KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The development of a Standard Evaluation Protocol and the dilemmas that go with it

TOPICS

- Rapidly changing context of current research

- Dilemmas in developing a Standard Evaluation Protocol (SEP)

- SEP 2015-2012: philosophy and architecture

- SEP criteria and indicators

- Main challenges

EVALUATION IN CHANGING CONTEXT

- The world inside (academic) research is changing: growing opposition against focus on publications only ('publish or perish'), San Francisco declaration, Science in Transition (SiT): wider focus

- The world outside research is rapidly changing: institutional differentiation , topsector policy, H2020 / GSC

    - CTOs of major companies (Shell, AKZO, ASML, Philips, Unilever) push the Dutch government to produce a common inspiring knowledge agenda: oriented towards European challenges, multidisciplinary, beyond topsector policy

- Valorization and RRI: Ethics, public engagement, gender equality, science education, open access, governance

TRANSEPISTEMIC COMMUNITIES

Cf. Karin Knorr-Cetina transepistemic arena concept

The world outside the research lab, where researchers have to negotiate their views and interests in research with government, industry and society at large.

Academic disciplines having different research cultures, (communication, interaction, organisation), not independent of the surrounding policy and society (publish or perish, 'top sector' policy, ethics, integrity)

Different goals, expectations, interests in society (big industry, SMEs, local/national/supranational politics, public debate)

RESEARCH IN CONTEXT [@ TILO PROPP]
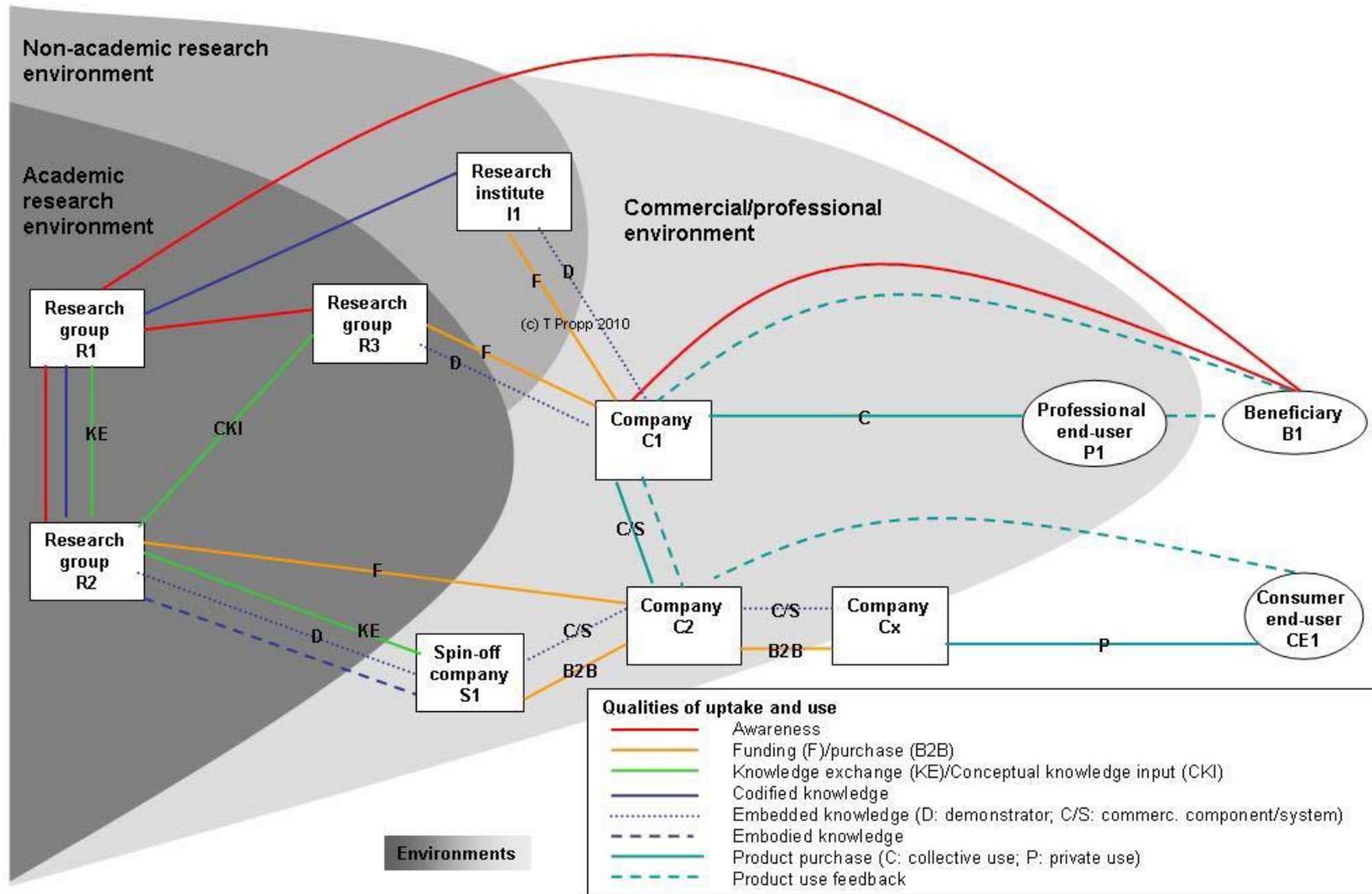
DILEMMAS FOR THE NEW PROTOCOL

How to make a assessment protocol that

- Is fit for all fields of research

- Has a balance between quantitative and qualitative indicators

- Is sensitive to variation of contextual variables

- Is fair and robust

- Is doable (meaning minimal administrative burden)

INDICATORS UNDER PRESSURE

- Bibliometric indicators (based on WoS) may have perverse effects ('publish or perish'), work better in natural and biomedical sciences than in social sciences, humanities, design and engineering fields, MIT research

- Indicators for societal relevance still in development, no robust data readily available, with few exceptions (patent statistics), lack of consensus about approach

- Indicators not fit for new (web based) ways of communication: open access, social media; but alternatives are in development, like webometrics, altmetrics, contextual response analysis

SOCIETAL IMPACT: METHODOLOGICAL ISSUES

- What is impact, societal impact? A linear concept, in reality interaction between stakeholders

- Attribution: who is responsible for which result, who is the sender and who is the receiver?

- Temporality : processes tend to continue over time, players trade places

- What is good result for the one is not necessarily good for the other

- → If knowledge develops in a network, network approaches are needed

- Gain knowledge about the network dynamics
- Different accountability and reward systems
- Different kind of data?

SOCIETAL IMPACT – SOCIETAL INNOVATION

- 'Impact' is the sum of many contributions by many different stakeholders

- Contributions vary from research articles to technical solutions to policy measures to end user preferences

- 'Impact' may refer to changes in human behavior, to organizational change, to conceptual innovation, to societal innovation

- Regards socio-economic, cultural, legal, political spheres of society

- Core of the 'impact' is in quality of life areas: food, health, housing, work, (permanent) education;

- Areas like food security, healthy aging, climate change, migration, urbanization, access to technology, opportunities for development

RIFT BETWEEN THEORY AND PRACTICE

The current state of the art in evaluation practice for measuring policy impacts does not match the concepts that are most common in the research literature to describe the connections between knowledge and policy.

The dominant concept in evaluation practice is linear, framed by logic models and the terminology of inputs, activities, outputs, and outcomes, sometimes with a loop back to planning. The policy process itself is a black box.

In contrast, the dominant concept in the research literature is the network or system, which is made up of many small conversations, interactions, and adjustments among a diverse set of actors; and complex concepts of the ebbs and flows of the policy process itself are incorporated. [Cozzens and Snoek, 2010]

GOVERNMENT NEW SPEAK: PLATEAU WITH PEAKS
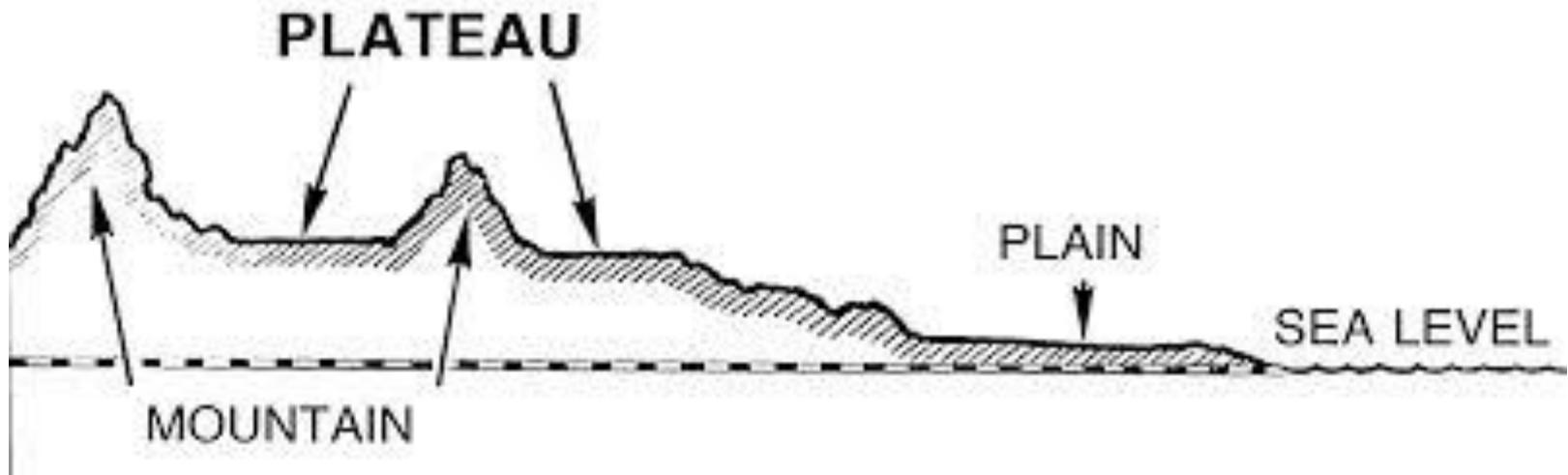
SEP 2015 – 2021: PHILOSOPHY

SEP 2015 – 2012: ARCHITECTURE

Self evaluation report including SWOT analysis

Goals of SEP
- Accountability to government and society
- Improvement of scientific quality, societal relevance, viability of research groups
- Verdict oriented (ex post) or strategic (forward looking): both

Focus of the SEP: research units of a reasonable size, not the individual researcher

Three main criteria: scientific quality, societal relevance, viability

Productivity no longer separate criterion (SiT discussion)

Societal relevance, valorization became more important

Review committees: allow for other expertise

SEP EVALUATION CRITERIA

o **Quality**: international recognition, innovative capacity

o **Productivity**: output in the SCI-journals, or SSCI, AHCI

o **Relevance**: societal impact

o **Viability** (flexibility, management aspects, leadership)

o **PhD programs** [no scores]

o **Research integrity** [no scores]

verdicts in four discrete categories: (1) world leading, (2) very good, (3) good, (4) unsatisfactory

used to be 5 (top), 1 (stop), verdicts were not so discrete

SEP PHILOSOPHY

o **Balance between criteria of scientific quality and societal relevance**

o **Productivity**: no longer separate criterion (SiT discussion)

o **Narratives:** allowed and stimulated for instances of societal relevance

o **Research integrity** new topic (fraud cases, but also data integrity)

o **New scores:** 4 discrete categories, top should be exceptional

o **Mixed review committees**

ASSESSMENT SCHEME FOR SEP RESEARCH

PEERS, OTHER EXPERTS and STAKEHOLDERS

Extended 'peer' review

EQUAL ATTENTION IN ASSESSMENT

Scientific quality

Societal relevance

INDICATORS BOTTOM UP

Output

Use

Recognition

SEP INDICATORS *[examples]*

	Scientific quality	Relevance to society
Demonstrable output	Sc. articles (refereed vs. non-refereed) Sc. books Other research outputs (instruments, infrastructure, datasets, softwaretools, designs) Dissertations	(policy) reports Articles in professional journals Other output (instruments, infrastructure, datasets, softwaretools, designs) Outreach-activitties, public lectures, exhibitions,
Demonstrable use	Citations Use of datasets, softwaretools, etc. by peers Use of research facilities by peers Revies in scholarly journals	Patents/licences Use of research facilities by societal partners Projects with societal partners Contract research
Demonstrable recognition	Scientific prizes Personal sc. subsidies Invited lectures Membership of sc. committees, editorial boards, etc.	Public prizes Valorisation funding Positions paid for by public parties Memberships of public advisory bodies

INDICATORS BOTTOM UP : HOW DOES THAT WORK

- Representative committees
- Focus groups, various stakeholders
- Pilot studies, institutes and programs
- Final report, proposal for indicators, inventory of challenges

➢ Quality Assessment in the **Design and Engineering** Disciplines (2010)

➢ Quality Indicators for Research in the **Humanitie**s (2011)

➢ Towards a Framework for the Quality Assessment of **Social Science** research (2013)

DIFFERENT RESEARCH FIELDS – SAME PROBLEMS

- Predominant bibliometric indicators (WoS based) not adequate for most of these fields

- Research oriented towards societal issues and stakeholders, not rewarded

- Lack of methods to assess their research production and communication

- Difficulties with visibility and recognition, not taken seriously in collaboration (top sectors, H2020)

DEVELOPING INDICATORS

Assessment criterion	Indicator development	Operationalisation
Articles, books, chapters, dissertations, ..	Journal ranking, publisher ranking, …	Institutional policy, attunement at national level
Other output	Experimental development	Who is going to do that? CWTS?

CHALLENGES

- Unbalance in development stage between indicators:

  - Between fields (N, M, vs. H and SS but also Engineering)
  - Between scientific and societal quality
  - Between hard and soft indicators

- Registration of data for indicators: many different systems

- The use of these indicators by review committees and policy makers: not very sophisticated (H-index still favorite)

- The fact that there is no national evaluation in most disciplines – comparability?

➢ Solution: bottom up, authorotative bodies within fields decide

MANY PROJECTS TO DEVELOP ALTERNATIVE INDICATORS

The three Academy studies

New SSH platform CWTS, Academy, Rathenau, others

European SSH platform (ENRESSH), UK research councils, USA (broader impact criteria NSF)

Evaluating Research in Context (www.eric-project.nl)

EU SIAMPI project on social impact measurement (www.siampi.eu)

→ No coordination, not even at the national level (some exceptions)

Thank you

Merci