising a greater diversity of research activities (such as peer review and supervisory work/mentoring), ii) following a more standardised and structured format that could be integrated with authoritative data sources, updated easily and reused for institute reviews, promotion panels, grant applications, etc. There are several initiatives trying to address exactly this issue that respondents highlighted that we could learn from, such as the NIH 'Biosketch' [46, 47], SciCV [45], the Researchfish® CV builder<sup>25</sup> and the Royal Society 'resumé for researchers' [48].

**Trends over time.** Respondents highlighted the importance of changes over time. Do the data shed light on the delivery of the strategy, have things got better or worse over time, and are there changes that can be shown to be subsequent to actions taken by the Institute?

**Impact pathways – the link from logic models to estimating societal and economic impact.** Several examples were identified of institutes and/or funders commissioning studies of economic and wider societal impacts of institutes. The results of these studies were agreed by respondents to the forum to be useful for accountability, but also served as an excellent communication tool, and for some facilities had increased researcher engagement in impact assessment. A good example was the European Bioinformatics Institute (EBI) economic impact study [52], which illustrated several interesting methodological aspects. Firstly, the increasing role of surveys and detailed interviews to capture new quantitative and qualitative information from stakeholders. Secondly, the importance of econometric models/cost benefit analysis to estimate the return on investment from research. Recent work by OECD [15] and the EU [16] have developed socio-economic impact methodology frameworks suited for application to research infrastructure and institutes.

---

<sup>25</sup> Since its inception, the Researchfish® system has included the functionality to reuse information harvested and input into a researcher profile in a CV-style document [44].

## References

---

- [1] Carter, J. and E. Stevens. 2020. 'UK Health Research Analysis 2018.' UK Clinical Research Collaboration. As of 7 December 2020 [Online]: <http://hrcsonline.net/wp-content/uploads/2020/01/UK-Health-Research-Analysis-2018-for-web-v1-28Jan2020.pdf>.
- [2] HM Treasury. 2018. 'The Green Book: Central Government Guidance on Appraisal and Evaluation.' As of 7 December 2020 [Online]: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/685903/The\\_Green\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf).
- [3] Parks, S., D. Rodriguez-Rincon, S. Parkinson & C. Manville. 2019. 'The Changing Research Landscape and Reflections on National Research Assessment in the Future.' As of 7 December 2020 [Online]: [https://www.rand.org/pubs/research\\_reports/RR3200.html](https://www.rand.org/pubs/research_reports/RR3200.html).
- [4] INORMS: Research Evaluation Working Group. May 2020. 'Better Decision Making through Responsible Research Evaluation.' As of 7 December 2020 [Online]: <https://inorms.net/wp-content/uploads/2020/03/scope.pdf>.
- [5] UNDP: Independent Evaluation Office. 2016. 'Evaluation and Independence: Existing Evaluation Policies and New Approaches.' As of 7 December 2020 [Online]: [http://web.undp.org/evaluation/evaluations/documents/Independence\\_of\\_Evaluation.pdf](http://web.undp.org/evaluation/evaluations/documents/Independence_of_Evaluation.pdf).
- [6] UK Prevention Research Partnership. June 2018. 'UKPRP Impact and Evaluation Framework.' As of 7 December 2020 [Online]: <https://ukprp.org/wp-content/uploads/2019/09/117669-UKPRP-UKPRP-Framework-Monitor-and-Evaluate-Bk-Proof-Sep-2nd.pdf>.
- [7] Wellcome Trust. 2020. 'What Researchers Think About the Culture They Work In|.' As of 7 December 2020 [Online]: <https://wellcome.org/reports/what-researchers-think-about-research-culture>.
- [8] Max Planck PhDnet. n.d. 'Home.' As of 7 December 2020: <https://www.phdnet.mpg.de/home>.
- [9] Guthrie, S., I. Ghiga & S. Wooding. 2020. 'What Do We Know About Grant Peer Review in the Health Sciences?' *F1000 Res.* 6 (1335): 1–23.
- [10] BetterEvaluation. 4 January 2016. 'Understand and Engage Stakeholders.' As of 7 December 2020: [https://www.betterevaluation.org/en/rainbow\\_framework/manage/understand\\_engage\\_stakeholders](https://www.betterevaluation.org/en/rainbow_framework/manage/understand_engage_stakeholders).
- [11] Holtrop, T., L. Hessels & A. Prins. 25 June 2020. 'Evaluative Inquiry I: Academic Value is More Than Performance.' *Leiden Madtrics*. As of 7 December 2020: <https://leidenmadtrics.nl/articles/evaluative-inquiry-i-academic-value-is-more-than-performance>.

- [12] EU-Life. n.d. 'Mission & Values | EU-LIFE.' As of 7 December 2020: <https://eu-life.eu/about/mission-values>.
- [13] InterAction. n.d. 'InterAction Collaboration Peer Reviews.' As of 7 December 2020: <https://www.interactions.org/interaction-collaboration-peer-reviews>.
- [14] HM Government. 2020. 'UK Research and Development Roadmap.' As of 7 December 2020: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/896799/UK\\_Research\\_and\\_Development\\_Roadmap.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896799/UK_Research_and_Development_Roadmap.pdf).
- [15] The EU-funded 'RI-PATHS' Project (Research Infrastructure imPact Assessment paTHwayS). n.d. As of 7 December 2020: <https://ri-paths.eu/>.
- [16] OECD. 2019. 'Reference Framework for Assessing the Scientific and Socio-economic Impact of Research Infrastructures.' As of 7 December 2020. [Online]: [https://www.oecd-ilibrary.org/science-and-technology/reference-framework-for-assessing-the-scientific-and-socio-economic-impact-of-research-infrastructures\\_3ffee43b-en](https://www.oecd-ilibrary.org/science-and-technology/reference-framework-for-assessing-the-scientific-and-socio-economic-impact-of-research-infrastructures_3ffee43b-en).
- [17] HM Treasury. 2011. 'The Magenta Book'. GOV.UK. As of 7 December 2020: <https://www.gov.uk/government/publications/the-magenta-book>.
- [18] Hinrichs, S. A. Kamenetzky, L. Borjes & J. Grant. May 2015. 'The Nature, Scale and Beneficiaries of Research Impact.' King's College London – The Policy Institute. As of 7 December 2020 [Online]: <https://www.kcl.ac.uk/policy-institute/research-analysis/nature-scale-beneficiaries-research-impact>.
- [19] Grant, J., S. Hinrichs & A. Gill. 2014. 'The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of the Research Excellence Framework (REF) 2014 impact case studies.' King's College London – The Policy Institute. As of 7 December 2020 [Online]: <https://www.kcl.ac.uk/policy-institute/assets/ref-impact.pdf>.
- [20] Fochler, M. & S. De Rijcke. 2017. 'Implicated in the Indicator Game? An Experimental Debate.' Engag. Sci. Technol. Soc. 3: 21–40.
- [21] De Rijcke, S., P. Wouters, A. D. Rushforth, T. P. Franssen & B. Hammarfelt. 2015. 'Evaluation Practices and Effects of Indicator Use – a Literature Review. Research Evaluation | Oxford Academic.' Res. Eval. 25 (2): 161–169.
- [22] ESFRI. 2019. 'Working Group Report: Monitoring of Research Infrastructures Performance.' As of 7 December 2020 [Online]: [https://www.esfri.eu/sites/default/files/ESFRI\\_WG\\_Monitoring\\_Report.pdf](https://www.esfri.eu/sites/default/files/ESFRI_WG_Monitoring_Report.pdf).
- [23] Kolar, J., A. Harrison & F. Gliksohn. n.d. Accelerate. 'Key Performance Indicators of Research Infrastructures.' As of 7 December 2020: <http://www.accelerate2020.eu/key-performance-indicators-of-research-infrastructures/>.
- [24] Clarivate. 2019. 'Institute for Scientific Information Launches Global Research Report – Profiles not Metrics.' As of 7 December 2020: <https://clarivate.com/news/institute-for-scientific-information-launches-global-research-report-profiles-not-metrics/>.

- [25] Curry, S. 2017. 'Peer Review is Essential to Good Science – It's Time to Credit Expert Reviewers.' *The Guardian*. As of 7 December 2020: <https://www.theguardian.com/science/2017/jun/01/peer-review-is-essential-to-good-science-its-time-to-credit-expert-reviewers>.
- [26] Understanding Health Research. n.d. 'Understanding Health Research · Common Sources of Bias.' As of 7 December 2020: <https://www.understandinghealthresearch.org/useful-information/common-sources-of-bias-2>.
- [27] Long, T. & M. Johnson. March 2000. 'Rigour, Reliability and Validity in Qualitative Research.' *Clin. Eff. Nurs.* 4 (1): 30–37.
- [28] Higher Funding Council for England. 2018. 'Clarivate Analytics Will Provide Citation Data During REF 2021.' As of 7 December 2020: <https://www.ref.ac.uk/news/clarivate-analytics-will-provide-citation-data-during-ref-2021/>.
- [29] Wilsdon, J. et al. 2015. 'The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. doi: 10.13140/RG.2.1.4929.1363.
- [30] NWO, VSNU, and KNAW. 2020. 'Strategy Evaluation Protocol 2012–2027.' The Hague. As of 7 December 2020: [https://www.vsn.nl/files/documenten/Domeinen/Onderzoek/SEP\\_2021-2027.pdf](https://www.vsn.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf).
- [31] Gadd, E. 'Dear REF, Please May We Have a SEP?' 2 August 2020. *Wonkhe*. As of 7 December 2020: <https://wonkhe.com/blogs/dear-ref-please-may-we-have-a-sep/>.
- [32] Ofir, Z., T. Schwandt, C. Duggan & R. McLean. 2016. 'A Holistic Approach to Evaluating Research.' *Research Quality Plus*. As of 7 December 2020 [Online]: <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/56528/IDL-56528.pdf?sequence=2&isAllowed=y>.
- [33] Research Quality Plus. 14 June 2018. 'Research Can Change the World.' IDRC – International Development Research Centre. As of 7 December 2020: <https://www.idrc.ca/en/research-in-action/research-quality-plus>.
- [34] Tremblay, G., F. Bertrand & M. A. Fraser. 2017. 'Using Rubrics for an Evaluation: A National Research Council Pilot | Semantic Scholar.' *The Canadian Journal of Program Evaluation* 32. As of 7 December 2020 [Online]: <https://www.semanticscholar.org/paper/Using-Rubrics-for-an-Evaluation%3A-A-National-Council-Tremblay-Bertrand/a70209d3af4081a3cade03accedc0d5fd59d728e>.
- [35] Gadd, L. 15 May 2019. 'The blind and the Elephant: Bringing Clarity to Our Conversations About Responsible Metrics.' *The Bibliomagician*. As of 7 December 2020: <https://thebibliomagician.wordpress.com/2019/05/15/the-blind-and-the-elephant-bringing-clarity-to-our-conversations-about-responsible-metrics/>.
- [36] National Research Council Canada. 2 February 2016. 'Evaluation of the National Institute for Nanotechnology.' 2020. As of 7 December 2020 [Online]: <https://nrc.canada.ca/en/evaluation-national-institute-nanotechnology>.
- [37] Administrative Data Research Institute UK. 2019. 'Administrative Data is an Invaluable Resource for Public Good. Let's Use It.' *Annual Report*. As of 7 December 2020 [Online]: [https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK\\_Annual\\_Report\\_2018-19.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf).

- [38] EMBL. 2020. 'EMBL Archive – Science and Archives – Workshops on Scientific Archives.' As of 7 December 2020: <https://www.embl.de/aboutus/archive/working-with-scientific-archives/workshop/index.html>.
- [39] Human Brain Project. n.d. 'Welcome to the Human Brain Project.' As of 7 December 2020: <https://www.humanbrainproject.eu/en/>.
- [40] Census of Marine Life Mapping & Visualization. n.d. 'The World of the Census — Census of Marine Life Maps and Visualization.' As of 7 December 2020: <http://comlmaps.org/>.
- [41] Lehman, D. R. 2011. 'Office of Science Peer Reviews 10': 32.
- [42] Guthrie, S., J. Krapels, C. A. Lichten & S. Wooding. November 2016. '100 Metrics to Assess and Communicate the Value of Biomedical Research: An Ideas Book.' As of 7 December 2020 [Online]: [https://www.rand.org/pubs/research\\_reports/RR1606.html](https://www.rand.org/pubs/research_reports/RR1606.html).
- [43] Snowball Metrics. n.d. 'Snowball Metrics – STANDARDIZED RESEARCH METRICS – BY THE SECTOR FOR THE SECTOR.' As of 7 December 2020: <https://snowballmetrics.com/>.
- [44] ResearchFish® system. n.d. 'Track Research and Evidence Impact with Researchfish by Interfolio.' As of 7 December 2020: <https://researchfish.com>.
- [45] SciCV. 'Welcome to SciCV.' Swiss National Science Foundation. As of 7 December 2020: <https://scicv.ch/>.
- [46] NCBI. n.d. 'SciENCv: Science Experts Network Curriculum.' As of 7 December 2020: <https://www.ncbi.nlm.nih.gov/sciencv/background/>.
- [47] NIH Central Resource for Grants and Funding Information. n.d. 'Biosketch Format Pages, Instructions and Samples.' As of 7 December 2020: <https://grants.nih.gov/grants/forms/biosketch.htm>.
- [48] The Royal Society. n.d. 'Résumé for ResearchersRoyal Society.' As of 7 December 2020: <https://royalsociety.org/topics-policy/projects/research-culture/tools-for-support/resume-for-researchers/>.
- [49] Elsevier. 112 October 2017. 'Report Compares UK's Research Performance with Key Nations.' Elsevier Connect. As of 7 December 2020: <https://www.elsevier.com/connect/report-compares-uks-research-performance-with-key-nations>.
- [50] Boyack, K., M. E. Andersson & A. Christensen. 'Topic Prominence in Science.'
- [51] Florio M. & C. Pancotti. 30 June 2020. 'The Economics of Physics: The Social Cost-Benefit Analysis of Large Research Infrastructures.' Oxford Research Encyclopedia of Physics. As of 7 December 2020: <https://oxfordre.com/physics/view/10.1093/acrefore/9780190871994.001.0001/acrefore-9780190871994-e-23>.
- [52] Beagrie, N. & J. Houghton. 2016. 'The Value and Impact of the European Bioinformatics Institute: Full Report January 2016.' EMBL, Charles Beagrie. As of 7 December 2020 [Online]: <https://beagrie.com/static/resource/EBI-impact-report.pdf>.
- [53] BBC Radio. 2002. 'BBC Radio 4 – Reith Lectures 2002 – A Question Of Trust.' As of 7 December 2020: <http://www.bbc.co.uk/radio4/reith2002/>.

[54] Guthrie, S., I. Ghiga & S. Wooding. 2017. 'What Do We Know About Grant Peer Review in the Health Sciences?' *F1000research*. 2017;6:1335. doi: 10.12688/f1000research.11917.2.



## Annex A. Summary of the discussion forum

---

This annex provides a summary of the main points raised by participants in the discussion forum.

### Introduction from Professor Leyser

**Professor Leyser has recently been appointed CEO of UK Research and Innovation.<sup>26</sup> Previously, Professor Leyser was Director of the Sainsbury Laboratory, Cambridge University. We asked her to comment on the use of indicators and metrics in the assessment of research institutes.**

We are all keen to find ways to measure success, to ensure that we can track progress and assess the impact of strategic decisions. However, I caution against the use of simple proxy measures in assessing research performance. This does not mean that metrics are useless. UKRI has a forum for advising on the responsible use of research metrics<sup>27</sup> and has positively contributed to debate and analysis on this topic.<sup>28</sup> The general principle that has emerged from this work is that metrics can be useful in the context of evidencing narrative statements.<sup>29</sup> They cannot reliably be used as stand-alone performance measures. A core value of research is novelty and innovation, and there is a wealth of evidence supporting the crucial role that diversity of all kinds, which is antithetical to standardisation, plays in driving excellence.

Of course, funders and taxpayers need assurance that their money is being well spent. Furthermore, fair, transparent and effective mechanisms to decide how to allocate limited funding are needed. But fair and transparent does not mean standardised and robotically objective. Fair and transparent assessment of research and researcher performance can be based on nuanced, holistic, high quality, and well-informed subjective judgment.

Where performance measures/indicators/metrics are used to support, not supplant, expert peer review there are several important questions to be addressed:

---

<sup>26</sup> Formed in 2018 and operating across the whole of the UK with a combined budget of more than £8.4 billion, UKRI brings together the seven Research Councils, Innovate UK and Research England. <https://www.ukri.org/about-us/>.

<sup>27</sup> The forum for advising on the responsible use of research metrics can be found here:

<https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Pages/forum-for-responsible-research-metrics.aspx>.

<sup>28</sup> Wilsdon, J., *et al.* 2015. 'The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.' doi: 10.13140/RG.2.1.4929.136.

<sup>29</sup> 'The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies Research Report 2015/01.' King's College London and Digital Science March 2015 <https://www.kcl.ac.uk/policy-institute/assets/ref-impact.pdf>.

*1. Do the metrics correlate sufficiently with the desired performance criteria (e.g. quality)?*

Many relatively easy-to-collect metrics measure quantity and not quality. Furthermore, it is not clear what the ‘best’ quantity is for each measure. More is certainly not uniformly better for these metrics. Fewer, but of better quality is often preferable, or fewer to ensure time for other activities might be better.

*2. What are the likely consequences of adopting the proposed metrics on researcher behaviours?*

While some metrics used at institutional level are reasonable, it is clear from experience that it is extremely difficult to separate institutional from individual level incentives. This is at least as important in considering the things not measured as it is in considering the measured list. For example, if the papers that an institution publishes in specific journals are counted, this will devalue time invested in making negative results and datasets openly accessible. Although funders insist that they value these other aspects, if they are excluded from the assessment, the perception will be that they are not important. The same applies to important activities such as participating in the peer review process and other services to the research community.

*3. What are the costs of collecting the data and are they appropriately compensated by the benefits?*

The response to concerns about metrics not measuring everything that matters is often to go on adding more and more things to the list to be collected. This misses the point and simply adds administrative burden. As with all public spending, assessment systems must be good value for money.

There are strong parallels to wider questions of accountability and trust, as eloquently set out by Onora O’Neil in her Reith Lectures.<sup>30</sup> To quote her ‘Intelligent accountability, I suspect, requires more attention to good governance and fewer fantasies about total control. Good governance is possible only if institutions are allowed some margin for self-governance of a form appropriate to their particular tasks, within a framework of financial and other reporting. Such reporting, I believe, is not improved by being wholly standardised or relentlessly detailed, and since much that has to be accounted for is not easily measured it cannot be boiled down to a set of stock performance indicators.’ ‘Serious and effective accountability, I believe, needs to concentrate on good governance.’ Careful consideration is warranted for assessment of the governance arrangements for delivering excellence against an appropriately broad set of criteria informed by core values.

This highlights another key point. Funding should be allocated to a diversity of organisations, with different aims and approaches, ensuring that the wide range of objectives for research funding can be delivered across the system. This means that

---

<sup>30</sup> Onora O’Neil’s Reith Lectures, can be found at this website: <http://www.bbc.co.uk/radio4/reith2002/>

performance measures will not and should not be directly comparative between institutions. League tables make no sense when the teams listed are playing different sports.

With the creation of UKRI, there is an opportunity to consider the whole UK research system and to ensure that it is delivering what the UK needs to thrive as an inclusive knowledge economy. The entirely admirable drive to accountability and quality assurance in funding allocation must be deeply thought through in this context.

## Section 1: Planning and scoping institute evaluations

In this section, we asked participants about the reasons for evaluating research institutes, considerations for how to effectively and responsibly carry out an evaluation, and the intended audiences for the results.

### *Section 1-1: Why evaluate research institutes?*

This question focused on the reason for evaluating research institutes. Participants were shown a summary of the ‘6 As of evaluation’<sup>31</sup> and asked to comment on which, in their view, was the most important. Table 1 provides an overview of the summary of the ‘6 As of evaluation’ presented to participants:

**Table 1. The 6 As of evaluation**

Analysis	To understand why, how and whether research is effective, and how it can be better supported
Advocacy	To demonstrate the benefits of supporting research, and enhance the understanding of research and its processes among policymakers and the public
Allocation	To determine how to distribute funding across the research system
Accountability	To evidence that money and other resources have been used efficiently and effectively, and to hold stakeholders to account
Acclaim	To compare and recognise the value of research institutions and the research conducted within them
Adaptation	To steer change in organisational structures, behaviours and cultures, and research activities and priorities

**Each of the 6As were noted as useful, although different participants had different perspectives.** Several participants (including one of the original study authors) highlighted that the 6As were overlapping and not meant to be considered separately. The context for an evaluation is important, and frameworks like 6As (or alternatives<sup>32</sup>) provide a helpful common language to describe this. The purpose is not narrow

---

<sup>31</sup> Adapted from Sarah Parks, Daniela Rodriguez-Rincon, Sarah Parkinson & Catriona Manville, The changing research landscape and reflections on national research assessment in the future. Research England, 2019. [https://www.rand.org/pubs/research\\_reports/RR3200.html](https://www.rand.org/pubs/research_reports/RR3200.html)

<sup>32</sup> Lizzie Gadd suggests six reasons for conducting an evaluation in ‘The Blind & The Elephant: Bringing Clarity to Conversations about Responsible Metrics.’ [35]

categorisation, more to highlight that a) there are lots of different reasons and that b) different reasons require different methods – one size does not fit all.

**Incentives and rewards were added as reasons for evaluation.** More than one participant described how evaluations could be used to incentivise researchers. Partly this was a recognition that the very fact that an activity is measured acts as an incentive to engage in the activity, so better to make this explicit; it was also discussed as a way to encourage behaviours seen as beneficial by funders or other stakeholders (e.g. open access publishing); and lastly as a way to build trust between individual researchers and an institute, or an institute and its funder(s). The need for trust and respect between an institute and its funders was highlighted several times.

Comments in this section referenced evaluation culture – ensuring that evaluation is seen as highlighting the positive rather than just uncovering the negative.

Perhaps we are putting too much effort and such into trying to get things right by assessment, evaluation, inspection, review, etc., and not enough into how to support and enhance the science and the people driving/doing it?

Evaluation to support learning and improvement (as an extension of adaptation) was flagged as important. In general, this was felt to be missing from the 6As model. Helping researchers to step back and consider the bigger picture brings benefits for both the institute and individuals.

Often researchers or scientific leaders look, understandably, at monitoring and evaluation and see it as more cumbersome red tape, but if it is owned and run by them, it can provide valuable insight for the RI [research institute], give them the ability to demonstrate their importance, and allow them to create a feedback loop that will also benefit the research in many cases.

**Participants were split on how to treat research impact.** Impact – the need to understand how an institute's work is contributing to wider society – was mentioned by many participants. Several explicitly highlighted that due to the longer-term nature of institute funding, institutes should be able to tackle longer term strategic challenges in a way not possible through the university system. As a key point of difference for institutes, therefore, it was an important focus for evaluation. Other participants took the view that impact assessment ran through each of the stated reasons for evaluation and it wasn't necessary to call out separately. One participant emphasised that impact was a long-term endeavour and therefore difficult to assess in the timeframe of many evaluations; another suggested that a focus on the culture of the institute and whether it supported researchers to adequately focus on long-term impact was as important.

**Institutes should be evaluated against their own specific vision, mission or goals.** Participants highlighted that research institutes were unique environments, and that any evaluation should seek to capture the added value of the institute model. Facilitating collaborations, interdisciplinarity, innovative opportunities for training the next generations of researchers and accelerating translation were all mentioned.

Helping the delivery of vision and strategy... must be the main reason evaluation occurs.

Participants were clear that institutes must be evaluated in context – individual institutes have differing goals, so rather than looking to compare across institutions, most evaluation should seek to assess progress against stated objectives, regardless of the purpose.

Participants drew a distinction between evaluation conducted by an institute for internal governance purposes, and evaluation for/by an external party.

Regular evaluation is a necessary and healthy function of governance.

Participants highlighted regular internal analysis and adaptation, focused on understanding progress against strategic objectives, as an important part of institute management. Institutes themselves may be well placed to assess culture, processes and behaviours that foster a positive research environment.

Externally driven evaluations were more likely to be focused on accountability, or on advocacy and acclaim. The degree to which an institute is accountable to its funders depends on its funding and governance arrangements; an institute owned by a single funder can expect a greater degree of accountability than one with multiple funding streams and more independent governance.

Advocacy and acclaim are not so much focused on research institutes themselves; it is more about ‘selling’ research and research institutes.

The nature of the funder may also affect the extent to which accountability, advocacy and acclaim are important. Government funding requires a high level of accountability; charitable funding often requires a focus on advocacy to support fundraising. All funders were also keen to understand what works and how to support institutes in conducting the best science (analysis and adaptation).

**The reason for an evaluation needs to be agreed at the beginning.** One participant commented there could be hundreds of reasons to conduct an evaluation; what’s important is not how we categorise it, but that all parties are agreed at the outset on both the purpose of the evaluation, and the vision and goals that are being assessed. For institutes with complex governance (and hence many stakeholders) and/or shifting strategic priorities, this isn’t always straightforward. Participants highlighted open and constructive dialogue between all parties as the key to a successful evaluation.

#### *Section 1-2: How should we evaluate research institutes?*

Participants were shown a summary of the SCOPE process and asked to use this as a starting point for a discussion on how research institutes should be evaluated, considering what makes an institute unique and how success might change over time.

**The ‘how’ should follow on from the ‘why’.** The purpose of an evaluation should inform how it is conducted, and the methods used. In this context, one respondent advocated for the Dutch Strategy Evaluation Protocol (SEP)<sup>33</sup>; another suggested that the SEP was very good for developmental or formative evaluation, but not suitable for funding allocation. More generally, acclaim and advocacy were less likely to require comprehensive evaluation. As above, agreement between all parties on purpose and scope was seen as a prerequisite for effective evaluation.

---

<sup>33</sup> For more details on the Strategy Evaluation Protocol, see [30].

**Evaluation should take an institute's goals as the starting point.** Many participants agreed that the first S of Scope – ‘start with what you value’ – was the required first step. This should be based on the institute's goals, vision and/or strategy, and again many respondents emphasised the need for this to be communicated and agreed between stakeholders.

Participants emphasised the distinct niche that institutes occupy and cautioned against unthinking external comparison based on people and/or outputs. Funders should be clear on their rationale for supporting institutes over other types of investment, and frame evaluation questions accordingly. Most thought that the right comparison was an internal one of progress against strategic goals, rather than another institution – which will likely have a different set of goals, vision and strategy.

**However, frameworks should not be so rigid as to stifle innovation and creativity.** Institutes, particularly those focused on earlier stage discovery science, should be free to pursue their goals as they see fit. One participant commented:

If research institutes do not take up opportunities the fulfil their vision in ways other than those set out in their strategy, that would be a shame.

Institutes are typically long-term investments and should (and do) change over time in response to both internal change – of leadership, strategy, etc., and external context. Evaluation systems need to recognise the experimental and changing nature of science and be agile enough to accommodate changes in both research questions and researchers.

**A collegiate approach to institute evaluation is crucial to success.** Participants emphasised the need for institutes, funders and external parties to work together to deliver institute evaluations. Evaluation criteria need to be discussed amongst all stakeholders, and funders should take note of institute suggestions on how progress can be evidenced.

The SCOPE principle of ‘evaluate with the evaluated’ was highlighted, and funders emphasised the importance of trust. Institutes should feel that they are participating in an exercise aimed at helping them to improve, rather than being subjected to an evaluation that is done *with* them rather than *to* them.

Responsible evaluation stems from mutual transparency in motivation, basis for assessment and clarity of consequence

One participant suggested that decoupling evaluations from funding decisions, if practical concerns allow, can provide the space for a wider, richer evaluation.

**Evaluation needs to consider interactions between groups.** Institutes can be evaluated at the level of individual research groups, in research themes, or at the level of the institute. Several participants highlighted the opportunity for inter-group collaboration as a differentiator for the institute model. Others highlighted the expectation from funders and others that institutes be *more than the sum of their parts* – maximising collaborative opportunities to deliver ambitious scientific goals. Evaluations need to be designed to look beyond individual research groups and assess these interactions.

Physically distant or separated institutes present different challenges, and for these a culture of transparency and a strong sense of identify are important areas of research culture to assess.

**Assessment of research culture is important for institutes.** Participants thought that definitions of excellence should be broad, encompassing organisational culture as well as academic and other output. Leadership is crucial and should form part of the evaluation.

Participants discussed culture as an enabler of scientific progress, and cited training and mentoring opportunities, transparency and openness to new ideas as examples of internal processes that enable success.

**Evaluations should lead to clear conclusions.** Many participants emphasised the evaluations need to be learnt from and conclusions put into action to be useful. To that end, a clear conclusion (whether positive or negative) is important.

**Evaluation has costs as well as benefits, and care is needed to reduce or avoid these.** Participants welcomed the focus on ‘unintended consequences’ in the SCOPE framework. Evaluations can be resource intensive for institutes, and the desire for assessment needs to be balanced against the effort required. The quality of the evaluation needs to meet the expected quality of the science, and participants emphasised the importance of proactively engaging expert reviewers, both for effective input of the evaluation and, less directly, for the institute to feel the evaluation has been properly and objectively evaluated. Site visits are an important part of many in-depth evaluations. Participants also highlighted the role that institutes’ own advisory boards can play in the evaluation process.

Robust advisory boards who can provide expert (and of course critical where needed) views can be extremely helpful.

The principle of *evaluator, evaluate thyself* was mentioned, and one participant proposed conducting a process similar to an equality impact assessment, with the aim of ensuring the evaluation does not discriminate, and to understand (and therefore minimise) the degree of disruption to both the institute and the reviewers.

### *Section 1-3: Who are the intended audiences?*

Participants were asked to comment on the audiences for institute evaluations, and how well current approaches address those audiences. We proposed the academic community, policy makers, patients and the public, sponsors or funders, and industry or commercial partners as potential audiences, and invited participants to suggest other communities.

**Three key groups were flagged as important audiences.** Most participants highlighted the funder as the primary audience for an institute evaluation; both to ensure that the institute is spending money effectively and efficiently and that it is contributing to the funder’s wider strategic goals. A similar number of participants listed the senior management of the institute as a key audience; evaluations were seen as an important learning opportunity and a window to adjust and improve. Policy makers and (to a lesser degree) patients and the public were also mentioned; these groups are more likely to be secondary audiences, with outputs packaged as highlights or stories.

**Different audiences require different types of evaluation.** One participant questioned whether we should seek to satisfy all stakeholder groups with all evaluations. Participant consensus was no – that different audiences had different needs, and that seeking to meet these in a single evaluation was likely to be prohibitively ‘complex, burdensome and resource intensive’. One participant highlighted the importance of adapting format and language to the audience.

Participants also highlighted the resource and effort costs of conducting multiple evaluations for different audiences, and the benefits of repurposing elements of an evaluation; for example, using the outputs of an expert review to develop case studies that can be used for policymakers. One participant described ‘the time spent addressing similar but slightly different reporting needs from the same funding organisation.’

**Institute staff are an important audience for institute evaluations**

Participants highlighted that institute staff are an important audience for institute evaluation. They said that:

The critical audience is the institute itself and the researchers within it.

Constructive criticism for members and leaders of an institute can be very helpful.

Several participants pointed out that internal audiences should not be limited to senior management – communicating success back to staff at all levels can boost morale and aid retention and productivity. Evaluations should be a learning opportunity for staff at all levels, and open communication of what is working (or otherwise) builds trust between the institute and its stakeholders, particularly funders.

**Evaluators are also an audience for evaluations.** In the spirit of transparency and improvement, the evaluation community is an important audience for evaluations. This ensures that quality remains high by introducing new techniques, highlighting best practice and allowing peers to comment on methodologies and results. By modelling openness, funders and evaluators can build trust with other stakeholders.

## Section 2: Which aspects to evaluate

In this section, participants were asked to discuss the inputs, outputs and impacts of research institute work. Below we summarise the main points from the discussion.

### *Section 2-1: Frameworks and Logic Models*

This subsection focused on frameworks and logic models and the extent to which they are helpful in organising an evaluation. Participants were asked to focus on when they are useful and what the pitfalls of using them are. Participants were given an example of a logic model used by the National Research Council of Canada's evaluation of its National Institute of Nanotechnology [36].

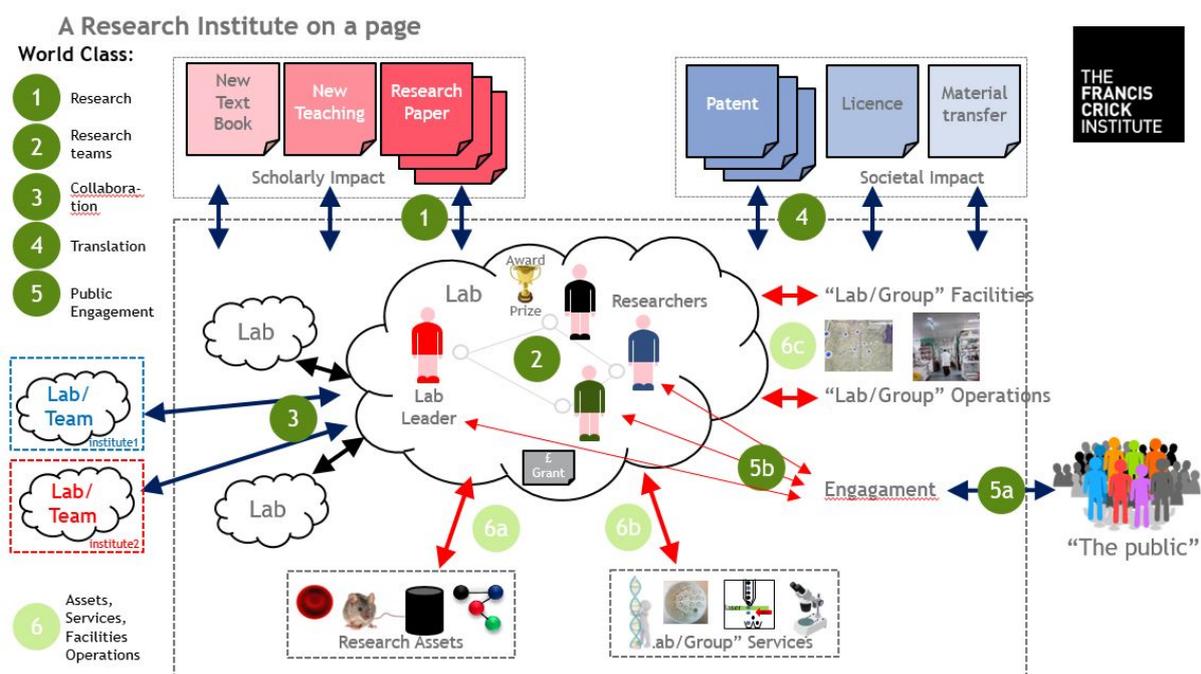
**Logic models can be useful to structure thinking.** Participants argued that logic models and frameworks are useful starting points for structuring thinking. They argued that logic models can provide structure to a messy web of interactions, help provide guidance to deliver certain outcomes, clarify and expose the assumptions that underpin impact pathways, and provide a solid framework for evaluations.

**Logic models are a useful communication tool to the diverse stakeholders involved in the evaluation of research institutes.** Participants felt that logic models provide one diagram that communicates the complex ways in which impact is generated. As a result, they create the basis for evaluations and provide grounds to prove the relative importance of different pathways using both qualitative and quantitative analysis. Within an institution, logic models can illustrate how different strands of work fit together as part of the overall objectives. They thereby act to ensure that everyone is on the same page when determining the scope of an evaluation and ensuring that there are no gaps.

**Many participants argued that frameworks should be simple.** Some participants felt that a simple logic model that focuses on activities, outcomes and impact, or that pulls out key strategic components is a good approach to avoid semantics around areas such as input, outputs and intermediate states. This can provide a helpful and simple starting point to collate evidence for each step in the logic model. Although simple models might not be applicable to the evaluation of a research institute as a whole, this can overcome some of the challenges of logic models that require a lot of investment to make work.

There was some disagreement on how simple logic models can be achieved. One participant said that the logic model should be simple enough to be described on a single page and suggested using six areas across the evaluation of different research institutes: i) research, ii) research teams, iii) collaboration, iv) translation, v) public engagement, and vi) operations (see Figure 1 for a visualisation of the diagram). The same participant also suggested a common language for logic models to provide quick and easy identification of different areas. Several others said that diagrams overcomplicate assessments and that assessments were simpler by only focusing on leadership, quality of science, quality of core facilities, output in terms of major discoveries and their impact, translation of research to commercial or clinical applications, training and career development, public outreach and funding. Figure 1 provides an overview of a model created by a participant to illustrate that models of research institutes should be simple enough to be described on a single page. They felt that six areas seem common: i) research, ii) research teams, iii) collaboration, iv) translation, v) public engagement, vi) operations. They also suggested that adopting a common set of words to describe a few things means that a framework can help with all levels of an evaluation (including institute self-evaluation and communication; researcher viewpoints; funder viewpoints, and policymaker viewpoints)

Figure 1. A Research Institute on a page



**Logic models should be developed at the start of the evaluation period rather than being retrofitted to justify activities.** Several participants argued that logic models should be prepared before the institute or programme is set up or before funding comes in. In this way, they can act as a reminder of the longer-term focus of the research institute.

Because logic models are specialised tools, different logic models should be bespoke for the aims of different institutes. A couple of participants felt that logic models should always be adapted to specific institutes and their needs.<sup>34</sup>

**Logic models should be used as a theory of change.** Participants also highlighted that frameworks should not be rigid; they should be developed as a theory of change. It should include a commitment to keep it ‘alive and to refresh’, ensure that research also considers outcomes at both the application and commissioning stage so that plans can be reviewed and changed, and account for the fact that research programmes and the external context change over time.<sup>35</sup>

**There might be some scale issues with the evaluation of research institutes using logic models.** Some participants felt that whilst logic models are useful for programme evaluation or specific interventions, they were unsure about the use of logic models when institutes are the unit of analysis. Reasons that were provided included the complexity of logic models, that pathways to impact are so diverse that they are

<sup>34</sup> Examples that were mentioned included Biotechnology and Biological Sciences Research Council (BBSRC), which has developed logic models that are bespoke for a subset of their investments in institutes that are strategically funded. The investments focused on national capability investments and logic models were then used to inform the development of unique quantitative and qualitative indicators for the different investments.

<sup>35</sup> Examples that were mentioned included UKPRP Impact and Evaluation Framework [6]. Another participant used a ‘theory of change’ to facilitate a workshop on Obesity Health Alliance Weight Strategy. They said that using a theory of change exposed where there were intellectual leaps at important junctures where change is meant to happen, and allowed assumptions to be laid out at the start, whilst recognising the wider ecosystem you are trying to influence.

difficult to fit into a single framework, that logic models require a lot of resource to put together, and that they can end up being overly simplistic, not robust enough and not fit for purpose. Examples of alternative frameworks that can be used include the maturity model used by UKRI funded institutes to organise group thinking, which was highlighted by one of the facilitators to the forum.<sup>36</sup>

### *Section 2-2: Research outputs and outcomes*

In this subsection, participants were asked to review some of the ways outputs and outcomes are referred to in institute reports, and to expand on the information provided and some of its merits. Participants were asked to consider what works well and not so well in the examples of inputs/outputs/outcomes referred to in the institute reports.

**Qualitative data, narratives, and the ‘so what?’ should be included when measuring outputs and outcomes.** Several participants felt that metrics on outputs and outcomes also have to include qualitative data. They said that it is necessary to create narratives, through i.e. case studies, to speak of the context of quantitative data, as well as the impact and meaning created by them. One participant suggested that institute records and archives are a helpful way of understanding the contribution and impacts of institutes in the long term.<sup>37</sup> Another participant said that it is necessary to investigate local datasets, involve stakeholders and to achieve a balance between quantitative and qualitative data to adequately measure outcomes at an institute level. Another participant said that an effective evaluation requires expert or peer review of the information included in quantitative metrics. However, one participant felt that a shift towards narratives and qualitative data can only happen with a culture change in the funding landscape that aims to understand both the successes and challenges faced by projects in the funding matrix.

Several important qualitative outputs and outcomes were missing from the ways outputs and outcomes were referred to in institute reports. Participants highlighted several qualitative outputs and outcomes that they thought should be included when measuring outputs and outcomes including: i) open, reproducible science, ii) diversity (beyond male and female ratios), iii) contributions to resources that benefit the wider research community,<sup>38</sup> iv) exit data related to career development, v) investment from diverse stakeholders, vi) initiatives to promote data reproducibility and sharing, vii) resources available to the wider scientific community, viii) research culture, ix) inclusive research, x) patient benefit, xi) progression of technology to preclinical space, xii) first-in-human trial, xiii) policy changes or implementation of new patient treatments, xiv) list of key discoveries in the past 5–10 years, xv) collaboration internally and externally with other institutes and other bodies, xvi) how teams are formed, xvii) impacts of research done in medical hospitals

---

<sup>36</sup> The Administrative Data Research Institute, funded by the ESRC, has also used a maturity model. Their annual report summarises a maturity model as a ‘way of considering performance against an agreed set of graded indicators where targets can be set for each (...) objective. Achieving or exceeding target levels of maturity becomes a measure of success for the programme’ [37].

The Health Data Research UK Institute, funded by the MRC, also used the maturity model to develop their evaluation plan. The maturity model framework was used in evaluation planning and as part of the evaluation process for internal assessment and external reporting.

<sup>37</sup> Further detail is provided here: [38]. It might be necessary with digital archives too, and annual workshops have been organised on this theme [38].

<sup>38</sup> Such as ‘consortia for sequencing and functional annotation of genomes and associated open data; curated databases or archives; novel animal models or materials’.

that results in changes in medical practices in hospitals, xviii) the extent to which an institute has enabled ground-breaking discoveries or pushed the frontiers of human knowledge.

For example, one participant wrote about their experience with capturing the extent to which their institute stimulates interdisciplinary approaches, bridges basic with clinical research, and contributes to other important collaborations in their research landscape as important ways to qualitatively measure outputs and outcomes. Another participant highlighted the administrative burden of collecting data on relevant qualitative indicators and that it is necessary to be selective about the right mix of indicators. One participant thought that guidance that shifts the focus from ‘suitable indicators’ to ‘how to ensure indicators are used/interpreted effectively’ could be helpful. One participant highlighted that different indicators are appropriate for different audiences. For example, scientific narratives can be used for the public, publications/citations for scientific review, and measures of economic and societal impact for government.<sup>39</sup>

**Participants also suggested that other metrics should be included** such as machine-readable metrics (such as reports from Scopus, precise financial information on funding for i.e. grants, training, engagement, open/collaborative science).

**It is important to be clear on attribution related to indicators and their associated narratives in the evaluation process.** Several participants felt that there is a challenge with attribution for narratives and indicators related to outputs and outcomes that should be made clearer during evaluation. In particular, participants felt that attribution is a challenge when an evaluation focuses on the specific outputs or outcomes of certain investments rather than the research institute as a whole. They also argued that significant resources can be spent on identifying attribution that could have been spent more effectively.

### *Section 2-3: Activities and processes*

In this subsection, participants were asked to discuss four questions: i) how can we measure efficiencies created by the institute model?, ii) how can we capture qualities, such as productive collaboration?, iii) what aspects of positive research culture do we currently capture, and how can we improve this?, iv) how can we assess good governance?

To measure positive research culture and good governance to ensure an inclusive research economy, there should a recognition of ‘voices on the ground’. Several participants highlighted that evaluations should engage with staff, students, professors, and those affected by the work of research institutes to adequately measure positive research culture and good governance. Participants said that it is necessary to involve the perspectives of PhD students to professors, technical and professional staff, support personnel, alumni, and internal and external consumers.

Suggested approaches to measure these aspects included anonymous surveys for staff asking about their perceptions of management and the research culture, organising meetings and informal chats with students and junior group leaders, having meetings with staff without management, having formal discussions with leaders, peer review and peer feedback, and carrying out ethnographies that combine interviews and participation in daily practice. Other suggestions included informal poster sessions and breakout groups

---

<sup>39</sup> Examples that were mentioned of work being done to define indicators for socio-economic evaluations of RI included the RI-PATHS work [16].

over lunch, receptions after plenary lectures, audits, exit interviews, and using Codes of Practice. Participants also highlighted that it is important to use a mix of online and face-to-face meetings to facilitate this ‘bottom-up’ approach to an evaluation.

Engagement with the diverse stakeholders involved in a research institute can also provide a strong indication of other factors such as openness, inclusion, the extent to which there is a supportive environment, mentorship, the sense of identity in the institute, facilitated collaborations and interactions throughout the institute, and the quality of the leadership.

Examples mentioned by participants included:

- MRC’s models of engagement with early career researchers (ECRs) at Quinquennial Reviews using poster sessions, breakout groups over lunch, and receptions after lectures.
- One research institute carried out an internal exercise to understand what they perceived their logic model to be, then created nine statements of ‘intent’. Users were then asked in their annual user survey whether they agreed or disagreed with the intent statements. These trends have been tracked since 2015.
- The CRICK real-time biography also recognised the ‘voices on the ground’, using a combination of qualitative and quantitative approaches.

(Productive) collaboration should be measured at different levels using a combination of qualitative and quantitative approaches. Suggested approaches to measure productive collaboration included measuring relationship capital, capturing collaborations at different levels through annual internal reports from group leaders, annual visits to the headquarters team, ResearchFish® [44] reports, and using collaborative network maps that can highlight different types of collaboration. Other suggestions include outputs and surveys linked to internal mechanisms that encourage collaboration across centres. One participant suggested using usual outputs such as joint publications, joint data platforms, additional joint funding, but to reflect on outcomes and outputs relative to peers. One participant highlighted that in quantitative terms one can explore outputs or outcomes per input (such as Work Load Models) and then benchmark relative to other Higher Education Institutions. Others suggested including process or culture aspects in the evaluation.

One participant argued that institutes are always embedded in a larger research ecology and that exchanges with the outside world are important when analysing research institute work. The participants shared a multimethod framework developed at the Francis Crick Institute to look at activities and processes in an integrative way, with a focus on collaboration, translation and public engagement. A real-time biography was written in collaboration with management, staff and scientists at the CRICK to capture how research can emerge and transform over time, and to understand the dynamics of research culture.

However, participants also highlighted that it is difficult to measure productive collaboration and that it will likely require extra data collection and create extra burden.

Examples that were mentioned by participants included:

- A pan-European distributed research infrastructure kept running a list of examples of positive things said by people in the context of relationship capital, e.g. ‘capacity-building effect across

different parts of the organisation’, ‘peer support’, ‘being part of something larger’, ‘opportunities to influence a pan-European strategy’. They then started designing perception surveys to capture evidence along these lines.

- The Human Brain Project will produce work relevant to collaboration by measuring collaboration across the network.<sup>40</sup>

**The evaluation of research institutes should measure governance structures in the evaluation.**

Participants argued that evaluations should recognise and measure governance structures, such as existing structures, transparent and effective procedures, functional supporting infrastructures, and standards for data or protocol sharing. One participant said that because the governance ecology is complex, it is important to address how governance happens and how different institutional cultures merge or create new governance ecologies. The same participant also felt that in larger international projects that the balance between autonomy and interdependence is important, and that scientists will value structures where they can provide their own input (albeit it might be difficult with infrastructures that have more of a research orientation).

However, measuring governance is challenging because it cuts across preexisting structures, power relations and hierarchies, and operates alongside different types of research. One participant said that it was necessary for there to be more learning on governance models across institutes through, for example, a comparative study, but to adapt the learning from the study to specific institutes and needs. One participant said that a way to solve some of these governance issues is to make country nodes (members at a national level) work together across jurisdictions using core funding, which makes institutes contractually bound to deliver collaborative projects internationally.

Examples of this approach that were mentioned by participants included:

- The Census of Marine Life works with local or regional nodes under the central government structures, which facilitates networking of projects that are separately funded into a coherent whole.<sup>41</sup>

**It is necessary with longitudinal approaches that can measure efficiencies and inefficiencies over time.**

Participants said that longitudinal approaches can provide insights into the strengths and challenges faced by research institutes and how they change over time. One participant said that longitudinal approaches can capture aspects such as the legacy agency, and another participant felt that the longitudinal approach can capture improvements in an efficiency metric over time for comparison and benchmarking. They highlighted that the evaluation should seek to explore both inefficiencies and efficiencies over time to ‘develop critical mass and deliver excellent science’. One participant highlighted that the counterfactual approach can be particularly useful to measure efficiencies. They also felt that there was an efficiency challenge with redundancy across research institutes as some techniques, tools, and approaches are replicated across research institutes in a discipline area.

---

<sup>40</sup> Further details can be found here: [39].

<sup>41</sup> Further details can be found here: [40].

**Evaluation should be delivered in a way that encourages collaboration.** Participants highlighted that evaluation should be delivered in a way that encourages collaboration. One participant said that both intra- and inter-mural collaboration should be valued in the evaluation process to ensure that individual groups do not feel that they have to deliver individual research papers for an evaluation. One participant also highlighted that it is necessary for evaluations to address how big institutes can support the diffusion of research excellence. For example, evidence of a positive research culture can be that equitable and place-based agendas move forward rather than institutes getting a bigger share every time.

*Section 2-4: Resources*

In this subsection, participants were asked to discuss the resources available to institutes and the cost of research. They were asked whether it was important to have clear information about inputs, why, and the aspects that are most important in terms of financial resources (whether core or externally funded), laboratory space, and number of front-line researchers. They were provided with two examples and were asked whether these approaches are helpful and how else resources should be taken into account when evaluating outputs and outcomes.

1. Staffing framework: A simple four-step career pathway with a second dimension to capture whether staff are on an academic or technology specialist (including facility management) career track, which could be used to provide a full description of the research workforce. Staff may of course move between these tracks. The staffing framework is provided in Table 2.

## Table 2. Experimental career stage/track framework for recording staffing data

**Problem** – The composition of the research workforce at a research institute is arguably an important part of explaining its productivity and achievement of key objectives. However, there is no simple framework for consistently describing the numbers of staff at various career stages and often specialist/technologist track staff including ‘team scientists’ are omitted from staffing statistics which focus on highlighting students, fellows and project leaders.

**Proposal** – A simple four step career pathway<sup>42</sup> is outlined with a second dimension to capture whether staff are on an academic or technology specialist (including facility management) career track, which could be used to provide a full description of the research workforce. Staff may of course move between these tracks.

---

<sup>42</sup> In consultation with other funders, the MRC developed a seven-stage career framework for biomedicine (<https://mrc.ukri.org/skills-careers/interactive-career-framework/#?funderview>). The simplification above assumes that ‘Education’ (e.g. undergraduate student) is not relevant to research institutes, and the ‘Consolidation/Exploration’ stages and ‘Independence/Leadership’ stages can be combined.

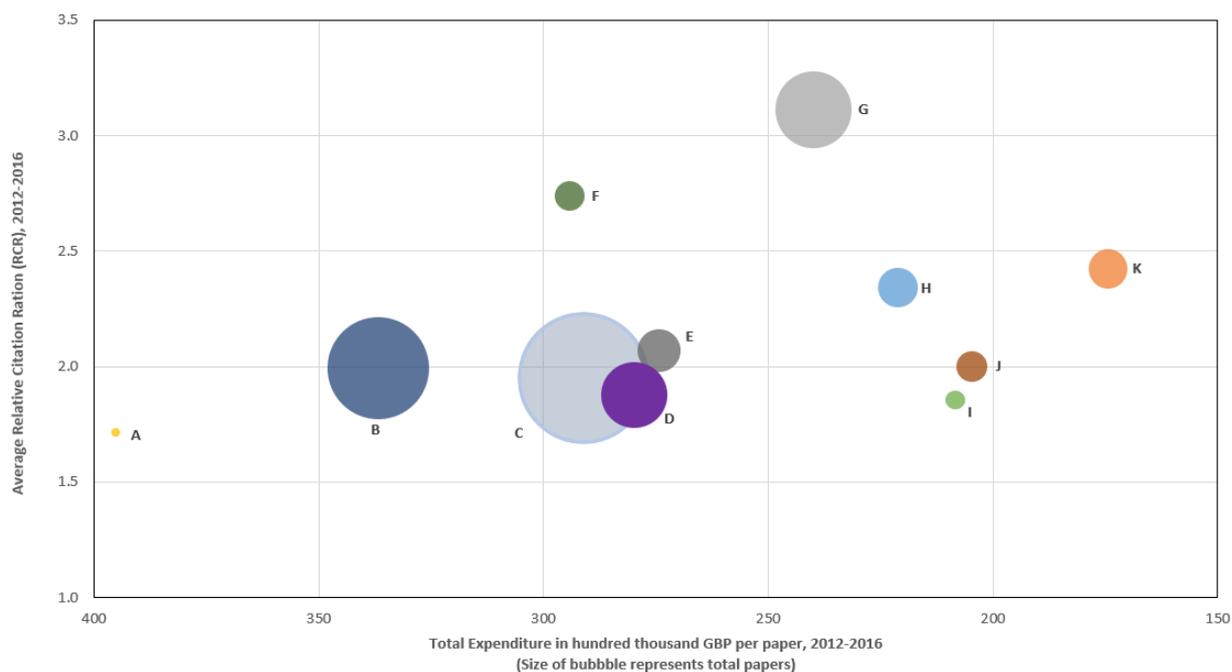
Career stage

		<b>Training</b>	<b>Consolidation/Exploration</b>	<b>Progression to/establishing independence</b>	<b>Transition to senior leadership</b>
		Acquisition of knowledge and skills, technical and transferable, through original research	Consolidation of research skills and exploration of aptitude to becoming independent	Leading independent research plans and establishing teams, start to lead and manage own programme/team/resources	Setting strategic direction and/or leadership and management of multiple programmes and/or teams and resources
<b>Career track</b>	<b>Academic track</b> Individuals currently pursuing a career path which could lead to appointment as a group leader, though they may move to other career paths in the future	<ul style="list-style-type: none"> <li>• PhD and MSc students</li> <li>• Clinical training fellowships</li> </ul>	<ul style="list-style-type: none"> <li>• Postdoctoral research staff</li> <li>• Junior fellowships (e.g. Sir Henry Wellcome Postdoctoral Fellowships)</li> <li>• NIHR academic clinical fellowships</li> </ul>	<ul style="list-style-type: none"> <li>• Externally funded fellowships (e.g. Career Development Fellowships, Henry Dale Fellowships)</li> <li>• Individuals referred to as 'Programme Leader Track' or '6+6 group leaders'</li> <li>• UKRI future leader fellows</li> </ul>	<ul style="list-style-type: none"> <li>• Senior group leaders and Director level roles</li> <li>• Permanent programme leaders and any equivalent long-term (3+years) seconded/externally funded group leaders, such as senior fellows</li> </ul>
	<b>Technician/Technology Specialist Track</b> Individuals that maintain and develop new and improved approaches to technologies and methodologies to better address research questions, including the organisation of facilities and services offered to groups across the institute	<ul style="list-style-type: none"> <li>• Research officers/technicians still receiving training in the role</li> </ul>	<ul style="list-style-type: none"> <li>• Research technology specialists (may include research nurses, analysts, junior informaticians, microscopy technicians, biological resource facility technicians etc.)</li> </ul>	<ul style="list-style-type: none"> <li>• Senior research technology specialist, laboratory/facility manager</li> </ul>	<ul style="list-style-type: none"> <li>• Research/technology specialist Director</li> </ul>

For completeness a further category of staff not related to the research career framework would include administrative (management, operations, IT support (not including front-line scientific computing) etc.) and any 'other' (e.g. non-scientific contractors such as cleaners, consultants) staff at the institute.

- Comparing publication productivity: Efficiency, or outputs per unit of input, has been included as part of institute assessment. Participants were given an anonymised example of a chart that compares one measure of input (total expenditure, i.e. financial resources) against two output measures (the number and influence of publications). Figure 2 provides the example given to participants.

**Figure 2. Comparison of expenditure and Relative Citation Ratio of publication output for comparator institutes (2012-2016)**



**These frameworks should be part of a wider evaluation and recognised within the full context.** One participant felt that the examples were useful as long as they were just used as aspects of a wider evaluation. Another participant addressed the publication productivity example and said that publications can be used as a metric as long as the trajectory of a project is accounted for as many research institutes have projects that fall outside of usual UKRI grant funding cycles. This was echoed by a participant who said that the frameworks and analytics were useful, but had to be considered within the full context, as some areas of research can more easily gain traction than others on plots that use metrics.

**Citation indexes and publication productivity comparisons have some limitations.** Many participants felt that there are clear limitations with a citation index as a function of expenditure. One participant said that it was inconsistent with delivering scientific excellence related to novelty, rigour and alignment with strategies at an organisational level. Another participant said that there are issues and biases embedded in both the citation index and expenditure data. They also felt that this form of analysis fails to account for diverse outputs and efforts to make outputs or outcomes reproducible and openly accessible. One participant said that some of these limitations can be addressed by balancing excellence and efficiency, i.e. using efficiency models, but then situating them within a wider context and set of analyses.

**A time to value framework is perceived to be more useful.** Participants argued that although it is more complex, a ‘time to value’ framework is more helpful. This framework looks at the economic impact or return of an investment for institutes over time. Another participant felt that it was a good way to share journeys.

**Staff or career tracking is a useful metric for evaluations.** Several participants felt that the staffing framework or staff or career tracking is a useful metric for evaluations. One participant said that developing staff who are highly skilled is a priority from an investment point of view. Participants found that staff or career tracking is particularly useful when measuring how research institutes contribute skills to the economy beyond linear progression academically. Participants valued that the staffing framework considers both academic and non-academic careers that are service-related.<sup>43</sup> Another participant suggested including the influence an institute can have on the careers of those who are not employed, but who gain training, experience or information that can help their career.

**Measures should capture how combinations of core and uplift funding to institutes have wider implications.** One participant said that many outputs arise from the combination of the two types of funding. A couple of participants felt that measures that capture how core funding to institutes can build collaborations and catalyse and leverage support from other sources is helpful. They felt that uplift or supplementary funding is often vital to facilitate the wide scope of research in a costly research infrastructure, but that it might be difficult to measure as it can be viewed as double-dipping for funding to meet certain objectives. However, others highlighted that external funding was also perceived as something that was leveraged by external funding and used as a way to increase impact.

Examples of recognising both in evaluations mentioned by participants included:

- BBSRC presents uplift funding per programme as a performance indicator to show how core support was leveraged.
- Novo Nordisk tracks funding and the expenditures of research centres by funding source, including both core funding and other private Danish grants. The outputs and outcomes of centre activities are then tracked through the Researchfish® [44] Research Centre Module.

**It might be useful to have a specific facility that collects data from all parts of the Institute.** Two participants agreed that it might be necessary to have one facility that can coordinate and collect data relevant to all parts of a Research Institute. They felt that the facility required significant investments, but had ‘enormous’ returns by providing time views, whole lab views across all data model areas, etc. They highlighted that funders can ‘help evaluation’ by providing funds that support the implementation and maintenance of this type of infrastructure.

Examples of specific facilities mentioned by participants were:

---

<sup>43</sup> One participant wondered if data and software-related metrics should be included in evaluations to also account for the performance of non-academic staff in the research infrastructure.

- The CRICK Data World (CDW): the CDW is supported by a cloud-based dashboard and reporting platform that holds copies of master data from key applications and systems and external sources<sup>44</sup> that is automatically refreshed from the sources.

---

<sup>44</sup> Such as Scopus, Microsoft Academic Graph, Nature, Altmetrics, etc.

### Section 3: Outcomes for stakeholder groups

In this section, respondents were asked to discuss how institute evaluation should include the perspectives of different stakeholders. Suggested stakeholder groups were the academic community, policy makers, patients and the public, research funders and commercial partners. Respondents gave their views on how to capture evidence about how these groups are affected by or interact with institutes and suggested other important stakeholder groups to include.

Respondents were prompted with an example 2019 MRC report featuring multi-stakeholder input into evaluation of their Translational Research portfolio. Here, evaluators interviewed 110 stakeholders from across academia, industry, funding agencies, investment firms and technology transfer offices. The aim of the evaluation was to explore: what MRC translational research had delivered; the national and international context for translational research; and issues to address in future funding. Stakeholder input was used to:

1. Provide richer context on the broader landscape of translational research in which the evaluation was being performed, including identifying major trends affecting the sector and any issues that future revisions to funding strategies should seek to address.
2. Identify how the portfolio contributed to the landscape, and how relevant and complementary the portfolio is to the past, current and ongoing needs and priorities of other stakeholders

While this evaluation did not focus on institutes, the benefits of stakeholder involvement are likely to be similar in institute evaluation.

**Involve diverse stakeholders during evaluation – but also before, and after.** Respondents emphasised that, rather than just acting as sources of evidence for the impact of an institute on different sectors, stakeholders can play more diverse roles in evaluation. For example, through consultation they can add valuable external perspectives to evaluation design, such as helping evaluators understand what are considered the most credible evidence of outcomes and impacts. This consultation could be gathered via an advisory board, or other novel routes such as a conference or online discussion forum. During evaluation, external stakeholders can inform the evaluator's interpretation of evidence. In the context of research institutes, this could take the form of including a diverse range of peer reviewers in the evaluation process, which is likely to be of particular value for evaluation of institute research, and which is closer to being applied. Once evaluation is complete, external stakeholders can inform communication strategies and support putting evaluation findings into practice.

**Evaluation should be a collaboration between funders and other stakeholders.** As the stakeholders are most likely to perform evaluation of institutes, respondents thought it was important that funders proactively reflect on, share and refine their evaluation methods and approaches with other stakeholders, particularly those being evaluated. Respondents felt that coproducing the most appropriate and proportionate approaches for evaluating institutes would promote engagement, best practice and progressive thinking in this area.

Others emphasised the importance of constant open dialogue between funders and institutional stakeholders during evaluation cycles, to shape and continuously improve their institutional evaluation processes. For example, ensuring that strategic objectives and evaluation approach

evolve alongside an institute's stage of growth and reflect the specific objectives and context of an individual institute. Participative methods such as Evaluative Inquiry, developed by CWTS Leiden, can be used by funders to ensure that they and the institute they evaluate have a shared understanding of what an institute is aiming to achieve, and when.

**Local stakeholders for evidencing local impact.** Respondents identified an additional stakeholder group – the local and regional connections of an institute. For many public funders, research funding, including funding for institutes, is part of an industrial strategy aiming to deliver economic growth. For example, in June 2020 the UK government announced a £400m investment in regional R&D projects expected to drive local economic growth as part of the UK R&D Roadmap. As government justification for research funding becomes more explicitly linked to economic growth, evaluation is likely to require more focus on evidencing economic impact, including the impact of institutes on local and regional stakeholders. Recent EU and OECD projects have developed frameworks for evidencing the impact of research infrastructure on local and regional stakeholders. There are many places where these conceptual models and indicators are likely to be relevant to evaluation of institutes.

**Academic stakeholders are an important and diverse group.** Respondents cautioned against an oversimplified view of how institutes are expected to impact academic stakeholders, given the diversity of this stakeholder group.

Where an institute's strategic objectives include goals such as improvements to research integrity, equity, diversity and inclusion and skills capacity building, respondents emphasised the importance of adequately monitoring impact on its internal and affiliated academic stakeholders, such as current and former staff and students, as well as external stakeholders. Funders are increasingly realising the importance of a creative, supportive, and inclusive research culture. As this could be an important support or barrier to sustainable institute impact, ensuring internal stakeholders are part of institute evaluation is important.

Other institutes were identified as another important academic stakeholder group, especially institutes from outside of the country of the institute being evaluated. Respondents felt that this perspective could be helpful to provide greater context on the impact of an institute on others within its field, and the extent to which it is enabling research by a wider circle of other academic stakeholders.

**Not all stakeholders are relevant to every institute.** Respondents cautioned against a one-size-fits-all approach, as the most relevant stakeholders for any given institute will vary depending on their strategic objectives, the aims of the evaluation and the likely users of different kinds of research. The SCOPE process for responsible research evaluation emphasises this step, with its recommendation to 'Start with what you value'.

## Section 4: Evaluation methods

In this section, respondents were asked to discuss the methodological approaches that could best be used to organise and collect the evidence to evaluate research institutes. The discussion emphasised direct involvement of the research institute and different ways to organise this process. It covered the merits of quantitative, qualitative and mixed methods, and dug into some of the general properties of metrics and indicators used.

**Involving research institutes in their assessment.** Ways to increase the involvement of research institutes in their own evaluation were discussed at several points in the forum. The example of the InterAction collaboration [13] was cited. This is an international network of particle physics laboratories focused on improving the communication of particle physics globally. Members of InterAction can request a peer review of their communications activities and this is then organised by the collaboration. Reviews are structured in a similar way to ‘Lehman reviews’ [41] (named after the US Department of Energy’s Director of the DOE Office of Science’s Office of Project Assessment), which are used to assess the status of large construction projects. This is very much the pattern for CRUK/MRC/Wellcome reviews of research institutes (an independent visiting expert subcommittee examines past progress and future proposals). The novel aspects of the InterAction collaboration are that a peer group of research organisations organises the review at the request of the institute, and the extent to which the terms of the review are coproduced with the member laboratory. It should be noted that the review is limited to communication activities.

Networks of research institutes to tackle operational/management in addition to/rather than scientific problems, such as the InterAction collaboration are not the norm, and the only life sciences network for research institutes that we are aware of is the EU-LIFE project [12]. EU-LIFE is a network of 14 research institutes across Europe that helps share good practice in the operation of core facilities, gender equality, open science, grants and funding strategies, technology transfer and other areas. Representatives from EU-LIFE participated in the discussion forum, and EU-LIFE does support a working group on indicators and publications that includes an annual internal benchmarking exercise.

**Frameworks for assessment.** Much of the research assessment of institutes is qualitative, expert judgement. Respondents highlighted that a rigorous approach can still be taken with qualitative research. While the core principles of rigour in qualitative research are debated, investigators aim for reliable and valid results by reducing sources of bias, increasing the diversity of views sought [27] and triangulating the results using independent sources of evidence. Quantitative data may support expert judgement, by highlighting areas to be examined in detail from the outset (metrics and indicators have been referred to as facilitators for more nuanced evaluative enquiry [21]) and contribute to triangulating views expressed as part of the assessment. Several assessment frameworks were highlighted that aim to capturing more diverse aspects of research, involve a mix of qualitative and quantitative approaches, and thereby reach a more holistic view of research performance. These approaches are discussed in more detail in section 2, however two aspects were highlighted in this section, i) self-evaluation and ii) rubrics.

**Self-evaluation.** The Strategy Evaluation Protocol (SEP – a joint effort of KNAW, VSNU and NWO in the Netherlands) [30] stimulated some discussion on the forum. Although the SEP was drafted for the evaluation of research groups and departments of universities respondents thought it could also be applied to research institutes. The SEP is meant to assess groups against their own strategic goals, not against a fixed

set of external criteria and to diminish the role of quantitative metrics. It still relies on an independent expert assessment committee to consider progress and future proposals, but the guidance for the process has received praise for its simplicity [35]. There is little experience of the impact of the recent update to SEP, but this will grow in the coming years. It offers the opportunity for tailoring evaluations and ensures that the objectives of the review are very clear, but respondents thought it could have the pitfall of groups/institutes setting unambitious strategies and not realising the potential that they would have achieved when challenged a bit more. Respondents questioned whether self-evaluation could provide enough accountability or credible assessment of value for money. In SEP this challenging role is with the board of the university, for institutes it would depend on the units evaluated, the governance structure of the institute, and the role that the sponsors took in agreeing the criteria for the review. MRC/CRUK and Wellcome all involve their institutes in agreeing the terms of reference and approach for their quinquennial reviews, but the SEP discussion may highlight that there could be opportunity for greater codevelopment of the assessment, not only the terms of reference, but the detailed assessment approach. A greater focus on self-assessment, and therefore the specific mission of an institute, may also diminish the interest in making comparisons between institutes.

**The use of rubrics.** Rubrics usually consist of agreed criteria and standards, which allow ratings to be assigned to complex qualitative activities. An example of a rigorously evaluated framework that included the use of rubrics was the 'RQ+ Assessment Framework' [33] launched by the International Development Research Centre (IDRC) in Ottawa, Canada, in 2012. The IDRC primarily funds and facilitates global south-based research for development. Research assessment at the IDRC has to accommodate the heterogeneity of its programmes, both in terms of the diversity of issues such as technology for food security, global health, climate change and inclusive economic growth, and in terms of the types of activities it supports, such as research, capacity strengthening, promoting use of research for policy and practice, constituency and network building, and its partnering as a peer, mentor or broker.

The RQ+ framework has three main elements i) key influences – contextual factors that may impact on research quality, which may include research capacity, political environment, etc., ii) research quality – dimensions include research design (integrity), importance, positioning for use, iii) evaluative rubrics – a set of criteria for each dimension in the first two elements is set out, so that projects can be rated. Once ratings have been produced these are aggregated in a way appropriate to the specific evaluation.

The use of rubrics for evaluation was also raised in discussion with the National Research Council (NRC) of Canada. The NRC supports 14 national research centres and has an independent audit and evaluation team responsible for the quinquennial review of these. The logic model used for the review of the National Institute of Nanotechnology [36] was included as an example in section 2. The NRC had piloted the use of rubrics [34] to:

- Deliver a clear verdict on programme performance based on well-defined performance levels and criteria.
- Create a shared understanding between programme managers and the evaluators of the criteria for each performance level.

This pilot found that development of rubrics was not as straightforward as expected and that the rubrics developed missed the interaction between different aspects of research centre work. This meant the process failed to support the drafting of an evaluation report. Although rubrics from this pilot exercise were not used in the final evaluation, the team learnt a lot from the work and plan to apply a modified approach in subsequent evaluations.

**Indicators and metrics.** Professor Leyser’s introductory remarks had highlighted that while we are all keen to find ways to measure success quantitatively, to ensure that we can track progress and assess the impact of strategic decisions, we should avoid the use of simple proxy measures in assessing research performance. Studies have highlighted that metrics can be useful in the context of evidencing narrative statements [18–19], but they cannot reliably be used as stand-alone performance measures. Many papers have been written on potential measures that could be chosen to provide insight into particular activities, e.g. in 2016 RAND working with the Association of American Medical Colleges collated an ideas book containing 100 metrics to assess and communicate biomedical research [42]. Some proposals for more regular and standardised measurement have gained traction, e.g.

- In 2019 the European Strategy Forum on Research Infrastructures (ESFRI) adopted a common approach across its infrastructures to monitor their performance based on KPIs. The proposed KPIs ranged from input to outcome indicators and were tested against the RACER criteria.<sup>45</sup>
- The Snowball Metrics ‘recipe book’ [43] provides a set of metric definitions agreed between research intensive universities for benchmarking.
- The Researchfish® [44] system common question set developed between 2007 and the present day, and used by UKRI, CRUK and Wellcome to collect annual feedback from all their investments, represents a consensus definition of a wide range of research outputs.

All these approaches emphasise that it is better to use independent facts to triangulate expert judgement, rather than rely on a single indicator. The use of a range of measures was emphasised by the head of the relaunched institute for scientific information at Clarivate in his 2019 paper subtitled ‘profiles not metrics’ [24].

Where performance measures/indicators/metrics are used to support, not supplant, expert peer review there are several important questions to be addressed:

1. Do the metrics correlate sufficiently with the desired performance criteria (e.g. quality)? Many relatively easy to collect metrics measure quantity and not quality. Furthermore, it is not clear what the best quantity is for each measure. More is certainly not uniformly better for these metrics. Fewer, but of better quality is often preferable, or fewer to ensure time for other activities might be better.
2. What are the likely consequences on researcher behaviours of adopting the proposed metrics? While some metrics used at institutional level are reasonable, it is clear from experience that it is extremely difficult to separate institutional from individual level incentives. This is at least as important in considering the things not measured as it is in considering the measured list. For example, if the papers that an

---

<sup>45</sup> RACER criteria, i.e. they had to be Relevant, Accepted, Credible, Easy to monitor, and Robust.

institution publishes in specific journals are counted, this will devalue time invested in making negative results and datasets openly accessible. Although funders insist that they value these other aspects, if they are excluded from the assessment, the perception will be that they are not important. The same applies to important activities such as participating in the peer review process and other services to the research community.

3. What are the costs of collecting the data and are they appropriately compensated by the benefits? The response to concerns about metrics not measuring everything that matters is often to go on adding more and more things to the list to be collected. This misses the point and simply adds administrative burden. As with all public spending, assessment systems must be good value for money.

Respondents highlighted that they wanted their 'evaluations to demand data'. Too often assessments began by cataloguing all the data available, rather than thinking hard about what the critical evidence might be to support the review.

Respondents also suggested that some of the more difficult to measure aspects of progress might be captured via funder involvement in institute advisory and governance boards (so that discussion of progress could be codified and reused) and institute staff surveys (to act as a barometer of research culture). Respondents also highlighted the crucial importance of the visiting expert subcommittee meeting.

**Researcher track records – CVs and biosketches.** Respondents highlighted the usefulness of CVs to list researcher achievements, capturing key input/output/outcome/impact information. These documents would benefit from i) emphasising a greater diversity of contributions from researchers, ii) following a more standardised and structured format that could be updated easily and reused for institute reviews, promotion panels, grant applications, etc. There are several initiatives trying to address exactly this issue that respondents highlighted that we could learn from.

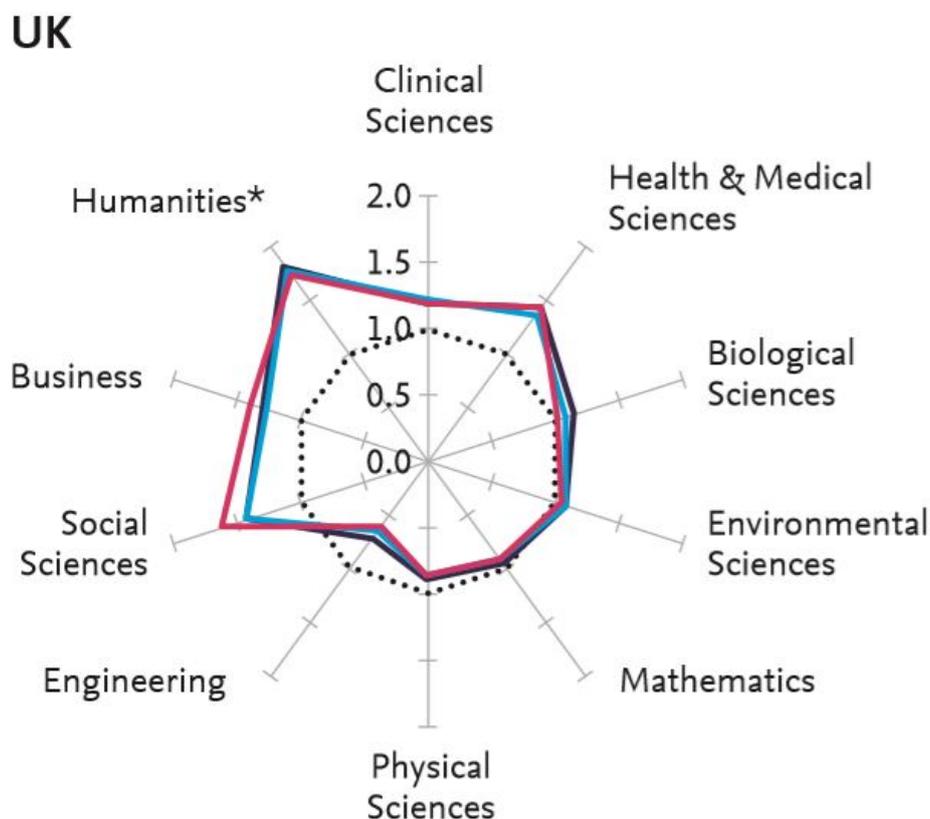
- **SciCV** [45] is a pilot project of the Swiss National Science Foundation (SNSF) to test a new CV format. The aim is to provide a way for researchers to compile their CV in a structured way (reusing data in systems such as ORCID) and to present their most important contributions to science in brief narratives, rather than only as lists of publications. This approach will help make other academic outputs, beyond publications, visible and valued and promote equal opportunities. It was mandatory for applications in the project funding call in biology and medicine in spring 2020 and the results are currently being evaluated. It may be helpful to discuss with colleagues at SNSF when the results of this evaluation are available.
- **PubMed Science Experts Network Curriculum Vitae (SciENCv)** [46] is an openly accessible electronic system built by the NCBI that helps researchers assemble the professional information needed for participation in federally funded research (e.g. US Department of Education, National Science Foundation and National Institute of Health applications), such as compiling NIH biosketches [47]. It gathers and compiles information on expertise, employment, education and professional accomplishments. Researchers can use SciENCv to create and maintain biosketches that are submitted with grant applications and annual reports. SciENCv allows researchers to describe and highlight their scientific contributions in their own words.

- **Researchfish\* CV builder.** The Researchfish system has since its inception included the ability to fill in a personal profile, attribute outputs already entered to a CV file and format this. The limitation has been working to an internationally agreed format for a CV, but funders should examine if this could be a route for further reuse of harvested and manually input information.
- **Royal Society resumé for researchers** [48]. The Royal Society has provided guidance for the narrative sections of a researcher CV. However, the concept is at an early stage of development.

**Bibliometrics.** Publication data is the most widely analysed research output, and the analysis of citations is the most widely used quantitative approach. For example the citations of publications submitted to the REF2021 exercise will be analysed and considered alongside other evidence by 11 of the 34 expert subpanels [28]. The principles of citation analysis and the responsible use of this data has been extensively reviewed elsewhere [29].

The International Comparative Performance of the UK Research Base 2016 report [49] prepared by Elsevier for the UK's Department of Business, Energy & Industrial Strategy (BEIS) was used as an example of bibliometric analysis. This report includes an assessment of the performance of the UK's research base compared with comparator nations using a range of bibliometric indicators.

The radar plot in Figure 3 below is of an activity index (a country's share of its total article output across different subject fields relative to the global share of articles in the same subject fields). Three years of data are plotted for the UK (2006 black, 2010 blue, 2014 red), and the world average of 1 is indicated by the dotted line. From this plot you can see that the UK has a greater than world average focus on social sciences, business, humanities, clinical sciences and health and medical sciences.

Figure 3. Radar Plot of Activity Index<sup>46</sup>

The Elsevier report illustrated the continuing importance of bibliometric analyses to research evaluation, and also one of very many ways to graph results. Analysis of the bibliographic information contained in journal articles allows the connection of information about funding, people, locations and subject matter. This may be used to study collaboration, changes in effort devoted to different research fields or the involvement of different funding sectors or research organisations (private, charity, hospital, etc.). The analysis of the citations these papers receive allows investigation of the attention that they secure in other papers, patents or policy documents. This may be used to track the diffusion of knowledge or its translation into intellectual property/new practice.

Respondents noted that there are aspects other than citation impact, that can bring valuable insight, including various aspects of network analyses and text mining of the content of papers.

The MRC shared an approach used in a recent portfolio review of research centre and unit investments. Expert panel discussion was supported with a dashboard that provided basic details of institute/unit mission statements, their expenditure and sources of funding, and some analysis of their publication output. As part of the dashboard we used topic clusters from Scopus [50] to provide a visualisation of the institute/unit portfolio. Importantly we used publication content as a rich source of data on the subject matter being researched; we were not analysing citations at all here. We are in the process of gathering feedback from experts involved in the review to determine how useful they found it.

<sup>46</sup> Taken from International Comparative Performance of the UK Research Base, 2016 [49]

Respondents again emphasised the triangulating power of independent sources of information, suggesting that bibliometric analysis can provide a starting point for review panels to understand priorities, research strategies and goals. This helps directors to reflect on whether they are moving in the right direction, or to spot new opportunities.

**Trends over time.** Respondents highlighted the importance of time. Do the data shed light on the delivery of the strategy, have things got better or worse over time, and are there changes that can be shown to be subsequent to actions taken by the institute?

Colleagues at the Francis Crick Institute explained that they had established a new data warehouse at the institute to hold all their administrative data ('Crick Data World'). The warehouse holds data from many different systems and institute process sources, and so it evolves and develops continuously. Having this kind of relational data means that it is possible to quickly examine where there is evidence of researchers interacting within the Crick and with external collaborators, where are the areas of growth in funding, and this could potentially be used to spot opportunities in future.

**Impact pathways – the link from logic models to estimating societal and economic impact.** Colleagues in Science and Technology Facilities Council (STFC) highlighted that there are challenges in demonstrating the impact of large, distributed infrastructures, but that several EU-funded projects were in the latter stages of projects to address this:

RI-PATHS – an EU funded project [16] in which the European Organization for Nuclear Research (CERN), ALBA, German Research centre for High Energy Physics (DESY) and ELIXIR participate. The aim of the project is to produce tools to better assess the impact of research infrastructures on the economy and society, with a focus on the long-term impact pathways of various infrastructures. Pilot projects have been undertaken across the participating centres and these have shown that infrastructures impact society in more ways than is usually recognised in procurement exercises, that context is important, and that capturing feedback from users is crucial to understanding impact. The project has developed a cost-benefit analysis approach that includes knowledge outputs, human capital, technological spillovers, information technology externalities, services provided to third parties and outreach activities [51].

Another EU project with research infrastructure participants is ACCELERATE. The Rathenau Institute, an ACCELERATE project partner, has drafted a societal impact protocol, which has been used to guide societal impact assessments of other partners (CERIC-ERIC, ELI-DC, European Spallation Source-ERIC, FRM II and HZG) [23].

The cost-benefit analysis approach from the RI-PATHS project has also been applied to the impact approach of the European Research Infrastructure for Heritage Science (E-RIHS), which is expected to become an ERIC<sup>47</sup> in 2023. The approach includes estimation of a willingness of potential users to pay, and a qualitative discussion of wider long-term impacts and their pathways. This mixed methods approach has been chosen to better describe the large variety of and multi-faceted potential impacts expected by E-RIHS ERIC.

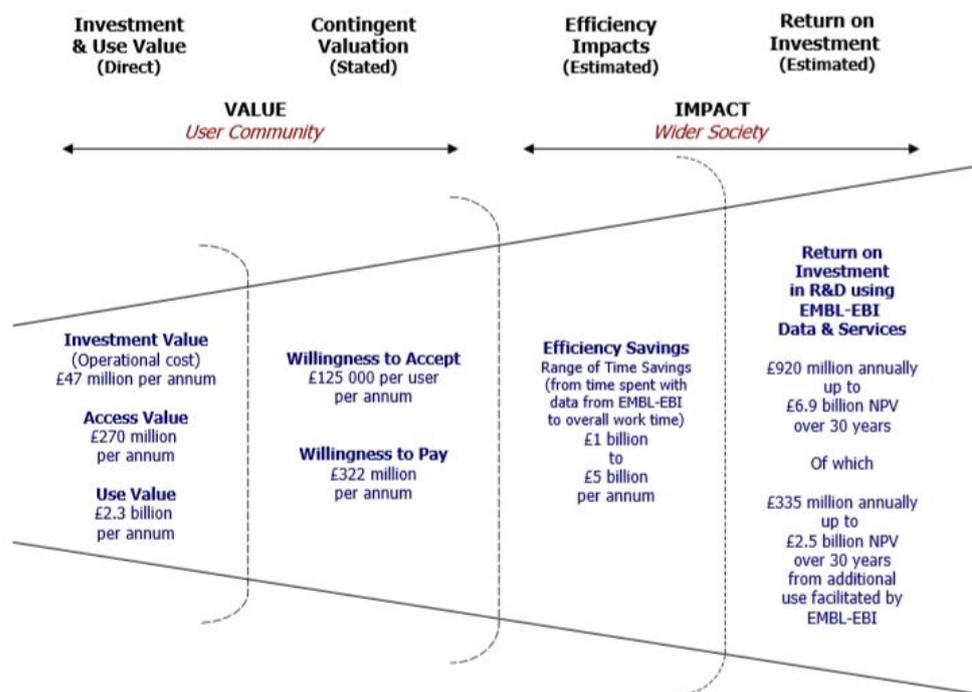
---

<sup>47</sup> The European Research Infrastructure Consortium (ERIC) is a specific legal form that facilitates the establishment and operation of Research Infrastructures with European interest.

The forum material included an example of an impact analysis:

The European Bioinformatics Institute (EBI) commissioned an economic impact study [52] that included a survey of over 4000 data-service users, 45 per cent of whom indicated that they could neither have created nor collected themselves the last data they used, nor obtained it elsewhere. Figure 4 below summarises the econometric findings, which emphasise the vital role of public databases in life-science research and indicate that for every million pounds invested in EMBL-EBI, roughly £20m is returned to the global economy.

Figure 4. Econometric findings from EBI’s economic impact study [52]



The EBI study illustrated several interesting methodological aspects. Firstly, the increasing role of surveys and detailed interviews to capture new quantitative and qualitative information from stakeholders. Secondly, the importance of econometric models to estimate the return on investment from research.

**Identifying good governance.** Back to Professor Leyser’s introductory remarks. ‘There are strong parallels to wider questions of accountability and trust, as eloquently set out by Onora O’Neil in her Reith Lectures [53]. To quote her “Intelligent accountability, I suspect, requires more attention to good governance and fewer fantasies about total control. Good governance is possible only if institutions are allowed some margin for self-governance of a form appropriate to their particular tasks, within a framework of financial and other reporting. Such reporting, I believe, is not improved by being wholly standardised or relentlessly detailed, and since much that has to be accounted for is not easily measured it cannot be boiled down to a set of stock performance indicators.” “Serious and effective accountability, I believe, needs to concentrate on good governance.” Careful consideration is warranted for assessment of the governance arrangements for delivering excellence against an appropriately broad set of criteria informed by core values.’

One director raised the issue of the frequency of reviews. It is notable that the REF exercises occur every six to seven years, that investigators are often funded for six years at a time (e.g. MRC investigators are typically six years in duration, Crick junior group leaders are appointed to 6+6 year positions), however institutes are

usually reviewed every five years. It is suggested that, subject to assurance concerning good governance, that a longer period between reviews could be considered.

With the creation of UKRI, there is an opportunity to consider the whole UK research system and to ensure that it is delivering what the UK needs to thrive as an inclusive knowledge economy. The entirely admirable drive to accountability and quality assurance in funding allocation must be deeply thought through in this context.

## Annex B. List of invitees to the discussion forum

---

150 researchers and representatives of research organisations and funders agreed to be invited to the discussion forum. These are listed below. 75 logged into the forum during the week and 38 added comments.

Invitee		Organisation
Marta	Agostinho	EU-LIFE alliance
Abiola	Aiyenigba	Centre for Capacity Research, Liverpool School of Tropical Medicine
Erik	Arnold	Technopolis group
Gert	Balling	Novo Nordisk Foundation, Denmark
Mary	Barlow	European Bioinformatics Institute, Cambridge
Peter	Barnsley	Francis Crick Institute, London
Henrik	Barslund Fosse	Novo Nordisk Foundation, Denmark
David	Bates	University of Nottingham
Jackie	Beesley	CRUK Beatson Institute, Glasgow
Michela	Bertero	Centre for Genomic Regulation, Spain
Veronique	Birault	Francis Crick Institute, London
Ewan	Birney	European Bioinformatics Institute, Cambridge
Kate	Bishop	Francis Crick Institute, London
Ekin	Bolukbasi	Wellcome, London
Frédérique	Bone	Sussex Policy Research Unit, Sussex University
Simon	Boulton	Francis Crick Institute, London
Sarion	Bowers	Wellcome Sanger Institute, Cambridge
Joshua	Brickman	University of Copenhagen
Martin	Bushell	CRUK Beatson Institute, Glasgow
Caroline	Cake	Health Data Research UK
Doreen	Cantrell	University of Dundee
Valerie	Conn	Science Philanthropy Alliance
Anne-Marie	Coriat	Wellcome, London
Henry	Cormack	UK Research and Innovation (UKRI)
Clemencia	Cosentino	US National Science Foundation
Susie	Crossman	Centre for Capacity Research, Liverpool School of Tropical Medicine
Rachel	Curran	European Bioinformatics Institute, Cambridge
Alexandra	Dagger	National Research Council Canada

Invitee		Organisation
Maggie	Dallman	Imperial College London
Anoushka	Dave	Technopolis group
Mary	De Silva	Wellcome, London
Bart	de Strooper	UK Dementia Research Institute
Franziska	Denk	King's College London
Hana	Dlouha	European Molecular Biology Laboratory
Kevin	Dolby	Medical Research Council (MRC)
Claire	Donovan	Brunel University
Megan	Dowie	Medical Research Council (MRC)
Jeanette	Dyson	European Molecular Biology Laboratory
Jacob	Edler	Fraunhofer Institute for Systems and Innovation Research
Amanda	Fisher	MRC London Institute for Medical Sciences
James	Fleming	Francis Crick Institute, London
Iain	Foulkes	Cancer Research UK (CRUK)
Elizabeth	Gadd	Loughborough University
Kathleen	Gerber	National Research Council Canada
Fiona	Goff	Natural Environment Research Council (NERC)
Jorge	Gomez Magenti	Alzheimer's Research UK
Jef	Grainger	Biotechnology and Biological Sciences Research Council (BBSRC)
Jonathan	Grant	King's College London
Petra	Hajkova	MRC London Institute for Medical Sciences
Bill	Hansson	Max Planck Society
Phillip	Hawkins	Babraham Institute, Cambridge
Adrian	Hayday	Francis Crick Institute, London
Caitlin	Higgott	Rosalind Franklin Institute
Caroline	Hill	Francis Crick Institute, London
Steve	Hill	Nottingham University
Steven	Hill	Research England
Laura	Himanen	International Network of Research Management Societies (INORMS)
Dan	Hodges	Innovate UK
Laura	Holland	Rosalind Franklin Institute
Chris	Holmes	University of Oxford
Michael	Hopkins	Sussex Policy Research Unit, Sussex University

Invitee		Organisation
Hazel	Housden	Rosalind Franklin Institute
Danielle	Hoyle	Babraham Institute, Cambridge
Simon	Hubbard	University of Manchester
Gareth	Inman	CRUK Beatson Institute, Glasgow
Robert	Insall	CRUK Beatson Institute, Glasgow
Adrian	Iverson	UK Dementia Research Institute
Austra	Jenner Parsons	Wellcome Sanger Institute, Cambridge
Chonnettia	Jones	Michael Smith Foundation for Health Research
Adam	Kamenetzky	National Institute for Health Research (NIHR)
Tim	Killick	Wellcome, London
Angela	Kingman	Imperial College, London
Giovanna	Lalli	UK Dementia Research Institute
Fiona	Larner	Science and Technology Facilities Council (STFC)
Nick	Lemoine	Queen Mary University London
Ottoline	Leyser	Sainsbury Lab, Cambridge University
David	Lomas	University College London
Robin	Lovell-Badge	Francis Crick Institute, London
Chris	Lowe	University of Cambridge
Jane	Luff	National Institute for Health Research (NIHR)
Laura	Machesky	CRUK Beatson Institute, Glasgow
Alison	Maloney	Francis Crick Institute, London
Corrine	Martin	ELIXIR
Ian	McArdle	Imperial College, London
Rowan	McKibbin	Biotechnology and Biological Sciences Research Council (BBSRC)
Matthias	Merkenschlager	MRC London Institute for Medical Sciences
Irene	Miguel-Aliaga	MRC London Institute for Medical Sciences
Jennifer	Moloney	Engineering and Physical Sciences Research Council (EPSRC)
Kevin	Moreton	MRC Laboratory of Molecular Biology (LMB), Cambridge
Effie	Mutasa-Gottgens	European Bioinformatics Institute, Cambridge
Paul	Nightingale	Economic and Social Research Council (ESRC)
Anke	Nooijen	Technopolis group
Jim	Norman	CRUK Beatson Institute, Glasgow
Frank	Norman	Francis Crick Institute, London

Invitee		Organisation
Paul	Nurse	Francis Crick Institute, London
Sarah	Parks	UK Research and Innovation
Rickie	Patani	Francis Crick Institute, London
Jonathon	Pearce	Medical Research Council (MRC)
Alex	Pemberton	Cancer Research UK (CRUK)
Lindsey	Pike	University of Bristol
Alexandra	Pollitt	Kings College London
Erica	Pufall	Wellcome, London
Robert	Pyke	Babraham Institute, Cambridge
Fiona	Reddington	Accelerating Detection of Disease (ADD)
Paul	Reeves	Biotechnology and Biological Sciences Research Council (BBSRC)
Caetano	Reis e Sousa	Francis Crick Institute, London
Emily	Richardson	Wellcome, London
Nathan	Richardson	Medical Research Council (MRC)
Anne	River	National Institute for Health Research (NIHR)
Erik	Sahai	Francis Crick Institute, London
Owen	Sansom	CRUK Beatson Institute, Glasgow
George	Santangelo	US National Institutes of Health (NIH)
Andreas	Schaefer	Francis Crick Institute, London
Susan	Simon	Medical Research Council (MRC)
Emma	Sinha	European Molecular Biology Laboratory
Cheryl	Smyth	Babraham Institute, Cambridge
Tony	Soteriou	UK Research and Innovation (UKRI)
Mark	Stevens	Roslin Institute, University of Edinburgh
Bruce	Stillman	Cold Spring Harbor Laboratory, USA
Gitta	Stockinger	Francis Crick Institute, London
Jonathan	Stoye	Francis Crick Institute, London
Tanja	Strom	International Network of Research Management Societies (INORMS)
Carole	Swan	MRC London Institute for Medical Sciences
Elly	Tanaka	Research Institute of Molecular Pathology (IMP), Vienna
Mike	Thelwall	University of Wolverhampton
Beverley	Thomas	Biotechnology and Biological Sciences Research Council (BBSRC)
Sarah	Thomas	National Institute for Health Research (NIHR)

Invitee		Organisation
Fiona	Tomley	Royal Veterinary College, London
Richard	Treisman	Francis Crick Institute, London
Ghislaine	Tremblay	National Research Council Canada
Victor	Tybulewicz	Francis Crick Institute, London
Janna	Van Belle	Technopolis group
Geert	van der Veen	Technopolis group
Henri	van Luenen	Netherlands Cancer Institute
Marleen	Vanstraelen	VIB, Flanders
Peter	Varnai	Technopolis group
Judith	Vermeer	Technopolis group
Niki	Vermeulen	University of Edinburgh
Eileen	Wall	Scotland's Rural College (SRUC), Edinburgh
Ludo	Waltmann	Centre for Science and Technology Studies (CWTS), Leiden
Bruce	Whitelaw	Roslin Institute, University of Edinburgh
Caroline	Wilkinson	CRUK Manchester Institute
Anthony	Wilson	Food Standards Agency
Dominic	Withers	MRC London Institute for Medical Sciences
Karen	Woodey	CRUK Cambridge Institute
Aaron	Yarmoshuk	Centre for Capacity Research, Liverpool School of Tropical Medicine
Sana	Zakaria	National Institute for Health Research (NIHR)