



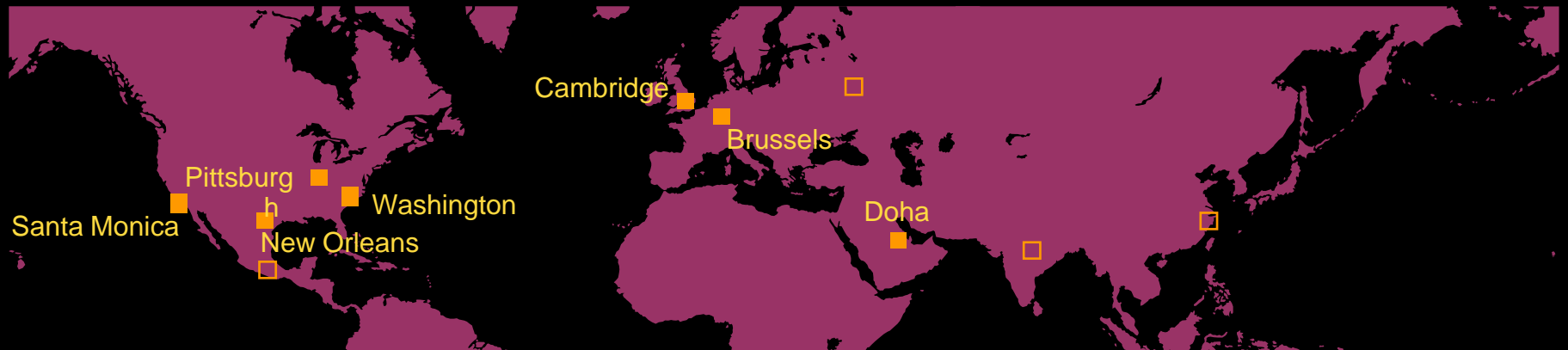
EUROPE

Capturing Research Impacts: a review of international frameworks

Sonja Marjanovic
IEA Meeting
November 2010
Brussels

RAND Europe

- Not-for-profit policy research organisation, part of the globally operating RAND Corporation
- Conduct objective research and analysis to improve policy-making for public benefit
- Multi-disciplinary
 - Innovation and Technology Policy, ICT, Governance and Regulation, Criminal Justice, Education, Population and Migration, Defence and Security, Transport, Drugs and Alcohol
- Body of research evaluation and science policy work for diverse clients: government and government agencies, charities, research councils nationally and internationally
- Combining academic rigour with a client-based approach



Outline

- Background to the review of international impact assessment frameworks, and approach
- Case-studies of four impact assessment frameworks
- In reflection: Key themes and messages

Outline

- Background to the review of international impact assessment frameworks, and approach
- Case-studies of four impact assessment frameworks
- In reflection: Key themes and messages

HEFCE required information to inform the development of the UK Research Excellence Framework (REF)

- Higher Education Funding Council for England (HEFCE) distributes quality-related public funding for teaching and research to universities and colleges in England. It allocates funding as a block grant to institutions, informed by the results of assessments.
- The REF is the new UK research assessment system which is to replace the Research Assessment Exercise (RAE) by 2014.
- The REF will:
 - inform the selective allocation of research funding to HEIs by funding bodies
 - provide benchmarking information and establish reputational yardsticks
 - provide accountability for public investment in research and demonstrate its benefits.
- REF process will be based on expert review of submissions by universities, informed by qualitative and quantitative indicators where appropriate.
- Institutions be assessed in terms of:
 - The quality of research outputs (bibliometrics and peer review)
 - The wider impact of research (pilots of potential methods taking place)
 - The vitality of the research environment

HEFCE felt there were a number of criteria which the REF framework would need to meet

- The framework needs to be:
 - be credible and acceptable to the academic as well as user communities
 - encompass a comprehensive range of benefits: economic, social, public policy, welfare, cultural and quality of- life benefits
 - within a single broad approach, be adaptable to apply to all disciplines
 - be practicable and not generate an excessive workload for the sector
 - avoid undesirable perceptions and incentives
 - complement other funding streams' approaches to increasing research impact
- HEFCE asked RAND to provide a review of the most relevant frameworks
 - Relevance was determined in consultations with HEFCE, and following an initial review of a large number of frameworks
 - We also felt the review should cover frameworks with different methodological approaches

Relevance was determined in consultations with HEFCE, based on their criteria and following an initial review of a large number of frameworks

- Australia – Research Quality Framework (RQF)
- Australia – Excellence in Research for Australia (ERA)
- Australia – Measurement of Research Impact and Achievement (MORIA)
- UK – Department of Innovation, Universities and Skills (DIUS) Economic Impacts
- UK – Library House Model
- UK – Arthritis Research Council (ARC) scoring
- US – Congressionally Directed Medical Research Program (CDMRP)
- US – Program Assessment Rating Tool (PART)
- Netherlands – Standard Evaluation Protocol (SEP)
- Netherlands – Leiden University Medical Centre (LUMC)
- Netherlands – Evaluating Research in Context (ERiC)
- Sweden – Swedish Governmental Agency for Innovation Systems (VINNOVA)
- Canada – Payback
- Japan – National Institute for Academic Degrees and University Education (NIAD-UE).

- scholarly and grey literature; interviews, workshops

Outline

- Background to the review of international impact assessment frameworks, and approach
- Case-studies of four impact assessment frameworks
- In reflection: Key themes and messages

THE CASE STUDY APPROACH WHICH ENGAGES DIVERSE STAKEHOLDERS



The Australian Research Quality Framework (RQF): Background and principles

- Proposed by previous government (2004) to measure:
 - the quality of all taxpayer funded research
 - the impact (economic, social, environmental and cultural benefits to end users and wider society regionally, nationally and internationally)
 - the quality of post-graduate research education and training
 - examine accessibility issues: infrastructure, ICT, libraries, results
- The RQF was by definition inequitable:
 - rewards excellence in research quality and impact
 - high impact institutions would receive proportionally more funding
 - weighting for impact element of assessment in the final funding formula was not been decided on (but expectation was app. 10%)
 - (HEIs would retain discretion on internal distribution of RQF driven funding allocations)
- The key units of assessment were to be ‘research groupings’ which can be virtual and are decided on by HEIs themselves (min size of 5 staff):
- Abandoned in 2007 due to political reasons, being replaced by efforts to develop a more metrics-based evaluation system (Excellence in Research for Australia)

The Australian Research Quality Framework (RQF): Methods 1

- **Case study approach:** research groupings submit examples of high- impact research: with qualitative and quantitative evidence
- **Context statement:**
 - a word-limited strategic case about the direction, focus and nature of research for a grouping, and how it relates to impact
- **Impact statement:**
 - **addresses four questions** at the level of a research grouping:
 - On stakeholder engagement to address social, economic, environmental and cultural issues
 - On new products, policies, legislation, paradigm, attitudes that have been implemented and adopted by end-users
 - One associated social, economic, environmental and/or cultural benefits from the above (outcomes)
 - On the magnitude of these benefits
 - **Is accompanied by up to four case studies** that illustrate and support the claims made, and **a list of end-users** which can be contacted for validation of evidence and claims made
 - Impact must have occurred during the 6 year assessment period, but could be informed by research conducted in the prior assessment period

The Australian Research Quality Framework (RQF): Methods 2

- A working group on impact assessment (appointed by the government) also provides HEIs with **a list of potential indicators they can draw on** to assist in making their statements, to help provide verifiable evidence:
 - Only as guidance, not prescriptive
 - Defining a list of all acceptable indicators could disenfranchise disciplines and fields as well as lead to collection of irrelevant data - so was not done
- **Expert Assessment Panels review and appraise entire portfolios submitted and produce ratings of research quality and ratings of impact.**
 - Panels of 3 international experts plus :
 - 3 discipline experts to rate research quality
 - and additional 3 end-users to be involved in rating impacts
- **High stakeholder involvement** facilitates credibility and acceptance by academic and end-user communities
 - Bottom- up approach: evaluatees provide the case study examples of impact
 - End users form part of the Expert Assessment Panels.
 - They could also act as 'auditors' for Research Groupings.

THE INDICATOR APPROACH THAT IS OF MINIMAL BURDEN



The UK RAND/ARC Impact Scoring System (RAISS): Background and principles

- The Arthritis Research Campaign (4th largest medical research charity in the UK) wanted to improve end of grant reporting and track impacts of research over time.
- The wanted a system that:
 - could be applied across all their grants to track outputs and impacts
 - was less burdensome than end of grant reports and easier to analyse information from
 - could provide an overview of their entire portfolio to inform portfolio-level strategy, influence future funding decisions and form a foundation for more detailed evaluations
- ARC implemented system in 2008 and it has replaced end-of-grant reports; MRC used it inform development of their own tool; NIHR is implementing the approach; other bodies interested in testing it (e.g. Canadian Institute for Musculoskeletal Health and Arthritis)
- As it stands, the framework was developed for a biomedical research context specifically, so it would need to be tailored to the specifics of different fields.

RAISS methods 1

- A web-based survey for impact mapping
 - Consists of 187 yes/no questions; 90% of investigators complete survey in less than 1 hour
 - Original pilot had 87% response rate and positive feedback from funder and researchers
- The questions cover different impact categories:
 - what further research has developed from the grant
 - what research tools have been developed
 - what dissemination activities have been carried out
 - what impacts on product development and health policy have occurred
 - how the grant has affected education
 - what health interventions / changes in service delivery or health advice produced
- System depends on grant-holder accuracy, recollection and honesty
 - No external audit (i.e. end-user validation)
 - Impacts pre-specified by funder (but the process of identifying impacts could involve recipients and end-users in principle)
- By analysing the data across the portfolio of research ARC funds, the tool allows ARC to understand the distribution and range of impacts arising from its research portfolio to inform strategic decision making

RAISS methods: the process and questionnaire

Researcher completes grant

Request to complete questionnaire emailed to

Researcher completes questionnaire

ARC analyses results

5.4: Dissemination - Other Dissemination

98. Dissemination to health professionals?

Yes No Not Known

If yes, has this included (If no, go to question 5):

99. Websites for health professionals?

Yes No Not Known

100. Printed material (such as booklets) for health professionals?

Yes No Not Known

101. If yes, what was the distribution of this printed material? If no, go to question 102.

	Yes
Local	<input type="radio"/>
National	<input type="radio"/>

102. Dissemination to patients/public?

Yes No Not Known

If yes, has this included (If no, question 9):

103. Websites?

Yes No Not Known

104. Printed material (such as booklets) for patients/public?

Page 40

8.1: Interventions/Products - IP

140. For the research cited by patents, or intellectual property protection, which stages were carried out during the grant, or have happened since?

	Yes	No	Not Known
Cited in a patent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Patent applied for/filed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Patent granted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Patent licensed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Page 41

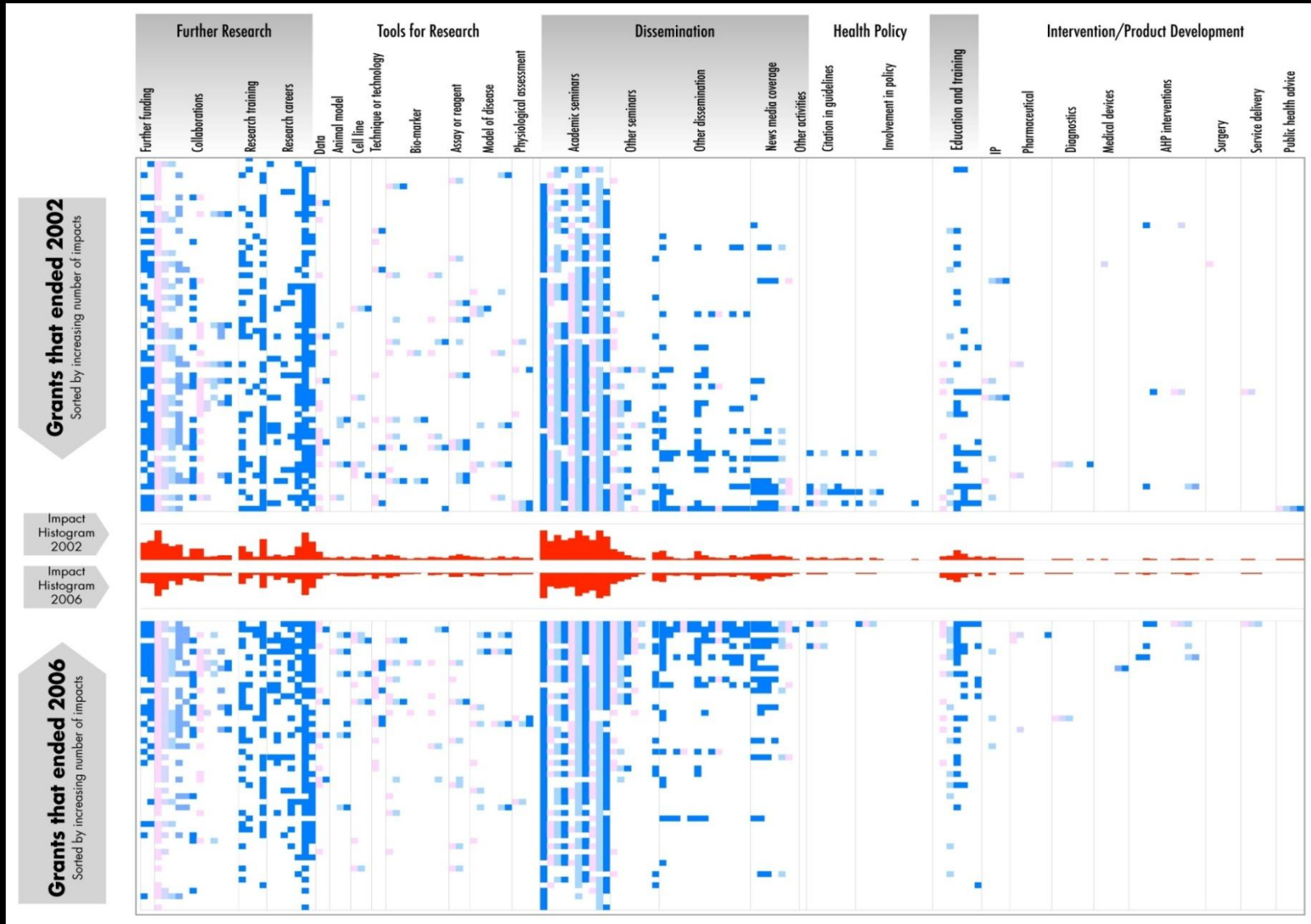
8.2: Interventions/Products - Pharmaceutical

141. For the therapeutic pharmaceutical product, which stages were carried out during the grant, or have happened since?

	Yes	No	Not Known
animal tests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phase 1 clinical trial	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phase 2 clinical trial	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phase 3 clinical trial	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
on the market	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
trials a current pharmaceutical for a new indication or with a new regime	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

RAISS outputs

- The display is in the form of impact arrays with indicators for different categories of payback
- Shading of levels of impact: early to late stage impacts; local to international impact



**THE SELF-EVALUATION DRIVEN APPROACH WHICH
CAN BE APPLIED TO DIVERSE FEDERAL
PROGRAMMES (NOT ONLY RESEARCH)**



US Program Assessment Rating Tool (PART): Background and principles

- PART's roots can be viewed as evolving out of the Government Performance Results Act (GPRA) of 1993, which mandated the development of a system for assessing performance of all government programmes
- PART was introduced in 2004 (results-oriented performance budgeting) . As of 2008, a circa 1000 federal programmes have been assessed.
- Programmes are assessed at least once every five years.
- The aims of this assessment system are:
 - **to improve government management:** provide decision makers with information they need to better allocate scarce resources
 - **to induce organisational change:** PART-driven funding decisions as an incentive for programme officials to introduce changes needed to improve performance, efficiency and effectiveness
- Developed and used by Office of Management and Budget (OMB) – mixed buy-in from Congress
- President Obama announced he would fundamentally reconfigure PART – establish a new 'performance improvement and analysis framework' that is more transparent and stakeholder inclusive

US Program Assessment Rating Tool (PART):Methods 1

- A diagnostic questionnaire, 25–30 general (Yes/No/NA) questions about each of the following topics:
 - Programme **purpose and design** (Yes/No/NA) (20%)
 - **Strategic planning**, e.g. targets, timelines, performance measures (Yes/No/NA) (10%)
 - **Programme management**, e.g. collection and review of progress by agency, collaboration (Yes/No/NA) (10%)
 - **Programme results**, e.g. achievement of annual goals, progress towards achieving longer-term objectives (Yes/ Large extent/ Small Extent, No) (50%)
- Each section carries a pre-specified weighting. Programme managers can alter weights within each category to emphasize key factors of the programme. To avoid manipulation of the total score, weights must be adjusted prior to responding to Qs.
- Based upon the numerical scores, OMB assigns a management and performance rating to each program, as well as an improvement plan

US Program Assessment Rating Tool (PART): The process

Programme officials provide answers to the questions together with explanations and evidence.

A budget examiner then reviews the materials submitted, and decides which answers to give for each of the questions. (Federal agencies can appeal if they disagree with answers)

On the basis of this OMB assigns an overall performance rating to the programme.

Each programme assessment includes an 'improvement plan' with up to three follow-up actions: management, policy, budget or legislative follow up.

PART used to inform funding decisions but a PART rating does not guarantee a specific level of funding. This can be influenced by factors such as duplication with other programmes, or national priority levels

US Program Assessment Rating Tool (PART):Outputs

- The website ExpectMore.gov is the primary public mechanism for reporting on Federal programme performance and what is being done to improve results.
- Shows well performing and poorly performing federal programmes with ratings of:
 - Effective
 - Moderately effective
 - AdequateOR
 - Ineffective
 - Results not demonstrated



THE CONTEXT- SENSITIVE APPROACH THAT COMBINES METHODOLOGIES



Dutch Evaluating Research in Context (ERiC): Background and principles

- Developed by a collaboration of higher education stakeholders in the Netherlands (ERiC project began in 2006). Adapted from a system developed by network of science policy makers in the Netherlands, Sci_Quest, in the 1990s. To be integrated into current system, Standard Evaluation Protocol (SEP)
- The aim is to address historical emphasis on research quality but neglect of impact.
- Assessment of research:
 - quality (international recognition and innovative potential)
 - productivity (scientific output)
 - relevance (scientific and socio-economic impact)
 - vitality and feasibility (flexibility, management and leadership).
- Key features:
 - Combines qualitative and quantitative data
 - Forward- looking rather than judging
 - Contextual: dependent on programme mission and stakeholder-inclusive
- Outcomes are unlikely to influence funding decisions: rather to guide improvements

Dutch Evaluating Research in Context (ERiC): Method – a four stage process

Stage 1: Self-assessment

- Institute establishes and articulates mission and goals, their view on their impact (e.g. can be done surveys, interviews, workshops etc...)

Stage 2: Data gathering to produce a Research Embedment Performance Profile (REPP): indicators and outputs

- Evaluators work with institute to develop indicators of impact and to map outputs in various categories
- A map of outputs for a research group based on goals and relevant social domains

Stage 3: Stakeholder analysis

- Based on REPP, identify relevant stakeholders (e.g. collaborators, end-users, social organisations, industry etc...)
- Use questionnaire and/or telephone interviews to establish role of stakeholder and consult on and verify research impact as mapped in stage 2.

Stage 4: Feedback and forward look

- Outputs of stages 2 and 3 are compared against stage 1 to illustrate match or mismatch between a group's self perception and its actual impact
- On this basis, future plans are made to build on established strengths and to address weaknesses

ERiC: Methods 2

- Stakeholder engagement:
 - Institute defines its goals and mission
 - System is run by an evaluator who conducts surveys / interviews and gathers REPP indicator and output data
 - End-users respond to survey/interviews
- Pilot studies being conducted across the social and natural sciences (agriculture, pharmacy, architecture, law)
- Also being piloted and developed as part of an collaborative EU project (SCIAMPI)
- To be done every 3 or 6 years?
- REPP outputs can take the form of a radar graph or a table. They could be benchmarked to give a ranking (currently a ++ to – ranking approach is being tested)
- Public dissemination likely to be through reports on institute website (but yet to be defined)

Science, certified knowledge	
relative citation impact	--
productivity scientific publications	++** (1)
international visibility and collaborations	=
representation in editorial boards	++
invited lectures	++
Industry, market	
non-academic/commercial citing environment	++
productivity professional publications	++*(1)
involvement in industry/market	-
advisory and expert roles in commercial domain	--
editorships professional journals	++**
Policy, societal	
involvement in policy domain	+
memberships and expert roles in governmental bodies	++
memberships of societal organisations: advisory/education	++*
production of public goods	+
additional grants from policy	+

Outline

- Background to the review of international impact assessment frameworks, and approach
- Case-studies of four impact assessment frameworks
- In reflection: Key themes and messages

Process of identifying key issues emerging from the case studies

- Stakeholder workshop
- HEFCE representatives and the RAND team worked together to identify and cluster key themes emerging from the case studies (These were themes which were felt to be important for developing the UK REF)
- On the basis of issues analysis workshops, 6 *key* issues with the four case studies could be identified:
 - Strategic intent
 - Definitions of impact (i.e. A common language)
 - Units of analysis
 - Attribution
 - End user roles
 - Burden (Administrative and time demands)

Strategic intent

RQF:

- To provide a consistent and comprehensive approach to assess publically funded research which will ultimately drive the distribution of research resources

RAISS:

- To provide a framework for assessment across an entire portfolio of ARC funding that could replace end of grant reports, make it easier to analyse information, and be used to inform strategic decision making

PART:

- To improve government management through providing evidence needed to better allocate scarce resources and to induce organisational change and learning through improvement plans

ERiC:

- To provide a way of assessing impacts that is sensitive to the unique goals of different disciplines and research units

Defining impact

RQF:

- Economic, social, environmental and cultural benefits to end users and wider society regionally, nationally and internationally

RAISS:

- Impact categories defined by funder, and include influences on further research, new research tools, dissemination, impacts on health policy, impacts on education, changes in health service delivery and health advice produced

PART:

- Impact is defined by evaluatee based on self-articulated programme goals , targets and performance measures

ERiC:

- Impact as context specific and stakeholder-specific . No single definition

Unit of analysis

RQF:

- Self-selected research groupings in HEIs

RAISS:

- All ARC grants

PART:

- Federal programmes of different types

ERiC:

- Research institutes (an organisational unit covering a more or less coherent area of research): assessment at HEI level where a range of different disciplines are included could be difficult except through aggregation of scores gained at department or 'institute' level

Scalability

RQF:

- The entire HEI system in Australia

RAISS:

- Easily scalable in terms of grant numbers but only developed for biomedical science

PART:

- All federal programmes – used across government

ERiC:

- Not known yet, but could be burdensome for first round

Attribution

Australian RQF:

- Explicit about it being a sampling approach

RAISS:

- Relatively simplistic lens on attribution, relies on researcher honesty about 'significant contribution'

PART:

- Programmes have to demonstrate evidence of impact attribution to get credited (otherwise a rating of 'effectiveness not demonstrated' can result)

ERIC:

- Not addressed, perhaps through user input and audit

Time frame

Australian
RQF:

- 6 year assessment period, impacts can be informed by research from the prior assessment period

RAISS:

- Arbitrary- decided by funder when to use it; relies also on recall

PART:

- To the extent that what counts is whether programmes meet their annual (outcome) targets, it does not matter how far back the drivers for changes go. Assessments at least once every 5 years

ERIC:

- 3 or 6 years?

Burden

RQF:

- not known as a full scale operational pilot was never conducted

RAISS:

- low , 30-60 min

PART:

- not known (not across programmes at least, but positive feedback from implementers)

ERiC:

- likely to be high in first round of assessments, and lower thereafter once contents for assessment (e.g. Indicators) and stakeholders are established.

End users

RQF:

- high levels of engagement of stakeholders

RAISS:

- only researcher and funder, no end-users

PART:

- the public and advocacy groups have access to information and can provide feedback (e.g. OMB Watch)

ERiC:

- high levels of stakeholder engagement

FACTORS TO CONSIDER IN THE DEVELOPMENT OF SYSTEMIC EVALUATION FRAMEWORKS: KEY MESSAGES FROM CONSULTATIONS WITH HEFCE



Key issues and themes identified from the case studies

- **Strategic intent:**
 1. Frameworks are used for different purposes with different degrees of emphasis (e.g. funding allocation, learning, advocacy, accountability)
 2. To be effective, a framework:
 - Should be very clear on its intent and on criteria for assessing impact
 - Must provide an unambiguous rating and comparison of submissions – *against the intent.*
 3. Allowing those being assessed some flexibility in weightings given to criteria for impact assessments can be beneficial, as it can increase stakeholder buy-in, minimise perverse incentives and recognise diversity in the sector
- **Definitions of impact**
 1. The meanings of 'impact' vary in different contexts
 2. Central to the attempt to evaluate impact is to define what is meant by impact (seeing that impact definitions can have strong behavioural implications)
 3. A clear definition of impact can be built through wide-spread consultation
 4. Qualitative and quantitative indicators of impact are needed (to address the imperfections of an either/or approach). E.g. Case-studies and narratives supported by proxy quantifiable measures

Key issues and themes identified from the case studies

- **Units of analysis**

- Evaluation practice favours higher levels of aggregation, but it is important to try avoid to break up existing formal and informal organisational structures.
- An approach which allows researchers and institutions discretion in how they organise into groupings – while still requiring groups to have some minimum size was felt to be appropriate

- **Attribution**

- Evaluations should acknowledge that the issue of attribution is complex: scientific progress is a cumulative and collaborative effort
- A case-study approach can help mitigate this by providing detail and demonstrable and verifiable evidence. A verification mechanism is important to build confidence and ensure objectivity.
- A focus on contribution rather than attribution is likely to be more realistic

Key issues and themes identified from the case studies

- End users:
 - The engagement of end users of research in impact assessments is critical
 - if assessment is about improving impact
 - If we need verification mechanisms
- Burden (and associated costs)
 - If review based mechanisms will be needed to assess impact, it is clear that there will be an administrative burden
 - This burden will fall to different stakeholders groups:
 - The HEI preparing impact statements
 - The panels assessing impact statements
 - End-users being consulted
 - Extensive operational piloting of a framework is essential for determining levels of burden and feasibility
 - However, frameworks should generally not rely on only one method (e.g. bibliometrics)



EUROPE