Stellenbosch, November 2016

Under-representation of research in the global south

Biases in mainstream journal indexing systems

Ismael Rafols, Diego Chavarro (*) and Tommaso Ciarli

Ingenio (CSIC-UPV), Universitat Politècnica de València

SPRU (Science Policy Research Unit), University of Sussex, Brighton, UK

Uses of indicators: Pressing demands of research management and evaluation --- Can indicators help?

Yes, indicators can help make decisions…

- Reduce time and costs
- Increase transparency and sense of objectivity
- Reduce complexity, accessible to managers

but do they lead to the "right" decisions?

Evaluation gap (Wouters): "discrepancy between evaluation criteria [implicit in indicators] and the social and economic functions of science"

Dominant missions:

***Academia – excellence *Innovation – economic growth**

Missions ***not well covered***: agriculture, public health, defence, development, social inclusion,…

often for populations in marginalised dimensions: geographic, social, linguistic

Indicators may have consequences on research system

- Incentives: indicators signal to stakeholders what is important.
  - Goal displacement: instead of mission, follow indicators
- 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)
  - Invisible / undone science (Hess, 1997)
    - Reducing the options of possible research
- Local knowledge important for:
  - Supporting local communities in specific contexts
    - Agriculture, health
  - Global challenges need local knowledge
    - Climate change, pandemics…understanding local conditions is crucial to explaining global effects and trends.

An agenda for more inclusive metrics

- **Inclusiveness in the inputs**
  - *Broadening out:* Create more diverse indicators
    - Indicators of open science, RRI, hidden, social innovation
    - Improve representation of SSH scholarship, languages other than English, the “South”,…

- **Inclusiveness in the outputs**
  - *Opening up:* develop toolkits that allow exploration of choices. New ways of presenting indicators
    - From indicators to “vector” tools
    - Interactive visualisations

- **Inclusiveness in the policy process (??)**
  - Develop new social processes on use of indicators
    - STI indicators as tools for interpretation and deliberation (R. Barré)

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.

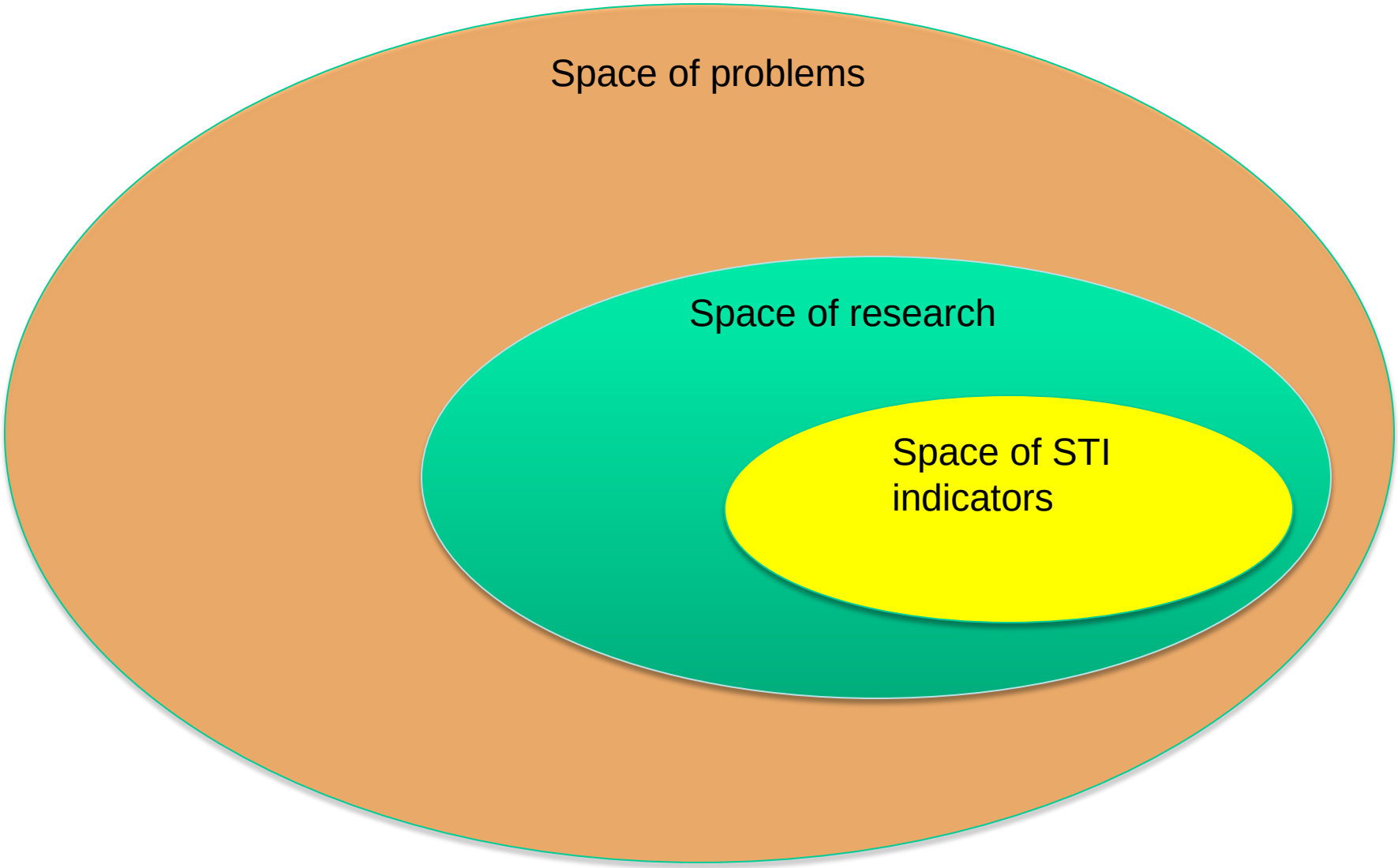
4. Keep data collection and analytical processes open, transparent and simple.
5. Allow those evaluated to verify data and analysis.

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.

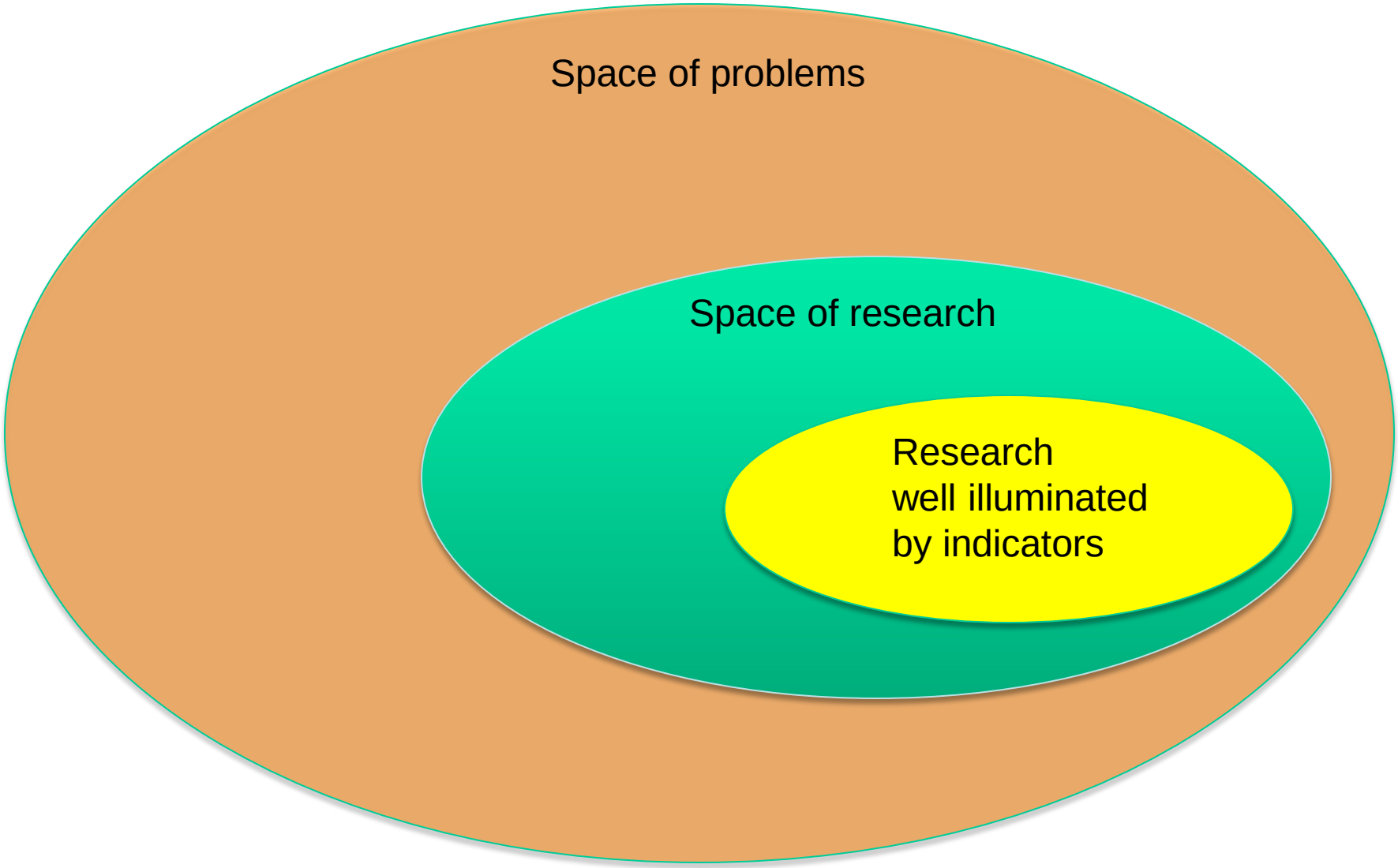
9. Recognize the systemic effects of and indicators.
10. Scrutinize indicators regularly and update them.

Hicks, Wouters, Waltman, de Rijcke and Rafols (Nature, 2015)

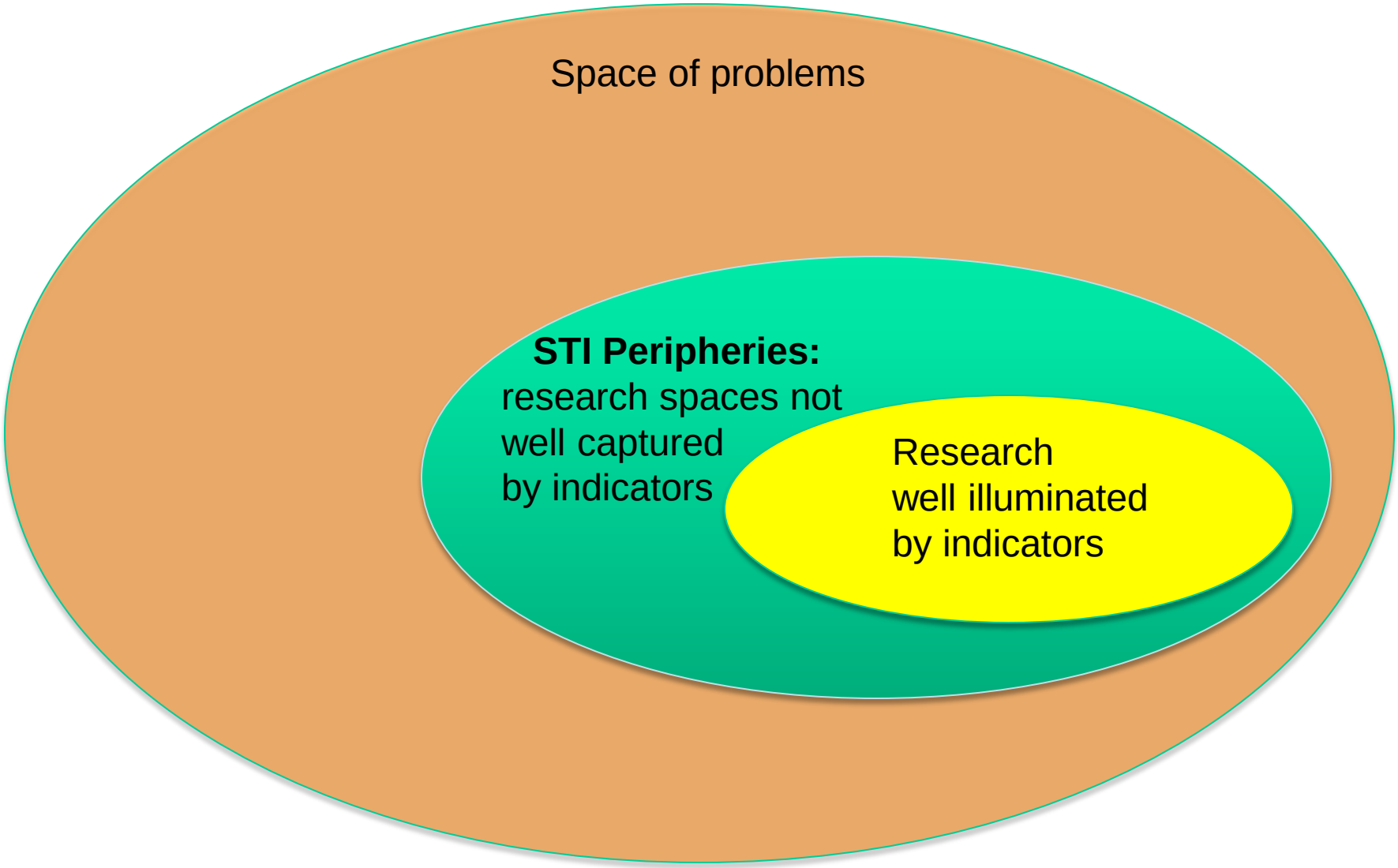
Problems, research, indicators and peripheries



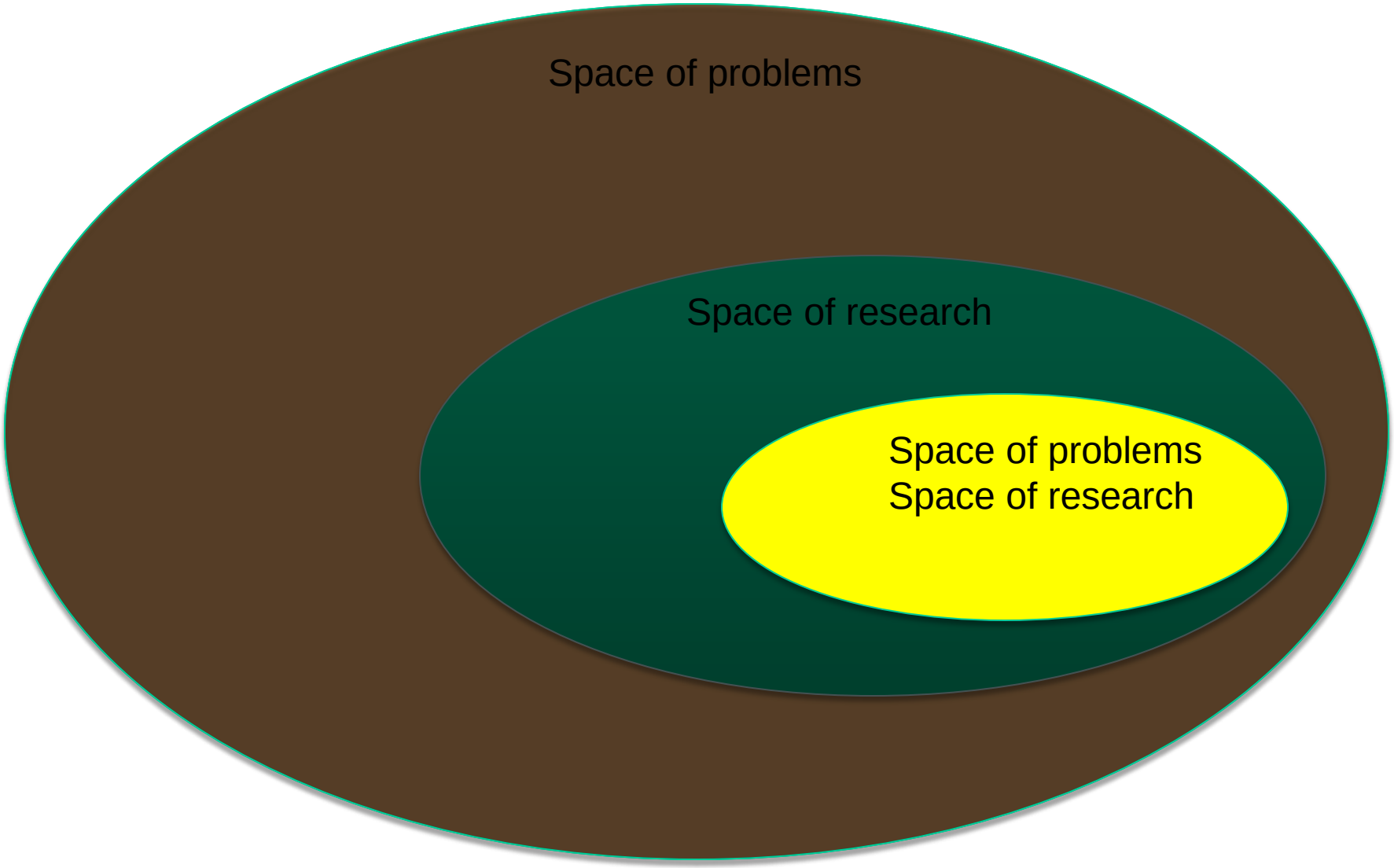
Problems, research, indicators and peripheries



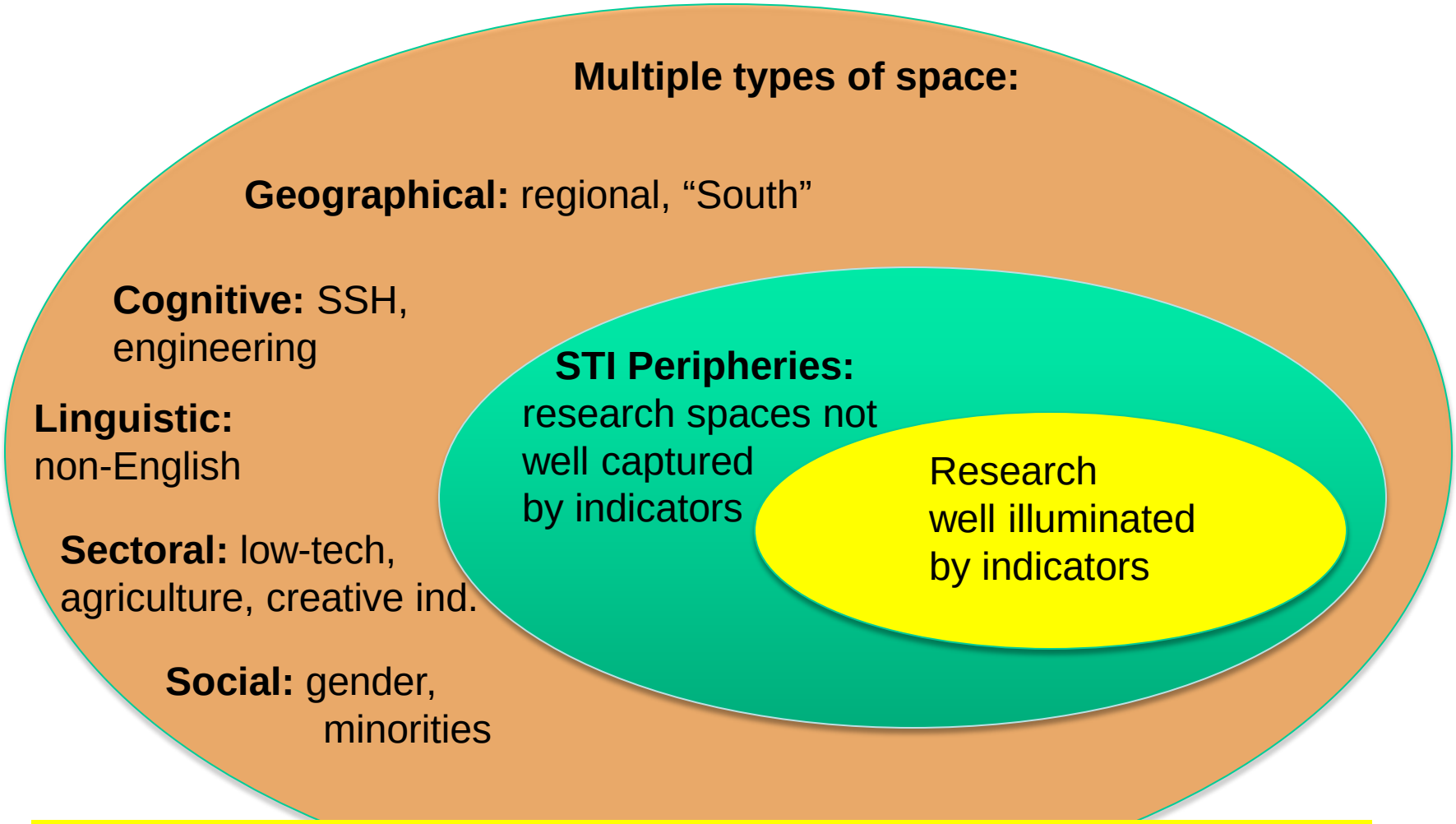
Problems, research, indicators and peripheries



Streetlight effect in indicators: mistaking light with “problems”



Problems, research, indicators and peripheries



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

Bias in bibliometric databases

- Web of Science is biased towards English-speaking publications and biomedical publications (Archambault et al., 2006).

- Scopus has a broader coverage, but similar ranks regarding country production over different fields,
  - ‘indicators of scientific production and citation at the country level are stable and largely independent of the database’(Archambault et al., 2009, p. 1320).

- In **international benchmarking,** major int’l organisations continue to use the main databases WoS (e.g UNESCO, 2010) and Scopus (e.g Royal Society, 2011).

- Recommendations have been made on the need to improve scientometric indicators in order to "properly evaluate global science" (Royal Society, 2011, p. 107).

What is the extent of bias?

1. Coverage in Journal Indexing Systems (JIS) according to country of publishing house?
2. Coverage of publications according to affiliation of authors (for rice research)?

The coverage of low and middle income countries is very low (xx%)

Does research quality explain publication of “non-mainstream” journals?

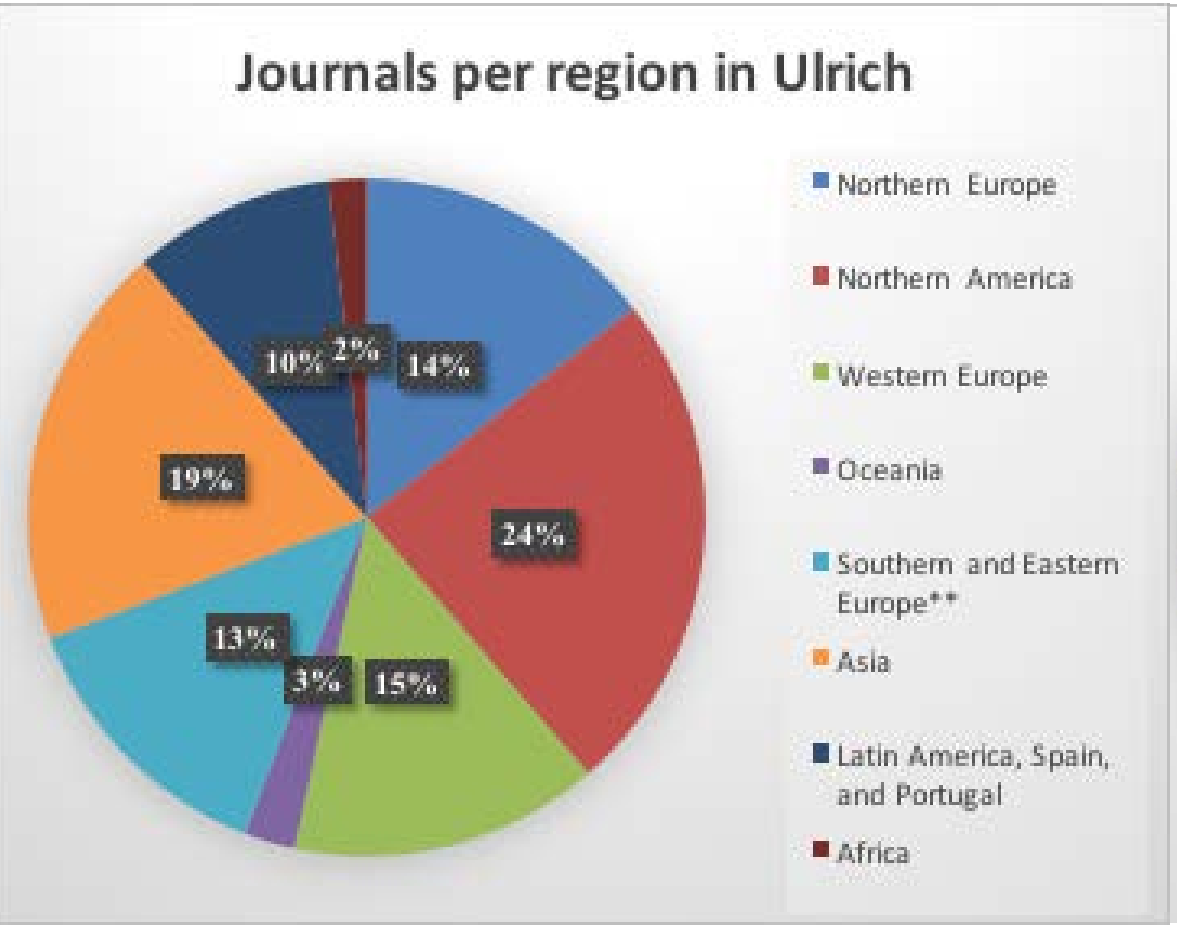
3. What properties explain inclusion in WoS for Latin American journals in the 2000s?

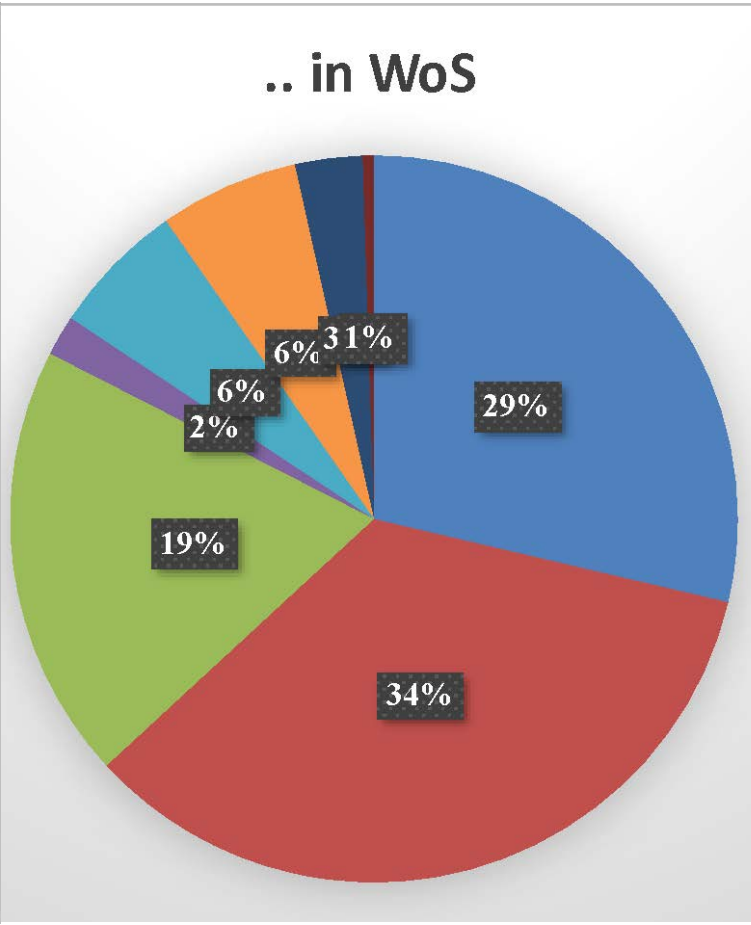
4. Why do researchers publish in non-mainstream in journals – the case of Colombia?

Non-mainstream journals play a role in science communication.

1. Coverage in Journal Indexing Systems (JIS) according to country of publishing house

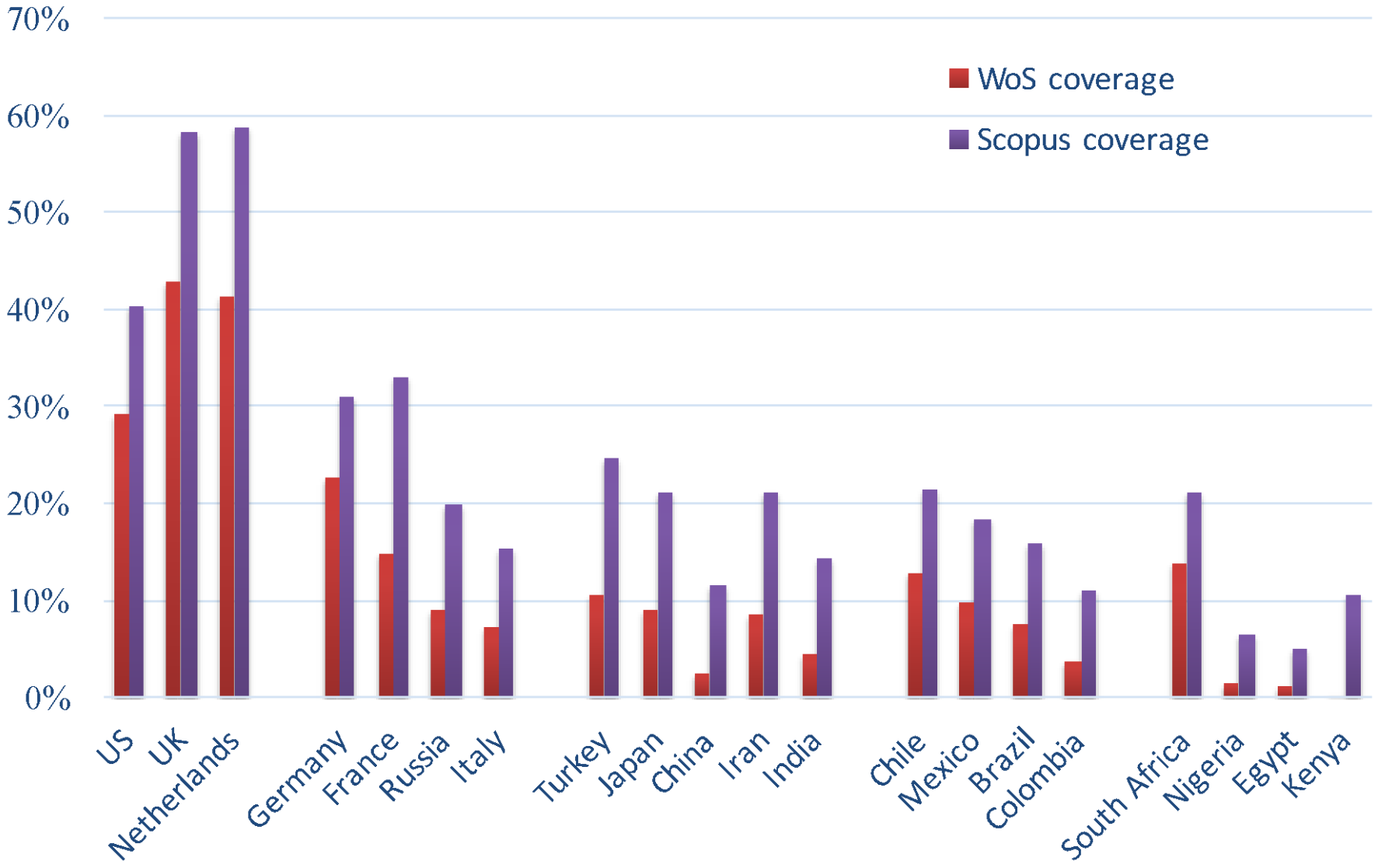
Percentage of journals in Ulrich and WoS (2012)



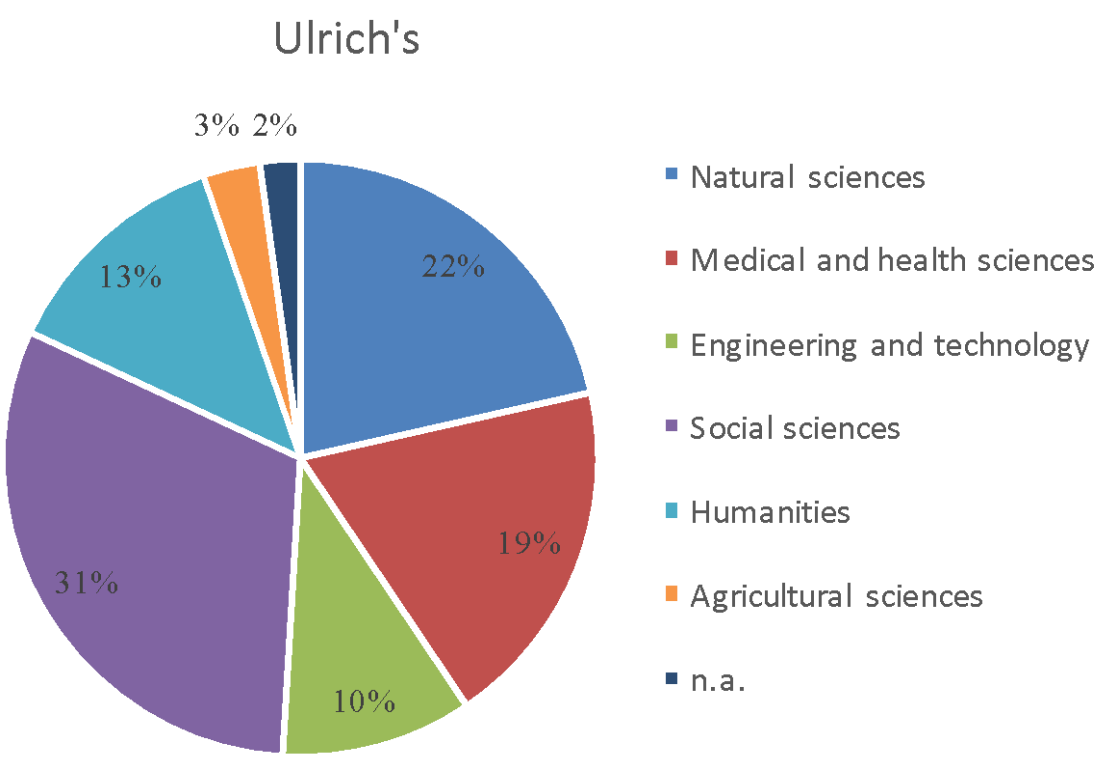


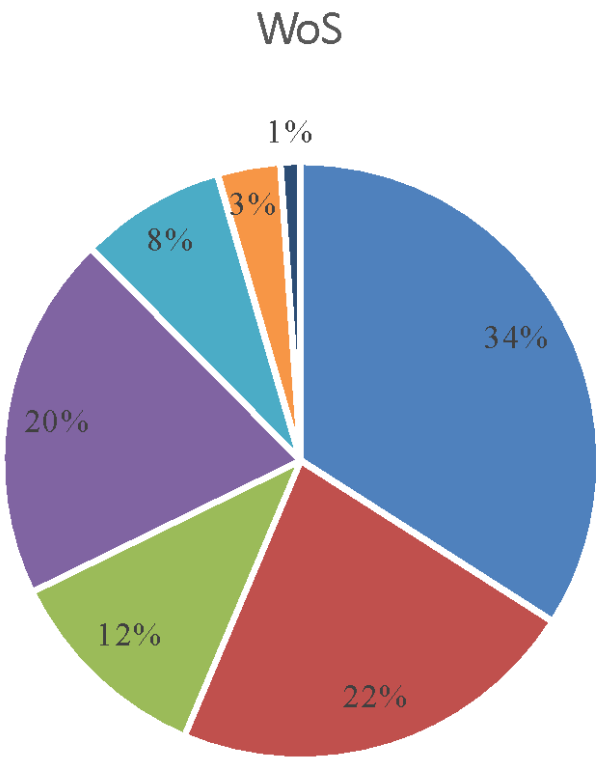
Coverage of WoS and Scopus in comparison to Ulrich

Country coverage of WoS and Scopus vs. Ulrich

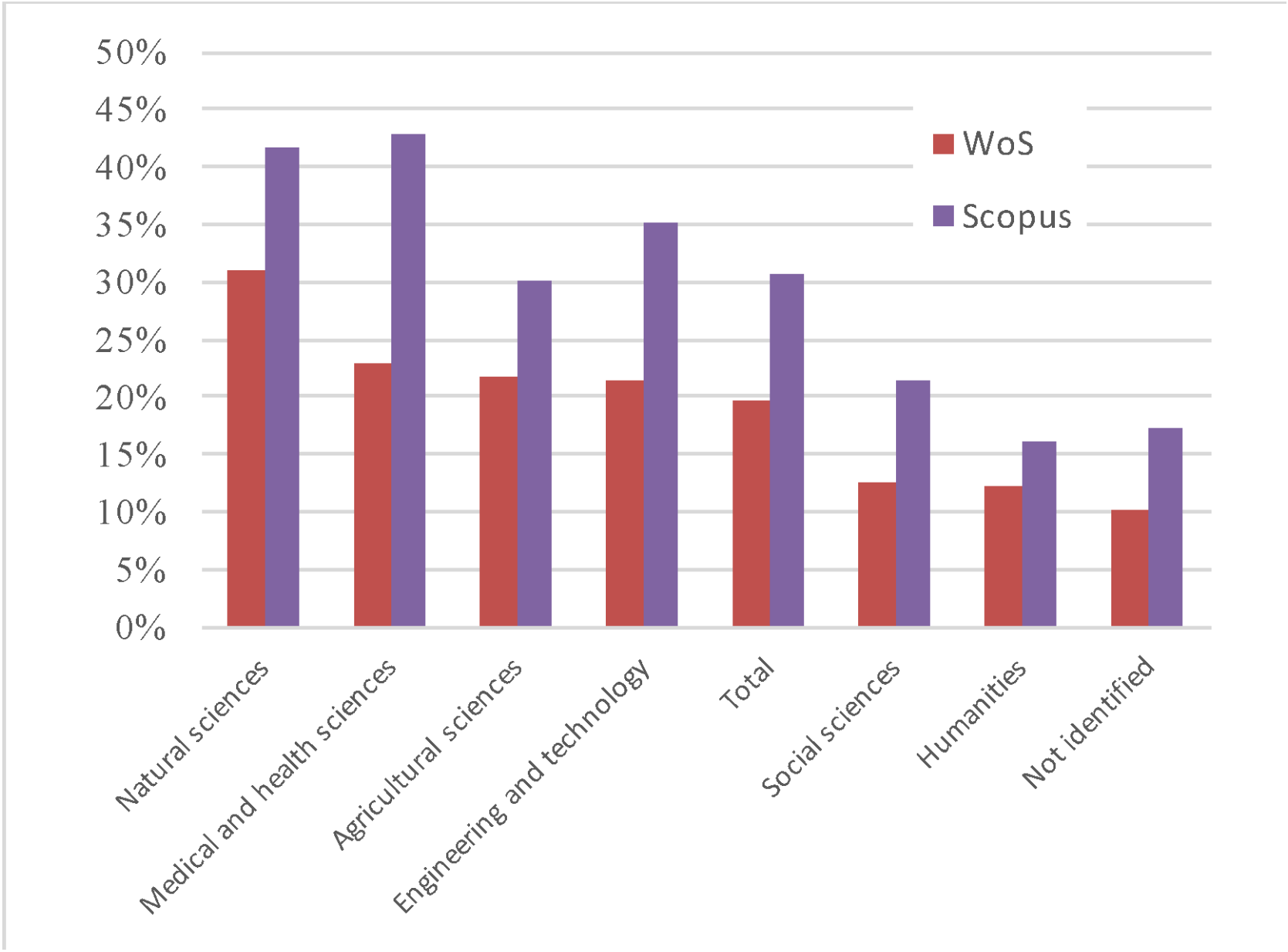


Percentage of journals in Ulrich and WoS (2012)





Disciplinary coverage of WoS and Scopus vs. Ulrich



2. Coverage of publications according to affiliation of authors (for rice research)

Comparison of WoS, Scopus and CABI

Databases compared

Publications on rice for 2003-2012 were downloaded from:

- **WoS** (including SCI-Expanded, SSCI, A&HCI, CPCI-S and CPCI-SSH) searching “rice” or “oryza” in the field “topic”.

  » **35,056 records**

- **Scopus** searching in **title, abstract or keywords**, i.e. TIT-ABS-KEY ("rice" OR "oryza").

  » **48,602 records.**

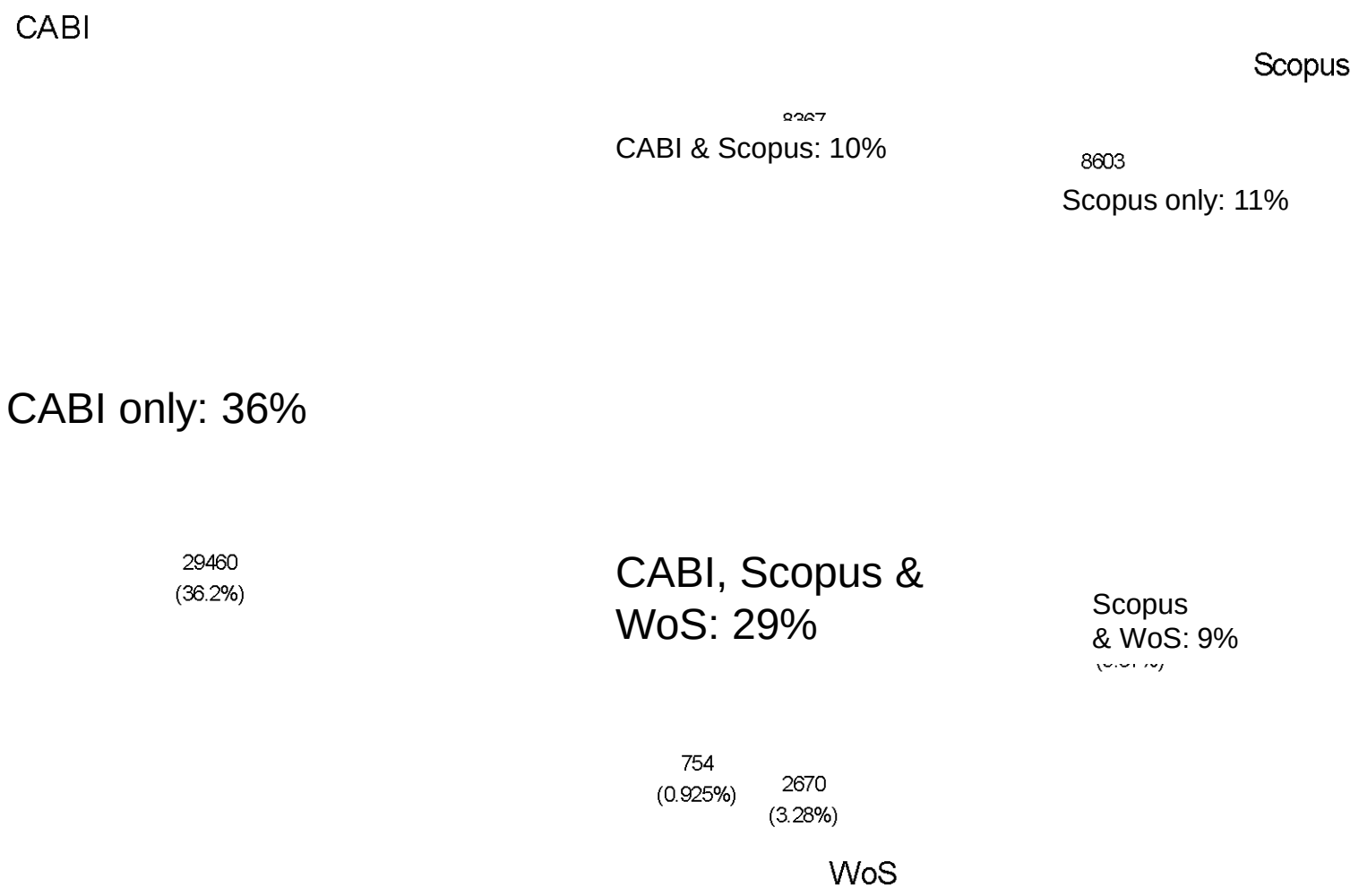
- Database **CAB Abstracts**, documents with “rice” or “oryza” were searched in **title and abstract**.

  » **62,574 records**

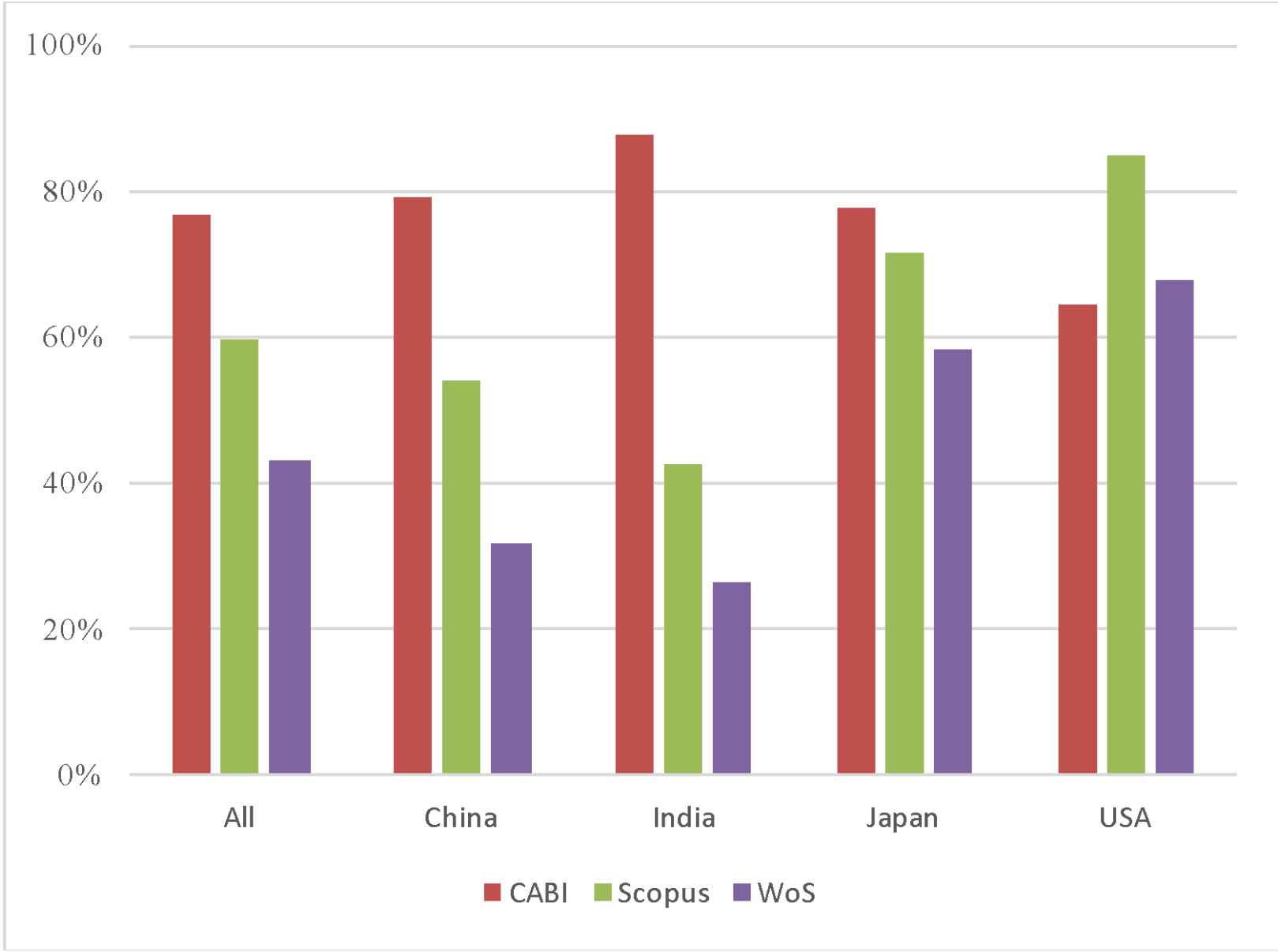
CAB Abstracts, CABI (Centre for Agriculture and Bioscience International) is a database focused on **environment and agriculture**.

Only address of first author.

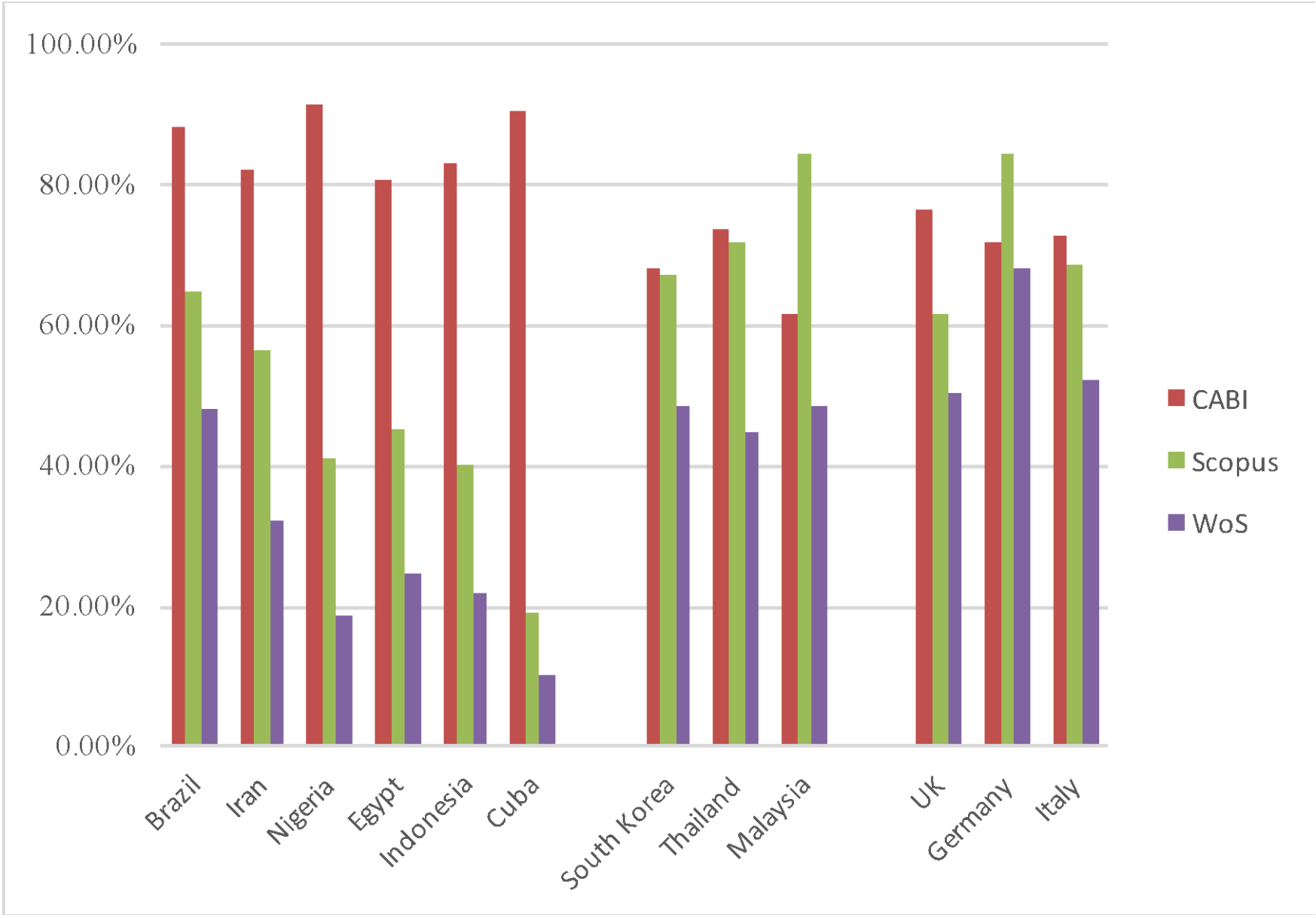
Relative coverage of databases



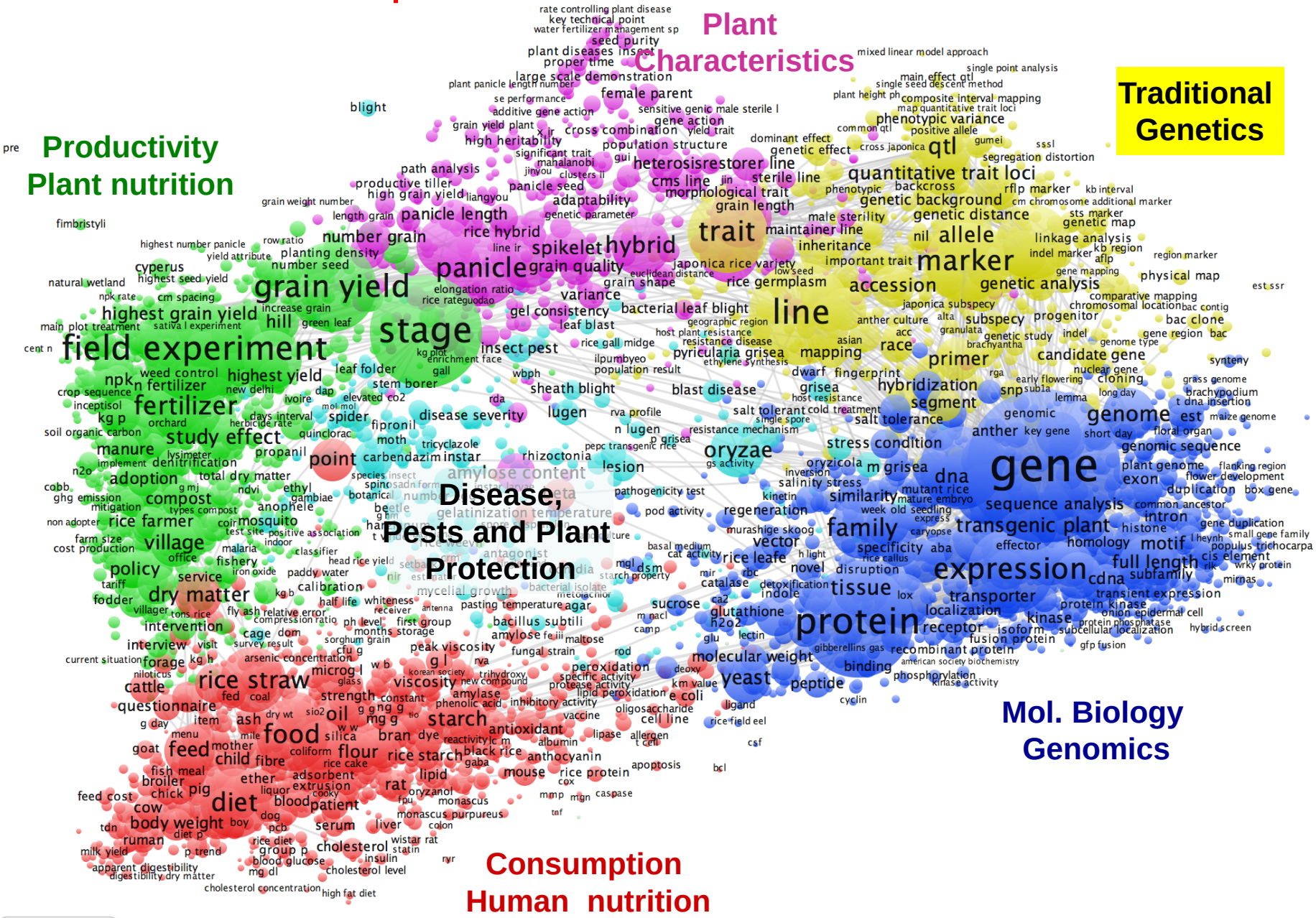
Coverage of largest countries in rice research



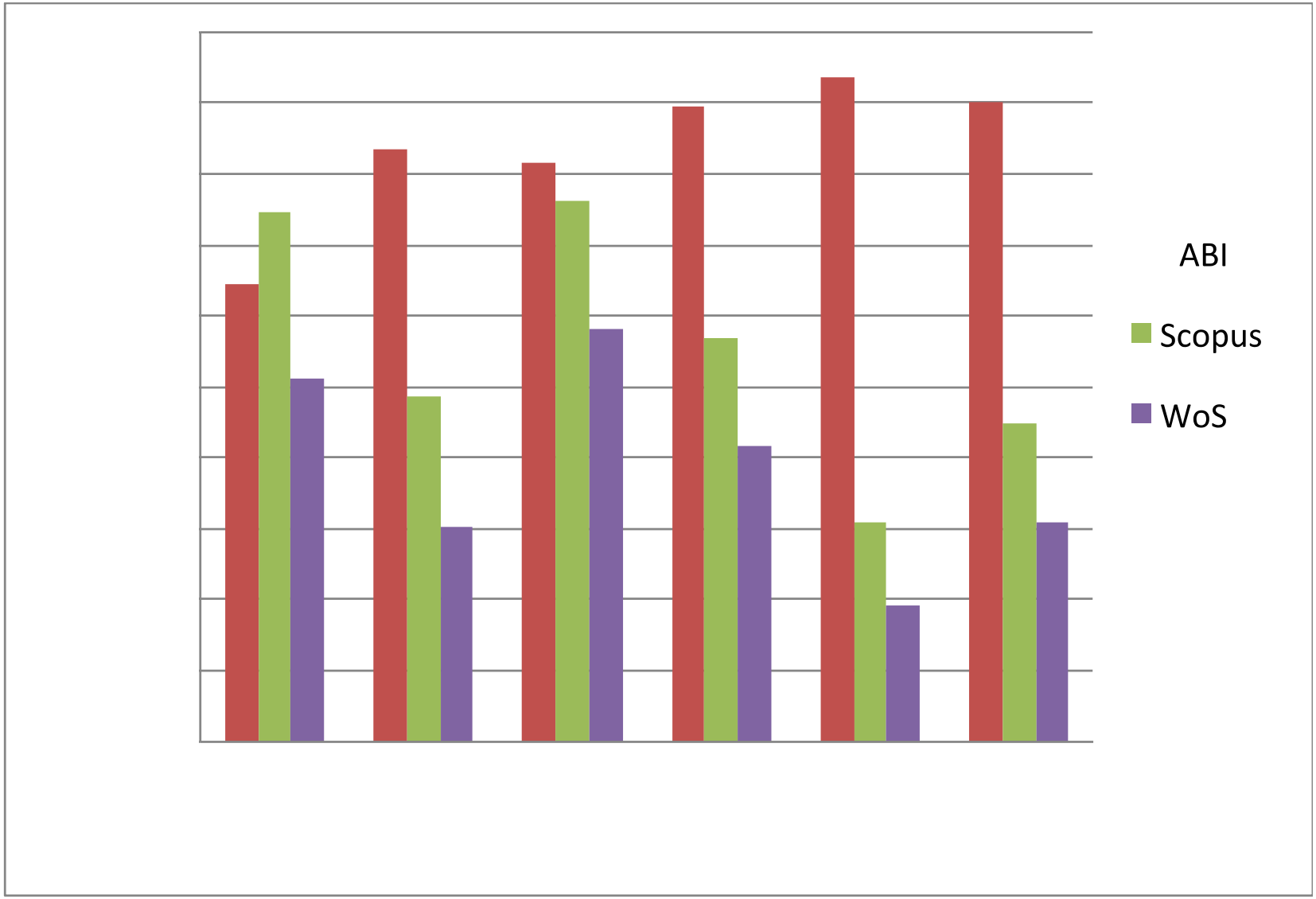
Coverage of other countries in rice research

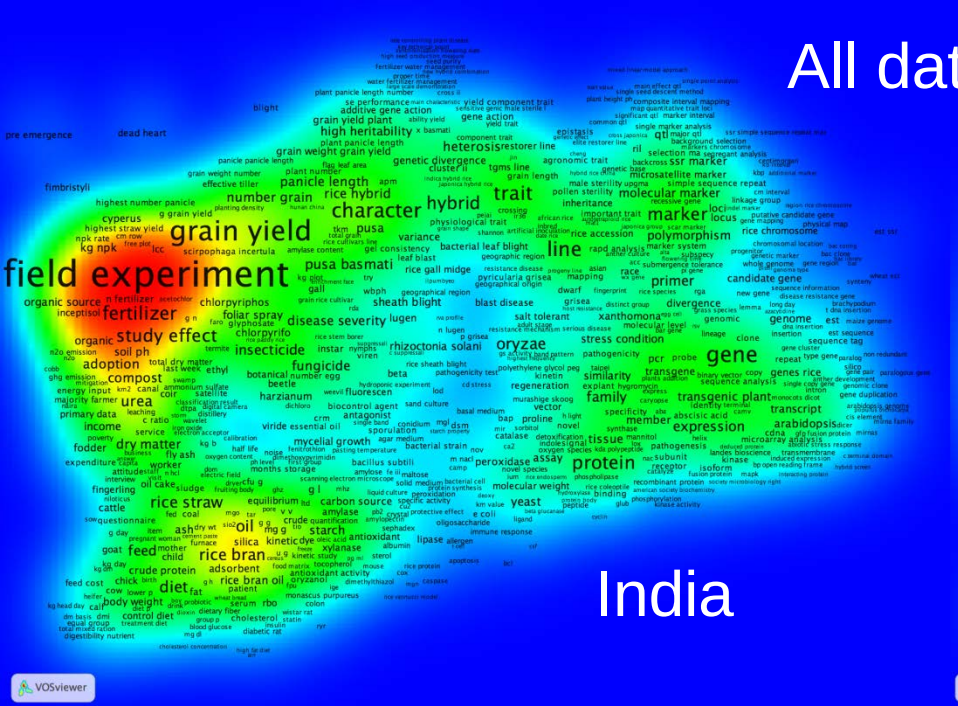


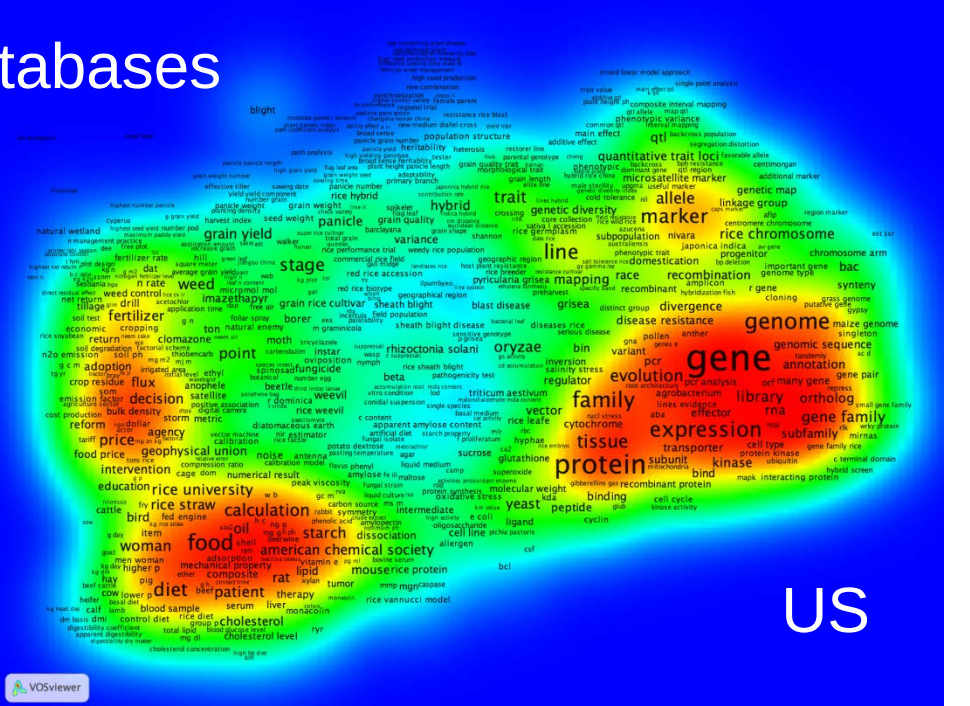
Rice research topics

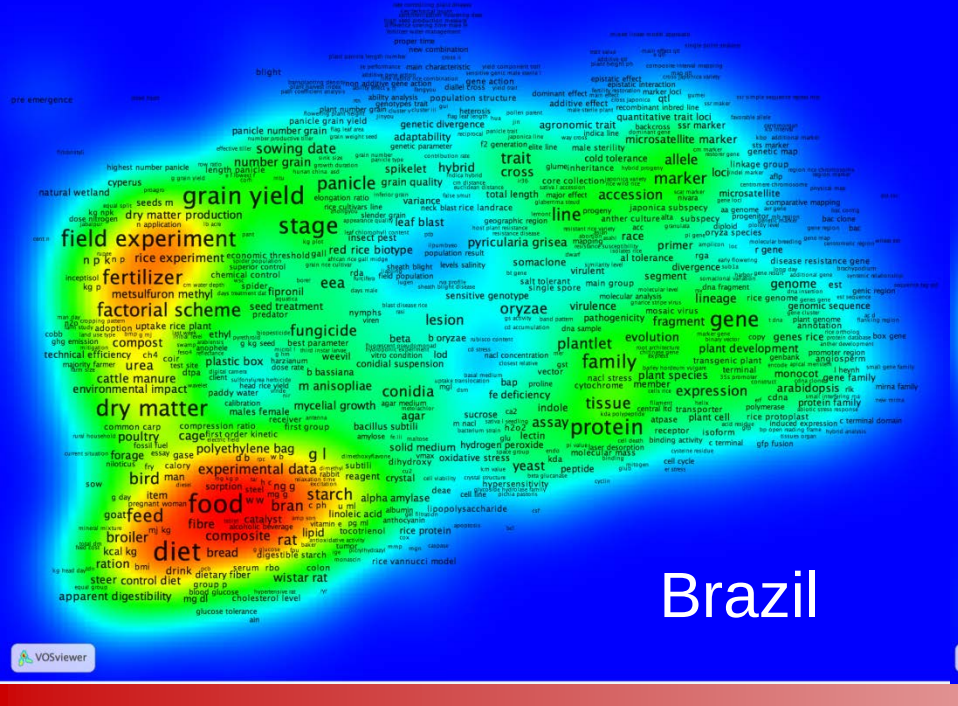


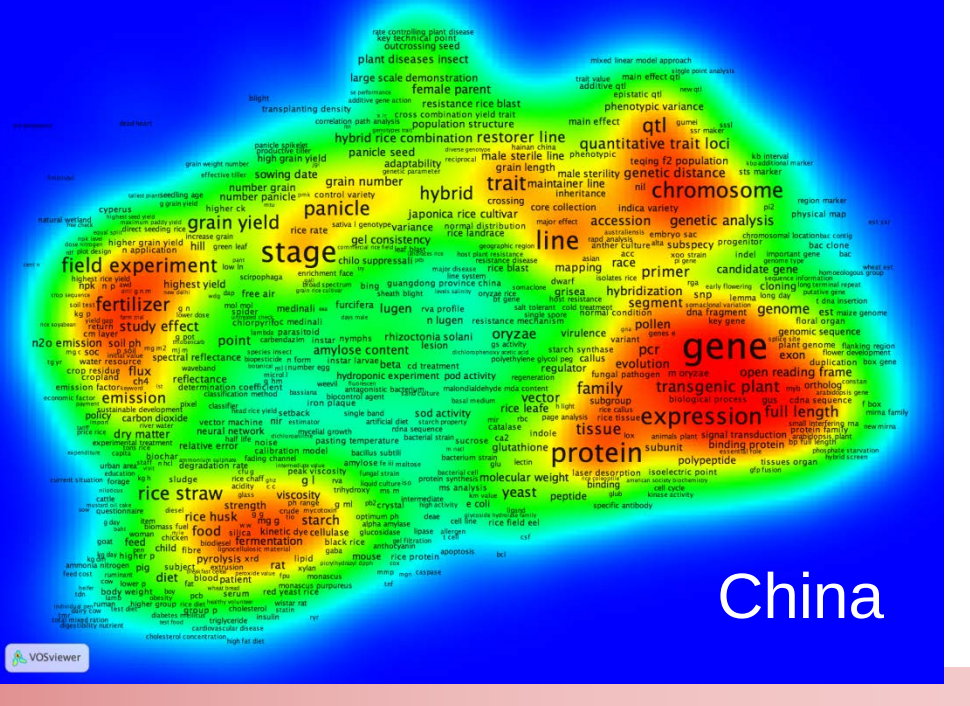
Coverage by database and research topic (2003-2012)

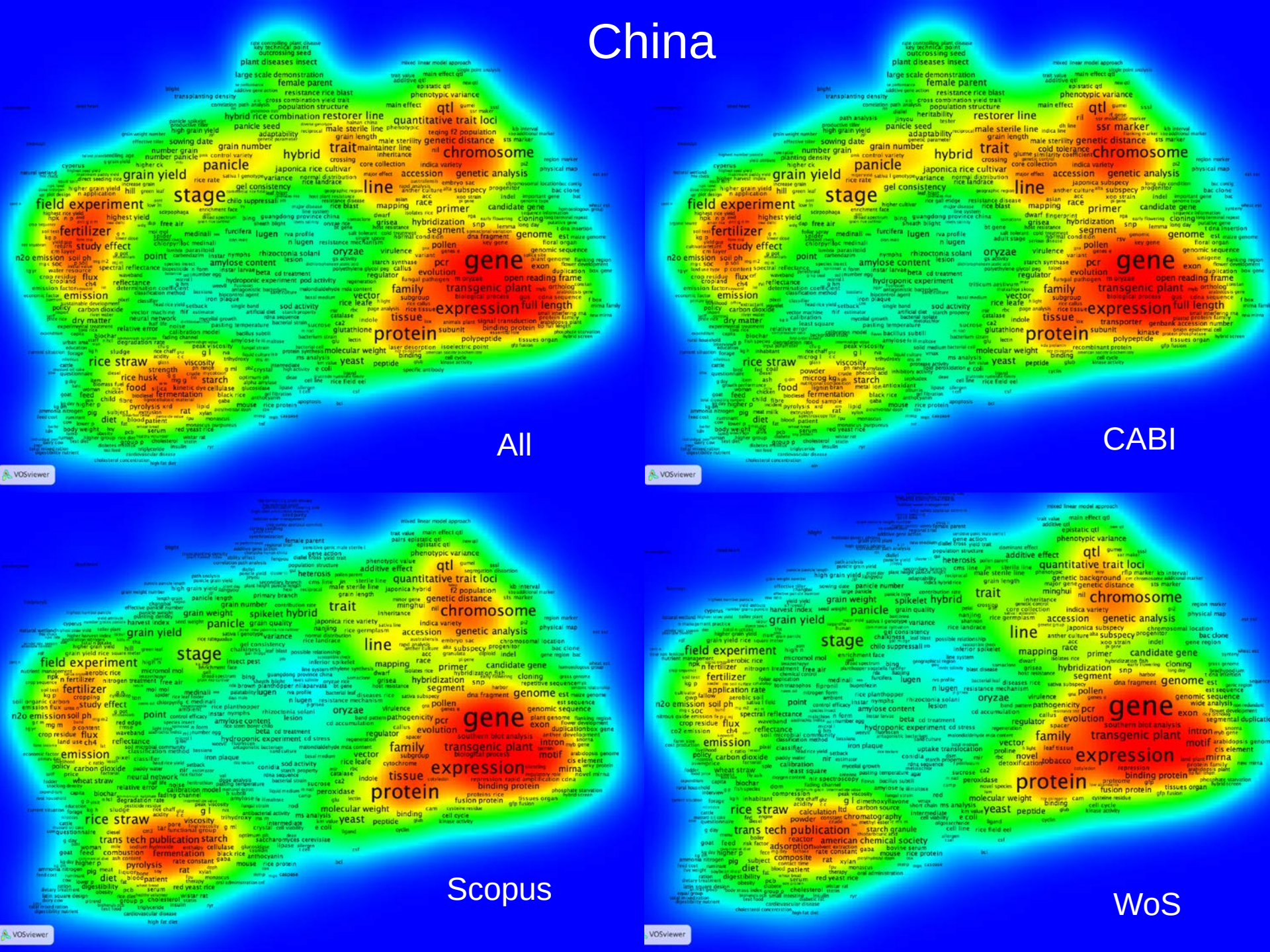



All databases
India


US


Brazil


China


China
All
CABI
Scopus
WoS

3. What properties explain inclusion in WoS for Latin American journals in the 2000s?

“Quality!??”

Modeling inclusion of journals in WoS in Latin America

What explains inclusion of WoS of LatAm journals between 2005 and 2012?

Latindex (a database with editorial standards)
Indexed between 2005-2012: 270 journals
Not indexed between 2005-2012: 1090 journals

Model #	Model: logistic regression
1	WoS ~ Editorial standards + h-Index + Journal age
2	WoS ~ Editorial standards + h-Index + Journal age + Language + Discipline + Country + Type of publishing organisation + Type of journal

Modeling inclusion of journals in WoS in Latin America

What explains inclusion of WoS of LatAm journals between 2005 and 2012?

Not only quality and age (**universalistic characteristics**)
but also discipline and country (particularistic characteristics) explain selection

Model #	Model: logistic regression
1	WoS ~ **Editorial standards + h-Index + Journal age**
2	WoS ~ Editorial standards + **h-Index + Journal age** + Language + **Discipline + Country** + Type of publishing organisation + Type of journal

- Medicine and Social Sciences less likely to be included
- From more to less likelihood : Spain, Brazil, Mexico, Venezuela, Portugal, Argentina and Colombia

4. Why do researchers publish in non-mainstream in journals – the case of Colombia?

Conventional explanation:

The research produced in middle and low income countries does not have enough quality to be published in journals indexed by mainstream JIS, especially in high impact factor journals.

Interviews on reasons to publish in (not) indexed journals

Methods

- 30 interviews in Colombia
  - Chemistry
  - Business and Management
  - Agricultural Sciences.
- Public and private universities and research institutes
- Triangulation with other data sources (pubs)

Findings

1. Training (13 respondents)
2. Knowledge bridging (11 respondents)
3. Knowledge-gap filling (15 respondents)

1. Training – learning towards WoS/Scopus

- Non-mainstream journals are perceived as “transit stations” to journals indexed by Wos.
- Publishing arena for new PhDs in order to get the necessary skills to publish in WoS-indexed journals.

  - "It is like when you go from elementary school, to high school, to college… one has to make that process in order to get to the big leagues.”
  - Another researcher referred to alternative systems as part of a ladder. At the top of the ladder are the mainstream indexing systems. A researcher on agricultural science defined them as a “stairway” to WoS.

2. Knowledge bridging

- Bridge between papers in mainstream journals and communities with limited access to mainstream literature
  - Due to linguistic or access issues

Mechanisms:

- **Education**: Articles in alternative indexing systems are used in the classroom.
- **Adaptation of knowledge**: knowledge Adaptation occurs when certain concepts or methods are transformed to adapt to a different original context.
- **Dissemination of knowledge**: when a concept that is not new in journals indexed by mainstream systems is introduced in journals indexed by alternative indexing systems and shared with a community.
- **Accessability over pay-wall**: this happens when a researcher publishes documents that incorporate references from closed-access journals into open-access articles.
- **Multilingual Referencing**: this is when researchers publish in languages other than English and incorporate references from journals in English.

3. Knowledge gap-filling

Publication of knowledge that is not well covered in Wos or Scopus.

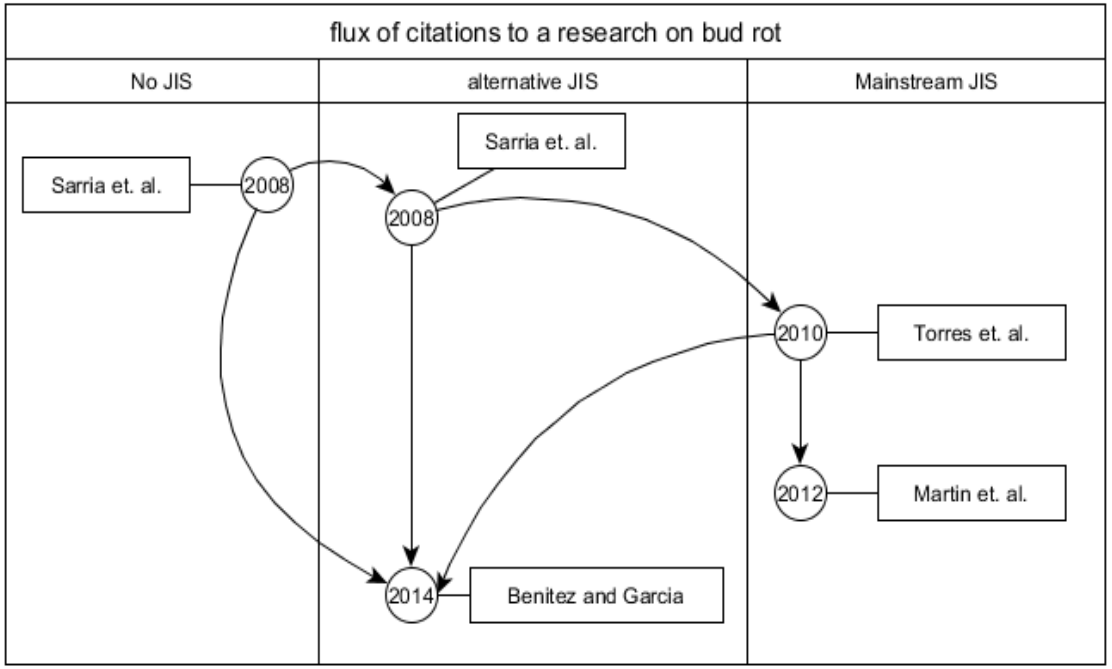
Research not relevant to context of mainstream journals.

Examples:

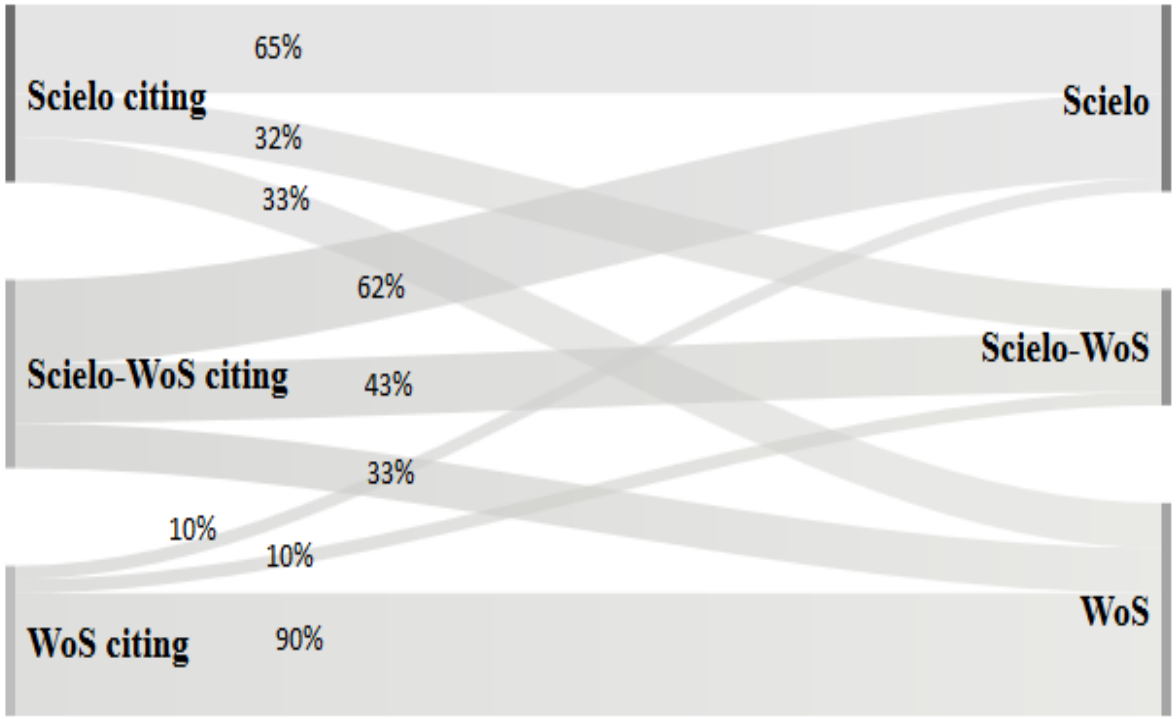
- Passion fruit production, oil palm diseases.
- Innov.n indicators in low and middle income countries

Discovery of treatment on bud rot (disease of palm oil):

from non-mainstream to mainstream journals



Pubs and Cits in Passion fruit



JIS of the paper	Papers Citing Scielo	Papers Citing Scielo-WoS	Papers Citing WoS	Unique citing papers
Scielo	169	84	86	260
Scielo-WoS	376	260	200	605
WoS	166	165	1495	1662
Total	711	509	1781	2527

5. Conclusions

i) Major under-representation of pubs by low and middle income countries

ii) Major under-representation of some topics relevant to global south in conventional databases

ii) Non-mainstream journal are a valuable part of scientific communication.

Findings regarding coverage

- Assumption on the stability of indicators of scientific production (WoS and Scopus) are incorrect (Archambault et al., 2009).

- Important result for international organisations such as FAO, IFRI or UNESCO (UNESCO, 2010) that aim to work on human development.

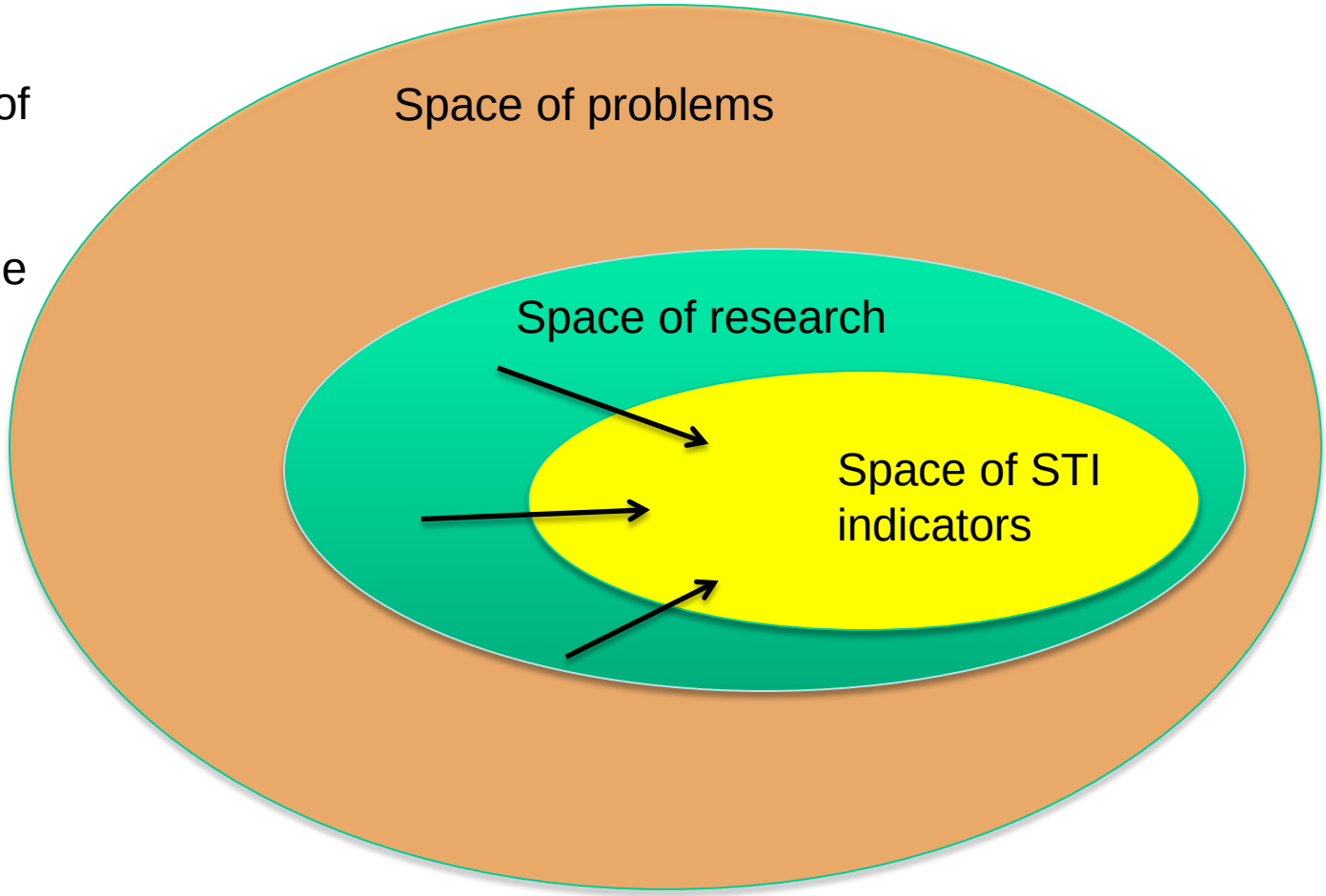
- Relevance of journal indexing systems at the regional level, such as Scielo or Redalyc that aim to provide visibility to local journals, often in languages other than English

Hypothesis: reduced indicator coverage may contract research space

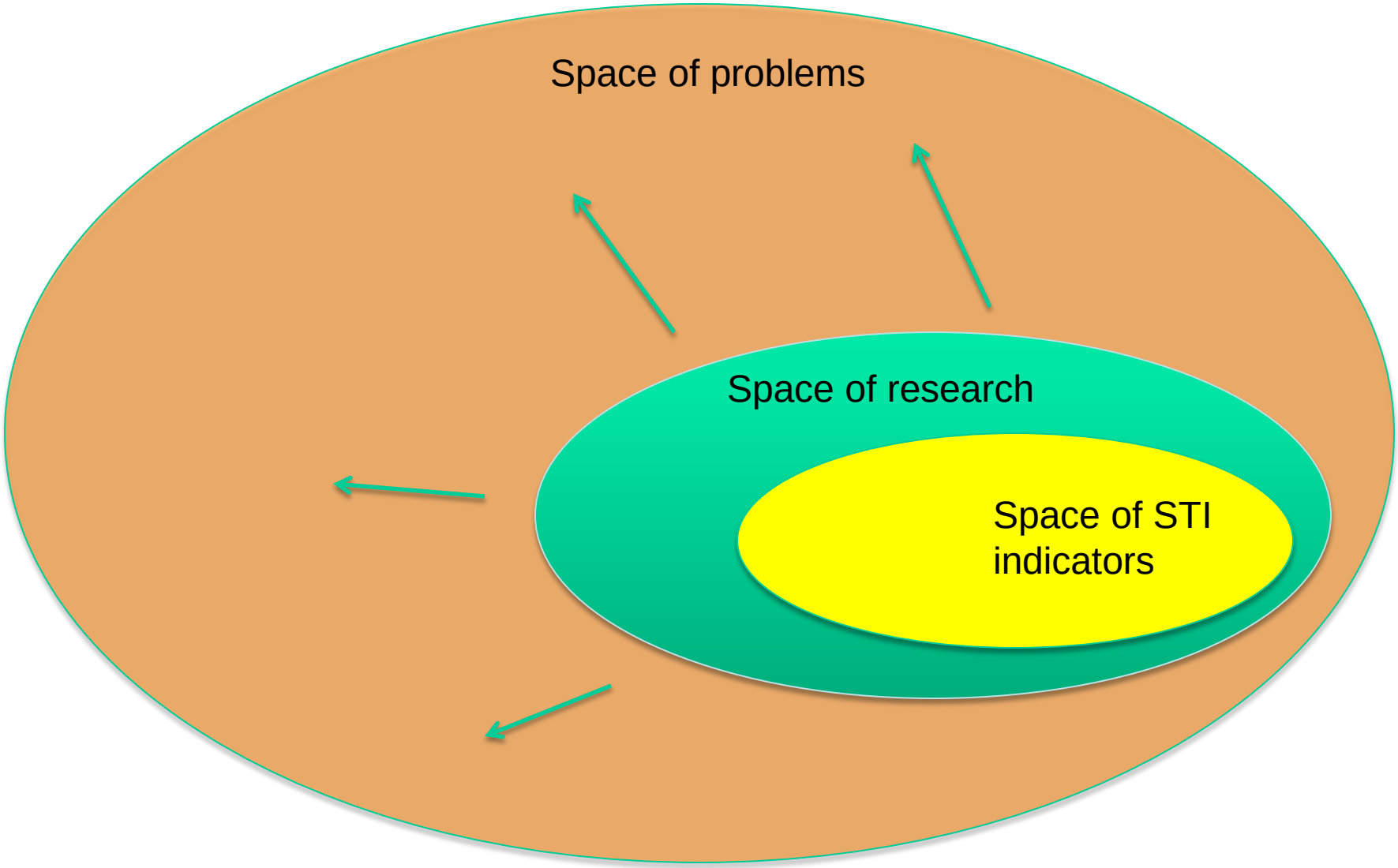
(No “hard evidence”)
The societal needs dealt by research that is under the streetlight effect, will be better rewarded.

Reduced diversity of research efforts...

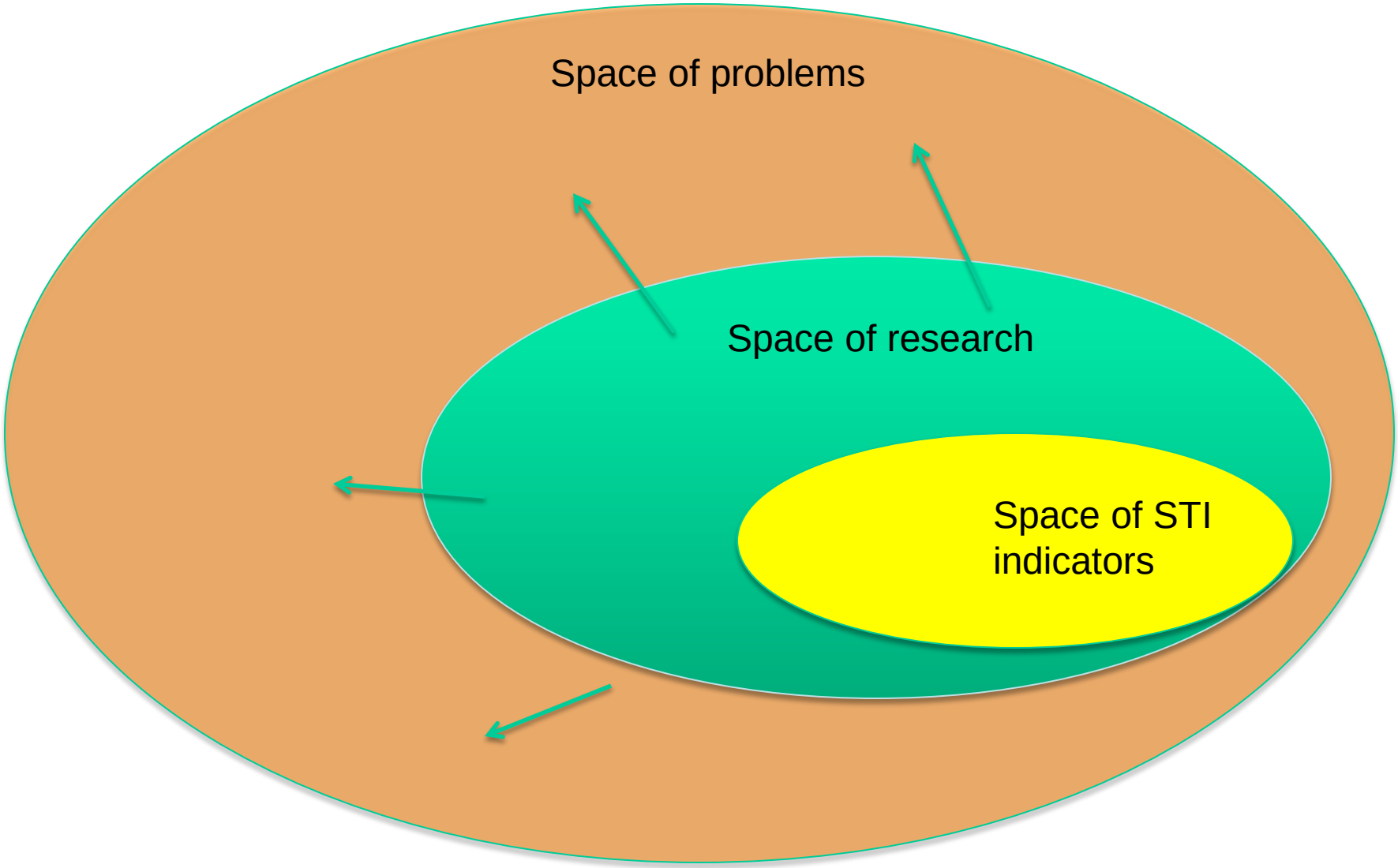
…reduced coverage of societal needs



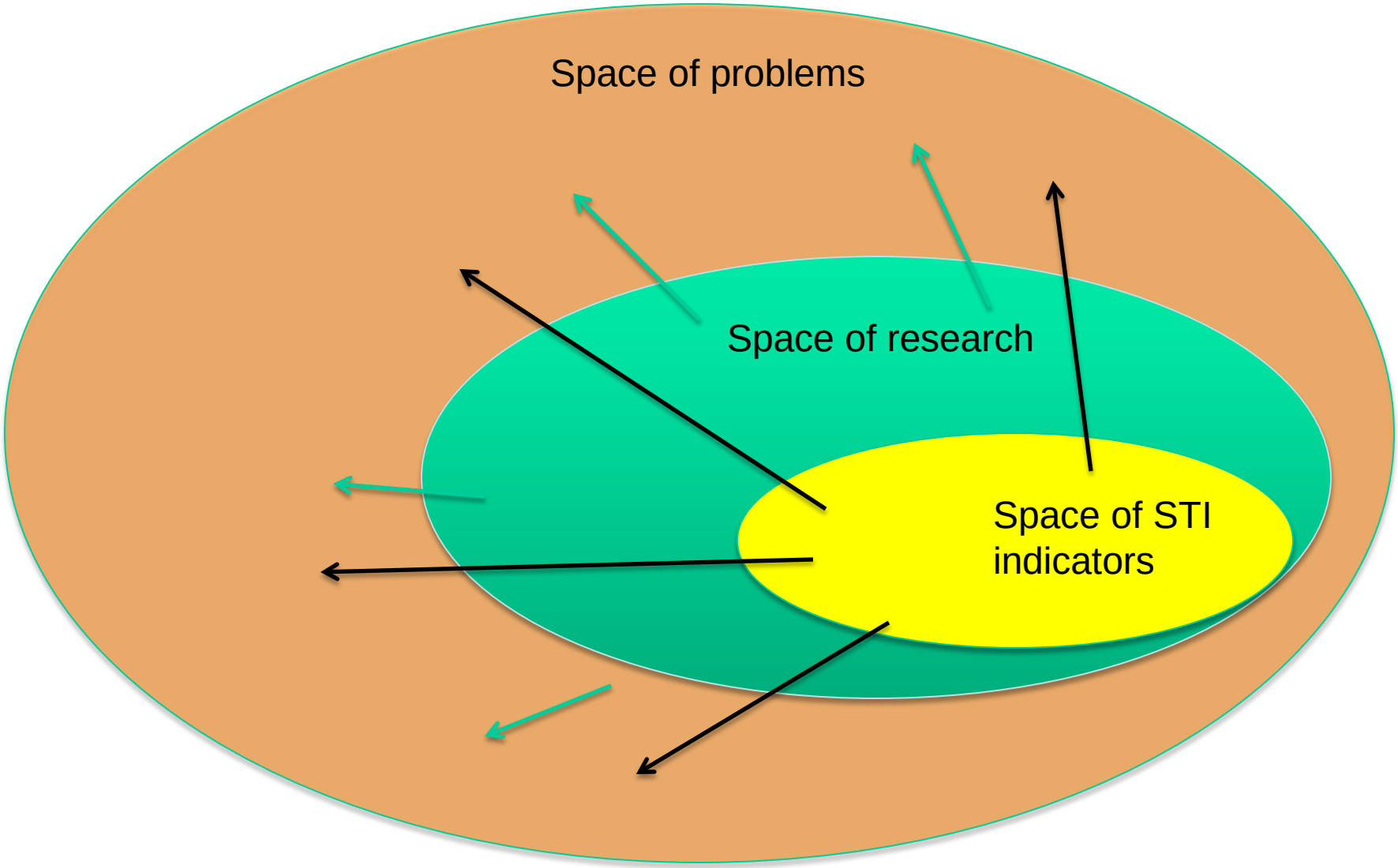
Demands for expanding role of science in society…



Demands for expanding role of science in society…



…may require expanded data sources



A case for comprehensive and inclusive research information infrastructure in the public domain.

Locally relevant knowledge

- Local knowledge important for:
  - Supporting local communities in specific contexts
    - Agriculture, health
  - Global challenges need local knowledge
    - Climate change, pandemics…understanding local conditions is crucial to explaining global effects and trends.

- Mapping research landscape of a topic We need a representation of ALL the knowledge on research topics relevant for a problem.

Indicators may have consequences on research system

- Incentives: indicators signal to stakeholders what is important.
  - Goal displacement: instead of mission, follow indicators
- 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)
  - Invisible / undone science (Hess, 1997)
    - Reducing the options of possible research
- Local knowledge important for:
  - Supporting local communities in specific contexts
    - Agriculture, health
  - Global challenges need local knowledge
    - Climate change, pandemics…understanding local conditions is crucial to explaining global effects and trends.

Findings regarding topic coverage

- Significant differences of coverage between research topics by database
  - Conventional databases (WoS and Scopus) have a larger coverage of molecular biology, traditional genetics and consumption
  - CABI has much better coverage about productivity, plant nutrition, plant characteristics and plant protection.

- High coverage appears to be related to
  - Resarch interests of actors in developed countries such as seed companies, food & industry
- Lower coverage appears to be related to
  - Potential interests of small farmers, local contexts.
  - Exception – nutrition? (to be confirmed)

- Need to contrast results with stakeholders.

Uneven coverage of databases

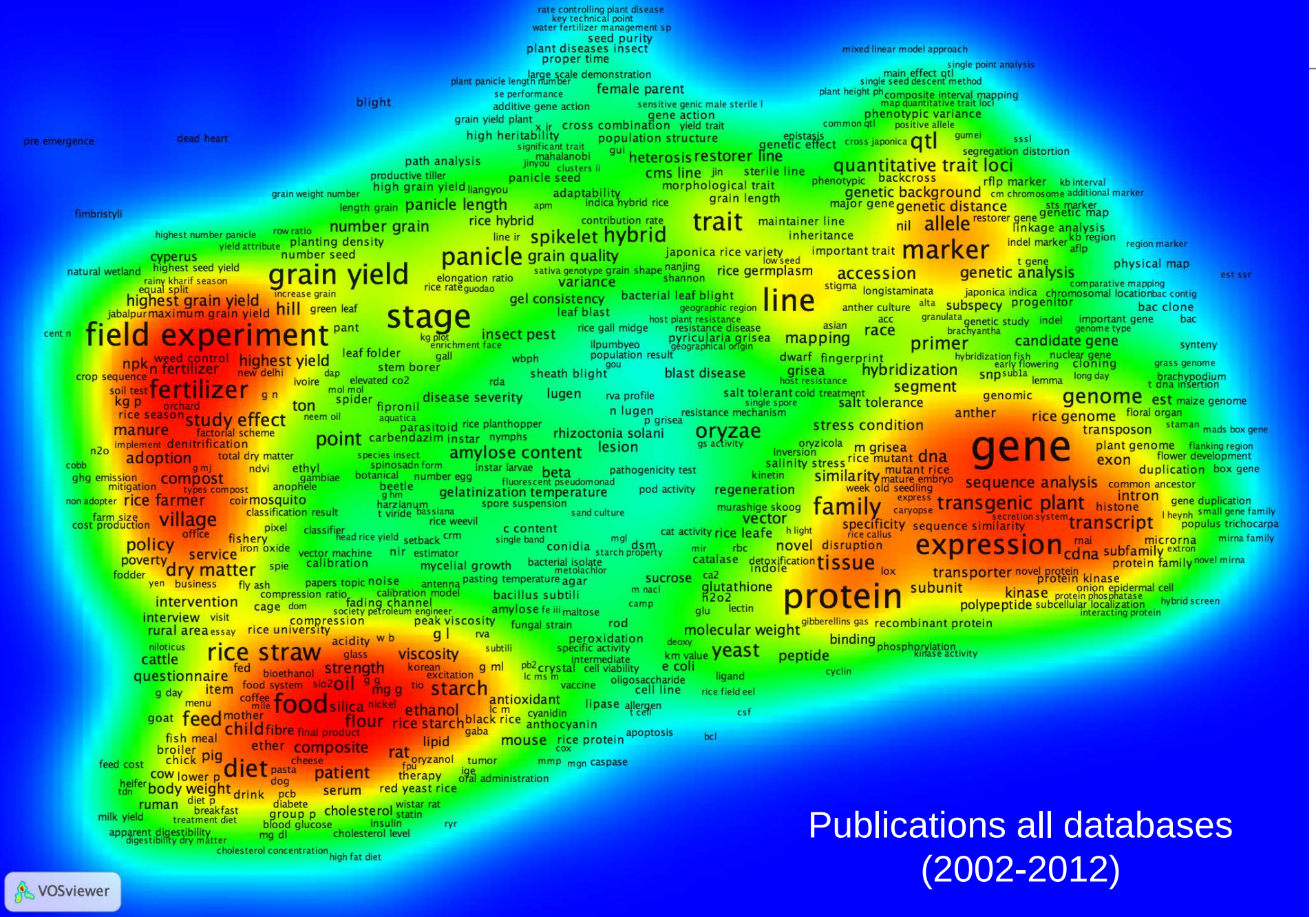
"When comparing databases one easily forgets that each database has a different purpose.

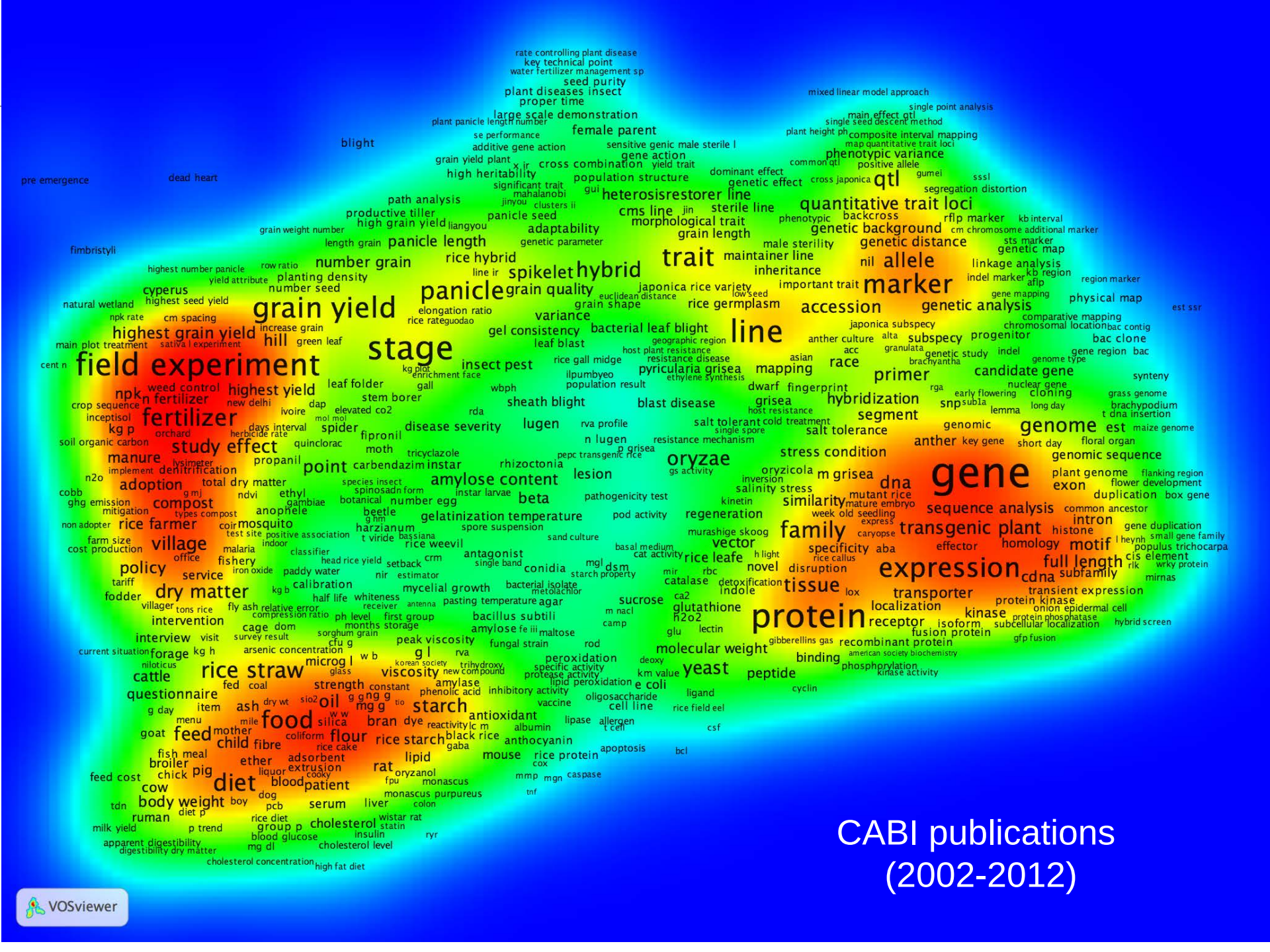
Thus, most of the subject specific databases (including CAB) aim for data completeness, whereas others like Web of Science, following Garfield's original idea, consider only the "core" journals, which are responsible for 80% of the citations in each discipline.

Thus, it is obvious that the coverage is biased in favour of journals published in industrialised countries, because these normally have a higher impact. (…)

Considering the conclusions, **it is alarming to see how often scientometric analyses are performed without even the correct choice of adequate data sources for the required purpose.** "

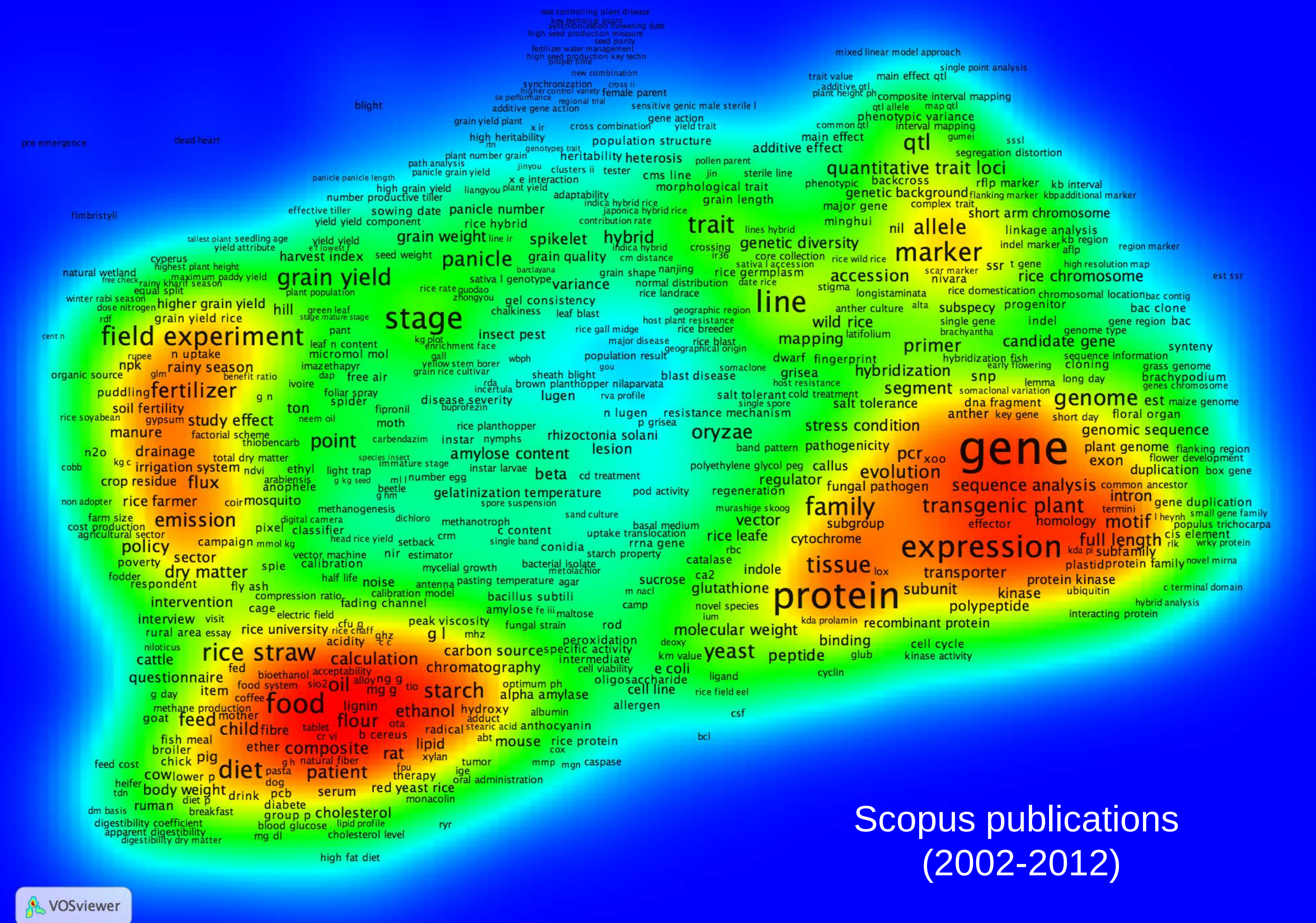
Reviewer of an earlier version of this paper


Publications all databases
(2002-2012)
VOSviewer

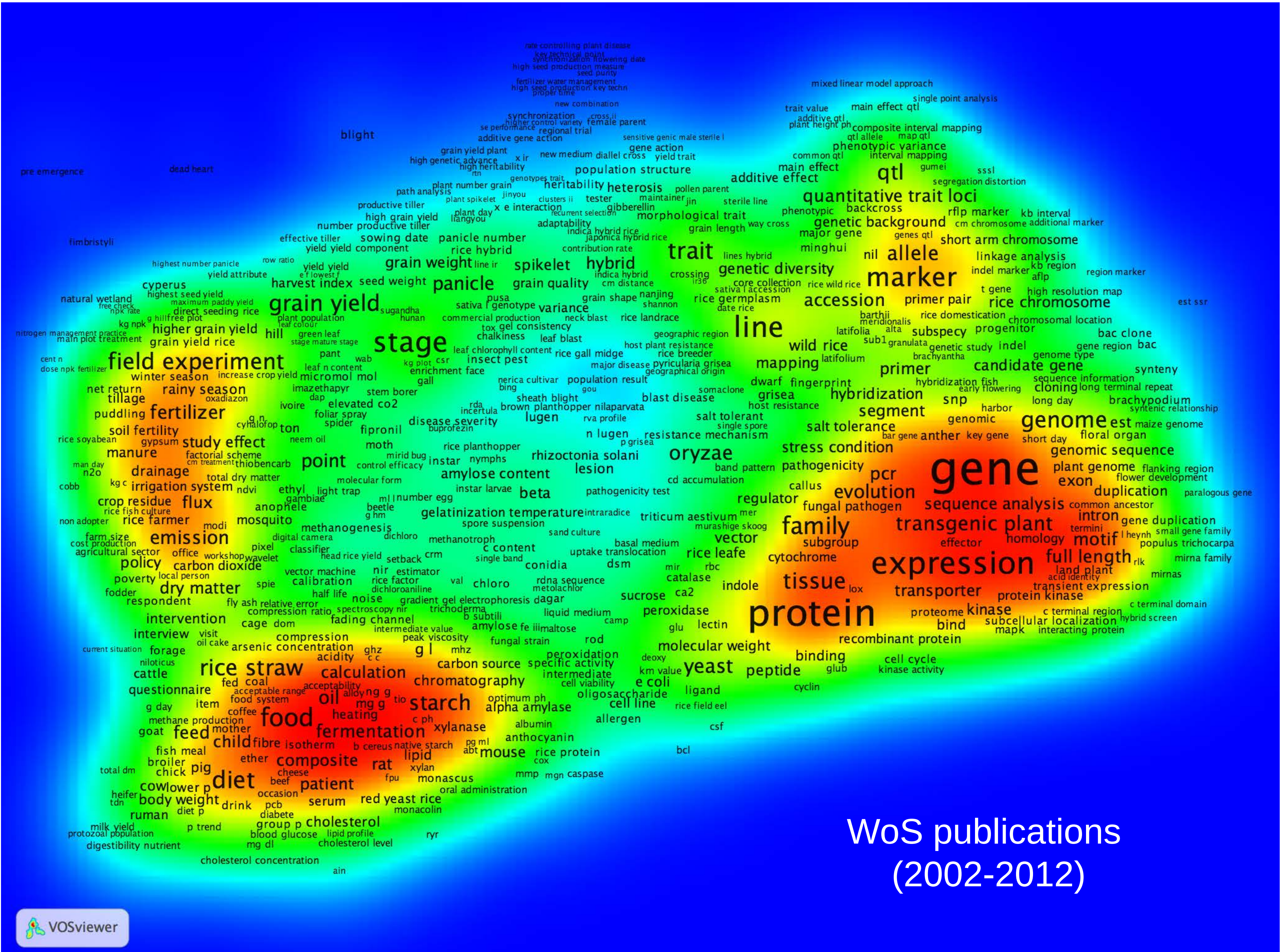


CABI publications
(2002-2012)

VOSviewer



Scopus publications (2002-2012)

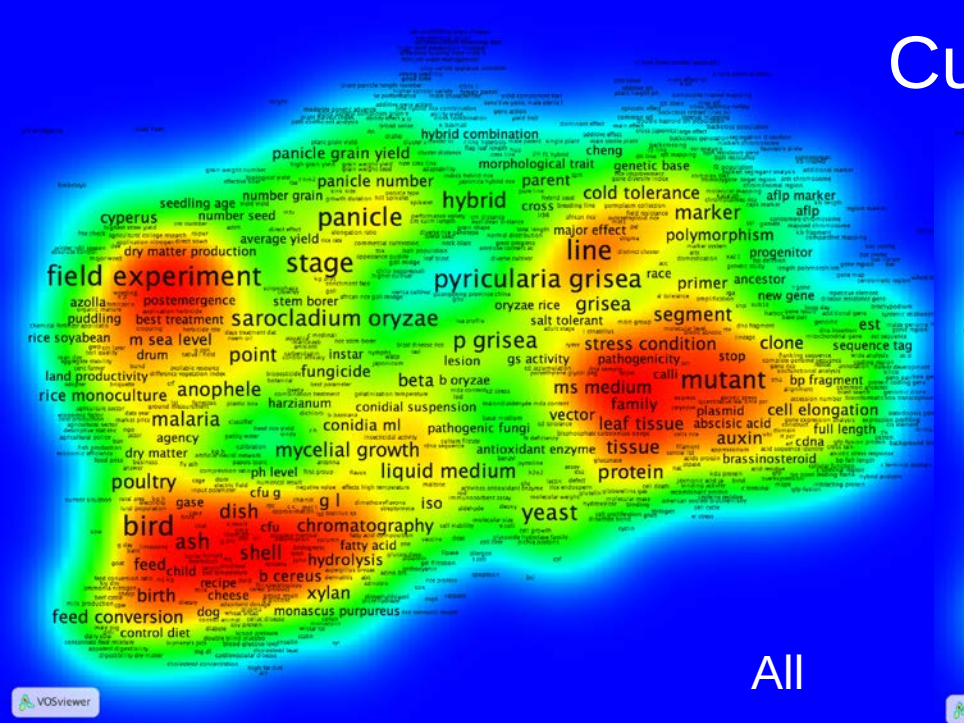


mixed linear model approach
single point analysis
trait value
main effect qtl
composite interval mapping
phenotypic variance
interval mapping
main effect
additive effect
qtl
segregation distortion
quantitative trait loci
backcross
rflp marker
kb interval
genetic background
major gene
short arm chromosome
allele
linkage analysis
marker
accession
primer pair
rice chromosome
est ssr
line
subspecy
progenitor
bac clone
wild rice
candidate gene
synteny
mapping
primer
cloning
hybridization
segment
genome
genomic sequence
salt tolerance
anther
stress condition
oryzae
pathogenicity
gene
plant genome
pcr
evolution
sequence analysis
duplication
family
transgenic plant
intron
gene duplication
motif
homology
full length
expression
tissue
transporter
protein kinase
transient expression
protein
kinase
recombinant protein
binding
cell cycle
peptide
yeast
molecular weight
blight
population structure
heritability
heterosis
trait
hybrid
genetic diversity
spikelet
grain weight
grain quality
panicle
harvest index
seed weight
grain yield
variance
stage
field experiment
higher grain yield
grain yield rice
fertilizer
rainy season
winter season
soil fertility
study effect
manure
drainage
irrigation system
crop residue
flux
emission
policy
carbon dioxide
dry matter
intervention
interview
forage
rice straw
calculation
chromatography
starch
food
fermentation
oil
feed
child
composite
diet
patient
rat
lipid
body weight
cholesterol
red yeast rice
mouse
rice protein
alpha amylase
amylose content
gelatinization temperature
rhizoctonia solani
lesion
lugen
blast disease
brown planthopper
disease severity
point
methanogenesis
chloro
noise
conidia
triticum aestivum
vector
rice leafe
cytochrome
indole
sucrose
peroxidase
lectin
allergen
oligosaccharide
cell line
e coli
digestibility nutrient
cholesterol concentration
WoS publications
(2002-2012)
VOSviewer

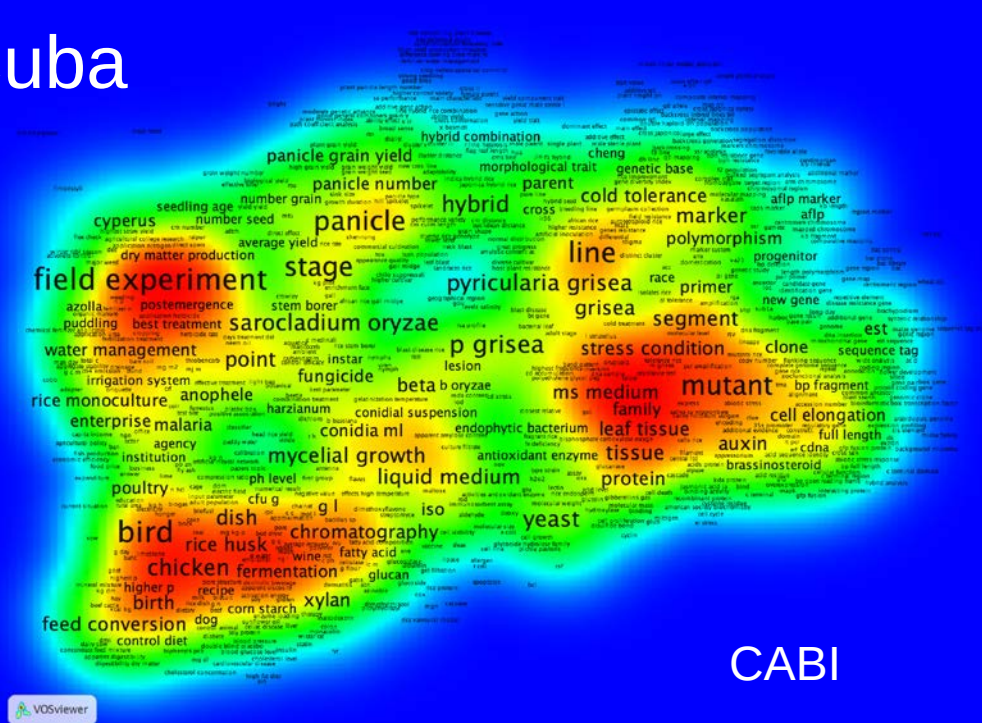
Comparison of document types

	CABI		WoS	
Language	# docs	%	# docs	%
English	**148577**	**71.84%**	**92554**	**94.93%**
Chinese	20544	9.93%	490	0.50%
Japanese	13844	6.69%	2032	2.08%
Portuguese	5356	2.59%	1015	1.04%
French	3942	1.91%	560	0.57%
Spanish	3320	1.61%	307	0.31%
Korean	3018	1.46%	31	0.03%
Russian	2396	1.16%	162	0.17%
Italian	1546	0.75%	22	0.02%
German	1462	0.71%	214	0.22%

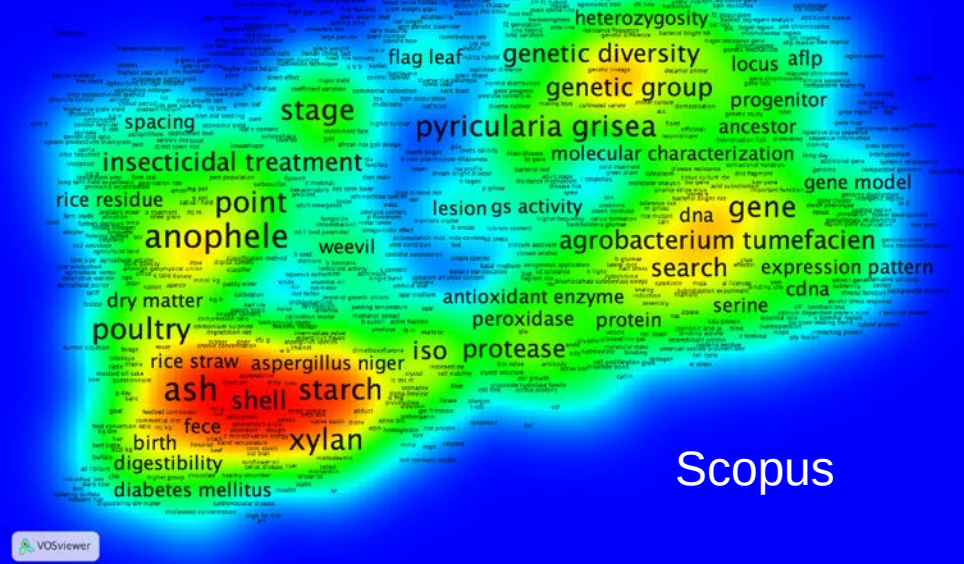
Cuba



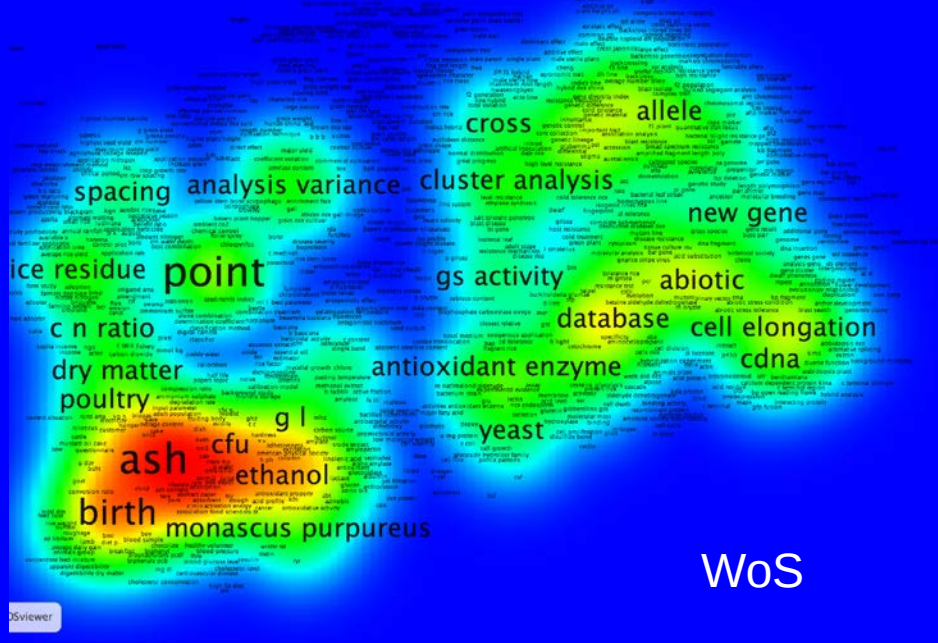
All



CABI



Scopus



WoS