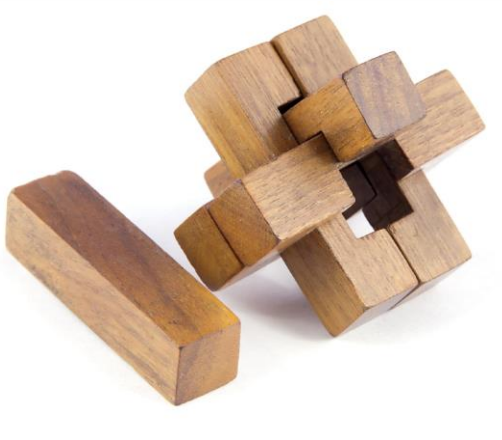


The International School
on Research Impact Assessment

Models and frameworks

Jonathan Grant
RAND Europe

Organised by:
Agència de Qualitat i Avaluació Sanitàries de Catalunya
AGENCY FOR HEALTH QUALITY AND ASSESSMENT OF CATALONIA
Generalitat de Catalunya
Departament de Salut
RAND EUROPE
Alberta Innovates Health Solutions

Learning objectives

- To review various research impact assessment frameworks that have been developed by others
- To assess the different characteristics and the strengths and weaknesses of different frameworks
- To provide the wear with all to develop bespoke, fit for purpose, frameworks for specific impact assessments

Outline

1. The art of conceptualization & organising information

2. Review of research Impact assessment frameworks

3. Characteristics of different frameworks

Outline

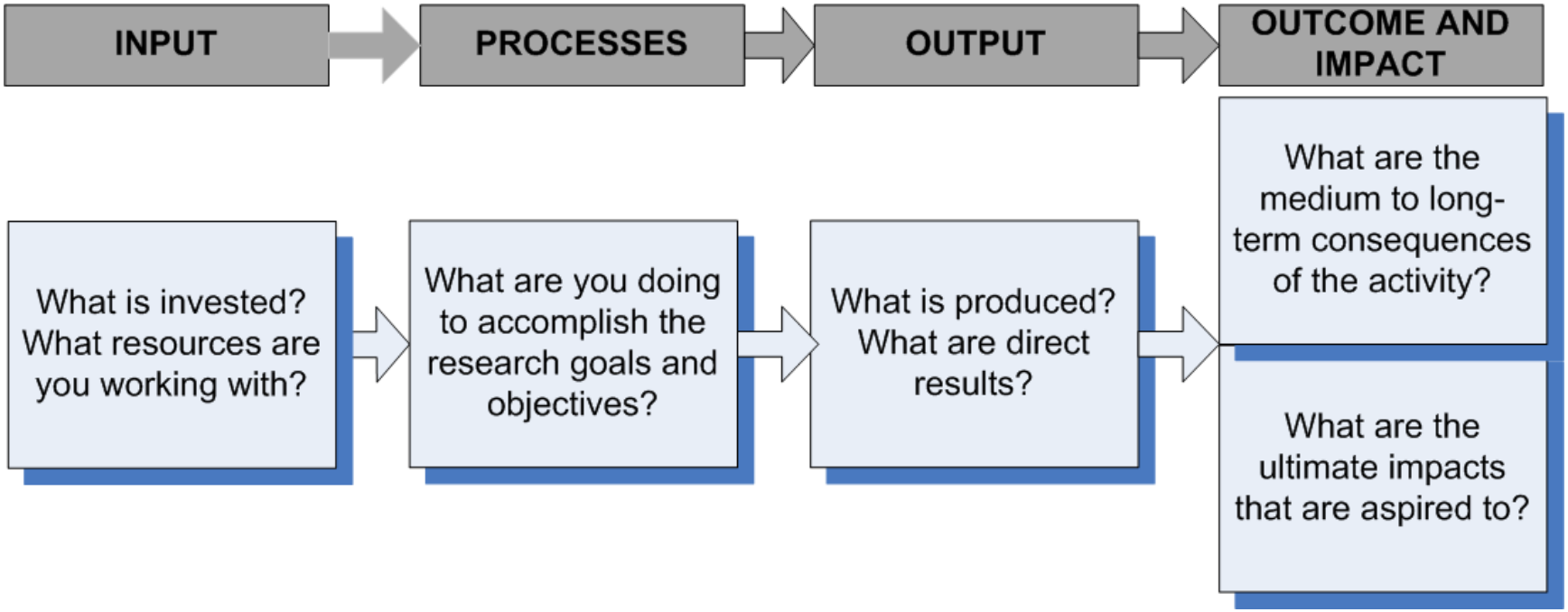
1. The art of conceptualization & organising information
2. Review of research Impact assessment frameworks
3. Characteristics of different frameworks

Four approaches to organising information

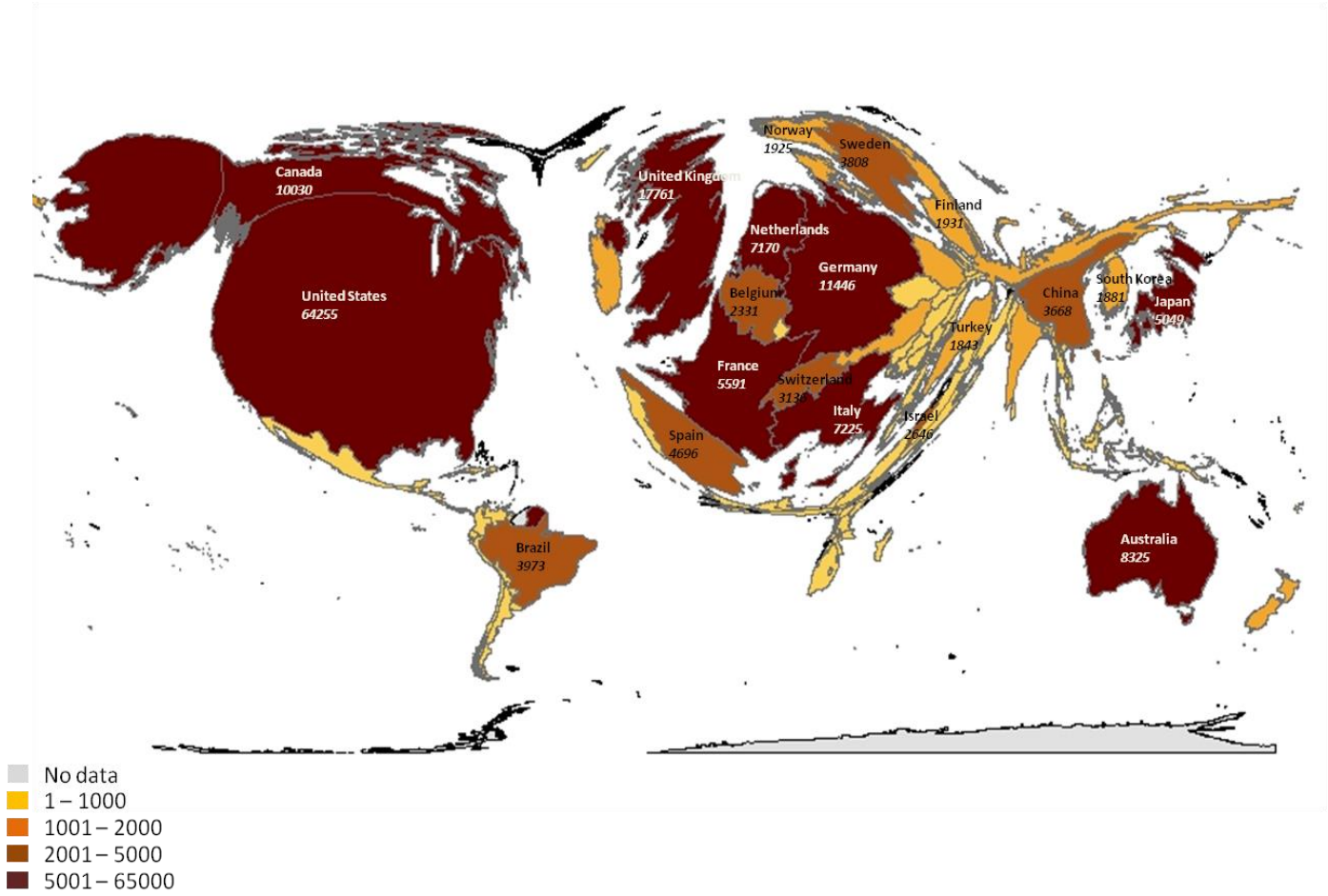
- By time
- By structure
- By rank
- By deductive reasoning

By time: Chronology

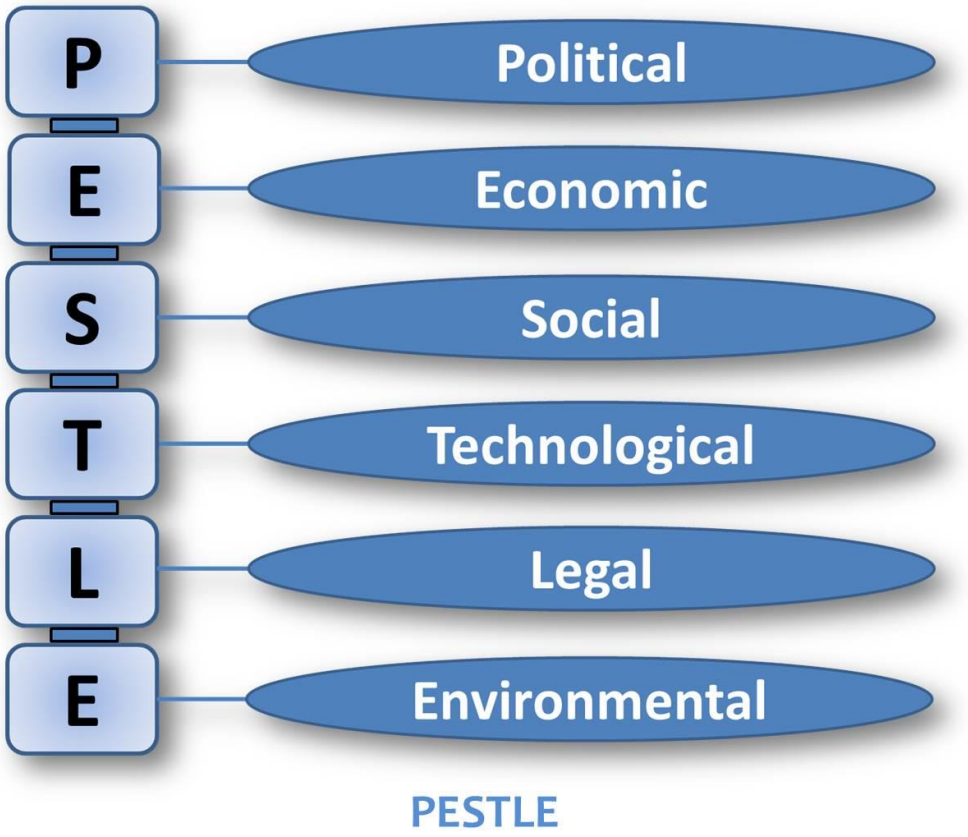
By time: Cause-effect



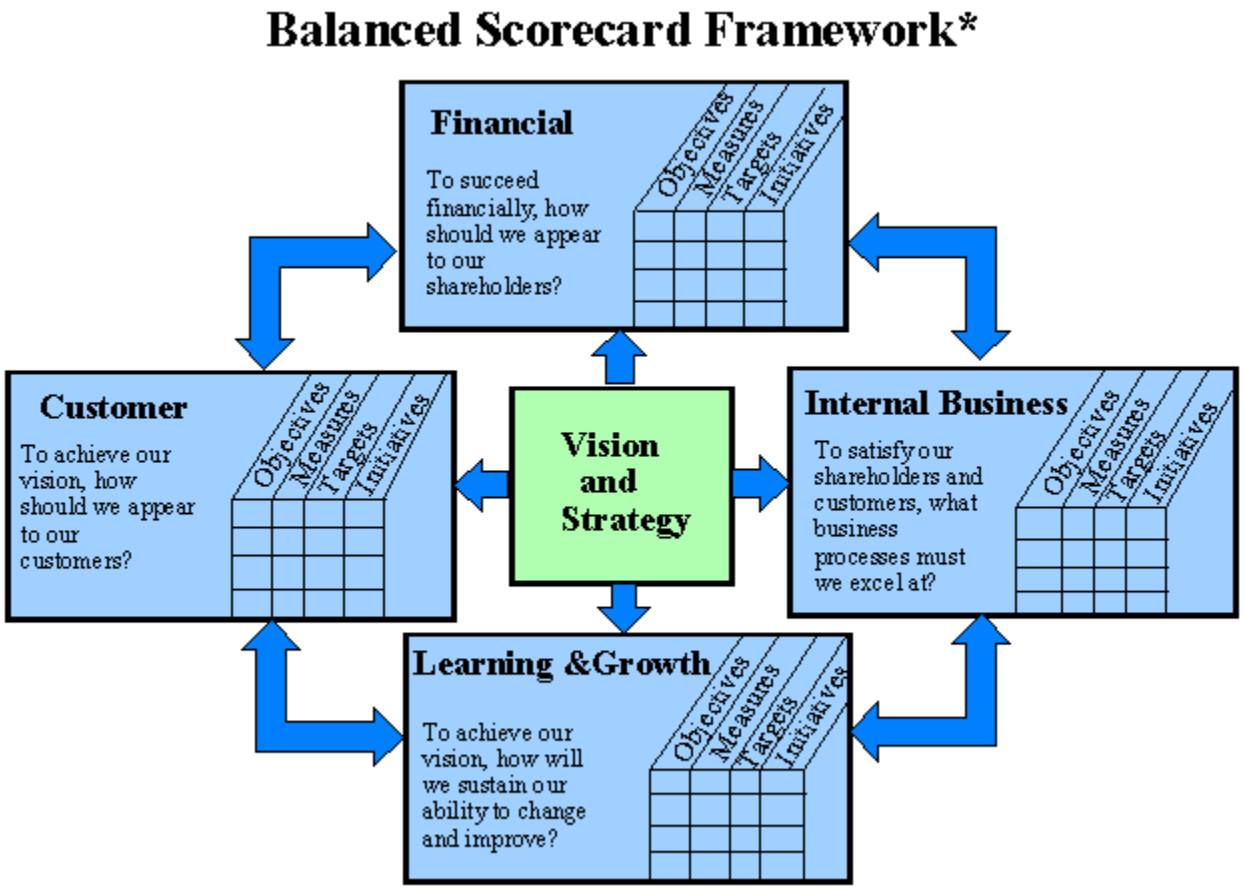
By structure: Geography



By structure: PESTLE

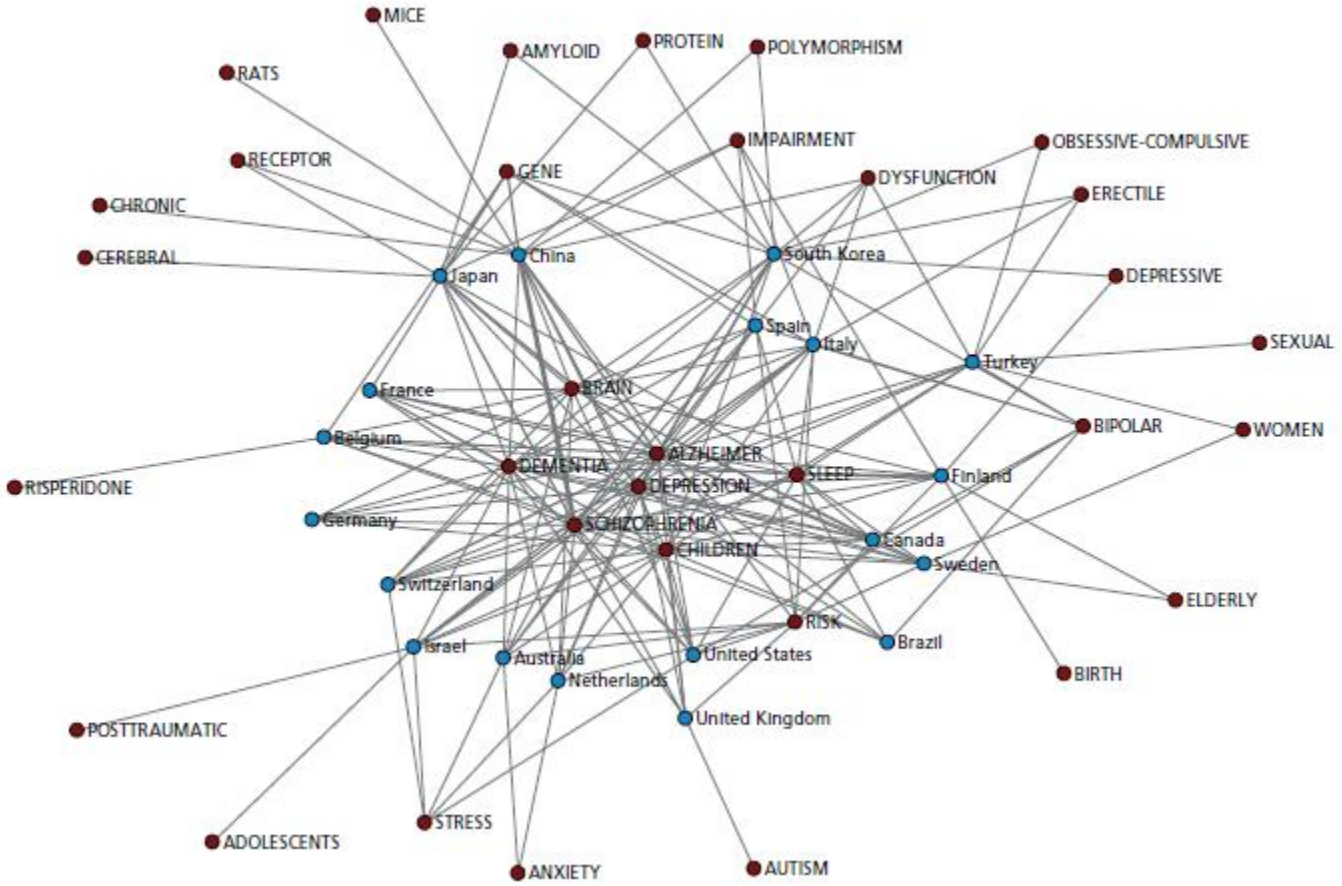


By structure: function

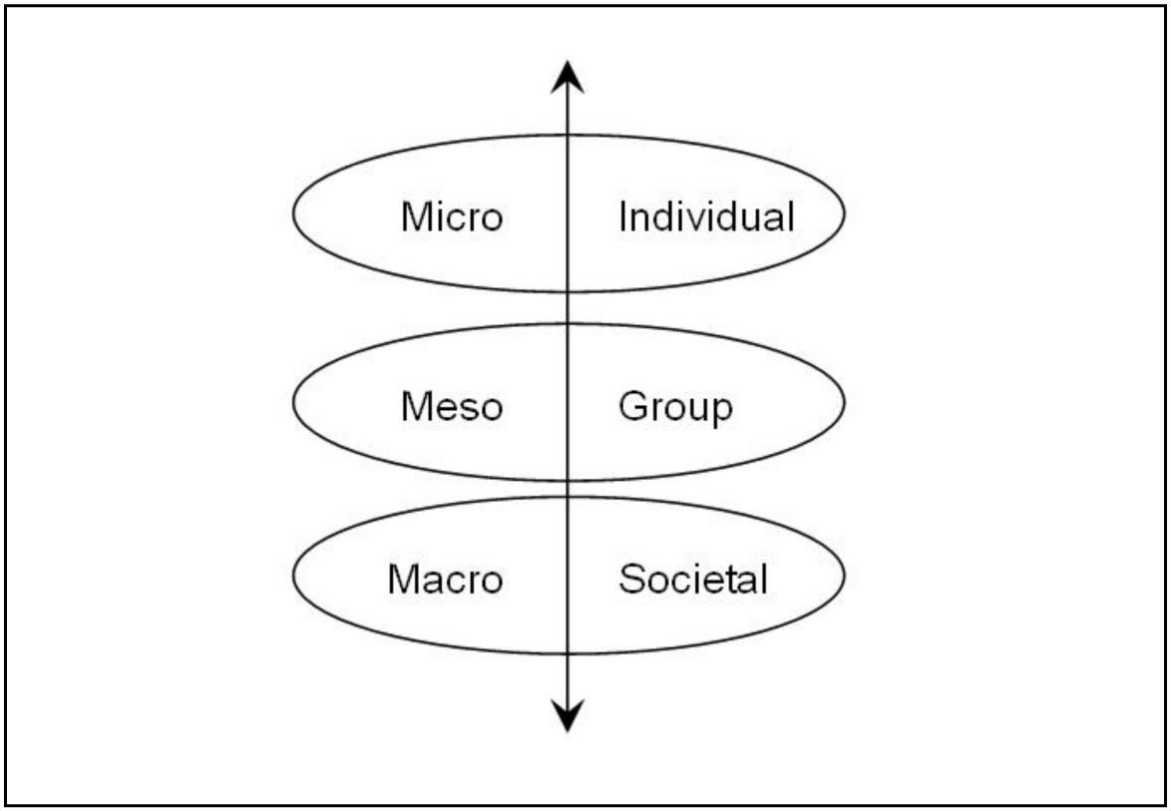


* Adapted from Kaplan & Norton 1996. *The Balanced Scorecard*. Harvard Business School Press: 9. Original from HBR Jan/Feb 1996, p. 76.

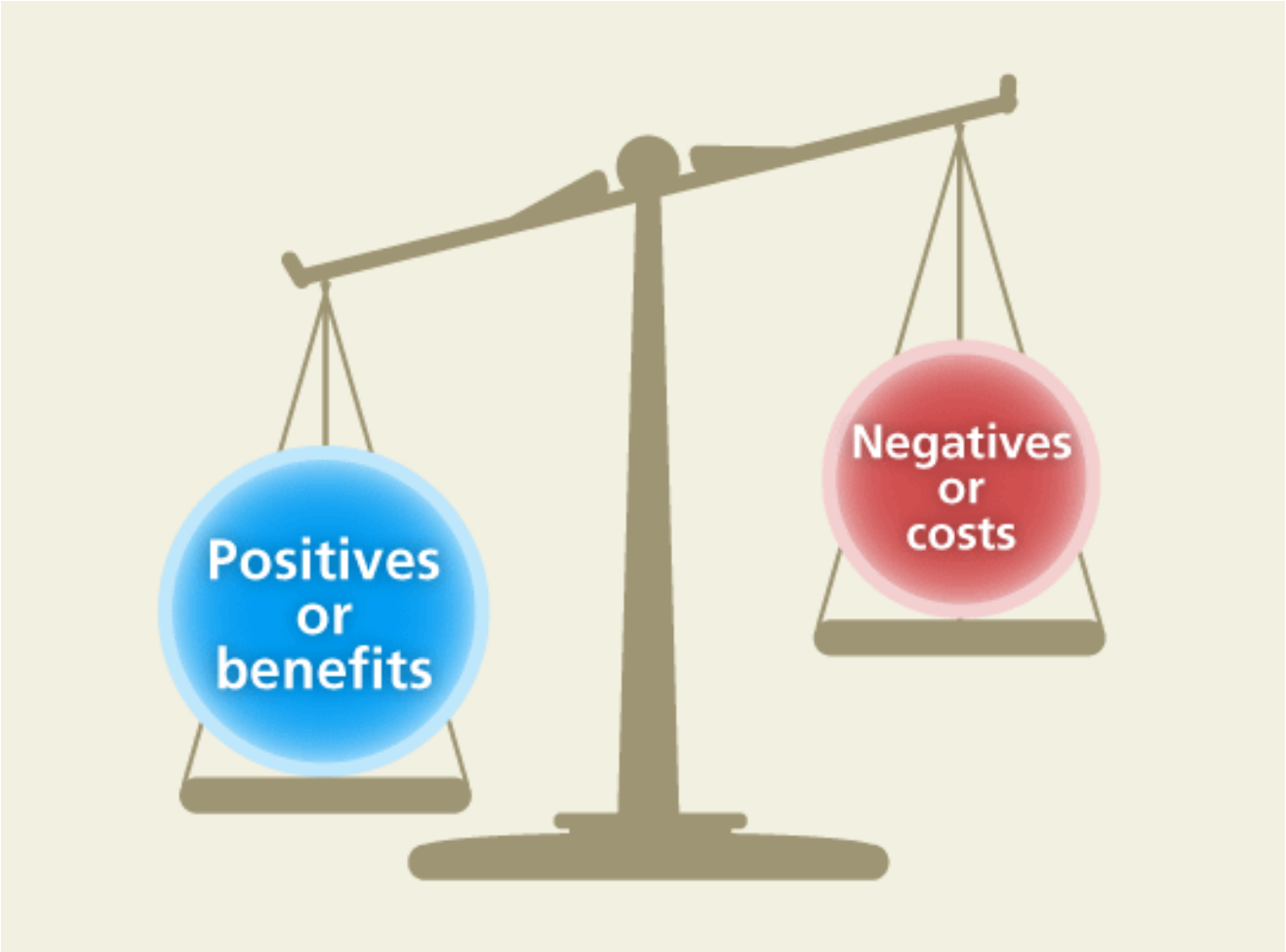
By structure: Interrelationships



By rank: Macro-micro



By deduction: Cost - benefit



Exercise

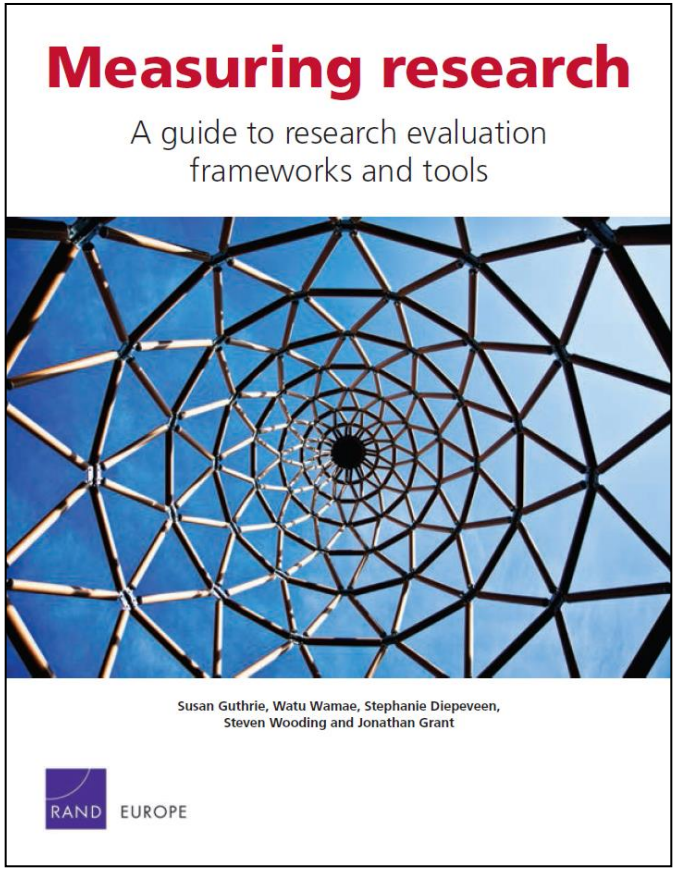
- In table groups discuss the organisation you work for
- Think of different ways that they could be grouped
- Think about different ways you could represent those groupings visually

Outline

1. The art of conceptualization & organising information
2. Review of research impact assessment frameworks
3. Characteristics of different frameworks

Aims of the study

- Act as a 'how-to guide' to evaluating research
  - Understand the challenges and trade-offs in evaluating research
  - Provide examples of frameworks and tools used for evaluating research internationally



Report available at: http://www.rand.org/pubs/monographs/MG1217.html
Webinar at: https://www.aamc.org/initiatives/research/348948/randreportrelease.html

Our approach

- Review of existing frameworks and tools for the evaluation of research
- Analysis of the characteristics of tools and frameworks using a factor analysis approach
- Developed decision tree to aid development of customised research evaluation frameworks

We reviewed six frameworks …

- Research Excellence Framework (REF), UK – assesses performance of UK universities to determine funding allocation

- STAR METRICS, US – uses data mining and other low burden methods to account for federal R&D spending

- Excellence in Research for Australia (ERA), AU – uses bibliometrics, and other quantitative indicators, to map R&D output

- Canadian Academy of Health Science (CAHS), CA – aims to provide consistency and comparability while retaining flexibility

- National Institute of Health Research (NIHR) Dashboard, UK – provides performance management information at various levels of aggregation

- Productive Interactions, EU – flexible approach to help institutions learn and improve their performance against their own goals

… and ten tools

- Bibliometrics
- Surveys
- Logic models
- Case studies
- Economic analysis
- Peer review
- Data mining
- Interviews
- Data visualisation
- Site visits
- Document review

We reviewed six frameworks …

- Research Excellence Framework (REF), UK – assesses performance of UK universities to determine funding allocation

- STAR METRICS, US – uses data mining and other low burden methods to account for federal R&D spending

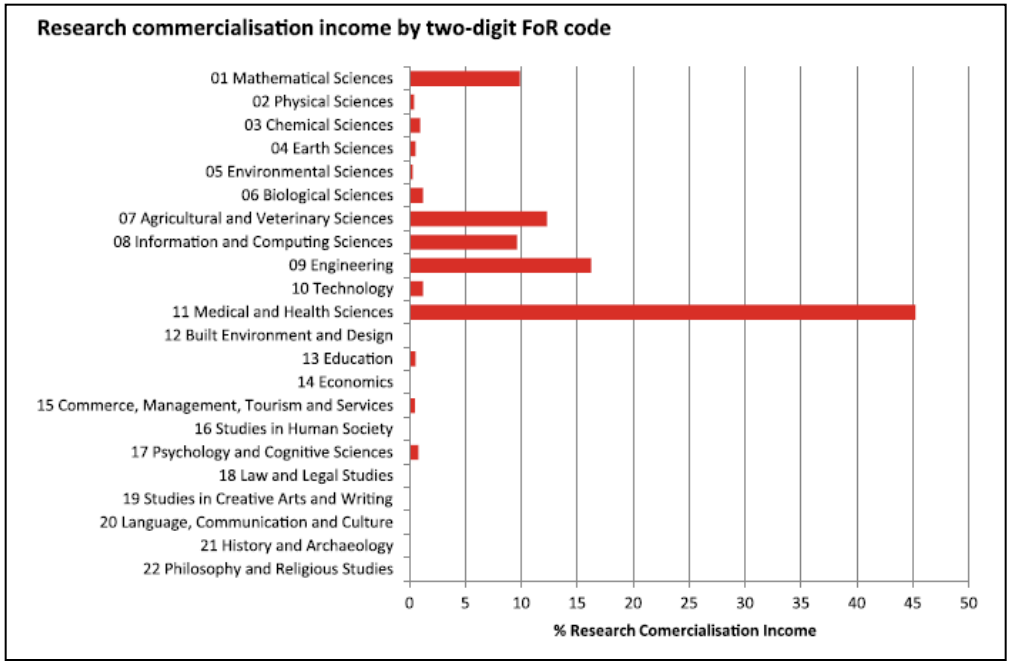
- Excellence in Research for Australia (ERA), AU – uses bibliometrics, and other quantitative indicators, to map R&D output

- Canadian Academy of Health Science (CAHS), CA – aims to provide consistency and comparability while retaining flexibility

- National Institute of Health Research (NIHR) Dashboard, UK – provides performance management information at various levels of aggregation

- Productive Interactions, EU – flexible approach to help institutions learn and improve their performance against their own goals

Accountability – to taxpayers & donors



Excellence in Research for Australia (ERA)

Origin and rationale:
Perceived need to include assessment of quality in block funding allocation (previously volume only). Advocacy purpose - demonstrate quality of Australian research

Scope:
Assesses quality, volume, application of research (impact), and measures of esteem for all Australian Universities at disciplinary level

Measurement:
Indicator approach, uses those appropriate at disciplinary level. Dashboard provided for review by expert panel

Application to date:
First round in 2010, broadly successful. Next round 2012, with minor changes. Intended for funding allocation, but not used for this as yet

Analysis:
Broadly positive reception. Meets aims, and burden not too great.

Limitation is the availability of appropriate indicators

Wider applicability:
Should be widely applicable, criticism limited in Australian context. Implementation appears to have been fairly straightforward

SWOT analysis for ERA

Strengths

- Acceptable to research community in Australia
- Burden on participants is moderate
- Indicator driven
- Produces a single performance indicator, which can be used for ranking
- Multi-disciplinary

Weaknesses

- Indicator driven
- Still moderated through peer review, reducing objectivity
- Not comprehensive – academic focus
- Summative
- Burden relative to return is high (not yet used for funding allocation)
- Requires some central expertise (bibliometric expertise on panel)

Opportunities

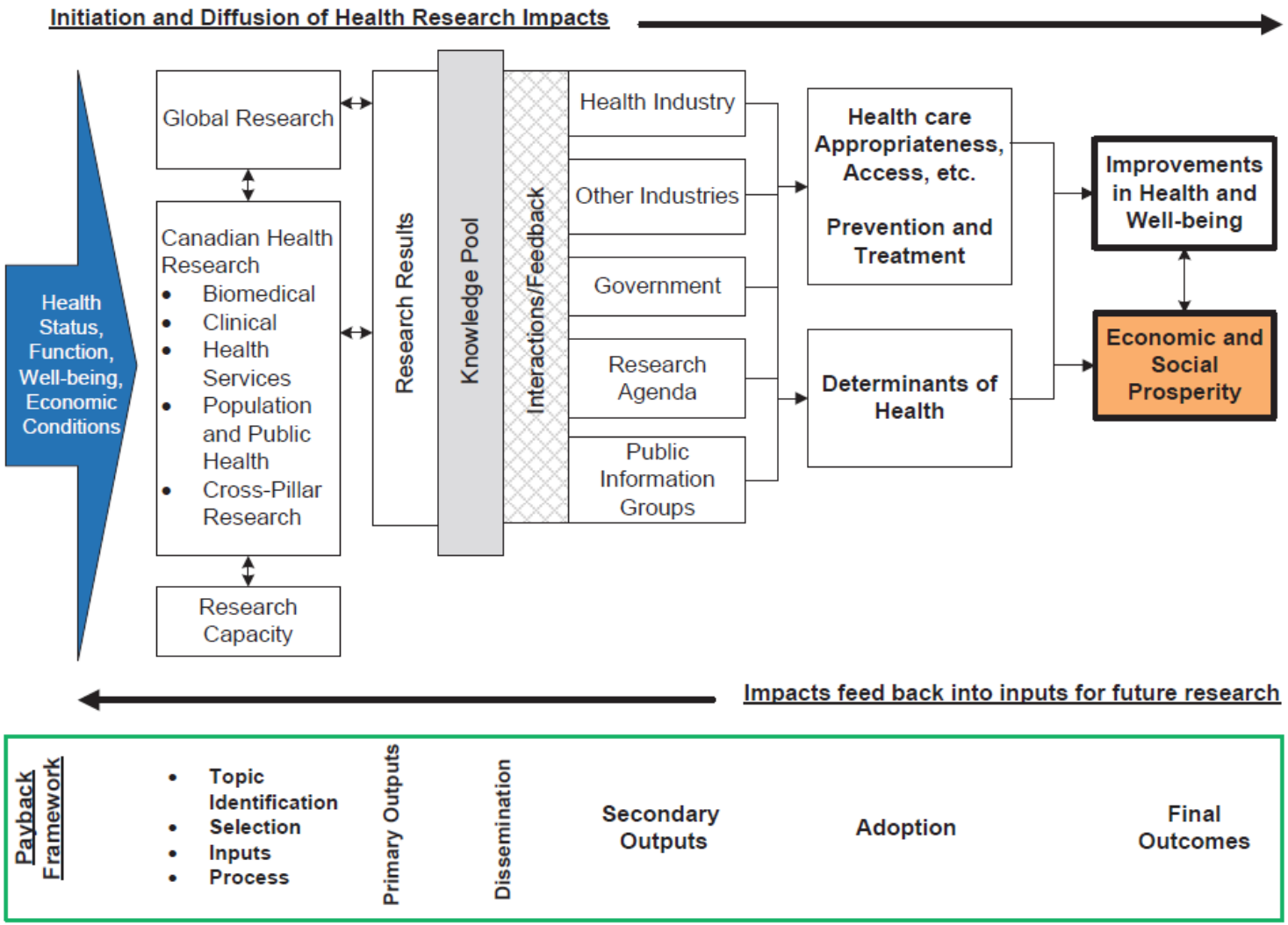
- Potential to add new indicators

Threats

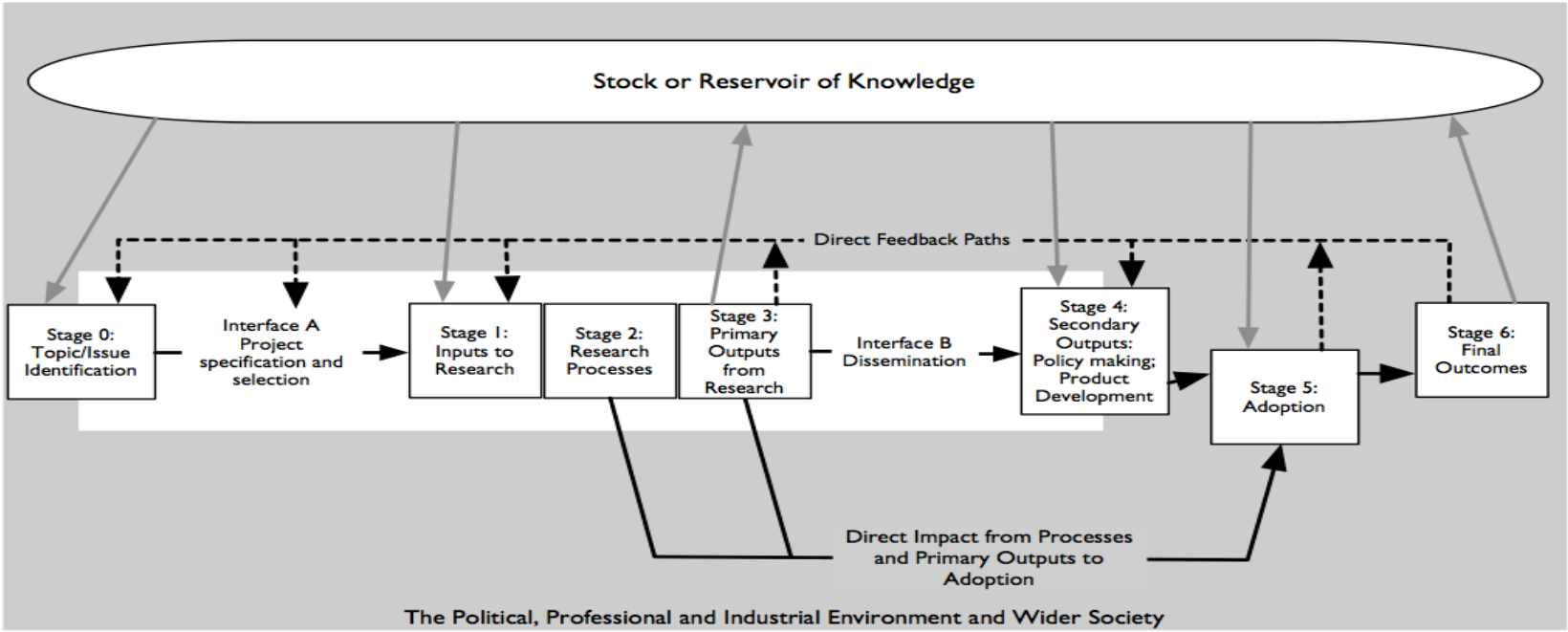
- No funding implications
- Politics informed its development
- Government and public appetite to include impact is limited in Australia (limits potential for development)

Broken Link designed by Stephen JB Thomas from The Noun Project

Canadian Academy of Health (CAHS)



The Payback framework



Adapted from Hanney S, Gonzalez-Block M, Buxton M and Kogan M, The Utilisation of health research in policy-making: concepts, examples and methods of assesssment. *Health Research Policy Systems* 2003, 1:2

Buxton, M., and S. Hanney. "How can payback from health services research be assessed?" *Journal of Health Services Research and Policy* 1 (1996): 35–43.

Payback categories

- Knowledge Production
- Research Targeting, Capacity Building
- Informing Policy or Product Development
- Sectoral Benefits (Social, Health, Environmental, Cultural)
- Broader Economic Benefits

Payback categories

- Knowledge production
  - Traditionally more academic focussed, can't be used for impact
  - Can provide useful starting points to trace impact forward
  - Indicators[1]: citation impacts; shares of publication
- Research capacity building
  - Elements which build future research capacity
  - Aids absorption of knowledge by the system
  - Indicators: Research resources; New methodologies; Career development of collaborators (outside academia); Leveraged funding

.

Payback categories

- Informing policy development or practice
  - Looks at impacts in both processes and policy outcomes
  - Policies and practice might change at multiple levels
  - Impacts include change in advice given by professional bodies; changes in professional practice within a sector; changes to training policies or guidelines
  - Indicators: Use of research in guidelines; Media citation analysis; Citations in advocacy guidance; Requests for research to support policy development
- Informing product development
  - Identify concrete steps in the commercialisation process
  - Trace proof of concept research through to clinical trials
  - Indicators: Citations in a patent, patent applications, contributions to a website

Payback categories

- Sectoral benefits (health, education, environment, cultural)
  - Identifies ways that sectors and user communities have gained from the research
  - Can include impacts from broader public knowledge creation
  - Indicators: More equitable access to services; Cost-savings within a sector; Health gains; Preservation of cultural heritage
- Socio-economic benefits
  - Economic benefits from the processes of product, policy, or professional development
  - Economic benefits from a healthier or more enriched society (eg increased productivity, lower crime rates, healthier society)
  - Impacts affecting the welfare, profits and revenues of individuals or organisations involved in the research
  - Indicators: improved efficiency or effectiveness of services due to research; commercialisation gains; well-being measures; gains in socio-economic status of communities

Canadian Academy of Health (CAHS)

Origin and rationale:
Draws on well established 'Payback' framework. Aims to improve comparability across a disparate health research system. Covers wide range of impacts

Scope:
Five categories: advancing knowledge; capacity building; informing policies and product development; health and health sector benefits; broader economic benefits.

Measurement:
Specific indicators for each category. Logic model has 4 research 'pillars': Biomedical; Clinical; Health services; Social cultural, environmental and population health

Application to date:
Used by public funders; predominantly CIHR (federal funder), but there has also been some uptake by regional organisations (e.g. Alberta Innovates)

Analysis:
Strengths: generalisable within health sector, can handle unexpected outcomes. But understanding needed at funder level - may limit uptake. Early stages hard to assess

Wider applicability:
Breadth, depth and flexibility mean framework should be widely applicable. However, it only provides a guide and needs significant work to tailor to specific circumstances

SWOT analysis for CAHS

Strengths
Very comprehensive
Flexible
Developed through engagement, and has strong buy-in
Formative
Looks at process as well as outputs and impacts
Concept of an indicator library
Aligned with main funders, framework

Weaknesses
Resource intensive
Complicated
Not easily comparable
Implementation challenging
Developed by committee
Requires participant expertise
Not ranking – hard to use to allocate funding
Large burden on participants
Not multi-disciplinary
Definitional ambiguity between outputs and outcomes

Opportunities
Unified but flexible approach
Potential to build an indicator platform and toolkit
Built on an internationally recognised framework - opportunity for international uptake and wider comparability

Threats
No implementing owner
Slow uptake
Dependent on CIHR endorsement

Broken Link designed by Stephen JB Thomas from The Noun Project

National Institute of Health Research Dashboard

El Turabi et al. Health Research Policy and Systems 2011, 9:13
http://www.health-policy-systems.com/content/9/1/13

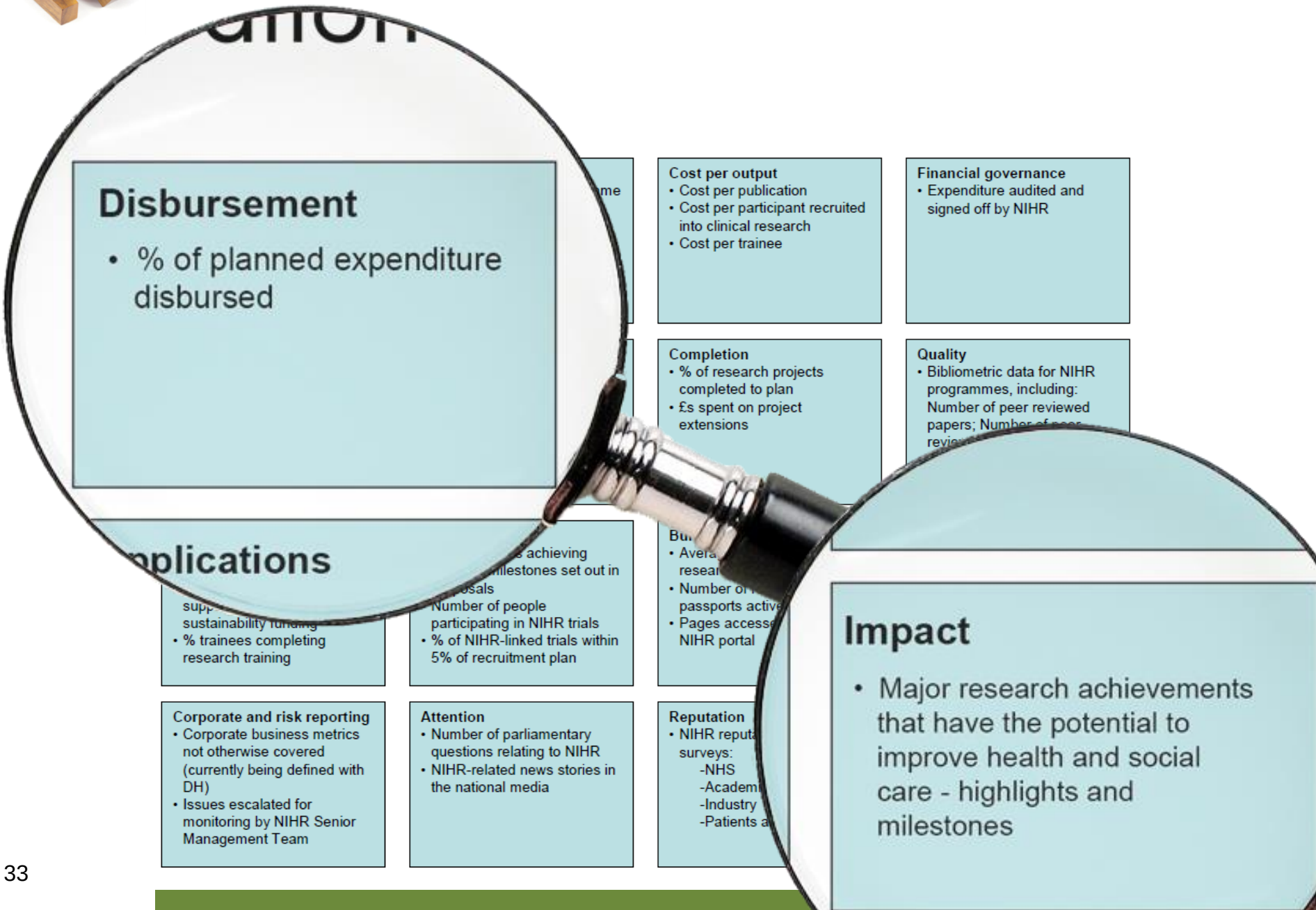
HEALTH RESEARCH POLICY AND SYSTEMS

RESEARCH Open Access

A novel performance monitoring framework for health research systems: experiences of the National Institute for Health Research in England

Anas El Turabi[1,2*], Michael Hallsworth[3], Tom Ling[2] and Jonathan Grant[2]

The International School on Research Impact Assessment



The Dashboard is incorporated into MIS

National Institute of Health Research Dashboard

Origin and rationale:
Aim is to develop a small but balanced set of indicators to support strategic decision making, monitoring performance on regular ongoing basis

Scope:
Data collected quarterly at programme level on inputs, processes, outputs and outcomes for 3 elements – financial, internal process, and user satisfaction

Measurement:
Programme specific data can be pooled to provide a system level dashboard. 15 indicators selected, matching core aims, collected quarterly

Application to date:
Launched July 2011 NIHR-wide, with data to be provided by the four coordinating centres, analysed and aggregated centrally

Analysis:
Designed to fit strategic objectives, so in that sense likely to be effective. However, only just launched, so detailed analysis premature

Wider applicability:
Should be applicable to other national health research funders. Performance indicators selected can be tailored to assessment needs

SWOT analysis for NIHR Dashboard

Strengths

Aligned with institutional goals
Bespoke
Formative
Can be used for monitoring (frequent assessments)
Wide applicability
Strong theoretical basis
Comparable
Focused and selective set of indicators
Indicator set is balanced
Continuous burden (not episodic)

Opportunities

Flexibility may allow use across multiple institutions
Useful at many levels

Weaknesses

High central burden
Bespoke
Reliant on information management systems
High up from burden
High level of central expertise required
Not comprehensive if incorrectly used – it only
monitors the indicators you select
Continuous burden (not episodic)
Not multi-disciplinary

Threats

Scalability across multiple institutions not demonstrated
New and not fully implemented

Broken Link designed by Stephen JB Thomas from The Noun Project

Exercise

- Take one of the remaining frameworks
  - UK REF, US Star Metrics, EC Productive Interactions
- Review its characteristics
- Identify its strengths, weaknesses, opportunities and threats
- Fill in the A1 sheet and be prepared to present back to the group

Research Excellence Framework (REF)

Origin and rationale:
Evolved from its predecessor, the RAE, and the RQF. Intended to be low burden, but pressure from researchers led to changes. Includes wider societal impact

Scope:
Assessment at subject level on 3 elements:
-Quality of research outputs
-Impact of research (not academic)
-Vitality of environment

Measurement:
Assessment by subject peer review panel of list of outputs, impact statement and case studies, and statement on research environment

Application to date:
Piloted 2009. First round of assessment 2014, results will determine funding allocation.

Analysis:
Burden not reduced, but adds wider impact to evaluation. Originally metrics based, but this was dropped as too unpopular

Wider applicability:
Suitable for similar cross institutional assessment of performance. High burden on institutions, arguably expensive, so best for significant funding allocation uses

38

SWOT analysis for REF

Strengths

Burden relative to return is low (determines significant funding allocation)
Acceptable to UK academic community as it uses peer review
Comprehensive (includes impact)
Multi-method
Multi-disciplinary
Successfully piloted, and many elements well tested
Produces a single performance indicator which can be used for ranking

Weaknesses

Cost
Total burden s high
Can discriminate against some types of researchers
Can discriminate against some types of institution
Summative
Scalability not demonstrated
Not transparent
Almost solely reliant on peer review – limits objectivity

Opportunities

Potential to move towards indicators
Move towards impact in UK and internationally
Increased focus on public accountability in UK

Threats

Non-participation
Political
Reductions in research funding may limit ability to fund
to match the quality demonstrated
Could result in research concentration

Broken Link designed by Stephen JB Thomas from The Noun Project

STAR METRICS

Origin and rationale:
Key aim to minimise burden on academics; Helps to meet US federal accountability requirements

Scope:
Two levels:
- *Level 1- number of jobs supported*
- *Level 2- range of research funded, researcher interactions, and wider impacts*

Measurement:
Data mining approach, automated. At present, only gathers jobs data.

Methodologies for level 2 still being developed

Application to date:
Level 1 rolled out to 80 universities
Level 2 still under development.
Voluntary participation - full coverage unlikely

Analysis:
Feedback generally positive, but feasibility of level 2 not proven

Wider applicability:
Potentially very wide depending on success of Level 2. There has been international interest, eg Japan, EC

SWOT analysis for STARMETRICS

Strengths

Data mining approach is relatively novel
Low participant burden once set up
Not a ranking approach – does not produce a single indicator of comparative performance

Weaknesses

Not fully developed and tested
High initial burden, and expertise required to establish
Approach beyond Level 1 not proven
Level 2 will depend on quality of data input
Level 1 focused entirely on jobs for money input (not comprehensive)
Summative (at present)
Not a ranking approach – does not produce a single indicator of comparative performance

Opportunities

Data mining
Harmonisation between funders
ARRA (Phase 1)
International interest

Threats

Non-participation (not compulsory)

Productive Interactions

Origin and rationale:
Measures productive interactions, defined as interactions with stakeholders that lead to change. Assessment against internal goals intended for learning.

Scope:
Intended to work in a wide range of contexts, best applied at research group or department level where goals are consistent

Measurement:
Three types interaction: direct personal contacts; indirect (e.g. via a publication), financial. Engages users, findings assessed against internal goals

Application to date:
Piloted across diverse disciplines and contexts in four European countries and at the EC level. No plans to roll out more widely at present

Analysis:
Tailored, so should help improve performance. No comparative ranking. Requires significant work from participants to generate their own set of goals and indicators

Wider applicability:
Indicators developed to meet goals, so widely applicable, but does not produce comparison between institutions, so not appropriate for allocation, and could be challenging to use for accountability

SWOT analysis for Productive Interactions

Strengths
Formative
Sensitive to institutional goals
Avoids perverse incentives
Comprehensive
Flexible
Some tools and ‘how to’ guides being developed
Avoids time lag interaction to impact
Avoiding time lag reduces bias against early career researchers
Multi-disciplinary

Weaknesses
High burden
Not comparable (between institutions)
Challenging to implement
Requires assessors to identify productive interactions
Assumes interactions are a good indicator of impact

Opportunities
Piloted in a range of countries and disciplines
Could support strategic thinking about impact

Threats
Scalability
No implementing owner
Needs to move from research to operationalisation
No developing owner – what will happen now FP7 grant funding has run out?

Broken Link designed by Stephen JB Thomas from The Noun Project

Outline

1. The art of conceptualization & organising information
2. Review of research Impact assessment frameworks
3. Characteristics of different frameworks

Key findings of analysis

- There is no silver bullet
- The framework should be designed based on the purpose of the evaluation
- Research evaluation tools typically fall into one of two groups
- There is a range of possible units of aggregation
- There are some perennial challenges to research evaluation that need to be addressed
- Research evaluation approaches need to suit their wider context
- Implementation needs ownership and the right incentives and support

There is no silver bullet

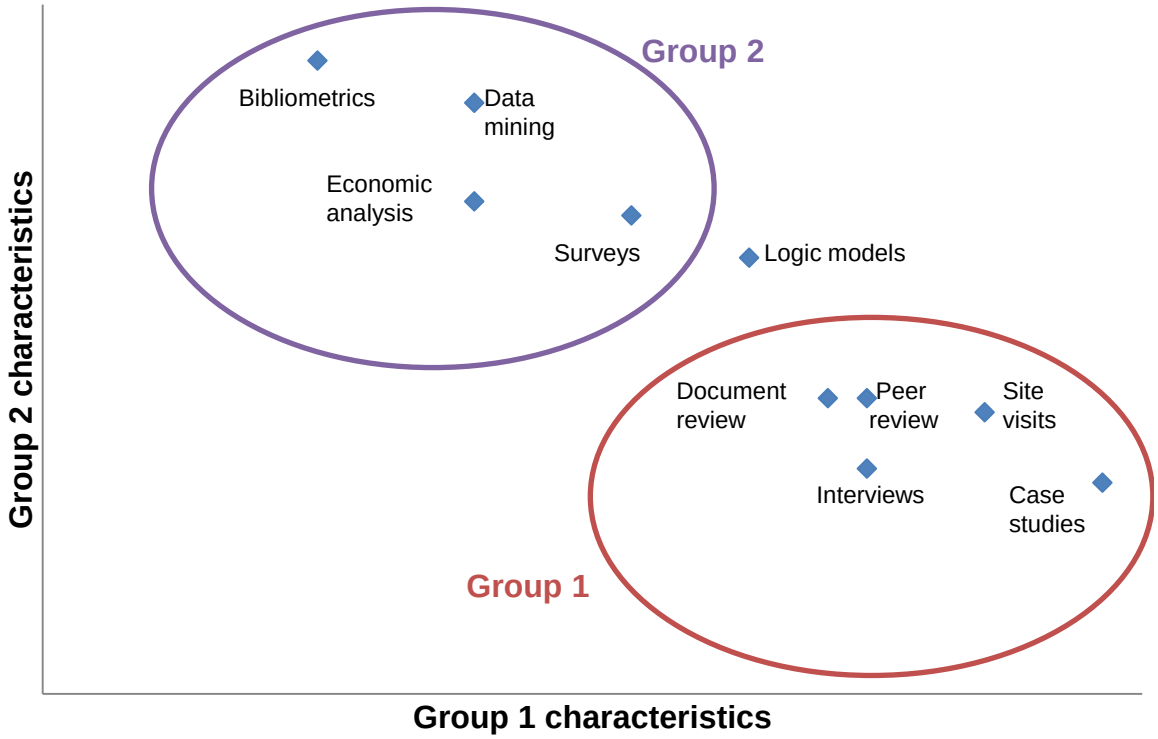
Designing a research evaluation framework requires trade-offs:

- Quantitative approaches tend to produce longitudinal data, do not require judgement or interpretation and are relatively transparent, but they have a high initial burden
- Formative approaches tend to be comprehensive, evaluating across a range of areas, and flexible, but they do not produce comparisons between institutions
- Approaches that have a high central burden tend not to be suitable for frequent use
- Approaches that have been more fully implemented tend to have a high level of central ownership
- Frameworks that place a high burden on participants require those participants to have a high level of expertise (or should provide capacity building and training to achieve this)

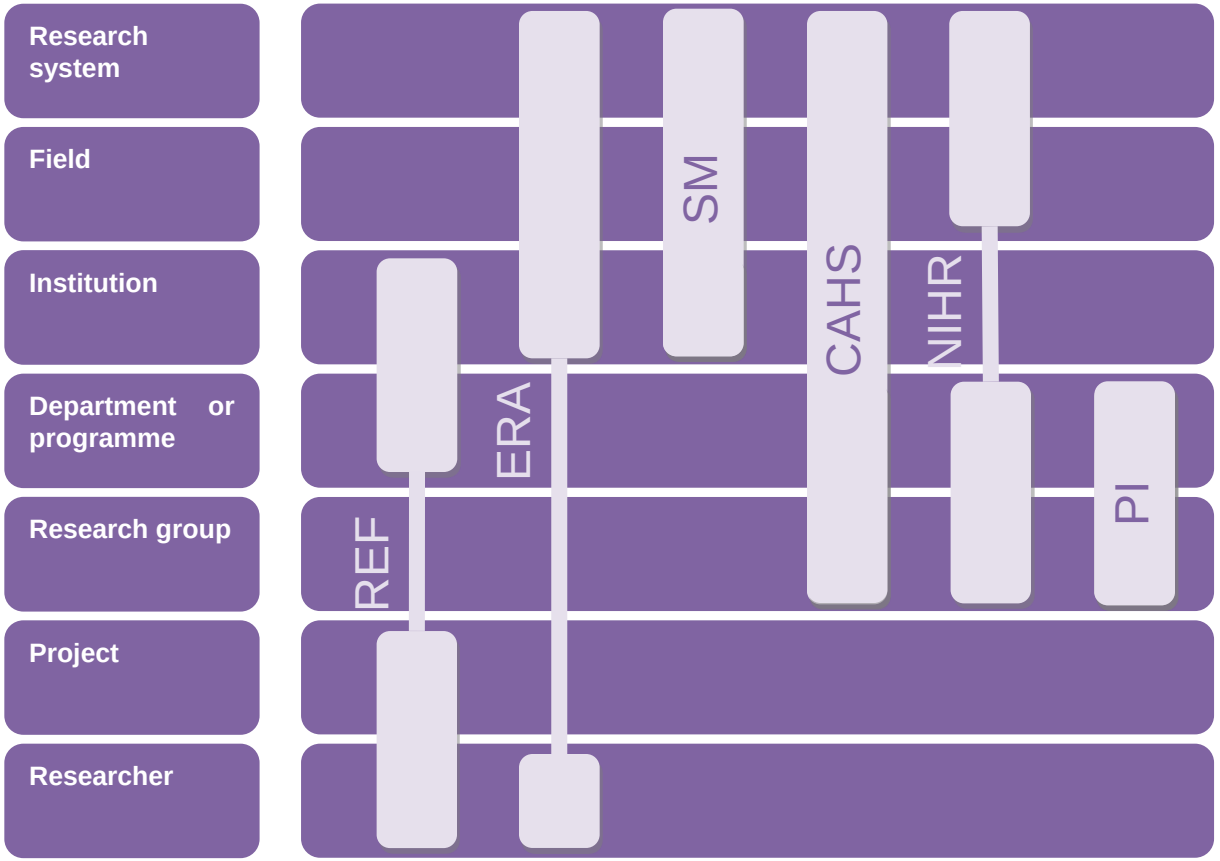
The framework should be designed based on the purpose of the evaluation

- **Analysis -** What works in research funding?
- **Advocacy -** 'make the case' for research funding
- **Accountability -** To taxpayer, donors, etc.
- **Allocation -** What to fund (institution, field, people …)

Research evaluation tools typically fall into one of two groups



There is a range of possible units of aggregation



There are some perennial challenges to research evaluation

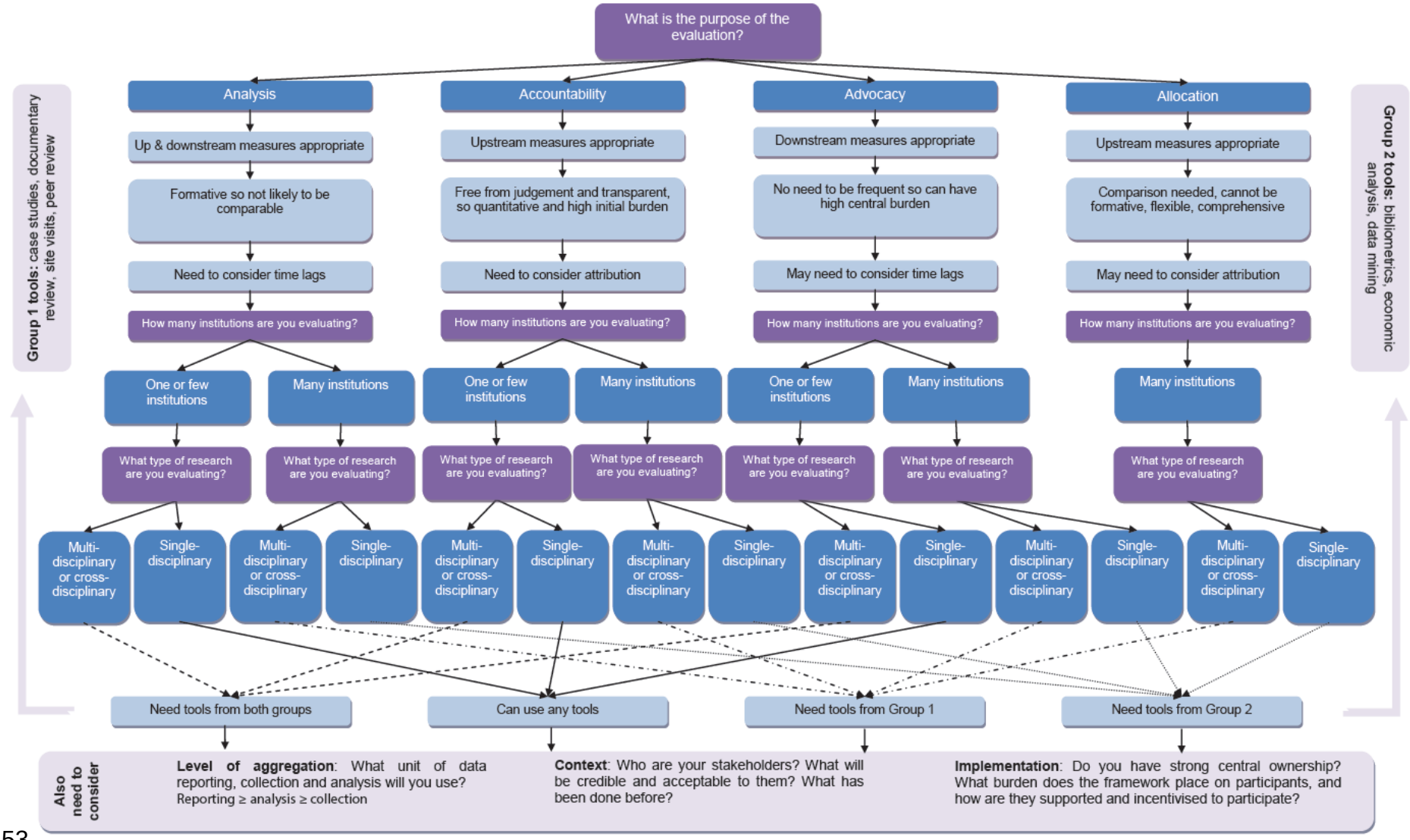
Research evaluation approaches need to suit their wider context

- Acceptability and credibility
- Differences between countries
- Need to ensure framework does not discriminate

Implementation needs ownership, the right incentives and support

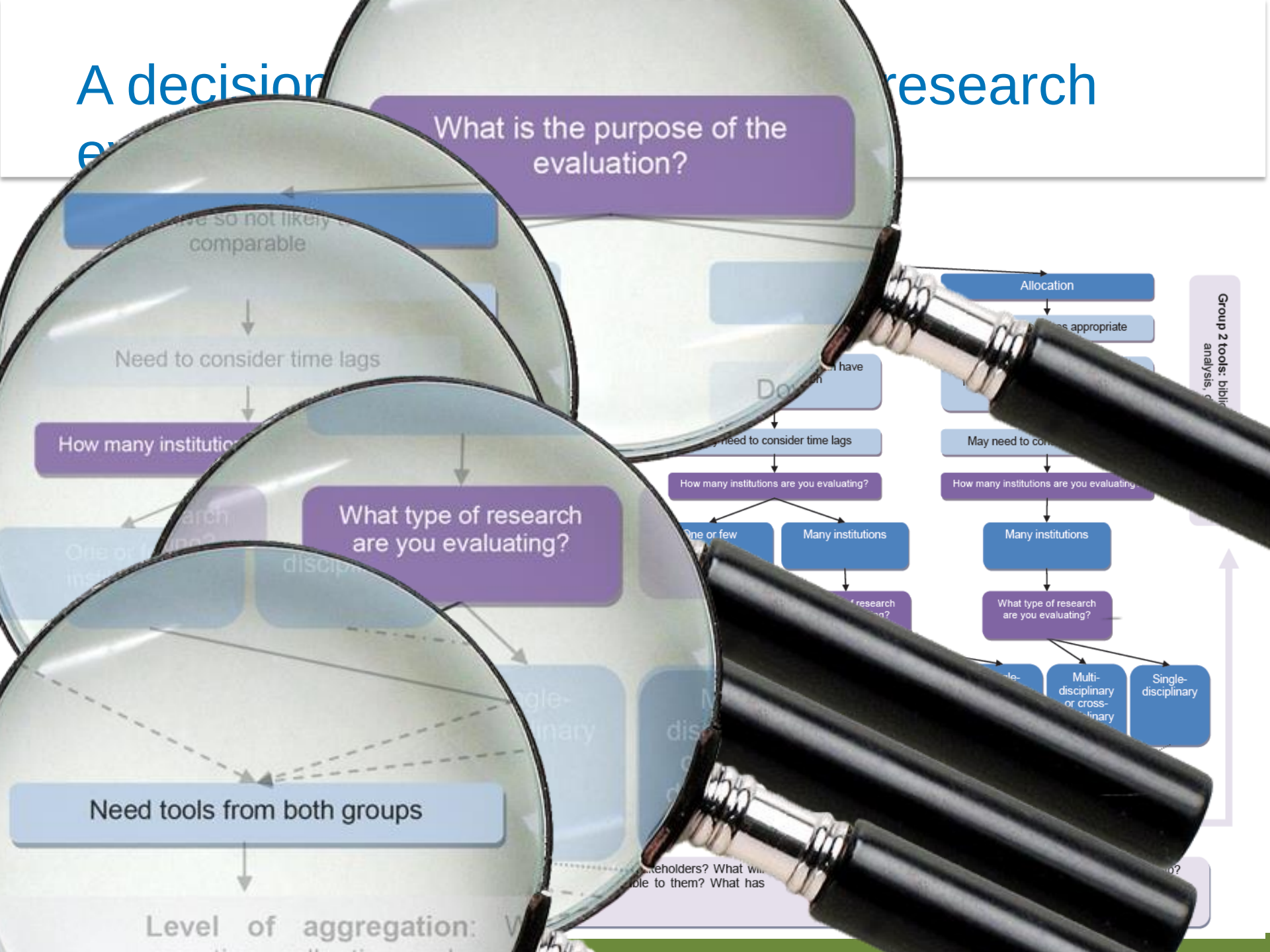
- Where compulsory, the challenge is to obtain support from the academic and wider community
- Where participation is voluntary, incentives need to be in place to promote and sustain uptake
- In both cases, participants need to be given the skills necessary for the process, through simplicity, training or a toolkit
- In all cases, strong central ownership is needed for effective large-scale implementation

A decision tree for developing a research evaluation framework



A decision

research



Key messages

- Know why you are measuring research
  - What is the objective of the research evaluation?
- Use a 'multi-method, multi-dimensional' approach
  - Don't rely on one method (e.g., bibliometrics)
- (Research) measurement is not easy
  - No (research) funder has the answer
- Need to move from advocacy to accountability
  - Need 'science of science' to understand what works
  - Need a practical evidence base for science policy

Questions and discussion