ETH zürich

What about next-generation assessments?
The future of assessment: Bottom-up

Michael Ochsner, ETH Zurich & FORS, Lausanne

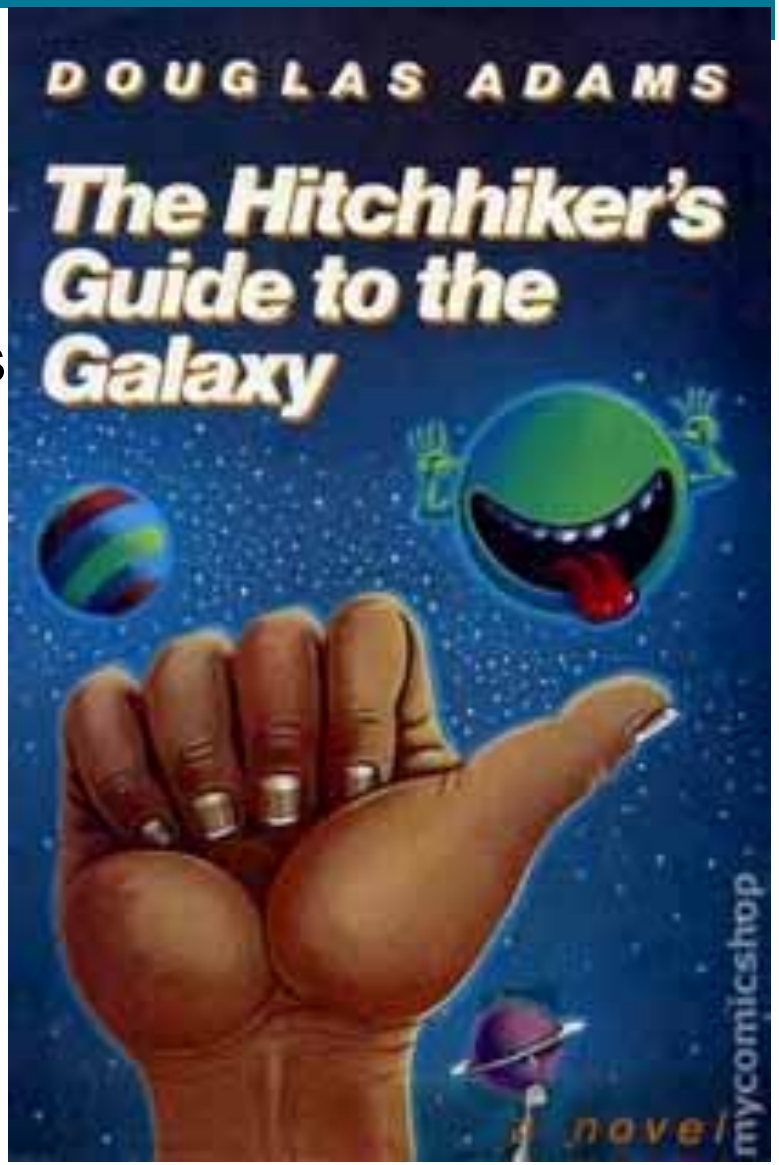
Conference: Beyond impact factor, h-index and rankings: Evaluate science in more meaningful ways, Bern, 21st November 2018

Presented work in collaboration with Sven E. Hug, Hans-Dieter Daniel, Mišo Dokmanović, Aldis Gedutis and Emanuel Kulczycki

Outline

- Validity or measuring implies Concepts
- Projects on quality, performance and potential in SSH
- 42 or Responsible Metrics?
- The Future: Bottom-Up Evaluation Procedures

42: Metrics and Concepts



- Novel/Radio play by Douglas Adams
  Hitchhiker's Guide to the Galaxy
  (1979)
- the ultimate question of life,
  the universe and everything
  - 7.5 million years to compute and check
  - The answer was.... **42**
- Deep Thought:
  answer is meaningless –
  because the question was stupid:
  - did not specify the form of answer
    nor did they really know what they asked for

Validity

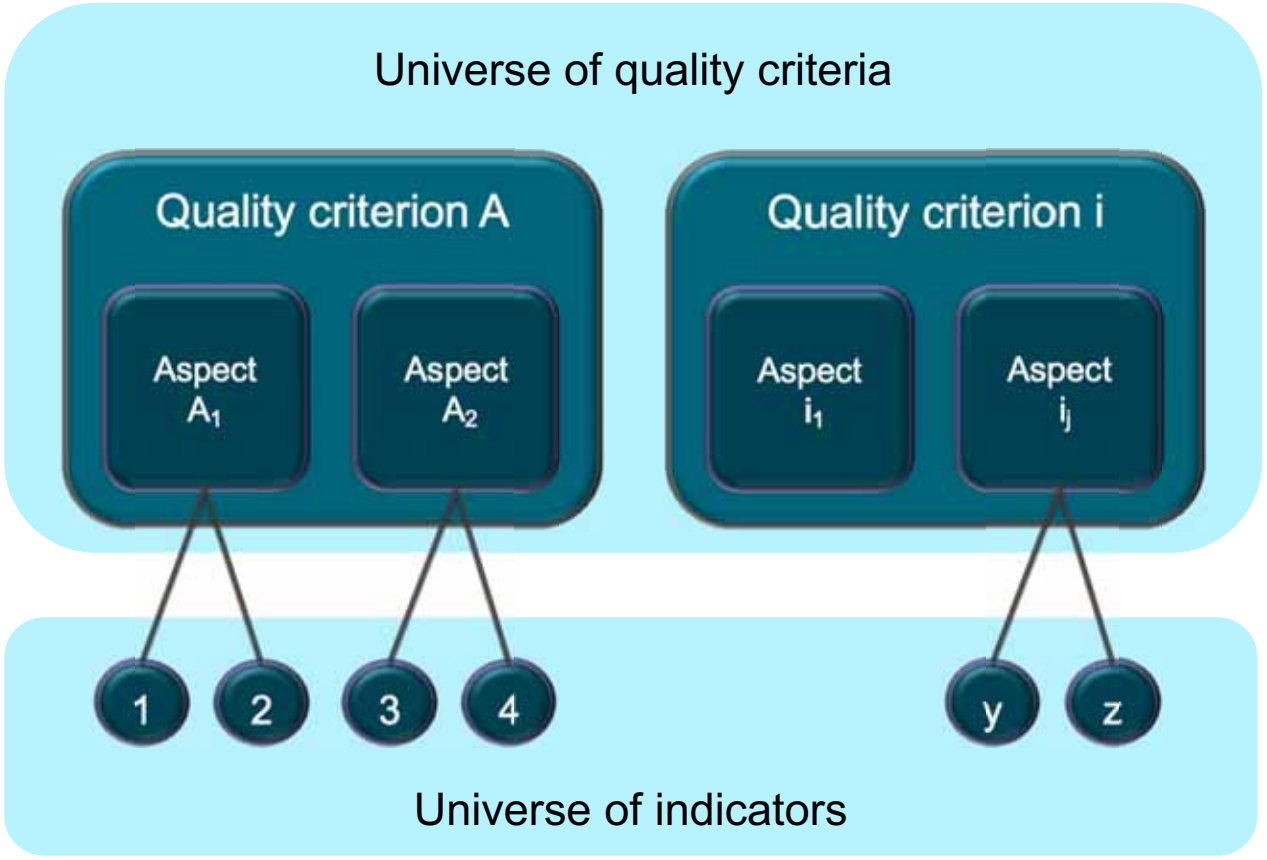
- The extent to which a measure (i.e., an indicator) actually measures what it purports to measure (i.e., a concept) (Borsboom et al., 2004, p. 1061)
- Data-driven: „measuring what can be measured“ endangers validity, mostly reducing it to correlation.
- Thunder correlates highly with lightning (and there is even a causal relationship). However, lightning cannot measure thunder.

Measurement in Scientometrics

- Bibliometrics and Scientometrics:
  - „[the assessments] often still make only a weak connection between theoretical definitions of quality and its measures“ (Brooks, 2005, p. 1-2)
  - „these metrics do not actually measure research quality. For example, research income is an input, rather than an output measure“ (Donovan, 2007, p. 586)
  - „It is [...] extremely difficult if not impossible to express what citations measure in one single theoretical concept [...]. Citations measure many aspects of scholarly activity at the same time.“ (Moed, 2005, p. 221)
  - Tahamtan & Bornmann (2017) review literature why authors cite. Quality of research is among a plethora of other reasons

Measurement Approach

Concepts



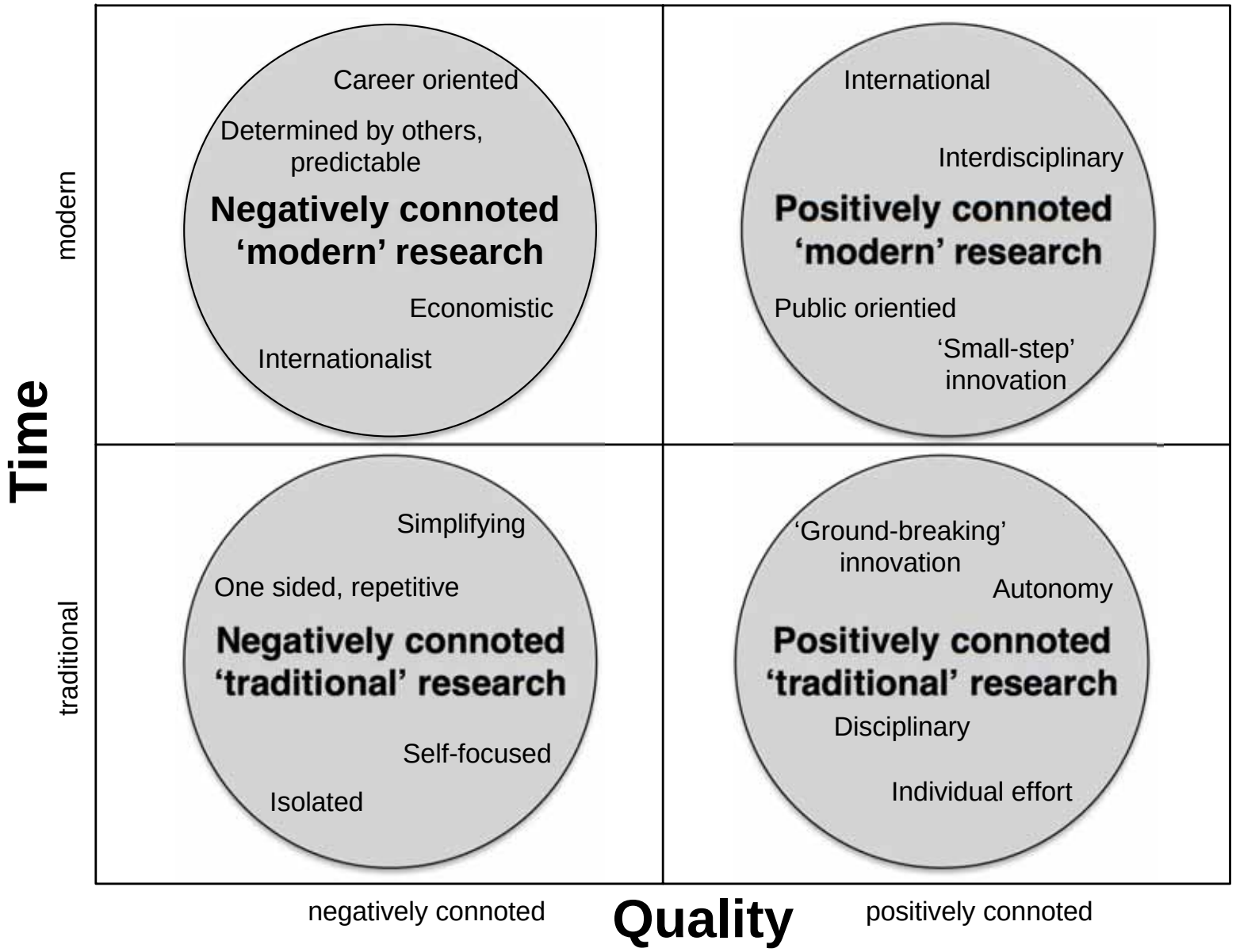
What?

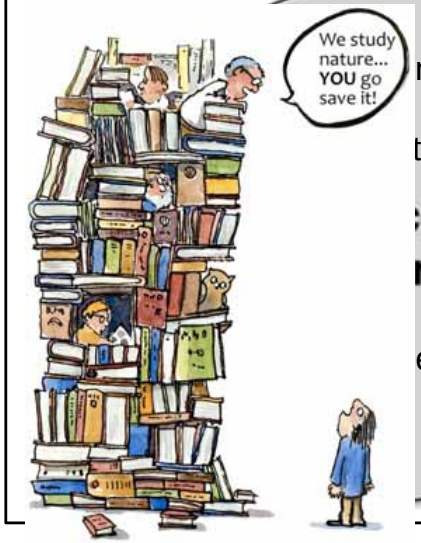
Indicators

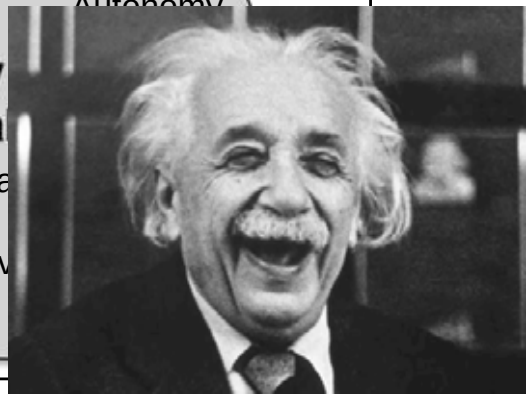
How?

ETH zürich

Projects on Quality, Performance and Career/Research Potential in the SSH


Career oriented
Determined by others, predictable
Negatively connoted 'modern' research
Economistic
Internationalist
International
Interdisciplinary
Positively connoted 'modern' research
Public orientied
'Small-step' innovation
Simplifying
One sided, repetitive
Negatively connoted 'traditional' research
Self-focused
Isolated
'Ground-breaking' innovation
Autonomy
Positively connoted 'traditional' research
Disciplinary
Individual effort
modern
traditional
Time
negatively connoted
Quality
positively connoted





Time

modern

traditional

negatively connoted

Quality

positively connoted

Criteria for Research Quality/Performance

- English Literature, German Literature and Art History
- Consensual Indicators (orange: all three; blue: in two disciplines)

1. Scholarly exchange
2. Innovation, originality
3. Productivity
4. Rigour
5. Fostering cultural memory
6. Recognition
7. Reflection, criticism
8. Continuity, continuation
9. Impact on research community
10. Relation to and impact on society
11. Variety of research
12. Connection to other research
13. Openness ideas and persons
14. Self-management, independence
15. Scholarship, erudition
16. Passion, enthusiasm
17. Vision of future research
18. Connection between research and teaching, scholarship of teaching
19. Relevance

Measuring Research Quality/Performance

- Which aspects can be measured by indicators?
  - 50% of the relevant aspects cannot be measured by indicators
- What do Indicators that are commonly used measure?
  - Are these the relevant criteria?

Table 1: Frequently used indicators and criteria they can potentially measure

Indicators	Criterion
Citations	Recognition; impact on research community; relevance
Prizes	Recognition; impact on research community; relevance
Third party funding	Recognition; impact on research community; relevance; relation to and impact on society
Collaborations	Scholarly exchange; recognition
Transfers to society and economy	Relation to and impact on society
Publications	Scholarly exchange; productivity
Board memberships	Scholarly exchange; recognition; impact on research community
Recruitment	Continuity, continuation

Criteria for Research Quality/Performance

- Measured by commonly used indicators (***bold and italic***)

1. ***Scholarly exchange***
2. Innovation, originality
3. ***Productivity***
4. Rigour
5. Fostering cultural memory
6. ***Recognition***
7. Reflection, criticism
8. ***Continuity, continuation***
9. ***Impact on research community***
10. ***Relation to and impact on society***
11. Variety of research
12. Connection to other research
13. Openness ideas and persons
14. Self-management, independence
15. Scholarship, erudition
16. Passion, enthusiasm
17. Vision of future research
18. Connection between research and teaching, scholarship of teaching
19. ***Relevance***

Validity Check for Commonly used Metrics

- Valid measures for research quality?
  orange: three disc.; blue: two disc.; ***bold and italic***: commonly used

1. ***Scholarly exchange***
2. Innovation, originality
3. ***Productivity***
4. Rigour
5. Fostering cultural memory
6. ***Recognition***
7. Reflection, criticism
8. ***Continuity, continuation***
9. ***Impact on research community***
10. ***Relation to and impact on society***
11. Variety of research
12. Connection to other research
13. Openness ideas and persons
14. Self-management, independence
15. Scholarship, erudition
16. Passion, enthusiasm
17. Vision of future research
18. Connection between research and teaching, scholarship of teaching
19. ***Relevance***

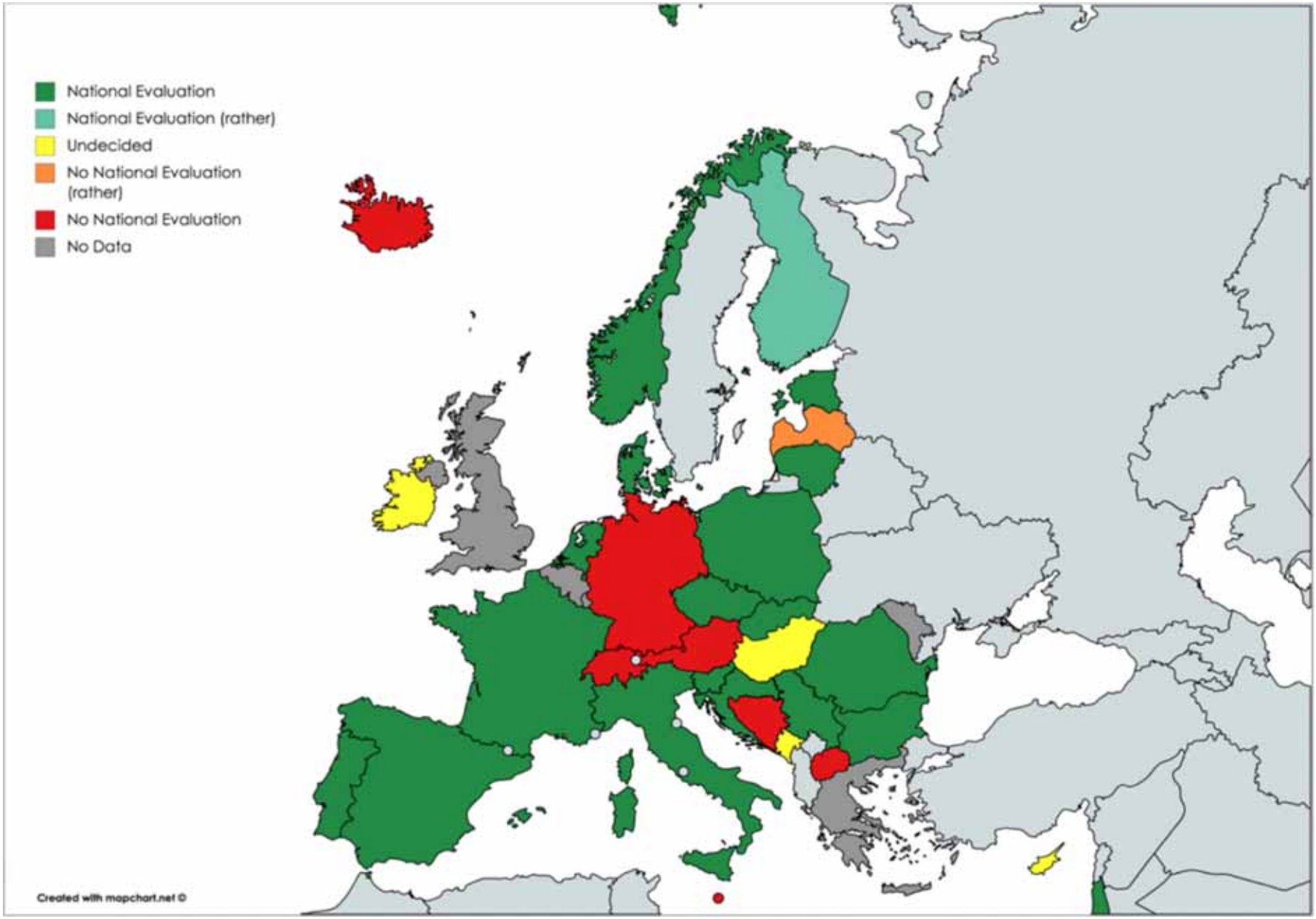
Other Applications

- Questions
  - How can the criteria be used in specific evaluation situations?
  - How do the criteria travel across more disciplines?
- Applications
  - Criteria for Grant Evaluation for Early Career Investigators
  - Social Sciences (Ochsner & Dokmanović, 2017)
  - Law (Lienhard et al., 2016), Theology (Mertens et al.)

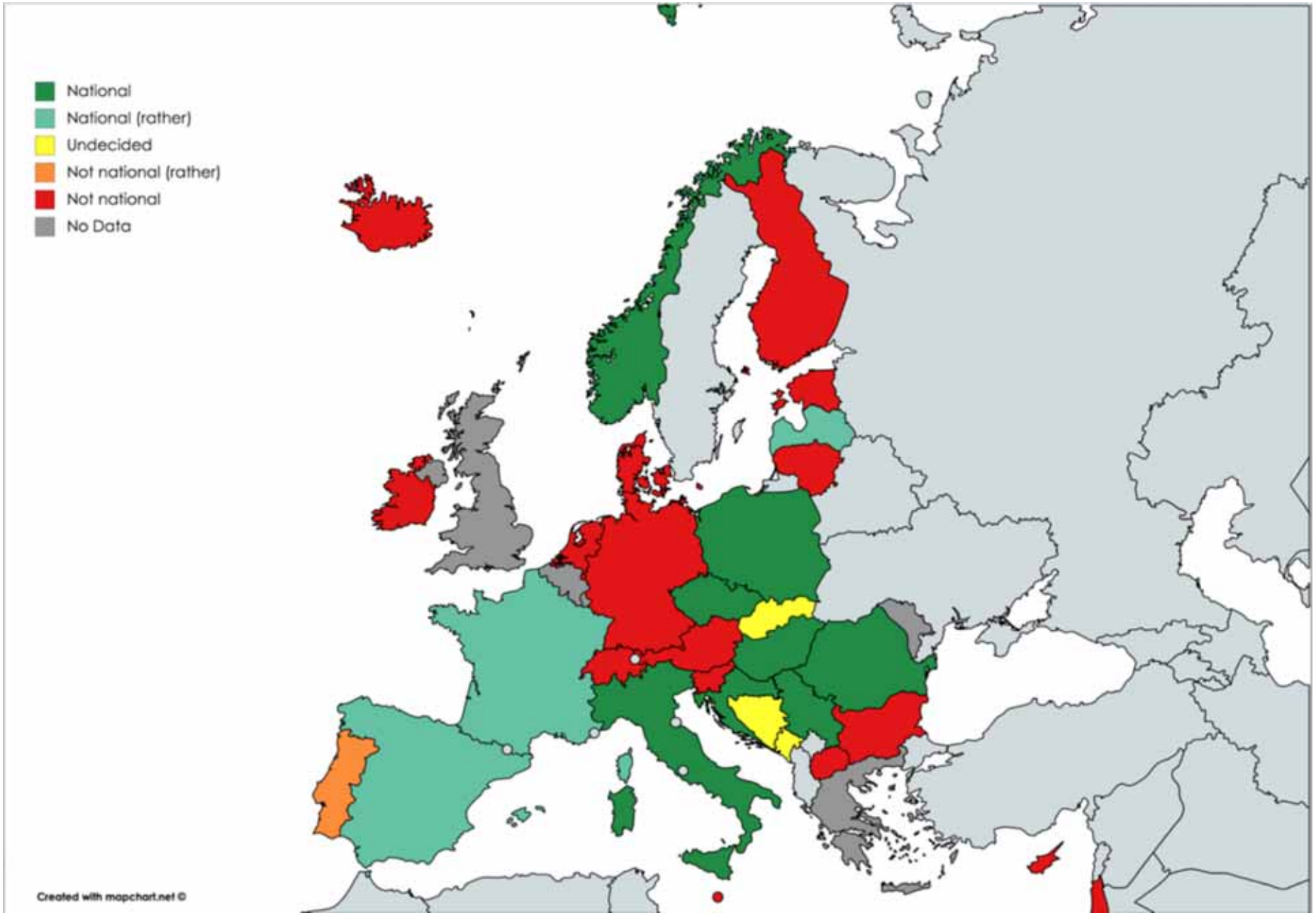
European Network for Research Evaluation in Social Sciences and Humanities

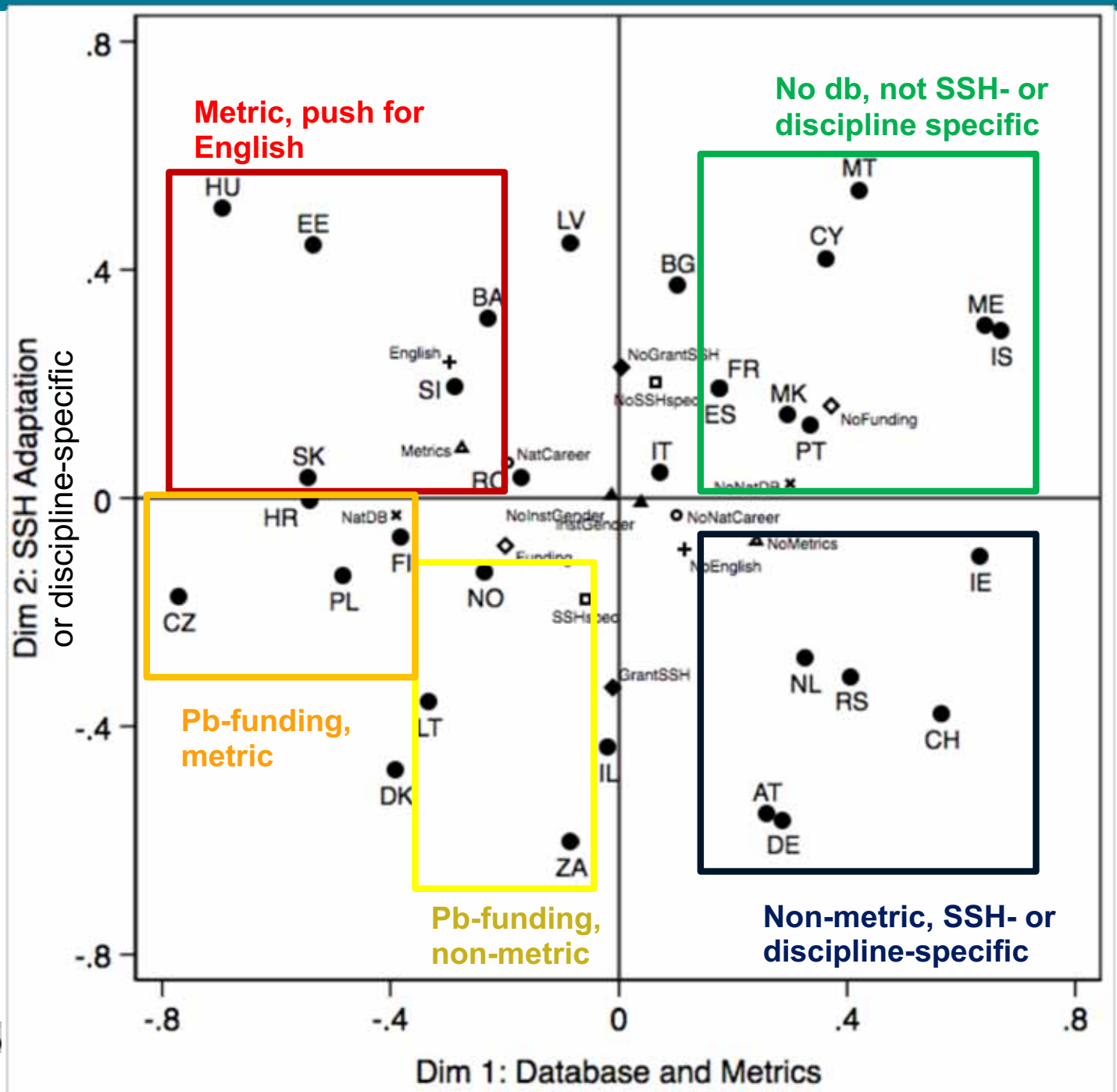
- COST-Action ENRESSH
- Set out to find responsible research evaluation procedures
- Including a metrics part but not limited to it
- www.enressh.eu

National Evaluation Procedure



National Career Promotion




Metric, push for English
No db, not SSH- or discipline specific
Pb-funding, metric
Pb-funding, non-metric
Non-metric, SSH- or discipline-specific
Dim 2: SSH Adaptation or discipline-specific
Dim 1: Database and Metrics
.8
.4
0
-.4
-.8
HU
EE
LV
BA
English
SI
SK
Metrics
NatCareer
RO
HR
NatDB
FI
PL
CZ
NO
LT
DK
IL
ZA
NoGrantSSH
NoSSHspec
BG
IT
MT
CY
ME
IS
FR
MK
ES
PT
NoFunding
NoNatCareer
NoMetrics
NoEnglish
IE
NL
RS
CH
AT
DE
GrantSSH

ETH zürich

What kind of Next-Generation Assessment Procedures?

Summary

- Most criteria travel well across disciplines and situations
  - SS & H have been investigated
  - STEM disciplines: most criteria seem to be valid as well
- Disciplinary differences exist
  - Most often only in the weighting (importance) of the criterion
- Societal Impact is not equal to Quality
- Science Europe: Best way to provide value to society is fostering **quality of research** (Science Europe, 2017)
- Metrics are not measuring quality comprehensively (<50%)

Conclusion: We know the answer anyway – 42

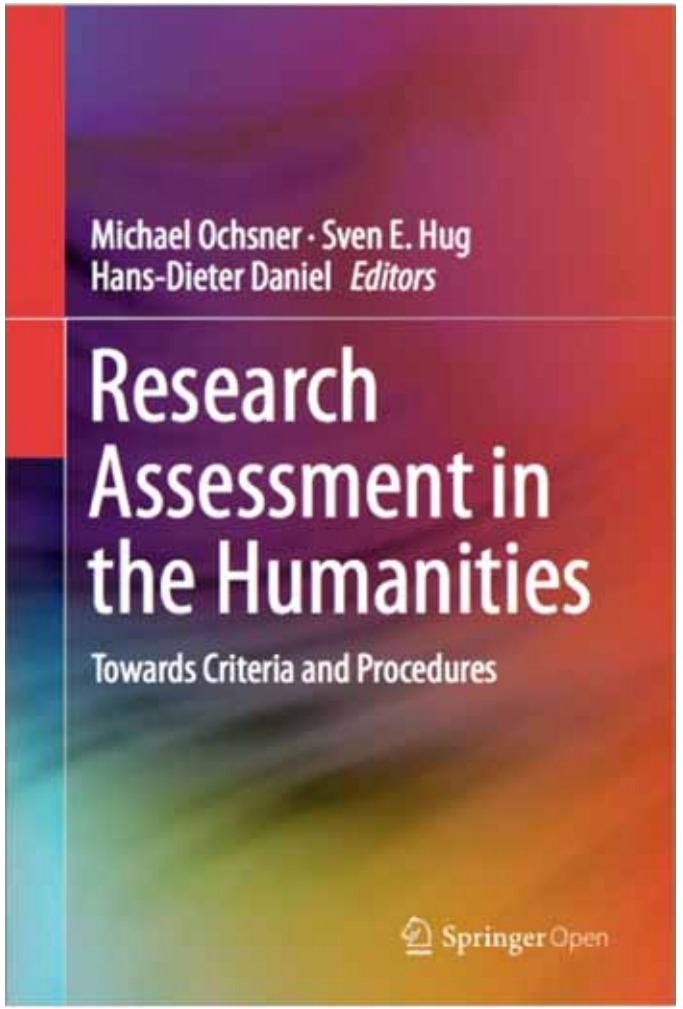
- Deep Thought created a new solution including beings that will resolve the question: Planet Earth, directed by white lab mice
  - Calculating time: 10 million years.
  - Earth destroyed before the result was ready by Psychiatrists who feared loss of their careers

Next-Generation Evaluation Procedures

- Importance of Criteria
  - Criteria lead to more robust and fair assessments (Thorngate et al., 2009)
  - Bottom-Up procedure leads to legitimacy (Hug, Ochsner & Daniel, 2014; Derrick & Samuel, 2017)
  - Metrics are never responsible, the users should be responsible
- Specificities of Bottom-Up Procedures
  - Respect disciplinary differences and levels of assessment
  - Based on research practices and knowledge production
  - Assure diversity and think of incentives metrics/criteria introduce
  - Transparency with decisions taken

Open Access Edited Volume on Bottom-Up Assessment Procedures in the Humanities

- Ochsner, M., Hug, S. E., & Daniel, H.-D. (Eds.). (2016). Research Assessment in the Humanities. Towards Criteria and Procedures. Cham: Springer Open. http://doi.org/10.1007/978-3-319-29016-4
- With contributions by i.a.: Wiljan van den Akker, Alfred Hornung, Wilhelm Krull, Michèle Lamont, Gerhard Lauer, Christian Mair, Ingo Plag, Björn Hammarfelt, Ingrid Gogolin, Gunnar Sivertsen, Elea Giménez-Toledo, Thomas König, Remigius Bunia



Literature on the projects (logical order)

- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. International Journal of Education Law and Policy, 10(1), 55–68.
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. Research Evaluation, 22(2), 79–92. http://doi.org/10.1093/reseval/rvs039
- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. Research Evaluation, 22(5), 369–383. http://doi.org/10.1093/reseval/rvt008
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012). Indicators for Research Quality in the Humanities: Opportunities and Limitations. Bibliometrie - Praxis und Forschung, 1, 4.
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2014). Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. Zeitschrift für Erziehungswissenschaft, 17(6 Suppl.), 111–132. http://doi.org/10.1007/s11618-014-0576-4
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2017). Assessment Criteria for Early Career Researcher's Proposals in the Humanities. *Proceedings of the 16th International Conference on Scientometrics and Informetrics of the International Society of Scientometrics and Informetrics.* 16.-20.10.2017 in Wuhan, China.
- Ochsner, M., Hug, S. E., & Galleron, I. (2017). The future of research assessment in the humanities: bottom-up assessment procedures. Palgrave Communications, 3, 17020. doi:10.1057/palcomms.2017.20
- Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems (pp. 1234–1241). Presented at the Proceedings of the 23rd International Conference on Science and Technology Indicators, Leiden.
- Peric, B., Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). AHRABi. Arts and Humanities Research Assessment Bibliography. Zurich: ETH Zurich. http://doi.org/10.3929/ethz-a-010610785

References

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. Psychological Review, 111(4), 1061–1071. http://doi.org/10.1037/0033-295X.111.4.1061
- Brooks, R. L. (2005). Measuring university quality. The Review of Higher Education, 29(1), 1-21.
- Derrick, G., & Samuel, G. (2017). The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability. Palgrave Communications, 3, palcomms201740. http://doi.org/10.1057/palcomms.2017.40
- Donovan, C. (2007). The qualitative future of research evaluation. Science and Public Policy, 34(8), 585–597. http://doi.org/10.3152/030234207X256538
- Lienhard, A., Tanquerel, T., Flu¨ckiger, A., Amschwand, F., Byland, K., and Herrmann, E. et al. (2016) Forschungsevaluation in der Rechtswissenschaft: Grundlagen und empirische Analyse in der Schweiz. Bern: Stämpfli.
- Moed, H. F. (2005). Citation analysis in research evaluation. Dordrecht: Springer.
- Ochsner, M., & Dokmanović, M. (2017). Quality criteria and research obstacles in the SSH in Macedonia. Book of abstracts of „RESSH 2017. 2nd International Conference on Research Evaluation in the Social Sciences and Humanities“ (pp. 69-70). University of Antwerp, Antwerp, Belgium. https://www.uantwerpen.be/images/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.pdf
- Tahamtan, I., & Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, *12*(1), 203–216. http://doi.org/10.1016/j.joi.2018.01.002
- Thorngate, W., Dawes, R. M., & Foddy, M. (2009). Judging merit. New York, NY: Psychology Press.

Contact

- Michael Ochsner, ETH Zurich & FORS, Lausanne, Switzerland
  ochsner@gess.ethz.ch
  http://forscenter.ch/de/about-us-3/staff/michael-ochsner/
  http://www.psh.ethz.ch/en/people/michael-ochsner.html

- COST- Action CA-15137:
  European Network for Research Evaluation in the Social Sciences and Humanities ENRESSH
  www.enressh.eu
- EvalHum:
  European Association for Research Evaluation in the Social Sciences and Humanities
  www.evalhum.eu