

The European Commission's science and knowledge service

Joint Research Centre



Step 1 & 2

Frameworks and Indicators

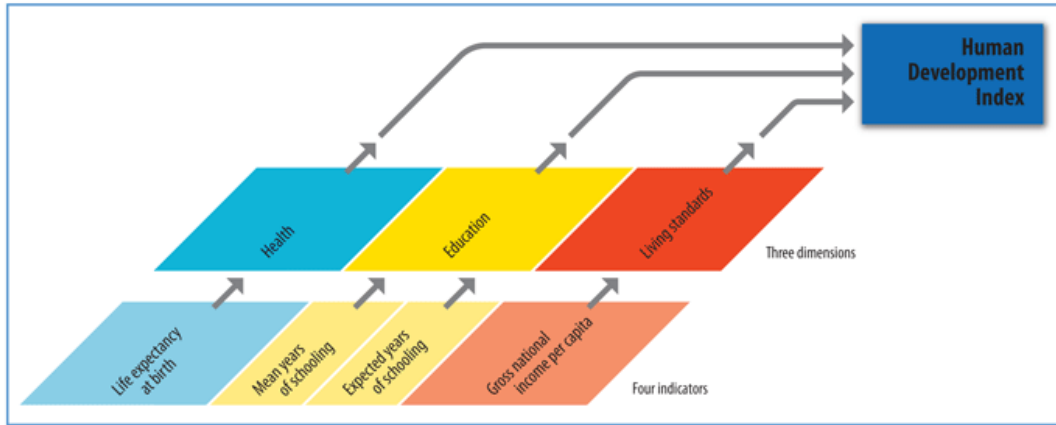
Dániel Vértesy

COIN 2017 - 15th JRC Annual Training on Composite Indicators & Scoreboards
06-08/11/2017, Ispra (IT)

Frameworks

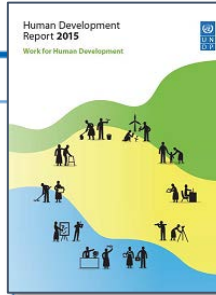
Components of the Human Development Index

The HDI—three dimensions and four indicators



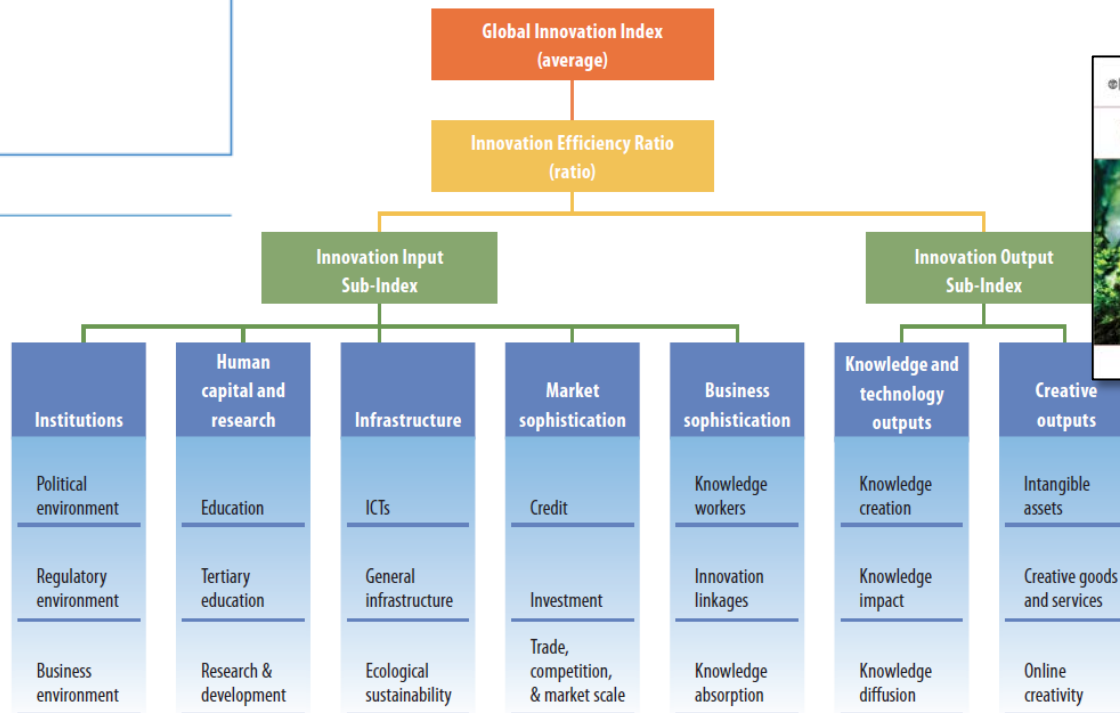
Note: The indicators presented in this figure follow the new methodology, as defined in box 1.2.

Source: HDRO.



Some indices seem to have a **simple framework...**
 ← Human Development Index (UNDP)
 4 indicators – 3 dimensions

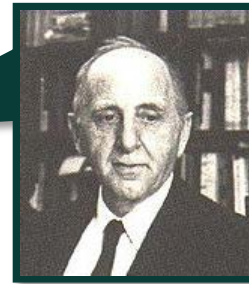
...Others look **complex**
 Global Innovation Index 2017
 (WIPO, Cornell, INSEAD) →
 7 pillars –
 21 sub-pillars –
 82 indicators



What framework to choose? What indicators?

I.e., what is **development**? How should we measure it?

- In terms of wealth (production)?
→ **GDP** is readily available (see S. Kuznets)



- Capabilities and choices (A. Sen)
→ **no single indicator**
 - What are the relevant indicators?
 - Are they universal?
 - Does strength in one dimension compensate weakness in another?

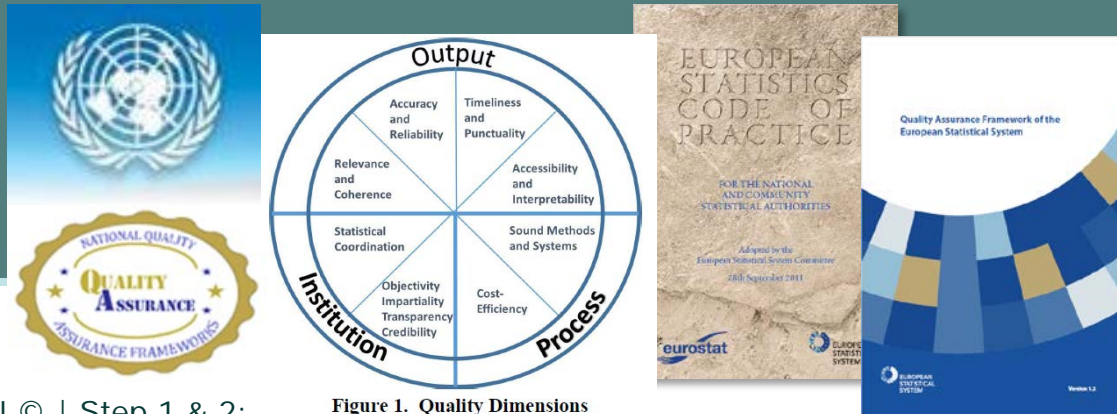


Quality profile of [composite] indicators

- Quality = accuracy?
[closeness of computation to the “exact true values”]
- Quality = fitness for use?
[depends on user needs, values, priorities]

- **Quality assessment frameworks:**

for Official Statistics



for Composite Indicator Frameworks



Assessment frameworks

1. The UNSD's Nat'l Quality Assurance Framework

- Refers to individual indicators, but relevant also for their combinations...
- Focus both on development process & results

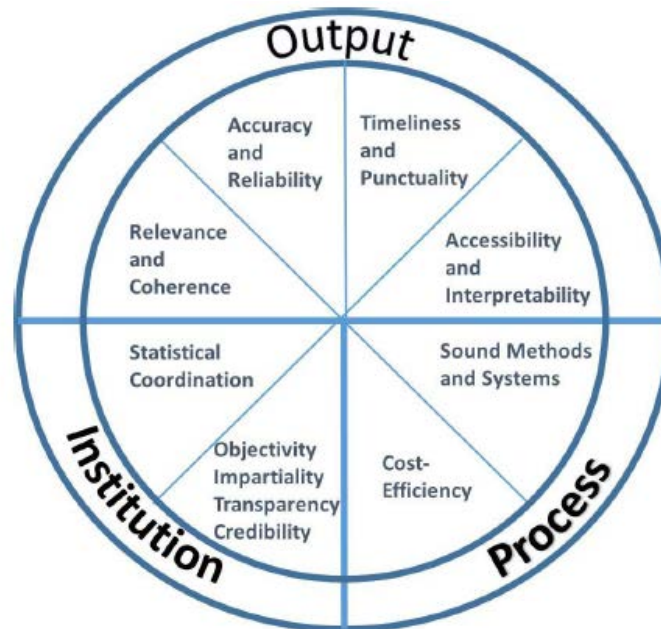


Figure 1. Quality Dimensions



Statistics Division

Towards a National Quality Assurance Framework: the UN Statistical Commission Initiative



Managing statistical processes

- Assuring methodological soundness
- Assuring cost-effectiveness
- Assuring soundness of implementation
- Managing the respondent burden

Managing statistical outputs

- Assuring relevance
- Assuring accuracy and reliability
- Assuring timeliness and punctuality
- Assuring accessibility and clarity
- Assuring coherence and comparability
- Managing metadata

2. European Statistics Code of Practice

Code of Practice (CoP) & ESS Quality Assurance Framework (QAF)

15 principles focusing on the

- Institutional Environment,
- Statistical Processes,
- **Statistical Output [11-15]**

PRINCIPLE
11

RELEVANCE



- 11. Relevance:** meet the needs of users
- 12. Accuracy and reliability:** accurately and reliably portray reality.
- 13. Timeliness and punctuality:** released in a timely and punctual manner
- 14. Coherence and Comparability:** consistent internally, over time and comparable between regions and countries; it is possible to combine and make joint use of related data from different sources.
- 15. Accessibility and Clarity:** presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

See: <http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>

3. The 10-step methodology

- Focus on development process
- Distinguish 10 steps in what is a very iterative process
- Has served as input for audits of indicators ([Saisana and Saltelli, 2010, 2012](#))

- Step 1. Developing a theoretical framework
- Step 2. Selecting variables
- Step 3. Imputation of missing data
- Step 4. Multivariate analysis
- Step 5. Normalization of data
- Step 6. Weighting and aggregation
- Step 7. Robustness and sensitivity tests
- Step 8. Back to the details (de-constructing composites)
- Step 9. Association with other variables
- Step 10. Presentation and dissemination

(JRC-OECD, 2008)



4. The “Gisselquist framework”



A framework of 10 questions to guide the development and evaluation of composite indexes

(Developed for governance indicators)

Focuses on:

- Concept, definition, operationalization
- + Data + Quality of methodology

Critical Questions:

1. What precisely does it aim to measure?
2. Content validity: does the operational definition capture the concept?
3. How good (reliable, valid and complete) are the data used?
4. Is the measure (including all of its sub-components) transparent and replicable?
5. How sensitive and robust is the measure to different data and design choices?
6. Does the measure allow the analyst to address key questions of interest?

Less Critical Questions:

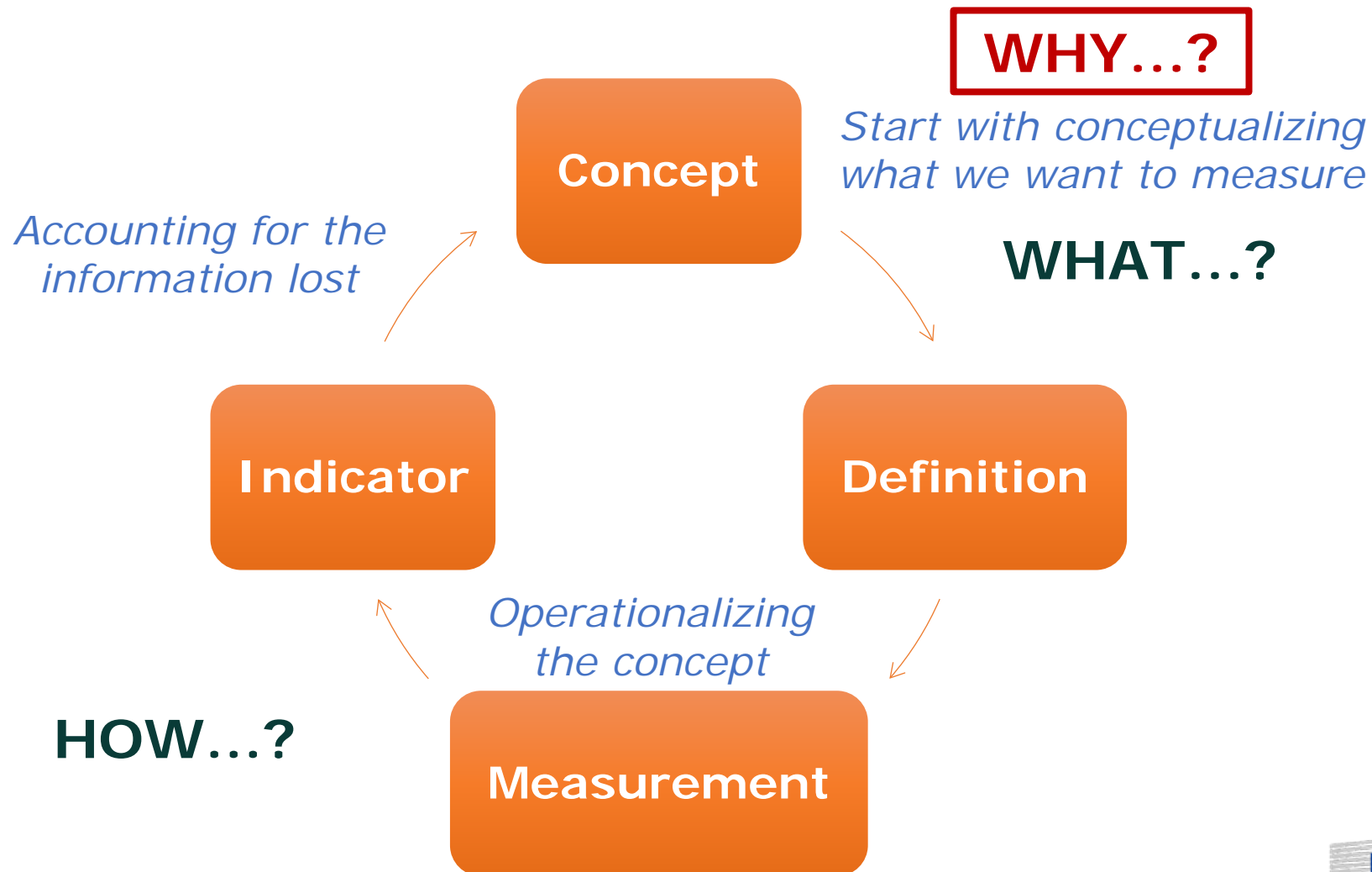
7. Does the measure fully capture [the concept of interest] in all its complexity? [descriptive compl.]
8. Does the measure behave as theory predicts? [theoretical fit]
9. How precise are index values and are confidence intervals specified? [precision of estimates]
10. Is the weighting 'correct'?

(Gisselquist, 2014)

Key message: 3 questions to clarify

- **What** do we want to measure?
- **Why** do we want to measure it?
- **How** do we want to measure it?

Operationalizing the Concept



An example: What is innovation?



Oslo Manual Definition: “the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organisation or external relations”

< - > **New combinations** [of knowledge, capabilities, resources, skills]?

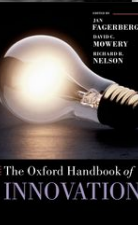
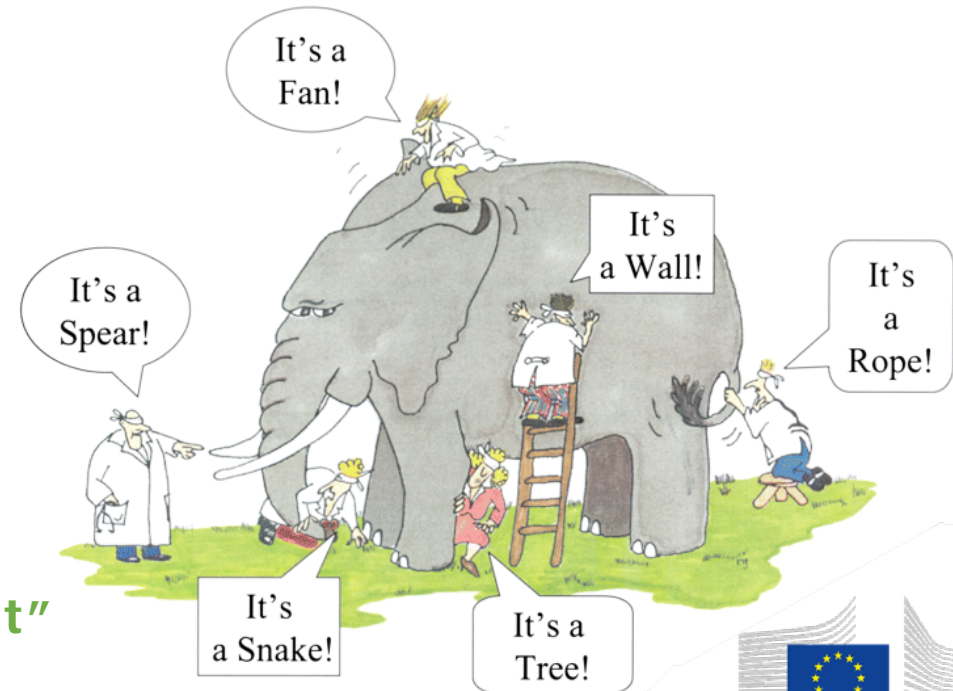
What type of innovations?

- Radical Innovations?
- Incremental Innovations?
- Diffusion of innovations?

Who defines what is innovative?

- The producer?
- The consumer?
- The researcher/expert?

J.G. Saxe's “The Blind Men and the Elephant”



Why do you need to measure innovation?

What is the **added value** of a new index or scoreboard?

- Intended purpose(s)?

Who would be the **main users**?

- Are they involved in the development process?
- Level of focus? (Unit of analysis =? Countries? Universities?)
- Any limitations?



An example: Measuring innovation

Innovation is a **complex process**, involves **multiple actors**, with different functions, values & interests

Expenditure on
innovation vs. R&D?

Interactions

New-to-world or new-
to-market products

Competitiveness
vs. jobs?

Whether you assess the **effort**, the **process**, the **outcomes**, or the **impacts** of Innovation, you will likely need different information; and attribute different importance to the elements

How do you need to measure innovation?

Measure the effort? [R&D target]

Measure the complexity? [GII]

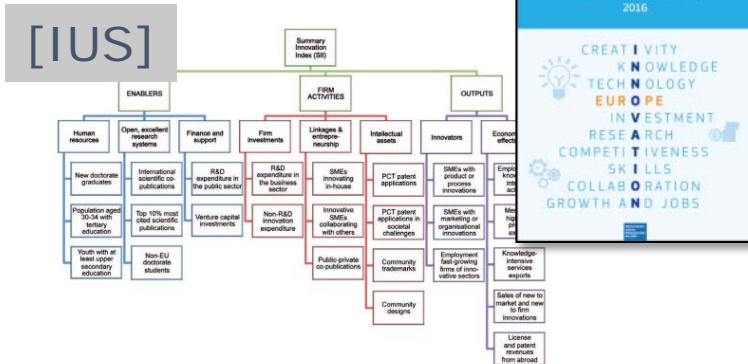


- I.e., The GII [aims to] to better capture the richness of innovation in society and go beyond such traditional measures of innovation as the number of research articles and the level of R&D expenditures

Measure the performance of innovation systems?

[IUS]

Measure the output? [IOI]



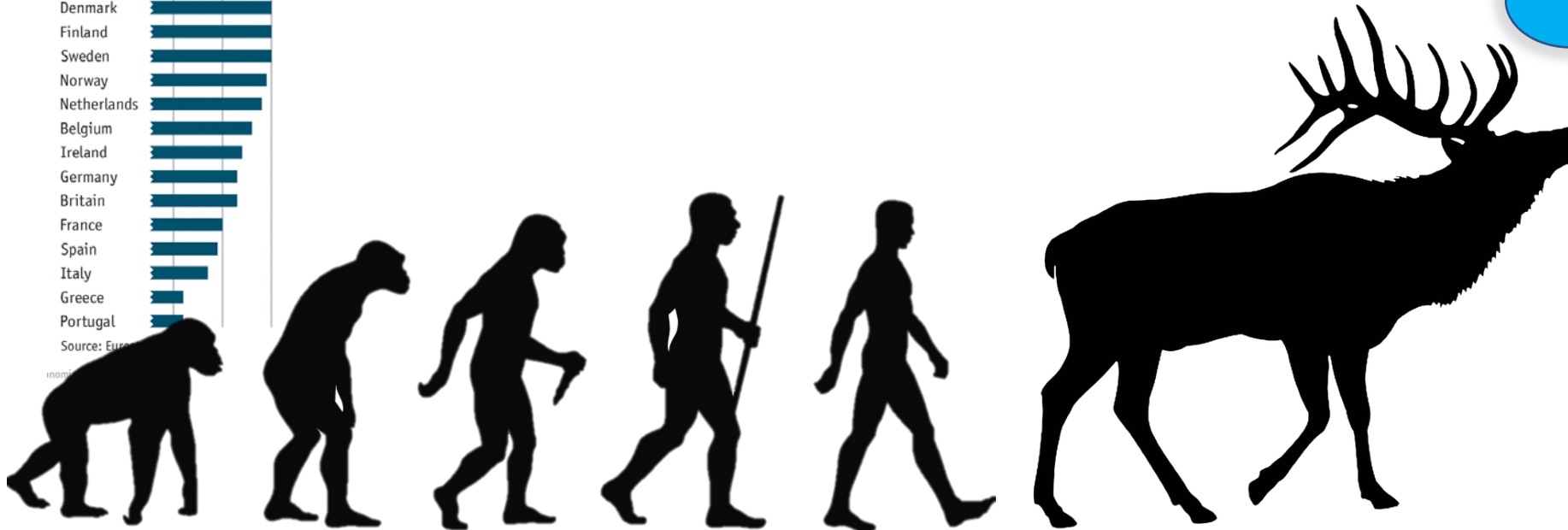
What's the added value of a new index?

Happiness in Europe

Self-reported, 2013, 10=most happy



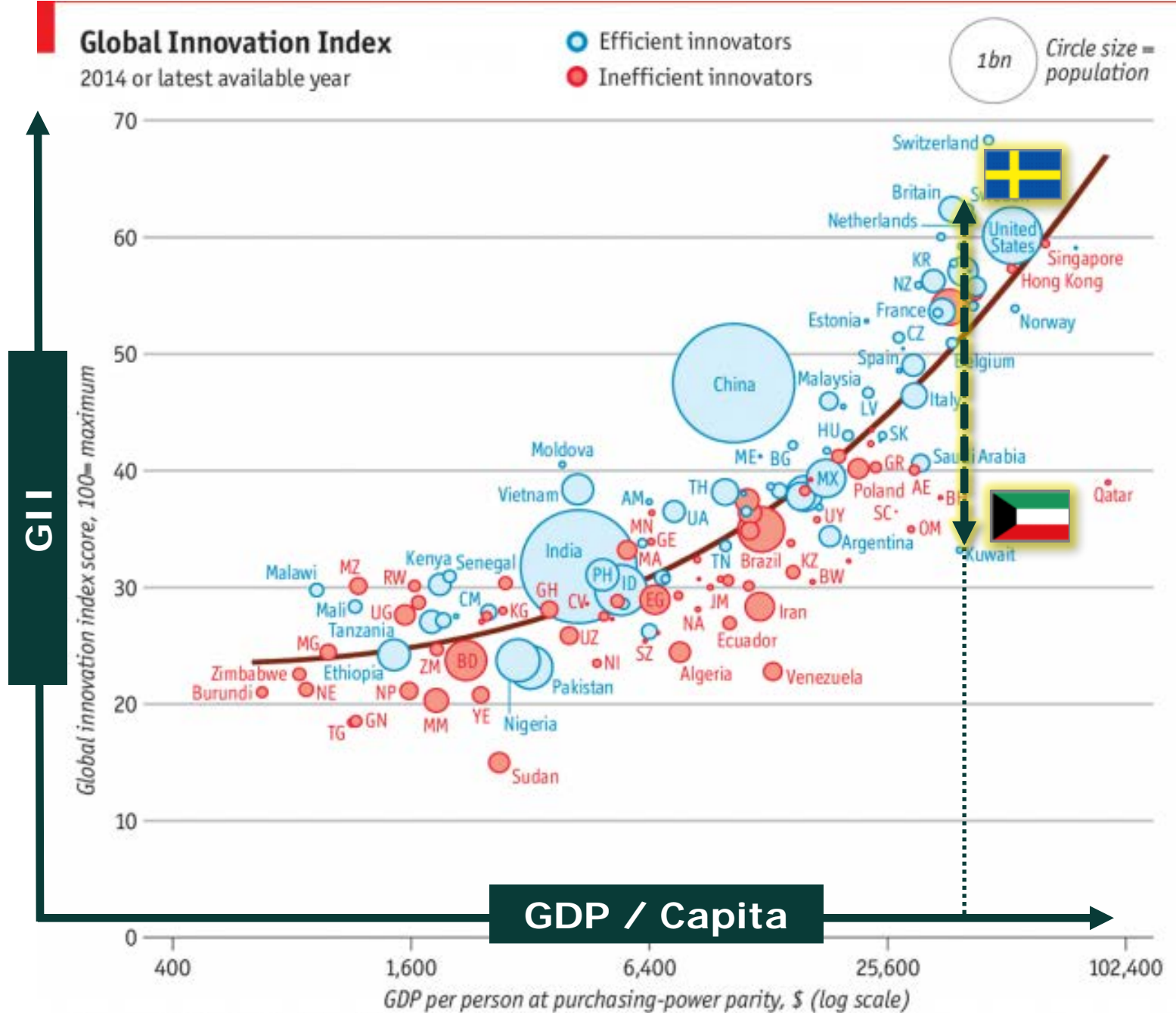
Source: Eurostat



= How
Scandinavian
are you?

Example for the added value / 1

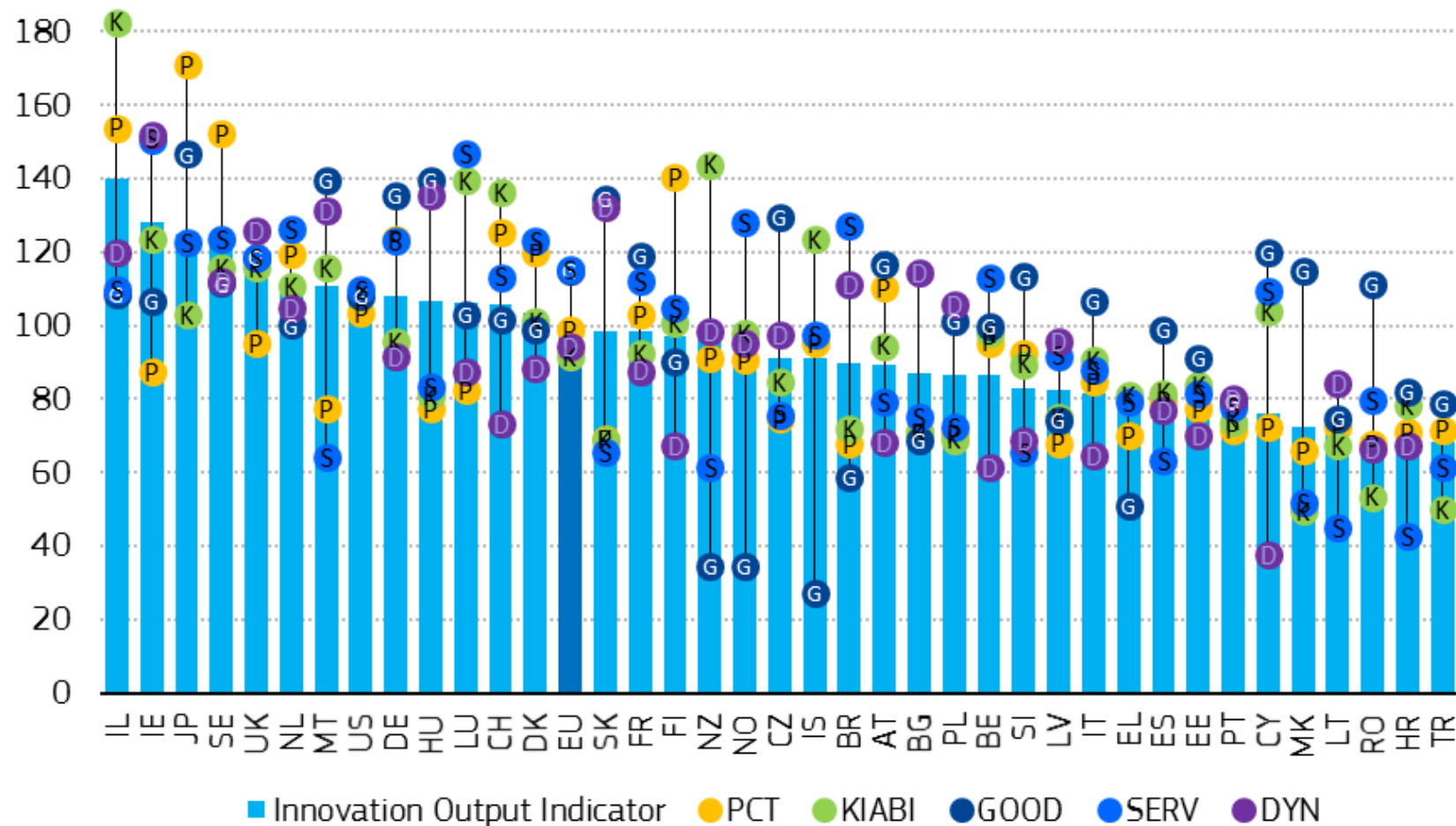
What GDP does not show...



Example for the added value / 2:

De-constructing indices:

similar performance,
different composition



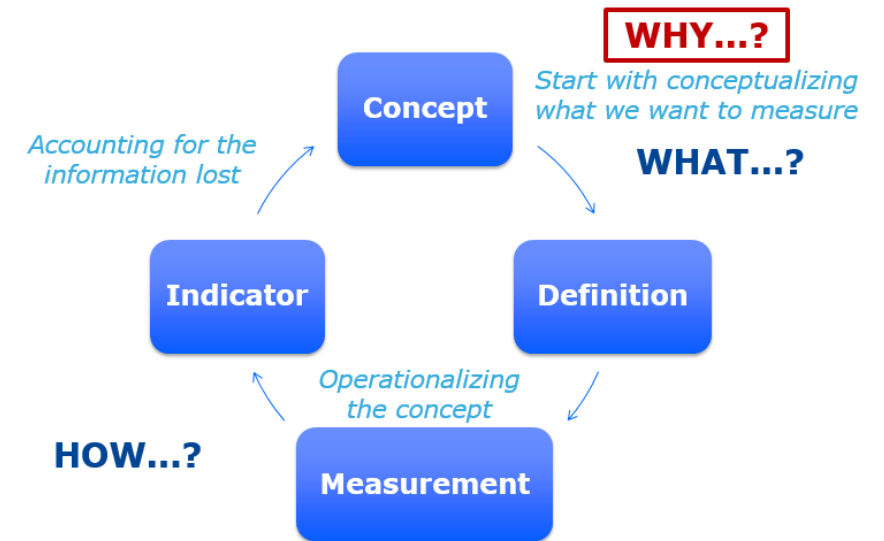
Quantification

A bit of theory

The Art of Quantification

- Quantification (modelling) involves making **normative choices** about...
- Normative choices affect:
 - the concept;
 - the operationalization;
 - accounting for information loss

⇒ Composite and stand-alone indicators are alike



Quantifying “essentially contested concepts”

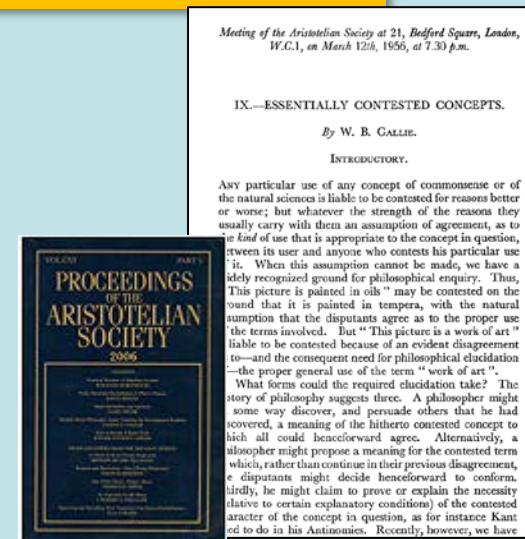
- Example: develop an indicator of “research excellence”

A complex framework,
with multiple dimensions of
engagement,
Including interactions



A simple framework,
with top-level output
measures only

- Debate with strong pro/con scientific arguments:
- “Excellence” is similar to concepts such as “a work of art”, “fairness”, “social justice” – groups of people disagree about the proper use of the concept; different uses serve different functions; even after the variety of functions is disclosed, disputes continue – these are **essentially contested concepts!**
- There is “**no one clearly definable general use** of ...which can be set up as the correct or standard use.”



W. B. Gallie (1956)

Quantifying “essentially contested concepts” – Implications

- Recognize that a concept is essentially contested:
 - Accept that “rival uses” of the concept not just likely, but represent a permanent critical challenge
 - If you consider “rival use as anathema, perverse, bestial or lunatic”, you will easily underestimate opponents’ positions
- Two sides of the coin:
 - *Marked raising of the level of quality of argument in the disputes of the contestant parties*
 - *Or, one contestant persist in the hope that they will ultimately persuade and convert all opponents by logical means.
“But once [we] let the truth out of the bag — i.e., the essential contestedness of the concept in question — then this harmless if deluded hope may well be replaced by a ruthless decision to cut the cackle, to damn the heretics and to exterminate the unwanted”*

The power of numbers



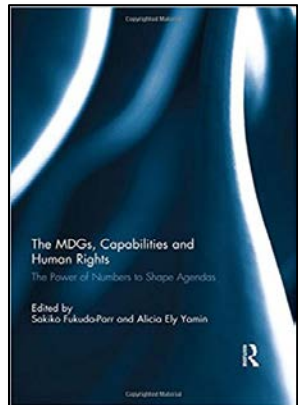
Numbers as targets: strong advocacy tool

- See the case of the **Millennium Development Goals (MDGs)**

*“While quantification is the key strength of global goals, it also involves **simplification**, **reification** and **abstraction**, which have far-reaching implications for redefining priorities”*

*Setting MDG goals/targets influenced policy priorities and had **normative effects** on development discourses;*

*All MDG goals/targets “led to **unintended consequences** in diverting attention from other important objectives and reshaping development thinking”*



(Fukuda-Parr, S., Yamin, A.E., Greenstein, J., 2014)

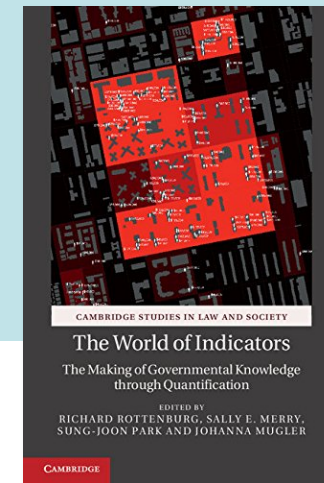
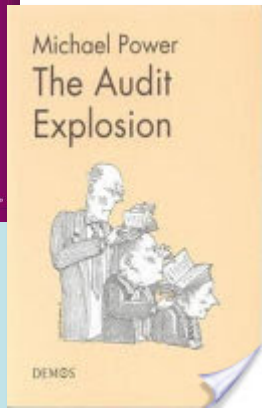
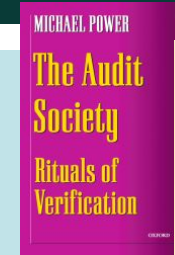
Indicators for Policy

Surge of indicators of policy

- Internationalization & globalization (need for comparison)
- Calls for accountability (the “audit explosion” – [Power, 1994](#))
 - legitimize interventions (*ex ante*) & demonstrate success (*ex post*)

- Indicators offer de-personalized accounts on the state of the world to allow for rational decisions to change the world for the better ([Rottenburg and Merry, 2015](#))

=> Aim to provide unbiased, comparable information on a single phenomenon of interest for different units

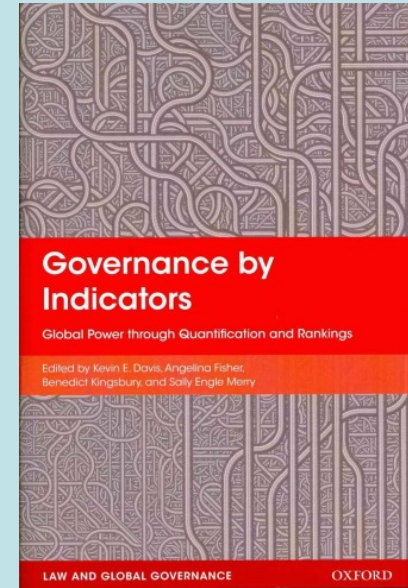


What are indicators, by the way?

numeric representations of the real world:

*"a **named** collection of **rank-ordered** data that purports to **represent** the past or projected performance of **different units**."*

(Davis et al. (2012))



Tools for benchmarking, evaluation = for policy purposes

Making diverse objects comparable = simplification, abstraction...

Or, tools for translation

Indicators are shaped by policy needs & discourse

Indicators, in turn, influence policy discourse

Indicators are embedded in a socio-political context

- Indicators & indicator frameworks are **value laden**; reflect policy discourse (Godin, 2002)
- Indicator developers & users better be aware of consequences:
 - The “agora model” (Barré, 2001, 2010): indicators are debating devices – it’s the process that matters!
 - Honest broker (Pielke, 2007)
 - “Indicators in the wild” – open up scope of data & expertise, process, framing of problems (Rafols, 2017)



Composite Indicators

Advantages:

- Support decision makers by summarizing complex or multi-dimensional issues
- Provide the “big picture”, highlight common trends
- Measure a latent phenomenon that is not directly measureable
- Attract public interest by benchmarking

The development process is crucial: it helps

- Better understand how a system functions
- Identify latent dimensions, redundancies or trade-offs between components

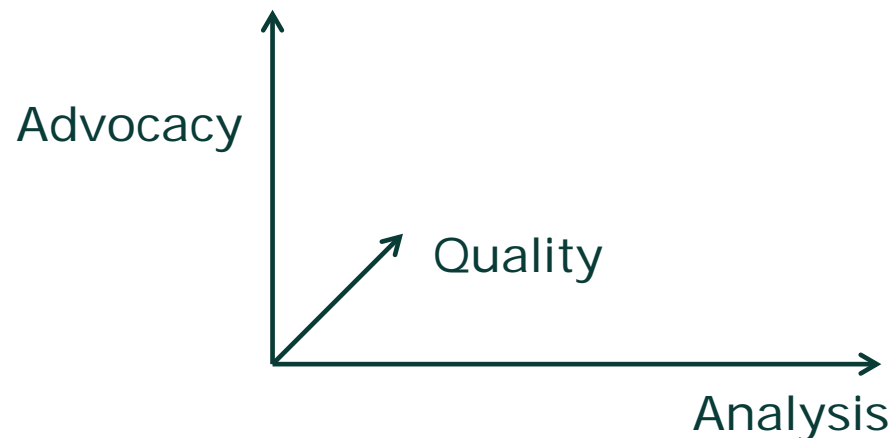
Pitfalls:

- Offer misleading, non-robust policy messages if they are poorly constructed or misinterpreted
- May invite politicians to draw simplistic policy conclusions
- Easier to “manipulate” than individual indicators; the selection of sub-indicators and weights could be the target of political challenge

Assessing their quality and validity is particularly relevant

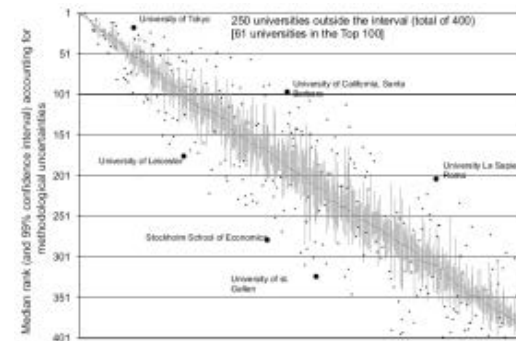
The Quality of Composite Indicators

“Composite indicators sit between analysis and advocacy, but **quality** discriminates the plausible from the rhetorical” (Saltelli, 2007)



Codified and continuously refined methodology

- The OECD-JRC Handbook (JRC-OECD, 2008)
- Audits – robustness and sensitivity analyses (i.e. Saisana et al, 2011; Paruolo et al, 2013)



Quality of Composite Indicators

- Statistical & conceptual – technical and normative aspects hard to separate
- Assessment frameworks help analyze developers' choices made with respect to:
 - the concept;
 - the operationalization process;
 - accounting for information loss...and thus help interpret results

=> audit of composite indicators and frameworks

- The **time** dimension and the **paradox** of assessment:
 - Indicator development is a **dynamic** process (see Barré, 2010; JRC-OECD 2008; refinement of indicators) ⇔ Assessment frameworks appear as **static**

=> Audits may need to be repeated in light of the evolution of an indicator!

A dynamic perspective on the assessment of composite indicators

- Co-evolution of...

The socio-economic context

Policy goals and strategies

Scientific understanding of the phenomenon, data availability, methods

...influence the conceptual / statistical coherence of composite indicators =>

➔ Assessment of the quality may change over time

In sum...

Developing a Conceptual Framework

Developing a Theoretical Framework

- Have a clear definition of the concept
- Identify the sub-groups of the multi-dimensional concept
- Set up the selection criteria for underlying indicators
- Take the time to document your choices...

Selecting Indicators

Selecting indicators

- Check the quality of available indicators
- Discuss strength and weakness of selected indicators
- Provide a summary table of key characteristics
 - Coverage (across time & space)
 - Source
 - Type (hard or soft measures; input / process / output?)

An example: selecting 'research excellence' indicators

- From an originally broad to a narrow framework
(see [Ferretti et al, 2017 forthcoming](#); [Hardeman and Vertesy, 2016](#) [Sorensen et al, 2015](#))
- Current use for monitoring progress to ERA:
 - how effectively do national research systems function?
- Research Excellence **Definition** = knowledge producing activities whose outputs are considered of high-end quality (based on [Tijssen, 2003](#))
 - Output oriented
 - Focus on quality, not quantity
 - Focus on top-end, not average
 - Distinguish Science and Technology (evaluated by peers vs. by market)

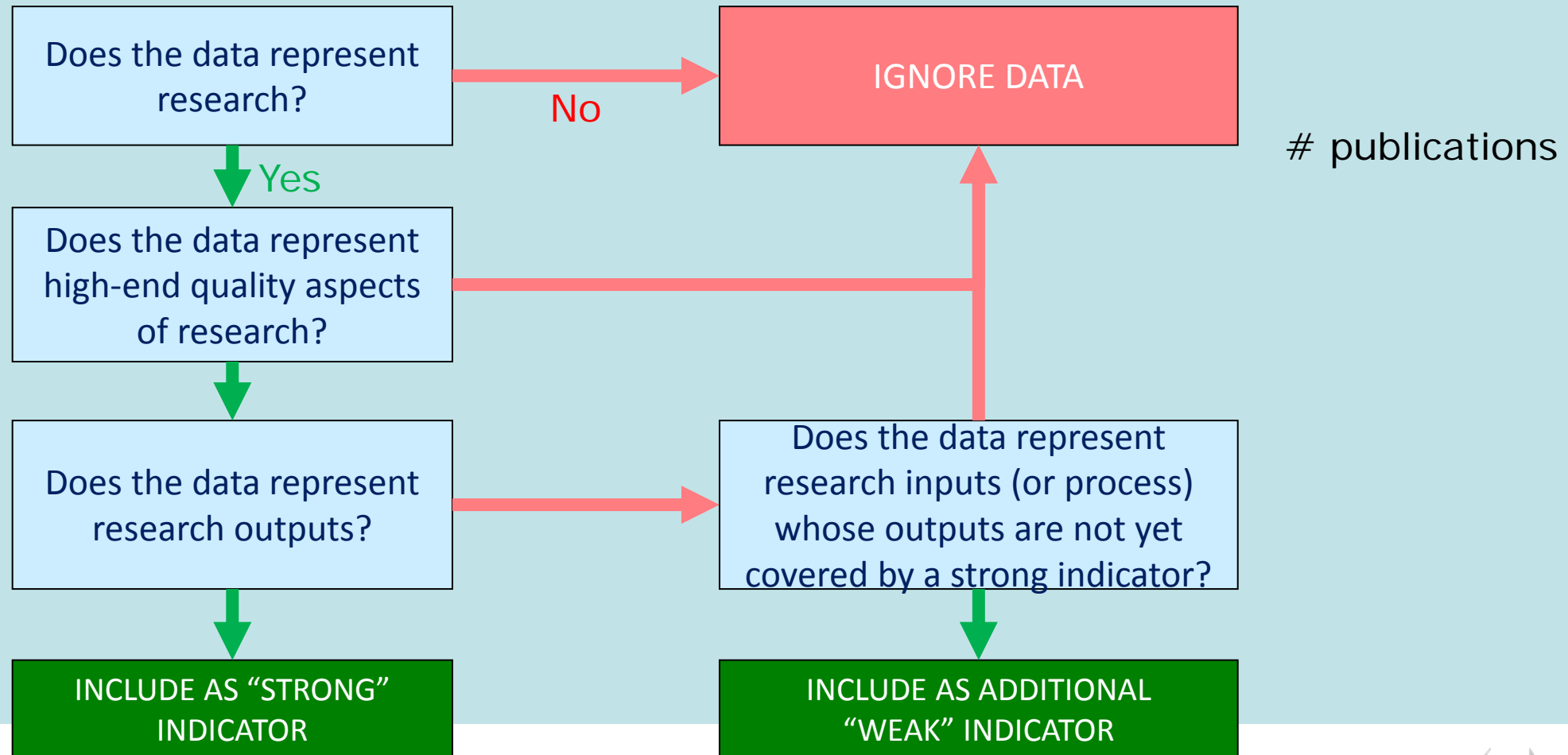


Main quality expectations from indicators:

- Extent: scale and scope of the data; type of activities covered
 - How many countries? 27 EU member states + int'l benchmark countries
 - At least two time points
- Reliability: Systematically collection data, consistent over time:
 - Ensure that definitions remain the same
- Validity: how well does the data correspond to the phenomenon to be measured?
 - see flow-chart

(see Hardeman et al, 2013; Hardeman and Vertesy, 2016)

Flow chart for selecting indicators:



Highly-cited publications

ERC grants

Selecting indicators: practical challenges

A. Use available data

- Was it collected for a similar purpose?
- Was it collected for another purpose but is relevant?

B. Generate new data

- Run surveys => costs; coverage; replicability
- Build from microdata => cost (also of replicability)
- Exploit Big Data (or admin data)
=> If desired indicators or desired granularity not available (cost, replicability)

C. Combine different sources

- Consider costs (€, time), ease of interpretation (intuitive?)
- Spell out the desired quality for the indicators to collect!
 - Can distinguish country performance? Missing data acceptable?

Selecting meaningful indicators



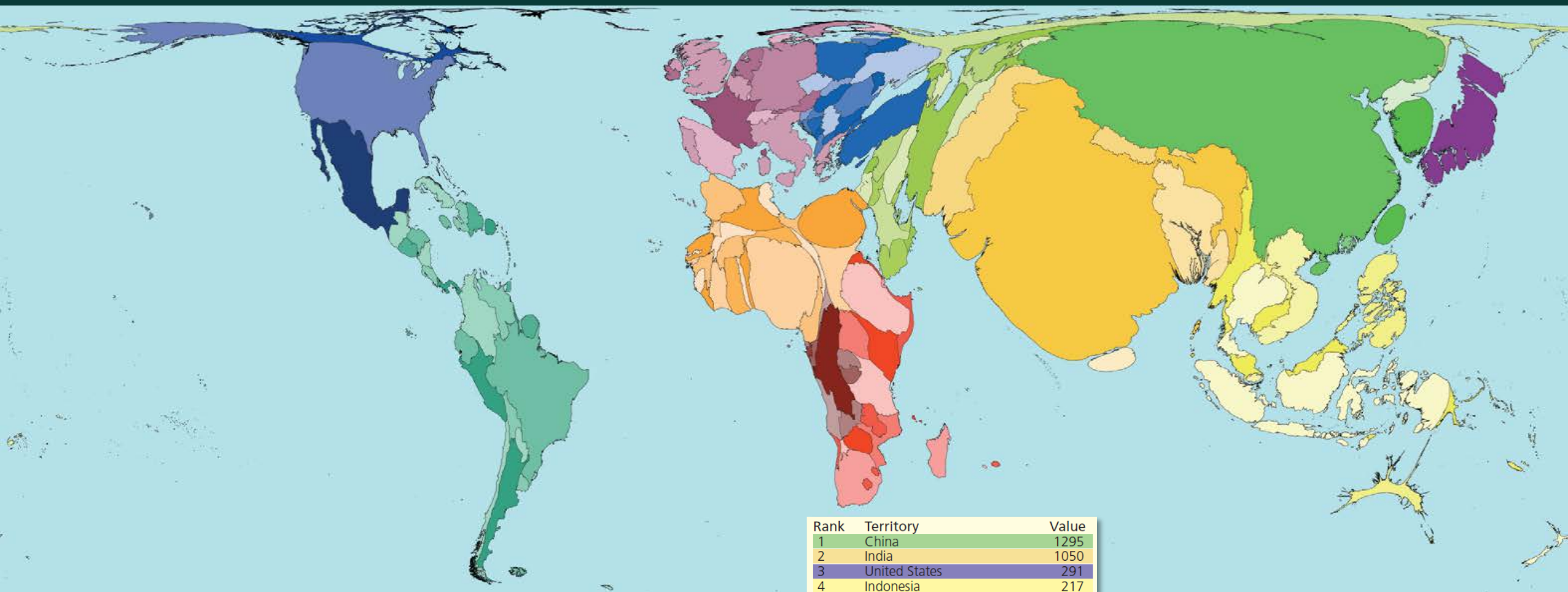
Is it a **meaningful indicator** for the purpose? Can it distinguish performance?

Typical trade-offs to tackle

- timeliness vs. completeness
- quality vs. breadth of coverage
- Novelty vs. acceptance
- Sophistication vs. intuition; reproducibility
- Changes in meaning at different levels of aggregation
- Dealing with units of different size:
 - Choice of denominator not straight forward!

	Years											
Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
HICIT	x	x	x	x	x	x	x	x	x			
TOPINST						?	?	?	?	?		
PCT	(x)	(x)	(x)	(x)	(x)	(x)	x	x	x	x	x	
ERC									(x)	x	x	x
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
HICIT	x	x	x	x	x	x	x	x	x			
TOPINST						?	?	?	?	?		
PCT	(x)	(x)	(x)	(x)	(x)	(x)	x	x	x	x	x	

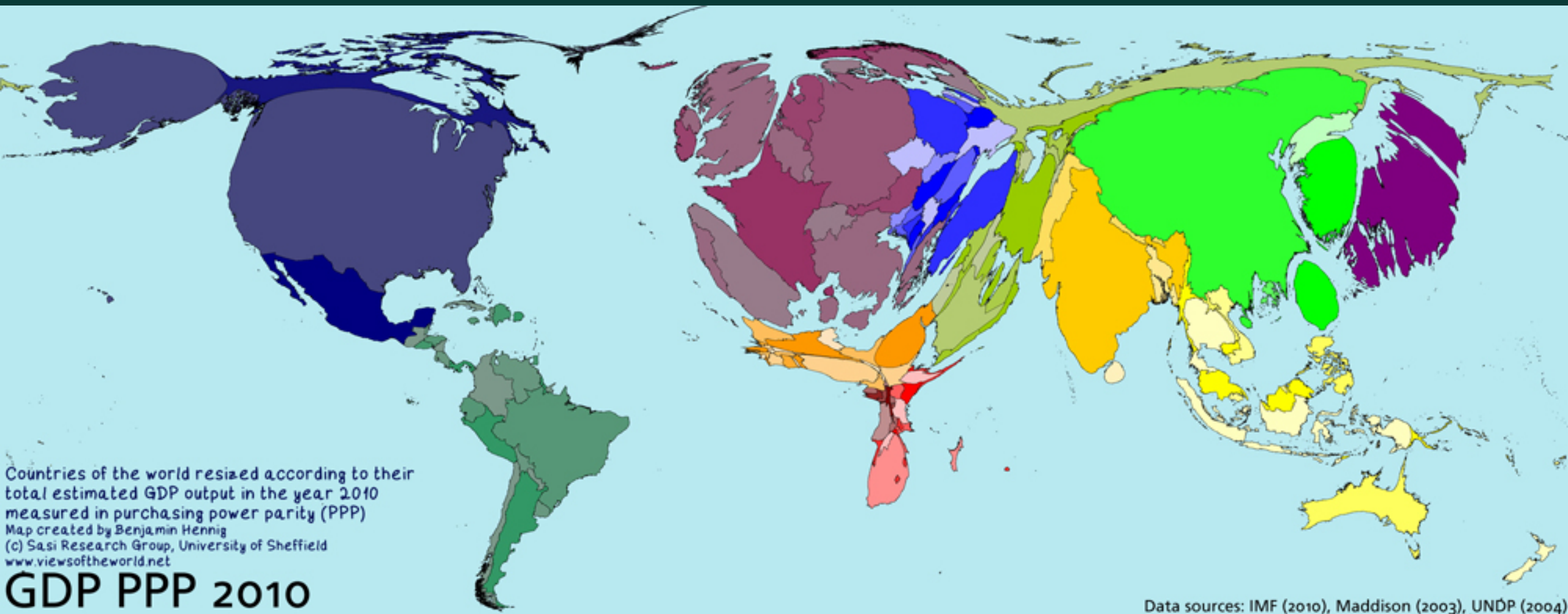
Normalize by Population...?



Rank	Territory	Value
1	China	1295
2	India	1050
3	United States	291
4	Indonesia	217
5	Brazil	176
6	Pakistan	150
7	Russian Federation	144
8	Bangladesh	144
9	Japan	128
10	Nigeria	121

millions

...or size of the economy (e.g. GDP)?



Populating the framework with indicators

An iterative process!

⇔ Conceptual framework

⇔ Statistical properties of the indicators

➔ *See sessions on Statistical Coherence*

- does the **correlation structure** reflect the conceptual framework?

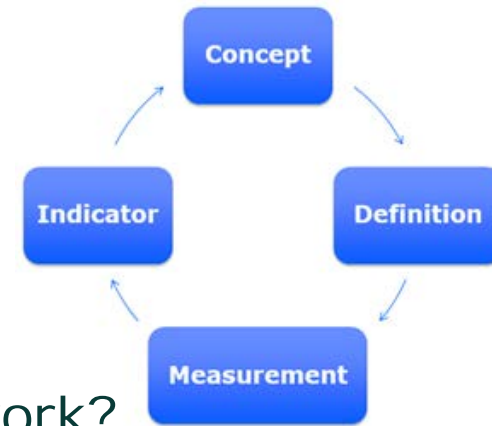
- If not, would changing the specification of the indicator (i.e., denominator) make a difference?

- Is data coverage acceptable? Is there another proxy with better coverage?

- Are the latent dimension(s) confirming the conceptual structure?

⇔ How can we interpret the correlation or principal component analysis (PCA) outcomes in light of the conceptual framework?

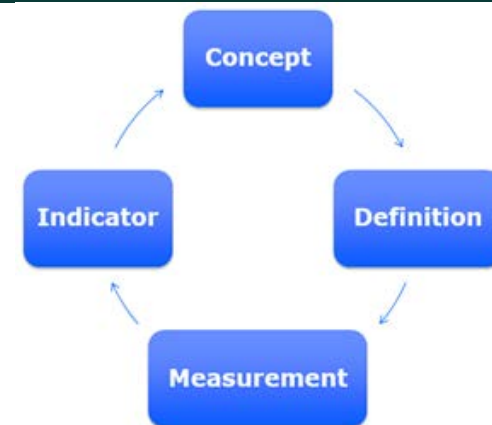
- Often a reason to refine/rethink the indicator framework (consider indicator development as a **learning process**)



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.00	0.46	-0.58	-0.06	0.23	-0.39	0.34	0.49	0.15	0.02	0.17	-0.07	0.37	0.44	0.45	0.46	0.43	0.45	0.46
	1.00	-0.22	0.00	0.09	-0.38	0.41	0.48	0.09	0.12	0.29	0.16	0.64	0.96	0.95	0.95	0.96	0.97	0.95
		1.00	-0.11	-0.24	0.52	-0.62	-0.61	-0.05	-0.14	-0.09	0.00	-0.22	-0.21	-0.26	-0.24	-0.23	-0.24	-0.24
			1.00	0.37	-0.32	-0.27	0.36	0.16	0.06	0.04	0.18	0.02	0.05	0.03	0.00	-0.01	0.02	0.01
				1.00	-0.19	0.72	0.57	0.08	0.01	-0.02	0.30	0.14	0.14	0.12	0.11	0.11	0.10	0.15
					1.00	-0.60	-0.81	0.88	-0.38	-0.54	0.83	-0.09	-0.32	-0.31	-0.19	-0.31	-0.39	-0.26
						1.00	0.95	-0.37	0.50	0.15	-0.06	0.40	0.41	0.48	0.38	0.40	0.38	0.45
							1.00	-0.73	0.19	0.56	-0.57	0.30	0.43	0.51	0.31	0.44	0.49	0.42
								1.00	0.11	-0.02	0.31	0.14	0.09	0.10	0.15	0.11	0.09	0.12
									1.00	0.12	0.00	0.01	0.13	0.09	0.10	0.09	0.11	0.10
										1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
											1.00	0.98	0.98	0.98	0.98	0.98	0.98	0.98
												1.00	0.98	0.98	0.98	0.98	0.98	0.98
													1.00	0.98	0.98	0.98	0.98	0.98
														1.00	0.98	0.98	0.98	0.98
															1.00	0.98	0.98	0.98
																1.00	0.98	0.98
																	1.00	0.98
																		1.00

Populating the framework with indicators

...an iterative process!



⇔ Does it meet the expectation of experts, analysts, policy users?

- if stakeholders have reservations with the framework, it is unlikely that the composite will be widely used



→ Make development participatory!

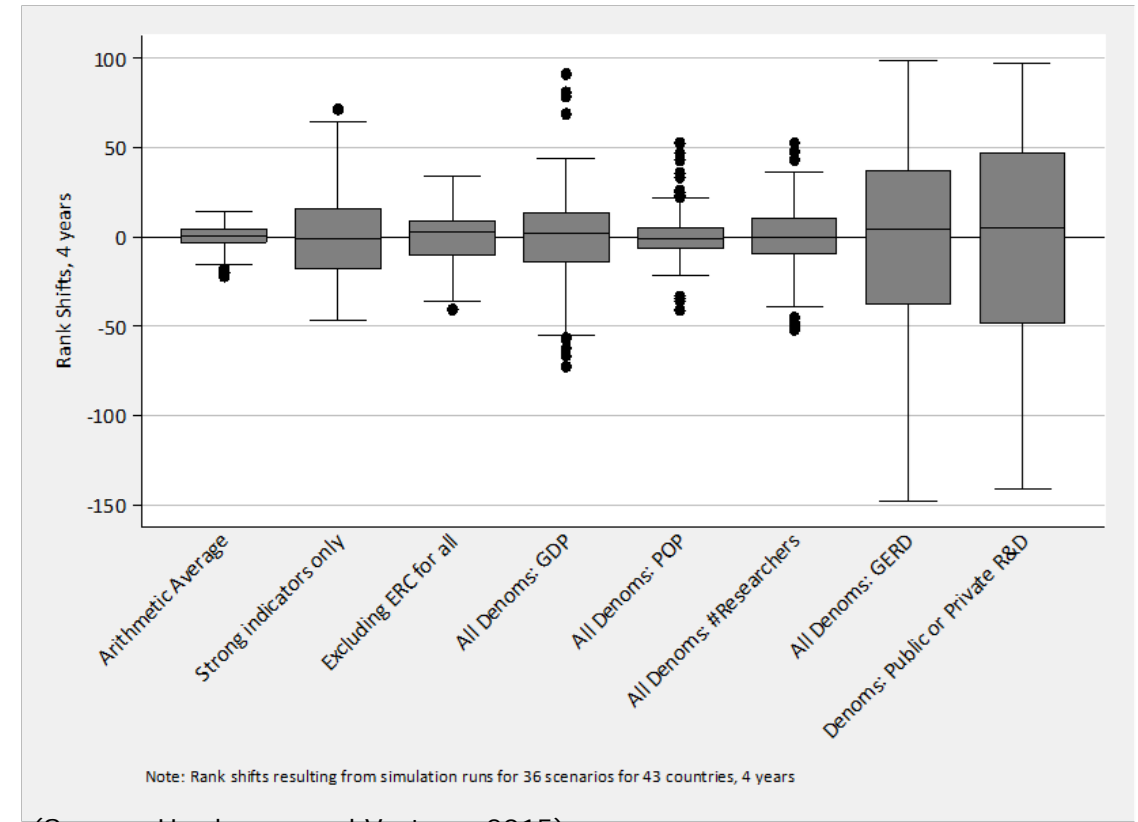
- Help articulate and refine different perspectives
- Compromises & normative choices unavoidable
=> these should be well documented!

Assumptions influence outcome

The impact of conceptual and methodological choices can be quantified

→ *see sensitivity analyses*

- How important are underlying assumptions?
- Identify key modelling choices
- Test their impact on the final composite scores and rankings



(Source: Hardeman and Vertesy, 2015)

Updating composite indicators

- Every new edition is an opportunity to refine framework & indicators
 - “Agora model”: changes and refinements help better understand the phenomenon of interest; discovery of new aspects
- Tradeoff between continuity & refinement [advocacy & analytical functions]
- For the developer: think of future updates at the start
 - Also in terms of data management

Choosing the IT platform



- Excel: → *see COIN Tool*
 - “WYSIWYG”: offers quick assessment of data quality profile; likely to share results (graphs); available everywhere
 - easy to lose track of manipulations – *make sure to use functions & keep dynamic links to original data sources*



MATLAB



- Statistical software, i.e. STATA, Matlab or R (or Excel VBA)
 - Less intuitive, high initial learning cost –
 - Easy to document choices in script languages (i.e., stata .do files)
 - Excel not ideal for some steps (PCA, simulations, etc...)
- Structuring data: *find layout most suitable for context*
- Downloads can be programmed (see `readSDMX` for R; `getdata` of STATA; etc.); bulk download preferable also in Excel
- *Keep track of numerators, denominators, different versions tested*

A final note

- Make your choices clear for ALL (including yourself!)
 - What, Why, How...?
 - = just as important: what ***not***, why ***not***, how ***not***...?
- Make your indicator time-resistant:
 - socio-political context may change!
- Clear documentation is essential





THANK YOU

Welcome to email us at: daniel.vertesy@ec.europa.eu
jrc-coin@ec.europa.eu

COIN in the EU Science Hub
<https://ec.europa.eu/jrc/en/coin>

COIN tools are available at:
<https://composite-indicators.jrc.ec.europa.eu/>

The European Commission's
Competence Centre on Composite
Indicators and Scoreboards



References / 1

- Barré, R. (2001). The Agora model of innovation systems: S&T indicators for a democratic knowledge society. *Research Evaluation*, 10(1), 13-18.
- Barré, R. (2010). Towards socially robust S&T indicators: Indicators as debatable devices, enabling collective learning. *Research Evaluation*, 19(3), 227-231.
- Davis, K. E., Fisher, A., Kingsbury, B. and Merry, S.E. (eds.) (2012) *Governance by Indicators: Global Power Through Quantification and Rankings*. Oxford: Oxford University Press.
- Ferretti, F., Guimaraes Pereira, A., Hardeman, S. & Vertesy, D. (2017) "Research Excellence Indicators: time to re-imagine the 'making of'?" Paper under revision for Science and Public Policy
- Fukuda-Parr, S., Yamin, A. E., & Greenstein, J. (2014). The power of numbers: a critical review of millennium development goal targets for human development and human rights. *Journal of Human Development and Capabilities*, 15(2-3), 105-117.
- Gisselquist, R. (2014) Developing and evaluating governance indexes: 10 questions. *Policy Studies* 35(5), 513-531
- Hardeman S., Van Roy V., Vértesy D. and Saisana M. (2013) *An analysis of national research systems (I) A Composite Indicator for Scientific and Technological Research Excellence* EUR 26093. Luxembourg: Publications Office of the EU; doi: 10.2788/95887
- Hardeman S. & Vértesy D. (2015) *An update of the research excellence indicator*. Luxembourg: Publications Office of the EU; doi: 10.2760/770103

References / 2

- Hardeman, S., & Vertesy, D. (2016). Evidence-based policy learning: the case of the research excellence indicator. Paper presented at the 2016 STI Conference in Valencia, Spain
- JRC-OECD (2008). *Handbook on constructing composite indicators: methodology and user guide*. Paris: OECD.
- Paruolo, P., Saisana M. and Saltelli A. (2013) "Ratings and rankings: voodoo or science?" *Journal of the Royal Statistical Society: Series A*. 176(3), 609–634
- Power, M. (1994). *The audit explosion* (No. 7). Demos Paper No. 7.
- Rottenburg, R., Merry, S.E., Park, S-J. and Mugler, J. (eds.) (2015) *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press
- Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1), 65-77.
- Saisana M., D'Hombres B., Saltelli A. (2011) "Rickety Numbers: Volatility of university rankings and policy implications." *Research Policy* 40, 165–177
- Saisana M., Saltelli A., Tarantola S. (2005) "Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators", *Journal of the Royal Statistical Society A* 168(2), 307-323.
- Tijssen, R. J. (2003). Scoreboards of research excellence. *Research Evaluation*, 12(2), 91-103
- United Nations (2012) *National Quality Assurance Frameworks*
[URL: <http://unstats.un.org/unsd/dnss/qualityNQAF/nqaf.aspx>]