

*Conference:
Future of the sciences in the light of authorship laws, researcher's ethical
codes and government evaluations of the quality of research.*

Contextualisation for Responsible Metrics in Evaluation

Ismael Ràfols

CWTS (Centre for Science and Technology Studies), University of Leiden

Ingenio (CSIC-UPV), Universitat Politècnica de València
SPRU (Science Policy Research Unit), University of Sussex

Summary

- **Indicator-based evaluation may lead to some negative consequences**
 - Goal displacements: publishing become an end in itself
 - Task reduction: Some research / teaching practices and topics may be marginalised
 - Reduction of diversity (language, topics, etc.)
- **Evaluation has different functions & research different purposes**
 - Monitoring, allocation, learning / improvement
 - Academic knowledge, professional training, solving societal problems
- **Need to develop more contextualised evaluation**
 - Specially for the purpose of reflecting on improving practices
 - Indicators inform, but do not substitute for expert judgement (P1)
 - In some contexts - particularly in SSH for societal impact / local issues / (P2/P3)
- **Only possible in more decentralised, autonomous processes?**
 - Not in large national allocation programs -- but at dept / lab level. (appraisal)
 - There are experiences in various agencies / countries

The parable of Prussian scientific forestry (*Seeing like a state*, J. Scott)

Forests in pre-modern Prussia

- Wild
- Uncontrolled
- Unpredictable
- Inefficient for timber production



Enlightenment and scientific forestry:

- Cut the wild forest
- Plant Norway spruce –reduce diversity
- Increase yield and predictability
- Loss of forest activities for peasants:
(fruits, hunting, medicinal herbs, mushrooms...)

The parable of Prussian scientific forestry (*Seeing like a state, J. Scott*)

Monocultures and Forest death

- Nutrient depletion leading to 20-30% production loss in 2nd generation
- Storm felling
- Pests due to loss of 'services' of insects, birds and animals.



Restoration forestry or forest hygiene:

- Artificial ant colonies & spiders
- Wooden boxes to provide bird nests
- The dangers of dismembering a complex set of relations and processes to isolate a single element of instrumental value

The parable of Prussian scientific forestry (Seeing like a state, J. Scott)

Task reduction

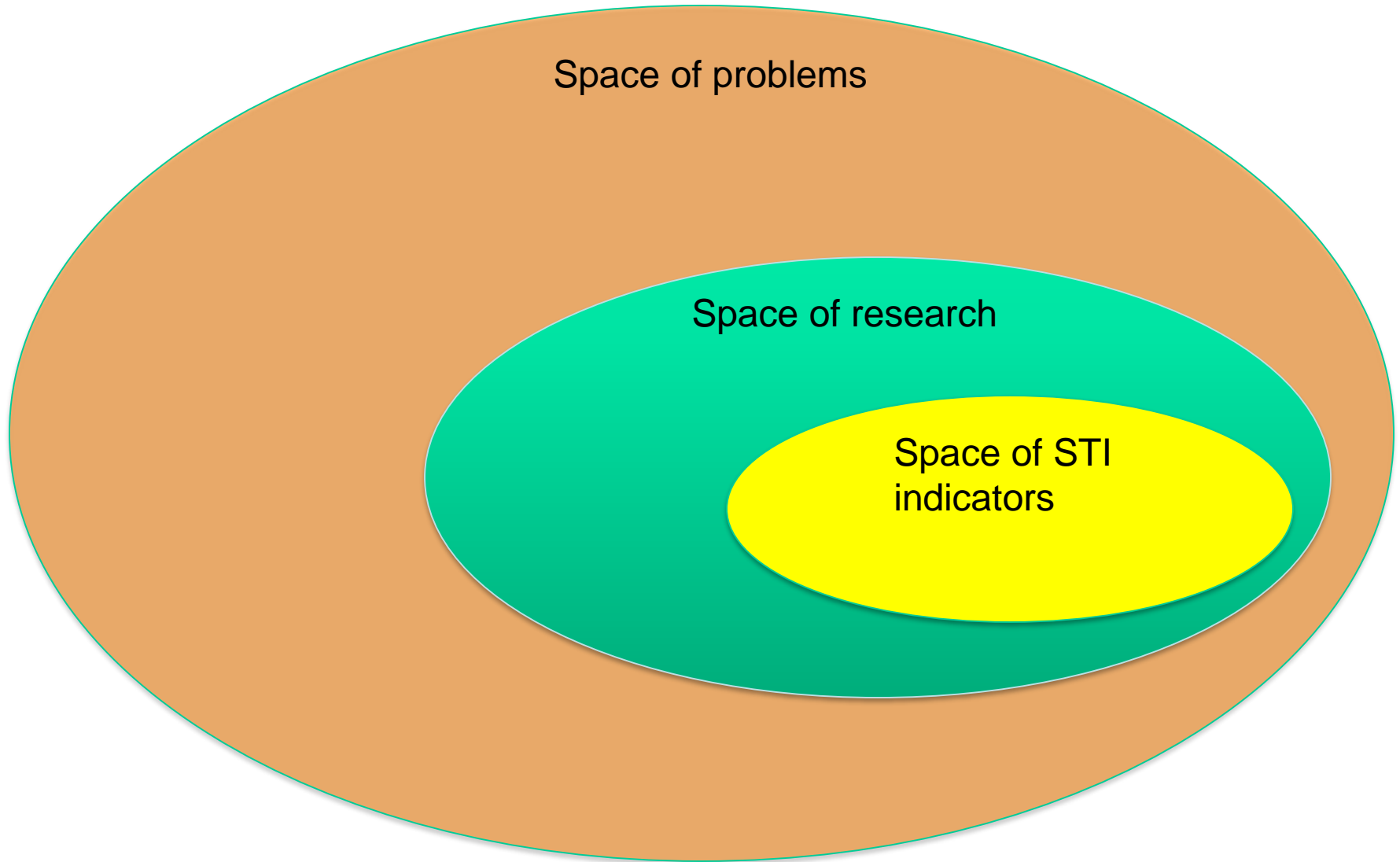
“Exaggerating only slightly, one might say that the crown's interest in forests was resolved through its fiscal lens into a **single number: the revenue yield of the timber** that might be extracted annually.. ”



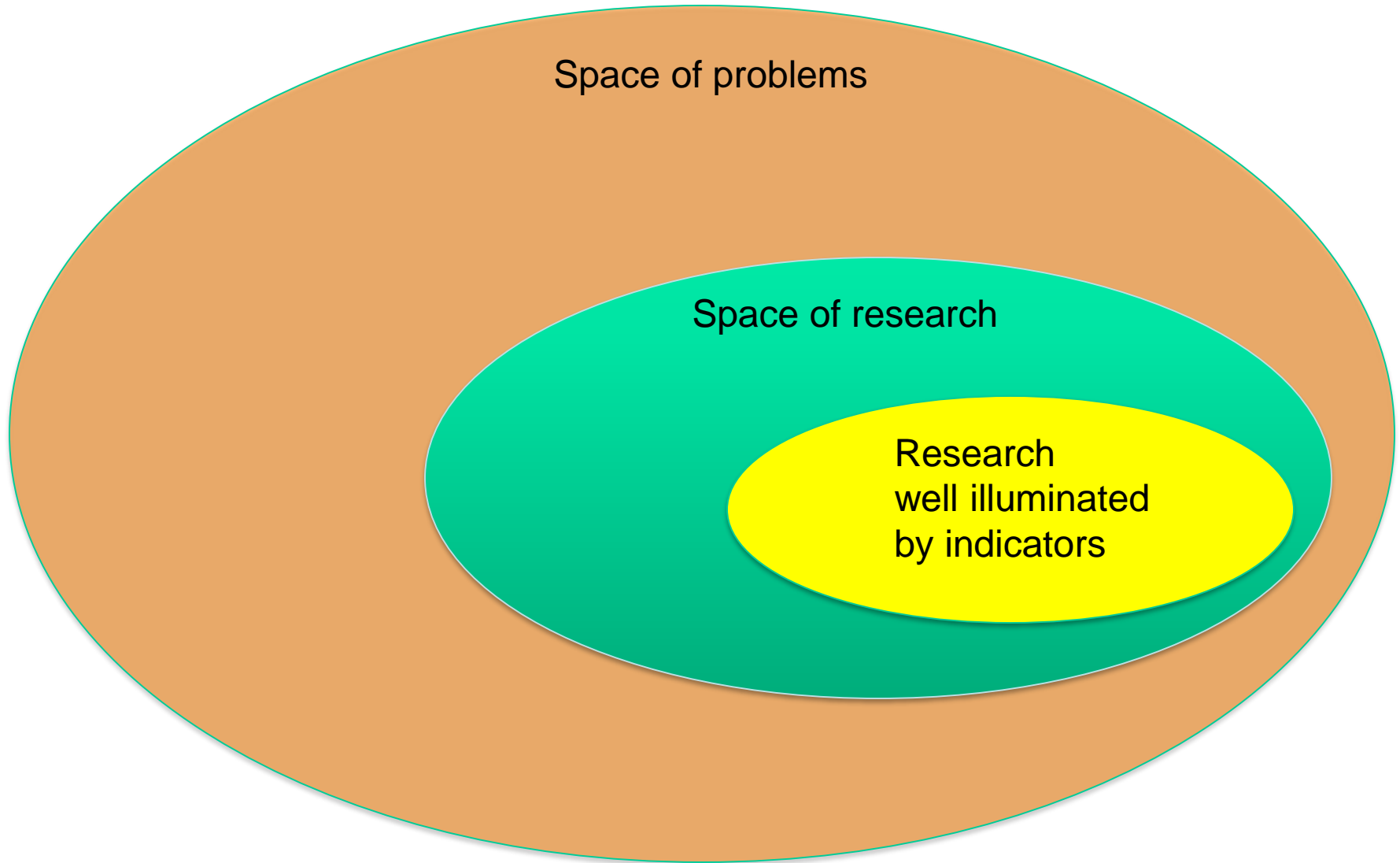
Performativity

“Backed by state power through records, courts, and ultimately coercion, **these state fictions transformed the reality they presumed to observe**, although never so thoroughly as to precisely fit the grid.”

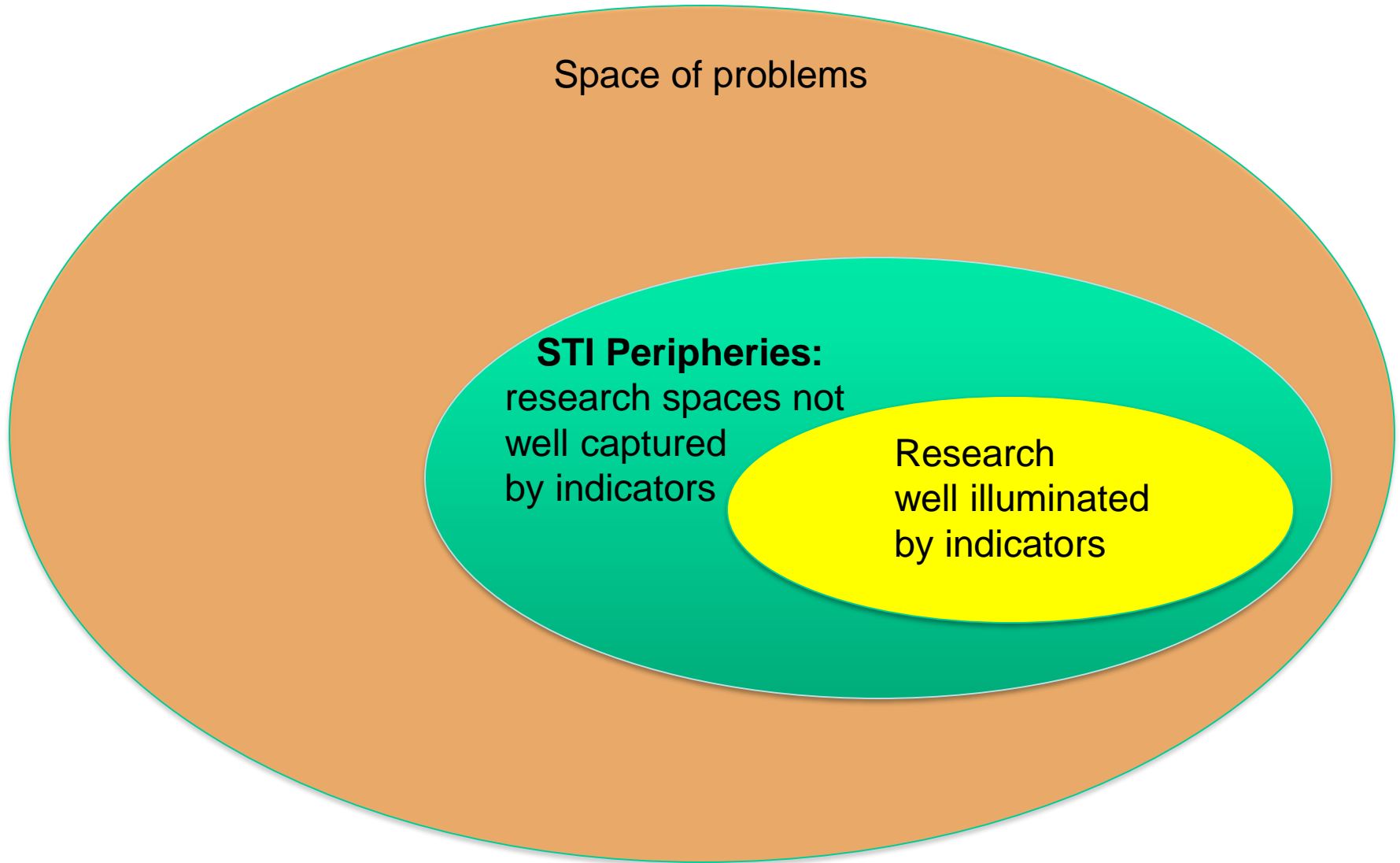
Problems, research, indicators and peripheries



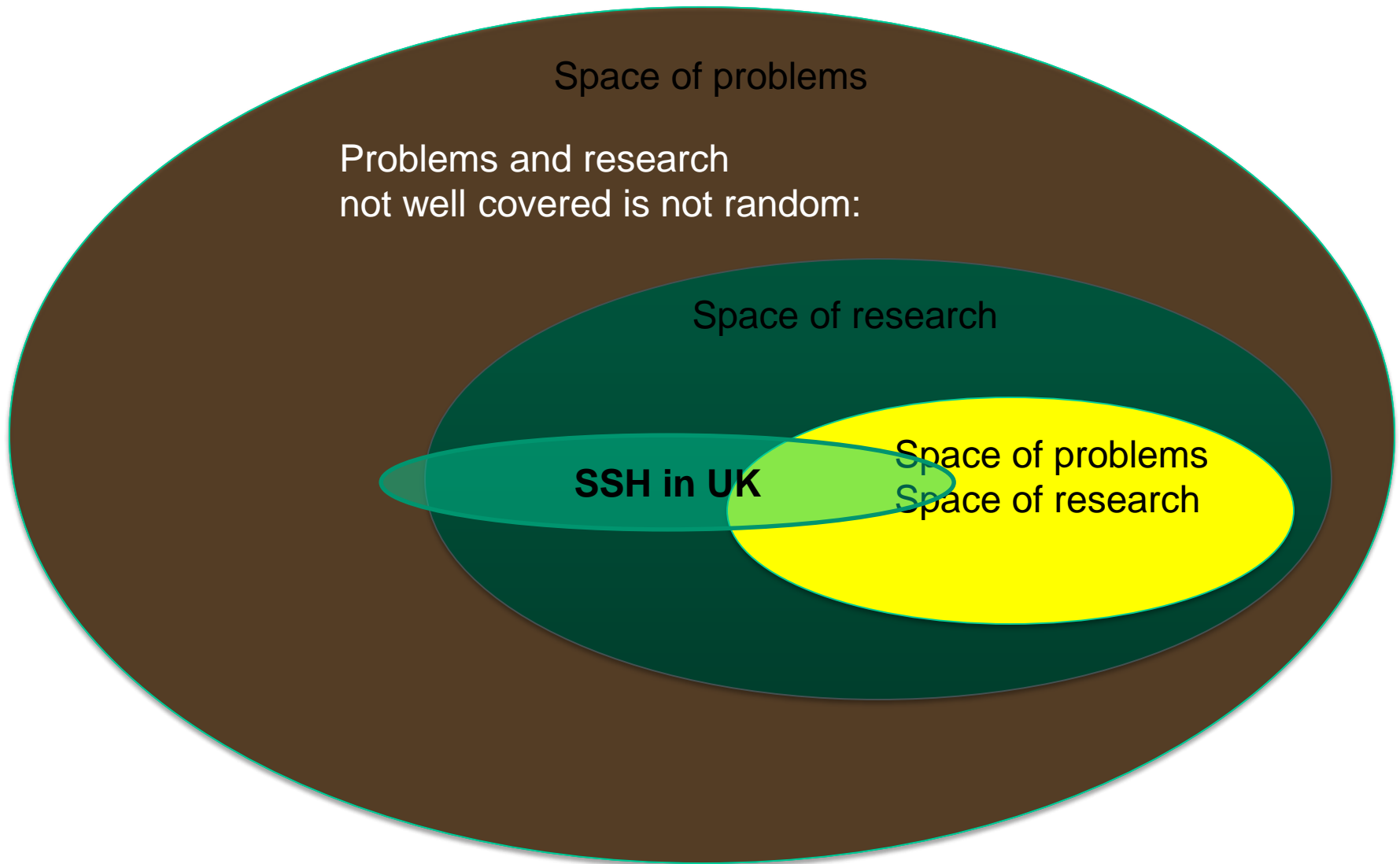
Problems, research, indicators and peripheries



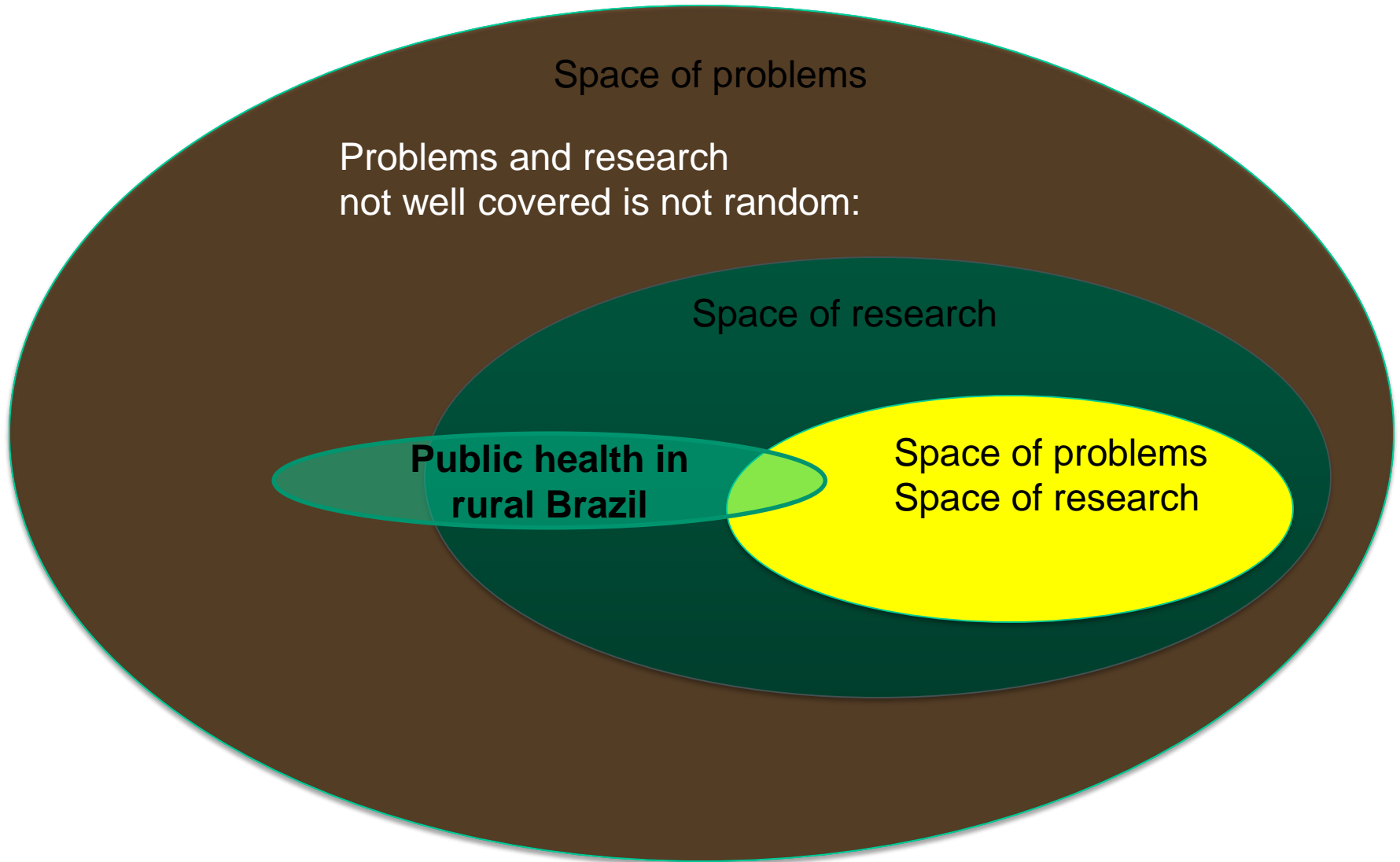
Problems, research, indicators and peripheries



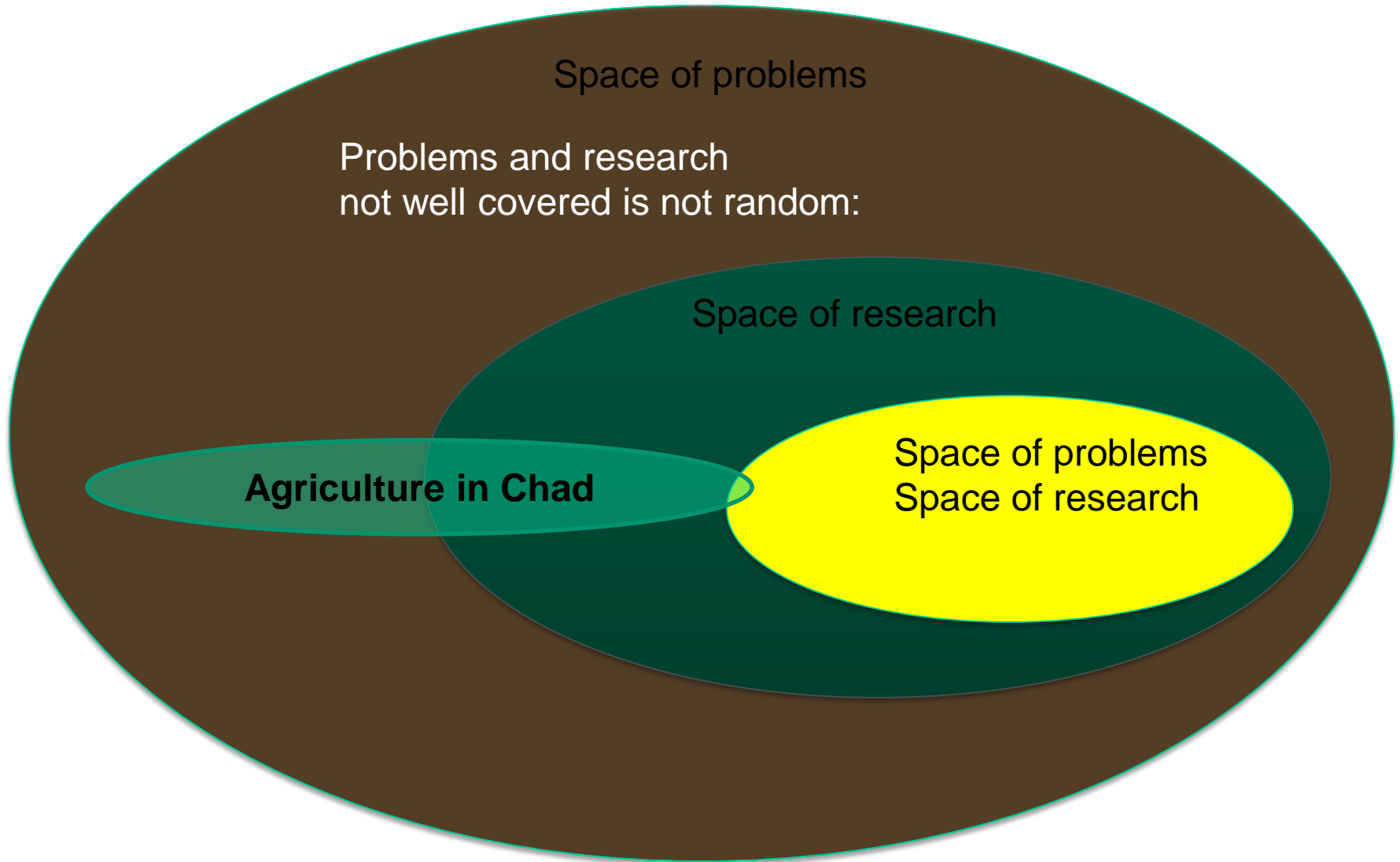
Streetlight effect in indicators: mistaking light with “problems”



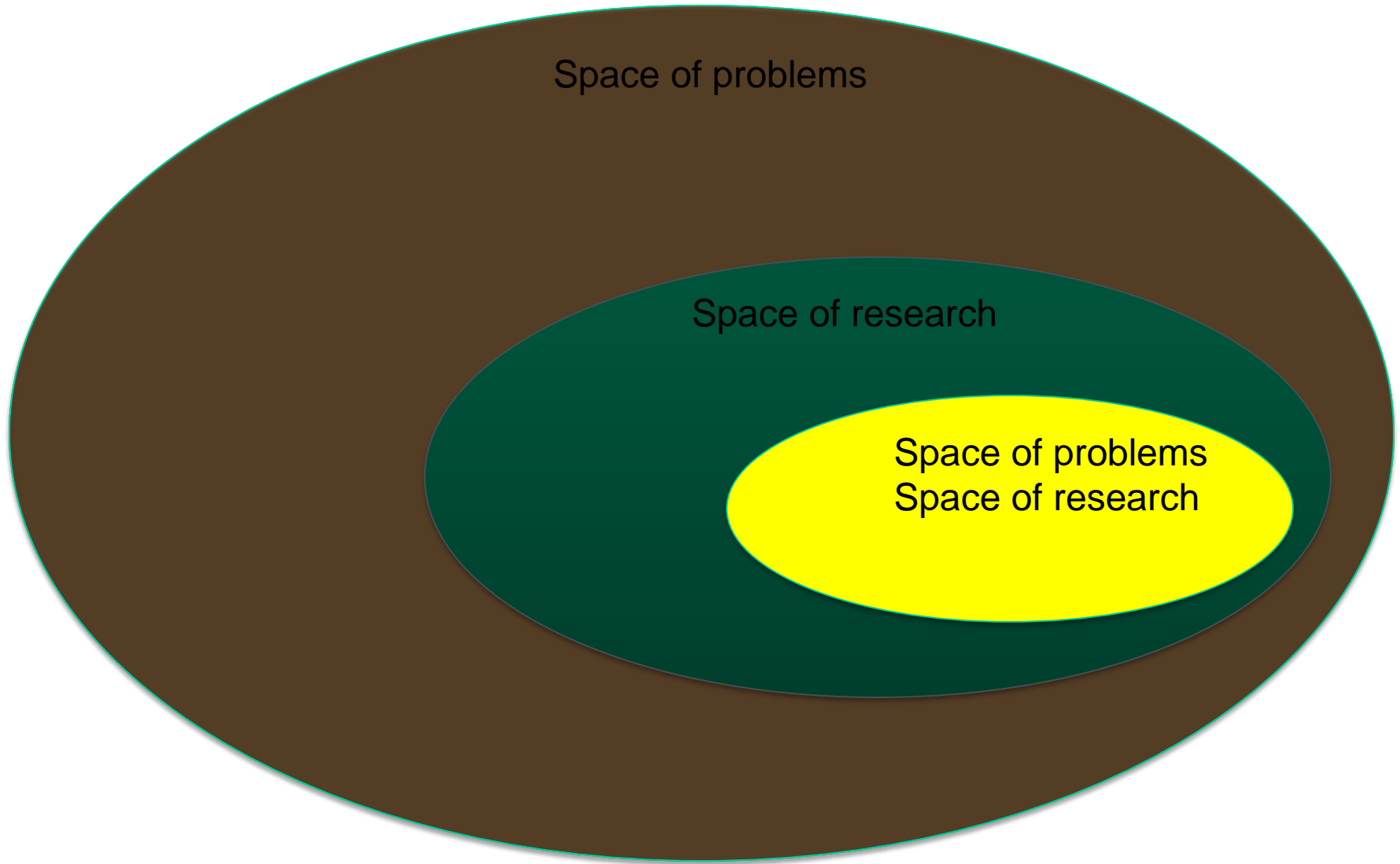
Indicators make “peripheral” spaces invisible



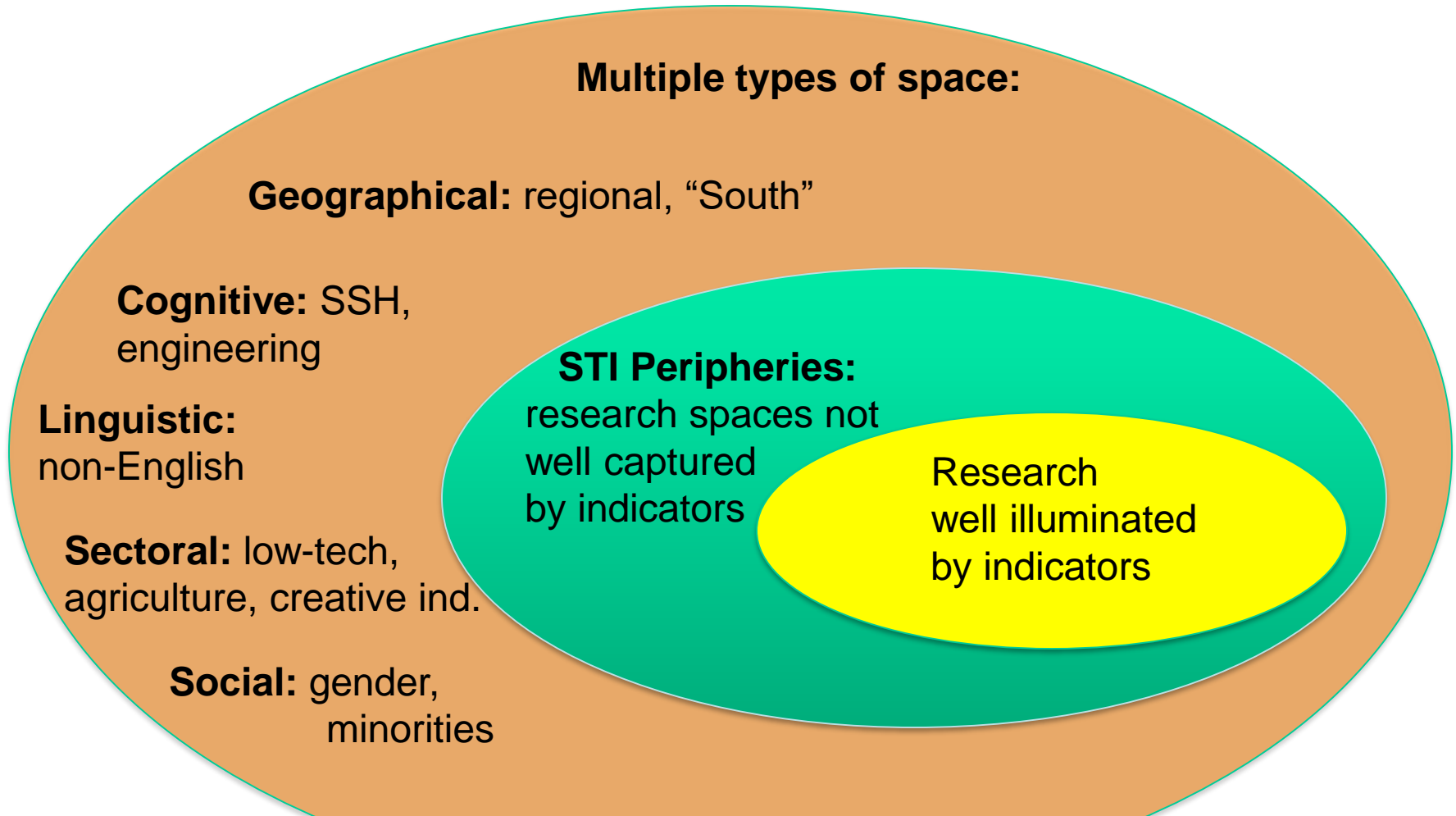
Indicators make “peripheral” spaces invisible



Streetlight effect in indicators: mistaking light with “problems”



Problems, research, indicators and peripheries



Usual multiple peripheries: primary health care of poor neighbourhoods in Salvador de Bahia, written in Portuguese (Cog, Geo, Ling, Soc.)

The streetlight effect and the performative role of indicators

- Use of indicators may have consequences on research system
- Incentive structure: indicators signal to stakeholders what is important.
 - Goal displacement: instead of mission, follow indicators
- Potential suppression of diversity
 - Shift towards English publications (Siversten, 2014)
 - Shift towards more technical / mainstream issues
 - Diversion of research away from local or national issues (Hicks, 2015)
 - Bias toward positive reporting (Ioannidis, 2005)
 - Invisible / undone science (Hess, 1997)

We cannot correct for these biases mathematically

→ expert values are needed

Goals of evaluation and functions of research

Evaluation (a dictionary definition): ‘To ascertain or fix the value or worth of an object against a certain specified criteria.’

Goals of research evaluation (Molas-Gallart, 2012; Adam, 2018)

- Monitoring or auditing (or control)
- Distribution of resources
- Improving or learning

Goals of evaluation and functions of research

Evaluation (a dictionary definition): ‘To ascertain or fix the value or worth of an object against a certain specified criteria.’

Missions of research (Molas-Gallart et al., 2003)

- Academic contribution
- Education & training
- Societal contributions

Bib. indicators vs. goals of evaluation and functions of research

- **Goals of research evaluation** (Molas-Gallart, 2012; Adam, 2018)
 - Auditing bib. indicators convenient
 - Distribution of resources bib. indicators useful for justification but not necessarily for setting strategy
 - Improving or learning bib. indicators not very helpful
- **Functions of research**
 - Academic contribution bib. indicators can be appropriate in some fields (natural sciences)
 - Education & training bib. indicators NOT appropriate
 - Societal impact bib. indicators NOT appropriate

Bibliometrics ONLY for auditing SOME academic contributions

However, in a bureaucratic system – auditing academia is main goal

Useful for managers but possibly harmful for the research system

Uses and abuses of bibliometric indicators tend to favour the ivory tower

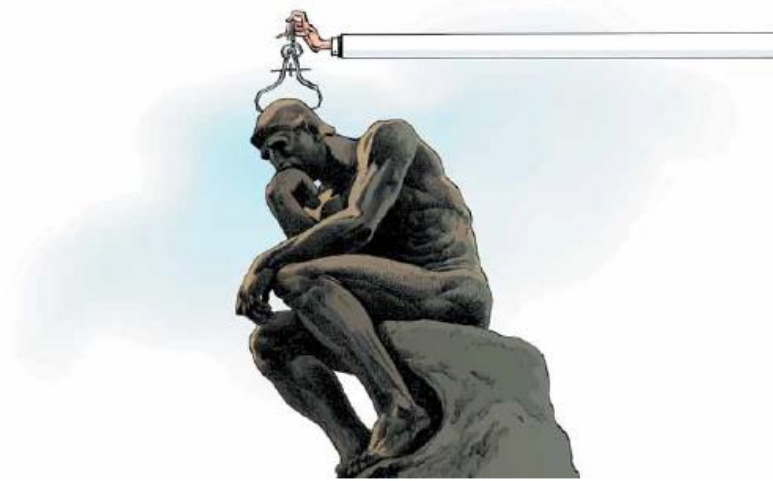
San Francisco Declaration of Research Assessment (DORA)

Don't use Impact Factors!!

ASCB (2013)

Conventional indicators perceived favour elitist research in the ivory tower.

Barrier to adoption of Open Science



The Leiden Manifesto for research metrics

- **Need to contextualise**
- **Indicator support (not replace) judgement.**

Hicks et al. (2015)

The Metric Tide

UK (Hefce) Report

Use indicators with

- **Robustness**
- **Humility**
- **Transparency**
- **Diversity**
- **Reflexivity**

Wilsdon et al. (2015)

Principles of the “The Leiden Manifesto”

1. Quantitative evaluation should support qualitative, expert assessment.
2. Measure performance against the research missions of the institution group or researcher.
3. Protect excellence in locally relevant research.
4. Keep data collection and analytical processes open, transparent and simple.
5. Allow those evaluated to verify data and analysis.
6. Account for variation by field in publication and citation practices.
7. Base assessment of individual researchers on a qualitative judgement of their portfolio.
8. Avoid misplaced concreteness and false precision.
9. Recognize the systemic effects of and indicators.
10. Scrutinize indicators regularly and update them.

Judgement

Sensitive to
Contexts

Transparency

Pluralism

Reflexivity

Evaluation informed by indicator frameworks

Dimensions to consider in evaluation

- Goal of evaluation
- Research mission
- Level of assessment
- Scientific field and methodological approaches
- Potential stakeholders, audiences and beneficiaries
- Research environment: Human and technical resources

For a given configuration of relevant dimensions, certain indicators will be relevant while others will not.

Attention:

- There cannot be general indicators of research 'quality' or OS

Between FULLY TAILORED – UNIVERSAL

EC Expert Group on Indicators for Researchers' Engagement with OS

(P. Wouters (chair), B. Holbrook, M. Jacob, Lynn Kamerlin, A. Oancea, I. Ràfols)

Science as a biosphere with dozens of diverse ecosystems



To evaluate means to value -- to make our values explicit

**It the wake of movements towards Open Science and RRI
We need:**

Forms of more 'open' research:

- More oriented to public good and close to society
- More diverse and pluralistic
- With the participation of social actors

This requires forms of evaluation that

Foster pluralization	→ Since S&T are uncertain and there are options
Sensitive to values	→ ¿What type of research is socially desirable?
Contextual	→ depending of specific socia-ecological spaces

Need to change (complement) the FORMS of evaluation:

- From centralised large evaluations to small (dept/lab) assessment
- By experts aimed at learning / improving
- Focus on the scientific and societal goals posed (not abstract 'quality')

What is research for?

***If indicators are the answer,
what is the question?***

“It is sometimes easier to develop quantitative ‘indicators’ of performance than to work out what the program has to accomplish.”

David Roessner (2000)

Evaluar significa valorar -- explicitando nuestros valores

Una investigación:

- más orientada al bienestar social de tod@s.
CyT más plural y más justa
- más cercana, más abierta a la sociedad
procesos de evaluación con participación de agentes sociales
- indicadores para informar la deliberación, abrir el debate

Qué tipos de evaluación?

- | | |
|----------------------|---|
| Pluralizadora | → el futuro de CyT es abierto, hay opciones diversas |
| Con valores | → ¿qué tipo de investigación es socialmente deseable? |
| Contextual | → depende de espacios sociales y naturales específicos. |

Propuesta para la bibliometría:

- No se trata de buscar nuevos indicadores sino de *pensar cómo los indicadores pueden contribuir a nuevas formas de evaluación*

Classic values of science

In 1942, Merton wrote about the aspirational values of science

- Communalism (shared knowledge)
 - Universalism (all humans can participate)
 - Disinterestedness (public good)
 - Organised Scepticism (scrutiny & transpar.)
- ...key values behind Open Science

His argument was prompted by use of science in authoritarian regimes (1930s-40s):

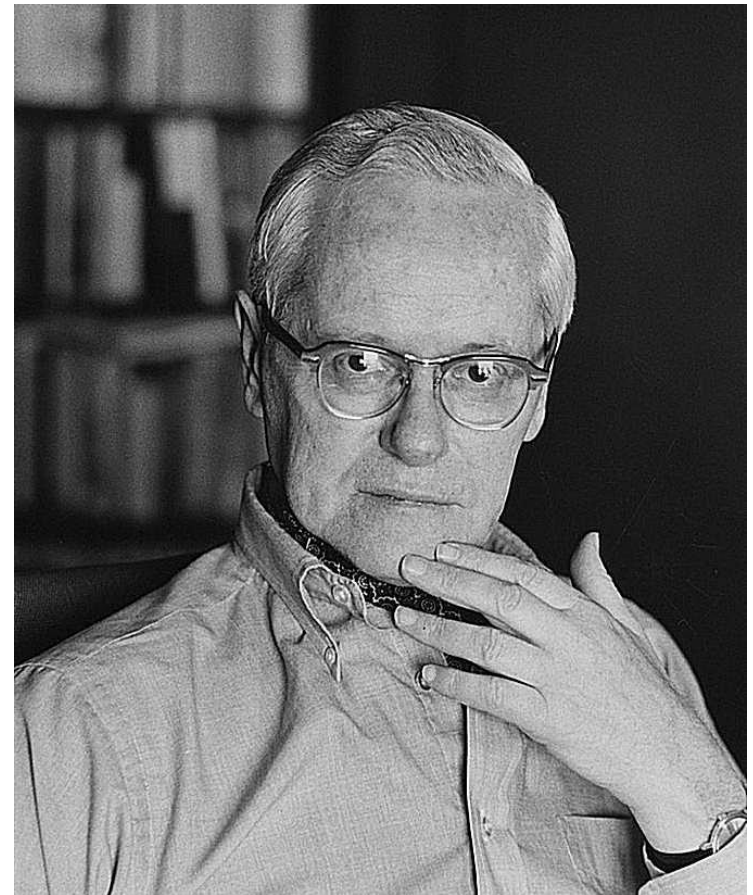
His answer: **‘Good’ science blossoms in pluralistic, democratic societies”**

- Authoritarian times
 - Suppression of rights and liberties, especially for migrants & minorities

Bureaucratic orders vs. public good
‘Open’ science in ‘closed’ societies?

Robert K. Merton

The normative structure of science



Evaluation -- about valuing -- about values -- Open science?

- OS in a time of increasing authoritarian govts.
- OS policies in China, Turkey, Russia? ... or some EU countries?



- Gov't decree to shut down websites without judiciary request
- Police requests to shut down of Git-hub application
 - Application for coordinating demonstrations at Git-hub. **Civil disobedience in non-violence actions**

A research evaluation process

What are the goals of this research?

What are the criteria to see alignment of activities with goals?

What indicators capture the criteria?

ISRIA Statement,
Paula Adam et al. 2018

Potential Opening up of Participation

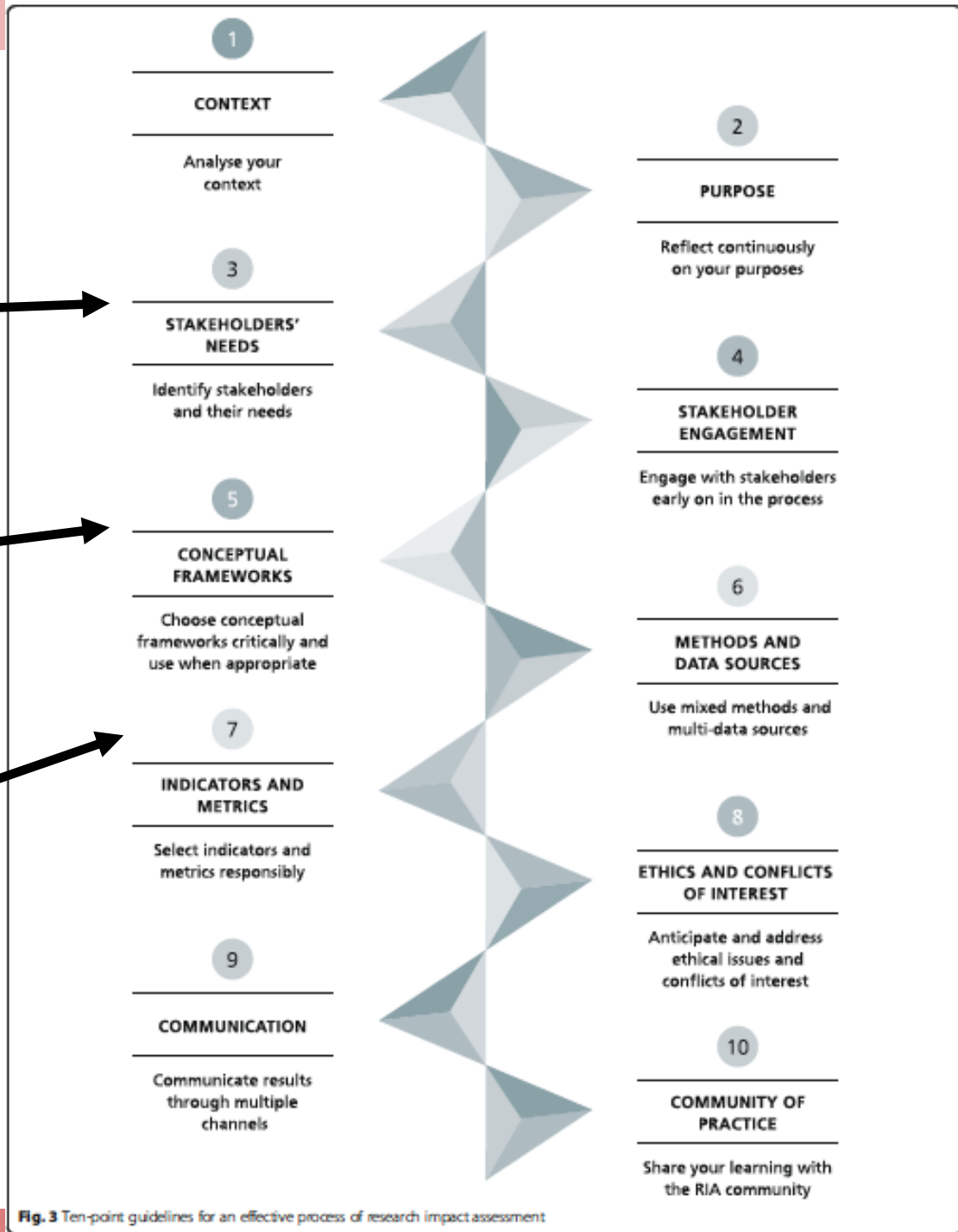


Fig. 3 Ten-point guidelines for an effective process of research impact assessment

References

- Adam (2018) Adam, P., Ovseiko, P. V., Grant, J., Graham, K. E., Boukhris, O. F., Dowd, A. M., ... & Sued, O. (2018). ISRIA statement: ten-point guidelines for an effective process of research impact assessment. *Health research policy and systems*, 16(1), 8.
- Bezuidenhout, Louise M., Sabina Leonelli, Ann H. Kelly, and Brian Rappert. (2017) "Beyond the digital divide: Towards a situated approach to open data." *Science and Public Policy* 44, no. 4: 464-475.
- Bornmann, Lutz. (2014) "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics." *Journal of informetrics* 8, no. 4: 895-903.
- Chavarro, D., Tang, P., & Ràfols, I. (2017). Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. *Research policy*, 46(9), 1666-1680.
- Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah De Rijcke, and Ismael Rafols.(2015) "Bibliometrics: the Leiden Manifesto for research metrics." *Nature News* 520, no. 7548: 429.
- Leonelli, Sabina. (2018) Without urgent action big and open data may widen existing inequalities and social divides. *Impact of Social Sciences Blog*.
- Merton, Robert K. (1942) "Science and technology in a democratic order." *Journal of legal and political sociology* 1, no. 1: 115-126.
- Molas-Gallart, Jordi. (2012) "Research governance and the role of evaluation: A comparative study." *American Journal of Evaluation* 33, no. 4: 583-598.
- Roessner, David. (2000) "Quantitative and qualitative methods and measures in the evaluation of research." *Research Evaluation* 9, no. 2: 125-132.
- Rottenburg, Richard, Sally E. Merry, Sung-Joon Park, and Johanna Mugler, eds. *The world of indicators: The making of governmental knowledge through quantification*. Cambridge University Press, 2015.
- Scott, James C. *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press, 1998.
- Siler, K., Haustein, S., Smith, E., Larivière, V., & Alperin, J. P. (2018). Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ*, 6, e4269.
- Sugimoto C. R., Work S., Larivière V. et al. (2016) 'Scholarly Use of Social Media and Altmetrics: A Review of the Literature', arXiv preprint arXiv: 1608.08112.
- Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, and P. Wouters. (2015) *The metric tide: report of the independent review of the role of metrics in research assessment and management*.

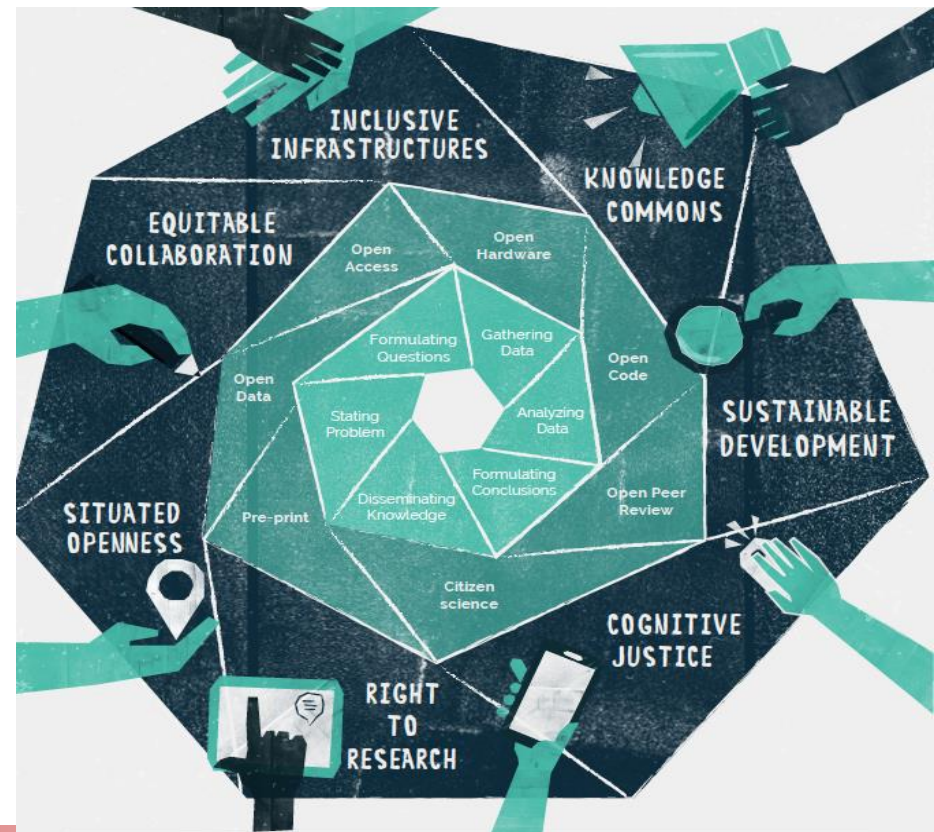
Manifesto of OCSDNet (Open and Collaborative Science in Development Network)

For all the claimed benefits of OS...

- ... current model is NOT making science a more **inclusive** practice.
- ... many scientists continue to be **underrepresented and excluded**
- ... new technologies exclude those with **limited digital rights** .
- ... citizens rarely get to **shape the research agenda**.

Principles

- knowledge commons
- cognitive justice
- situated openness
- right to research
- equitable collaboration
- inclusive infrastructures
- use knowledge as a pathway
to **sustainable development**



Evaluation informed by indicator frameworks

Dimensions to consider in evaluation

- Goal of evaluation (monitoring, allocation, learning)
- Research mission (academic, training, social contribution)
- Level of evaluation (system, institutional, individual)
- Scientific field and methodological approaches
- Potential stakeholders, audiences and beneficiaries
- Research environment:
 - Human and technical resources
 - OS capabilities, infrastructure

Between FULLY TAILORED & UNIVERSAL → **Prêt-à-porter**