technopolis |group|

Assessing research in SSH vs other fields: A practitioner's perspective

Erik Arnold
ENRESSH, Prague
19 January 2017

Who? Why?

- Me: from humanities to social science, and a life lived with the hard sciences
- Ex-SPRU, Booz.Allen & Hamilton and a quarter of a century of Technopolis
- Technopolis: 120 people, 8 countries, primarily research and innovation policy and evaluation
- Strong fan of mixed methods that include use of peer review

- SSH: traditionally evaluated, if at all, via peer review, but increasingly being pushed into evaluation systems intended to cover all disciplines
- Discussion here is based on RAEs in Latvia and Lithuania and field evaluations in Norway

Why is ex post evaluation in SSH now problematised?

- Intensified interest in accountability, ultimately stemming from the change in the research-society 'contract' staring in the 60s
- Correspondingly, growing interest in value for taxpayers' money and the need to justify SSH activities compared with science and technology
- Globalised competition among universities and growth of institutional ranking systems
- Growth of national research assessment and performance-based funding systems
- De-fragmentation of research in science and technology focuses attention on the structure of the research enterprise also in SSH
- Efficiency and comparability: desire to reduce the extent to which SSH is a 'special case' in evaluation

The changing agenda in performance-based research funding systems (PRFS)

- A PRFS is a two-part machine
  - *An assessment or evaluation process*
  - *A funding formula that uses the assessment as a guide to allocating institutional funding for research*
- More or less three generations of PRFS. The first two are only concerned with scholarly performance
  - *First generation: peer review (1986 on …)*
  - *Second generation: metrics (about 2000 on …)*
  - *Third generation: superposition of innovation and impact issues in the assessment process (about 2005 on …)*
- All PRFS have to cope with inter-field differences
  - *Which metrics systems do clunkily, with numbers, which have notorious imperfections and are less available in SSH*
  - *Which peer review systems do fairly easily, though they have to live with the well known problems of peer review*

Liv Langfeldt on bias in peer review

- Selection of panel in/excludes 'schools' of thought
- Time limits set by the organisers affect outcomes
- Tacit negotiations and compromises affect decisions - disagreements among peers get swept under the carpet
- Those who feel they have less knowledge rate more positively
- Division of labour within panels means some judgements are made by individuals, not the full panel

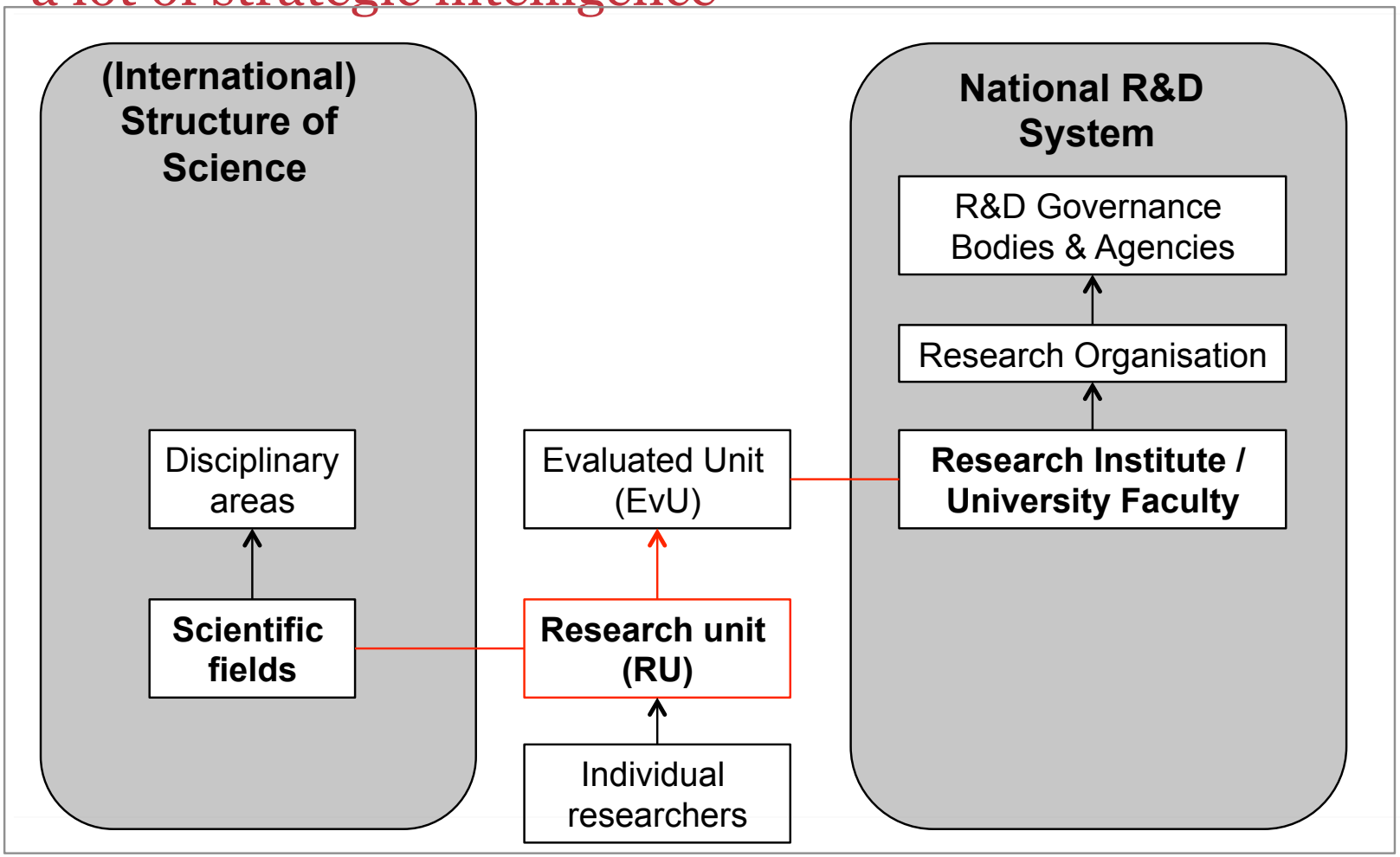
Liv Langfeld (2004) "Expert panels evaluating research: decision-making and sources of bias," ***Research Evaluation***, 13(1), pp51-62

Why countries say they use Performance-Based Research Funding Systems

- To enhance the quality of research and the country's research competitiveness
- To steer behaviour in order to tackle specific failures in the research system
- To strengthen accountability
- To provide strategic information for research strategy at institutional and/or national level

technopolis |group|

In principle a full research assessment system can give us a lot of strategic intelligence



Objectives of the Latvian Research Assessment Exercise (RAE)

- The **overall objective** of the assessment of the research performance of Latvian scientific institutions/structural units is
  - *To provide the Latvian public, policy-makers and decision-makers and the academic community with the most objective picture possible of the excellence and competitiveness of Latvian science in comparison with the global practice in the respective area of science*

- The assessment will produce analytical material that will describe the scientific excellence and competitiveness of Latvian science and the capacity of its scientific institutions. This material will
  - Provide evidence for science policy making at different of levels
  - Enable the scientific institutions involved in the process to improve their operations

technopolis |group|

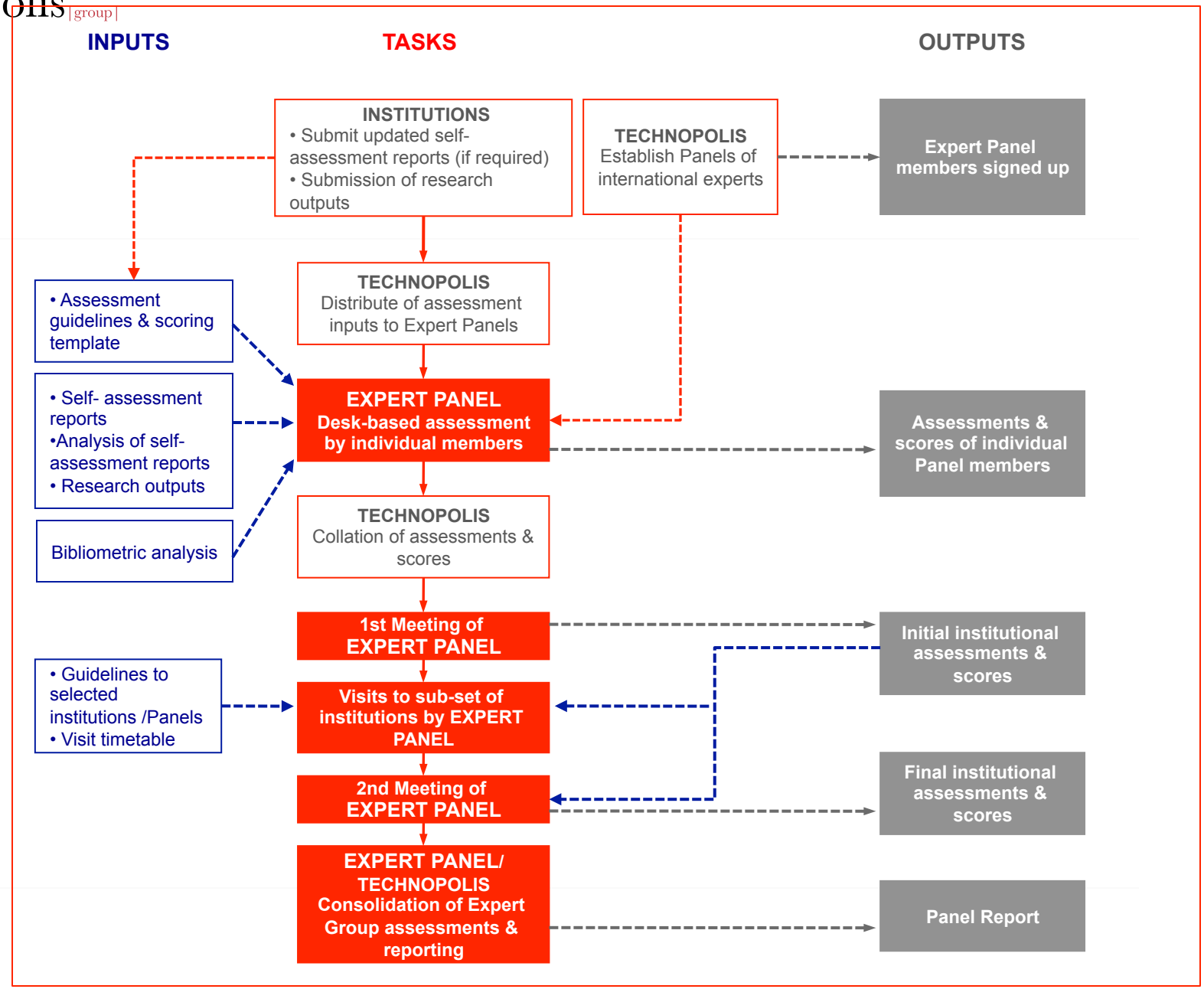
Expert Panels

Panel (full title)	Panel (abbreviation)
Agriculture, Forestry & Veterinary Science	Panel A
Engineering & Computer Science	Panel E
Humanities*	Panel H
Life Science & Medicine	Panel L
Natural Sciences & Mathematics	Panel M
Social Sciences*	Panel S

* Included a Latvian speaker

technopolis | group |

RAE Process

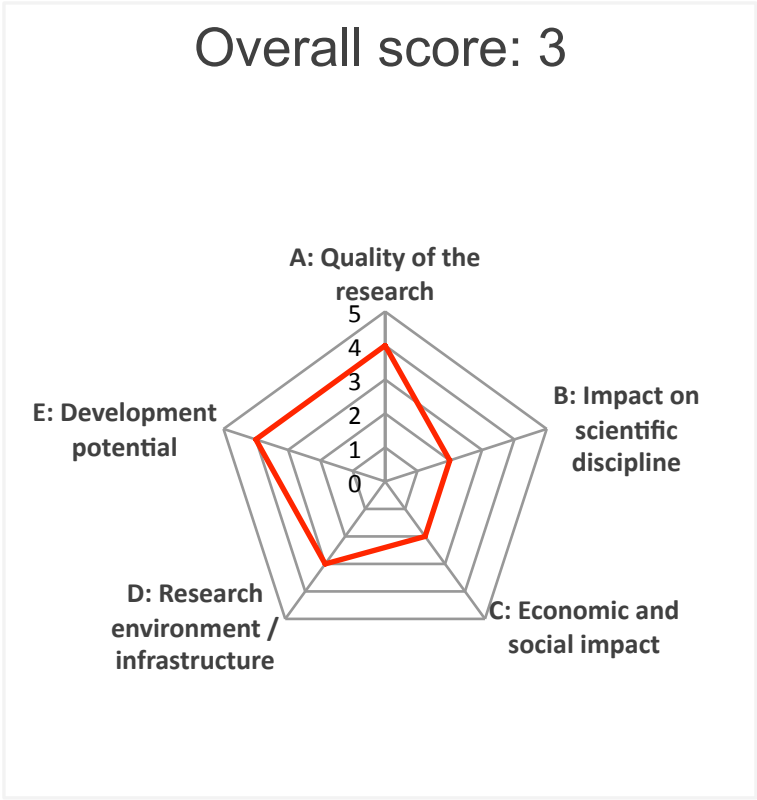


technopolis |group|

Panel report – institution level

Reports for each institution included

- **Overall score for Research Performance** and explanatory text
- **Scores for each of the five sub-elements**, with explanatory text for each
- **Recommendations** for the future development of the institution in the context of their area of research and the national science and innovation system. This may include:
  - *The potential evolution of the research environment and infrastructure, including strategic management and operational issues, composition of research staff etc.*
  - *Opinions regarding the potential for collaboration with other institutions and for interdisciplinary research*



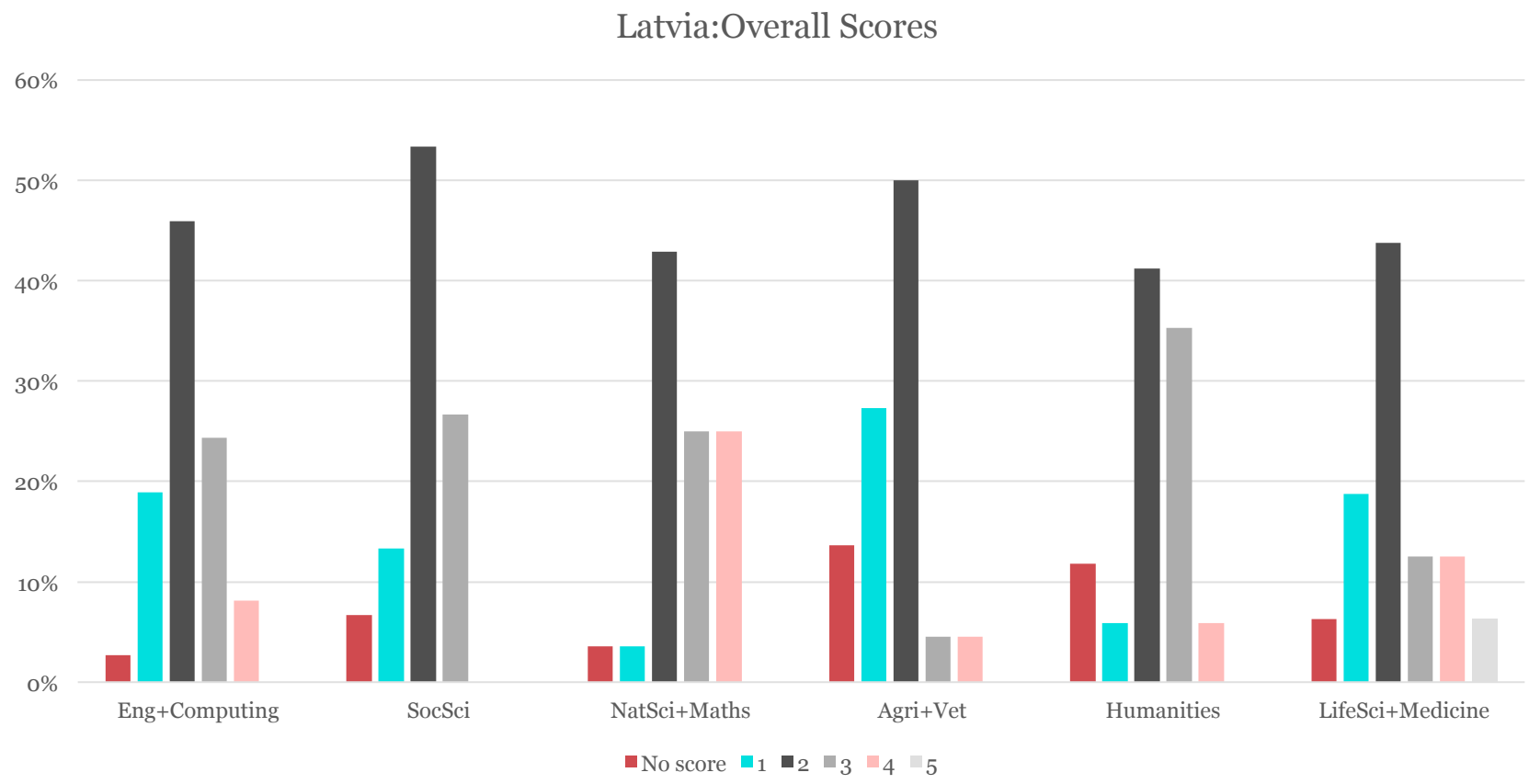
technopolis | group |

Quality criteria: trying to map research against global standards

A: QUALITY OF THE RESEARCH		
SCORE	**DEFINTION**	**Description**
	Particular factors to take into account	*Pure and applied research shall be evaluated as being of equal significance*
5	**Outstanding level of research**	In terms of the quality, the research output of an institution is **comparable with the best work in the same area of research**. The research possesses the requisite quality to meet **highest standard in terms of originality, significance and accuracy**. Work at this level should be **the primary point of reference** in the respective area
4	**Very good level of research**	Research by the institution possesses a very **good standard of quality** in terms of originality and importance. Work at this level can **arouse serious interest in the international academic community**, and **international publishers or journals with the most rigorous standards of publication** (irrespective of the place or language of publication**) could publish work of this level**.
3	**Good level of research**	**The importance of research by the institution is unquestionable** in the experts' assessment. Internationally recognized publishers or journals could publish work of this level.
2	**Adequate level of research**	The international academic community deems the significance of the research by the institution to be acceptable. **Nationally recognized publishers or journals could publish work of this level**.
1	**Poor level of research**	Research by the institution contains **new scientific discoveries only sporadically**. **The profile of the research by the institution is expressly national**, i.e., the institution is not involved in international debates of the scientific community. It focuses on introducing international research trends in Latvia.

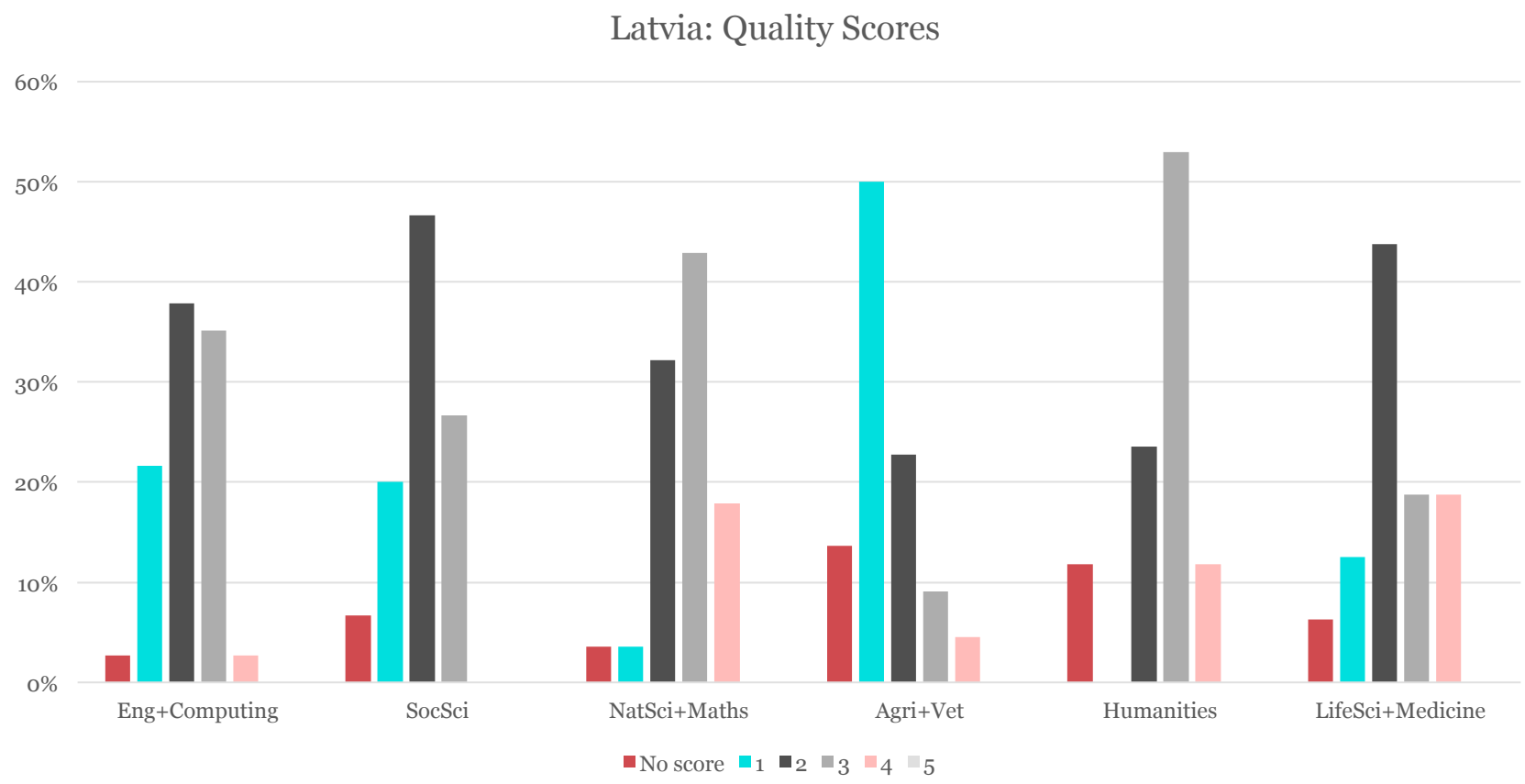
technopolis |group|

Latvia: overall scores



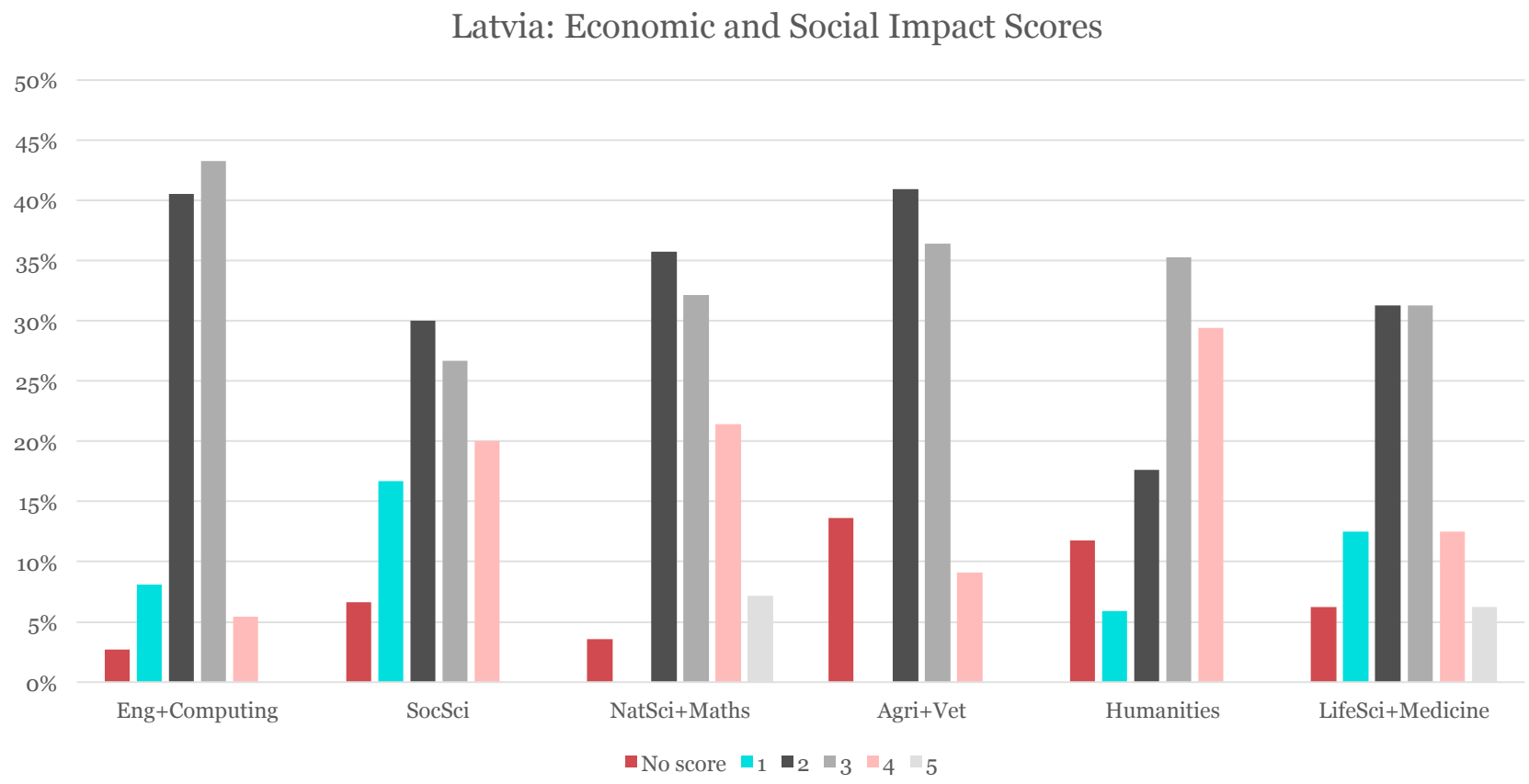
technopolis |group|

Latvia: Scientific quality scores



technopolis |group|

Latvia: impact scores



Norwegian SocSci institutes – evaluation dimensions (too many?)

- Relevance of the institute to its intended users
- Institute capability and quality of the research output
- Ability to recruit, retain and develop personnel
- Role of the institute in the Norwegian research structure
- Spatial reach of the institute's collaborations
- Social and economic impact of the institute's activities
- Quality and realism of the institute's strategy
- Appropriateness of framework conditions to the institute's mission
- Overall score and feedback to the institute

The quality scale

- 5 = Quality of an internationally leading standard, equal to the few best in the world
- 4 = Strong outputs, good enough to make the Institute visible and respected on the international scene, or equivalent to this level
- 3 = A good level nationally, though not necessarily likely to make the Institute more than occasionally visible outside Norway, or equivalent to this level
- 2 = Quality acceptable at the national or regional level, though not among the very best in Norway
- 1 = Unacceptably poor quality

SocSci panels, some observations

- Reluctance to score and particularly to publish scores
- Reluctance to work within a pre-defined evaluation process, preferring to devise this themselves
- Skepticism about quantitative evidence (user surveys, bibliometrics) on methodological grounds
- Desire to go to first principles in defining 'quality' and 'impact'
- Uncertainty about how to assess impact
- Discomfort with the use of international standards to criterion-reference scoring system – concern that this means a priori that 'international is good'
- Stress on the importance of reporting in the vernacular as well as English
- 'Two tribes': economists versus the rest

Humanities panels, some observations

- Varying degrees of acceptance of the imposed evaluation criteria and scoring systems
- Caution with scoring and reluctance to publish scores
- Disagreement about the use of an internationally criterion-referenced scale, occasionally leading to 'civil disobedience'
- Caution in use of publication indicators (eg Norwegian Publication Indicator scores as well as WoS, Scopus), in some cases partly based on unfamiliarity
- Two camps in regard to assessing the use of English vs the vernacular
- Uncertainty about how to assess impact

Interpretation

- Notion of 'quality' is contested but insufficiently defined within the discipline groups evaluated always to establish consensus
- 'Impact' a partly unfamiliar concept – hard to conceptualise or to compare across units of analysis
- Uncertainties compounded by evaluees' inexperience with being evaluated and with central evaluation concepts
- Implicit objections to evaluation per se, as an interference with academic freedom
- Willingness to use a defined evaluation and scoring system depends in part upon the employment/power relationship between the research funder and the panelists
- Early stage of developing 'evaluation culture' – is this where S&T was 30 years ago?

technopolis |group|

Observer effects: RAE/REF submission patterns by disciplinary groups

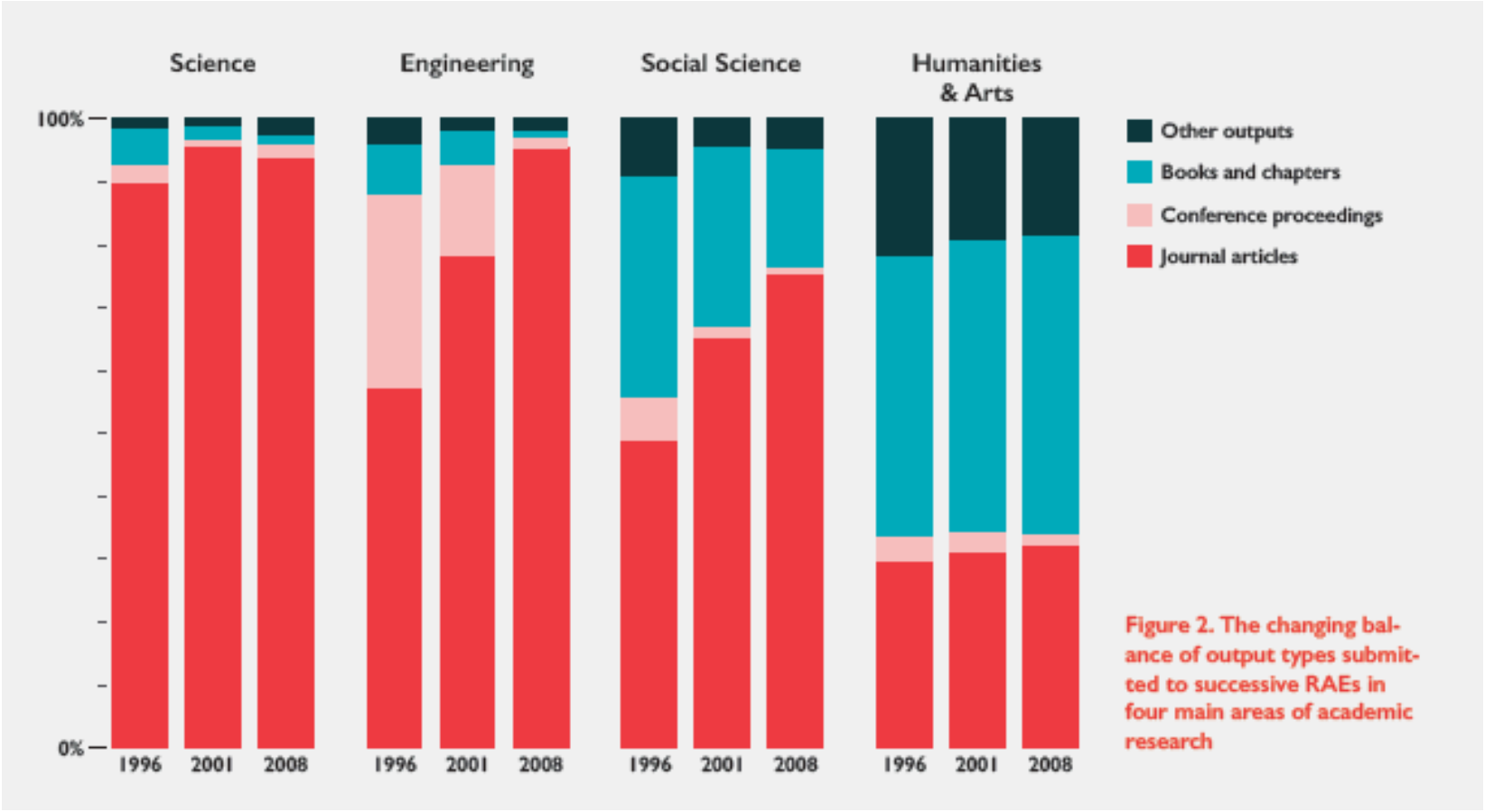


Figure 2. The changing balance of output types submitted to successive RAEs in four main areas of academic research