Pitfalls and promises of BIR in science studies: A case study of mapping scientific articles to SDGs

Frédérique Bordignon

Ecole des Ponts & Laboratoire LISIS, Paris, France

margin frederique.bordignon@enpc.fr





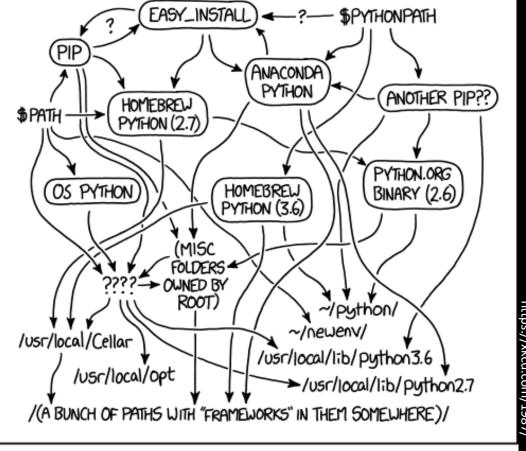








An end-user = not a computer scientist!



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.



Outline

The United Nations Sustainable Development Goals (SDGs)

The hopeless quest for comprehensiveness

The strive for nuance

Discussion with the BIR community



- UN: "SDGs are the blueprint to achieve a better and more sustainable future for all. They address the global challenges we face, including poverty, inequality, climate change, environmental degradation, peace and justice."
- Set in 2015 and to be achieved in 2030 -> research









































- Knowing and showing how to contribute to the SDGs
- Growing interest in mapping scholarly communication to SDGs
 - Dimensions, Scopus/SciVal, WoS/InCites
 - THE impact Ranking









































The hopeless quest for comprehensiveness





Research assessment Bibliometrics

Research on research Scientometrics

Comprehensiveness is required

uired Representativity is acceptable

Mostly self-centred

Contextualised

Precision & recall

Precision

Disparate sources

Corpus delineation

Success of international rankings & ready-to-use solutions

Beware of potential biases



similar corpora/files

Research assessment Bibliometrics

Research on research Scientometrics

Comprehensiveness is required

Representativity is acceptable

Mostly self-centred

Contextualised

Precision & recall

Precision

Disparate sources

Corpus delineation



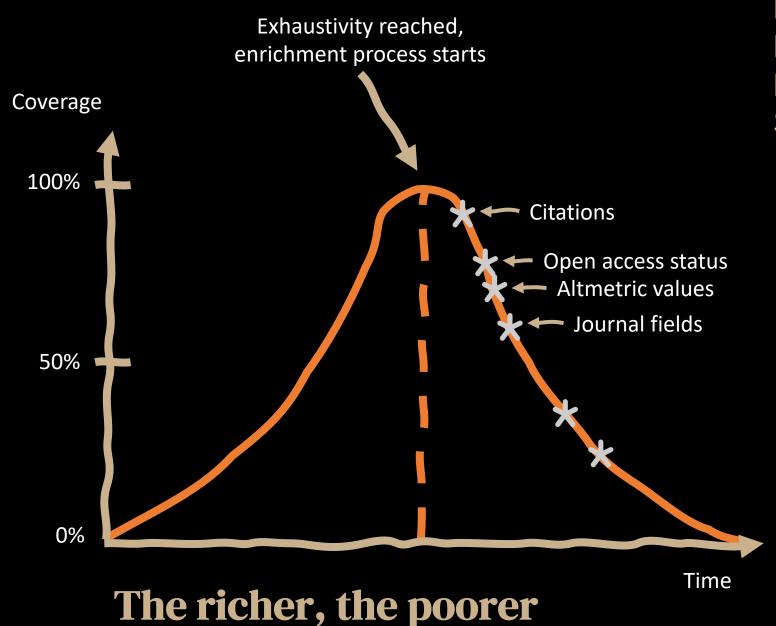
Success of international rankings & ready-to-use solutions

Beware of potential biases

similar corpora/files



Endless possibilities to enhance and enrich data.

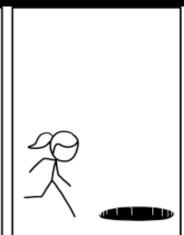


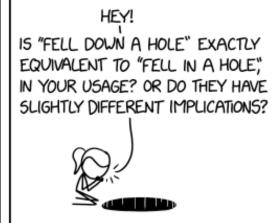
Data(bases) heterogeneity Data availability DOI dependance Skills

The strive for nuance

https://xkcd.com/2390/







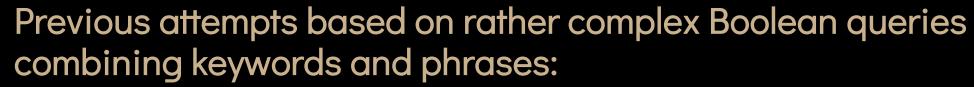
THERE'S A MYTH THAT LINGUISTS ARE PEDANTS WHO LOVE CORRECTING PEOPLE, BUT THEY'RE ACTUALLY JUST ENTHUSIASTIC ABOUT UNDERSTANDING LANGUAGE IN ALL ITS INFINITE VARIETIES, WHICH IS MUCH WORSE.



- Co-word analysis -> clusters...
- Named Entity Recognition -> locations...
- Citations polarity -> communities...

New method to map articles to SDGs

Bordignon, F. 2021. Dataset of search queries to map scientific publications to the UN sustainable development goals. *Data Brief* 34, (2021), 106731–106731. DOI:10.1016/j.dib.2021.106731



too narrow/specific because they stick to the UN text



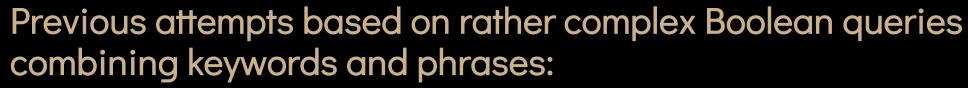
SDG13

AF-ID("University of Leicester" 60033125) AND TITLE-ABS-KEY("Climate action" OR "Climate adaptation" OR "Climate change planning" OR "Climate change policy" OR "Climate early warning" OR "Climate hazards" OR "Climate impact" OR "Climate mitigation" OR "Climate refugees" OR "Climate related hazards" OR "Climate resilience" OR "COP 21" OR "COP 22" OR "Extreme weather" OR "forest fire" OR "drought" OR "cyclone" OR "hurricanes" OR "Global warming" OR "Greenhouse gas" OR "Greenhouse gases" OR "Ice loss" OR "Low-carbon economy" OR "Natural disasters" OR "Ocean warming" OR "Paris Agreement" OR "Rising sea level" OR "Sea level rise" OR "Warming" OR "rising sea" OR "global temperature" OR " carbon emissions" OR "methane emissions" OR "nitrogen oxide emissions") AND (LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014))

climate model/modelling/simulation/projection Intergovernmental Panel on Climate Change/IPCC climate AND "ice-ocean interaction" "permafrost thaw"

Mistry, A., & Sellers, H. (2020). Scopus SDG Search Query codes. Dataset. *University of Leicester data repository*. 10.25392/leicester.data.12839552.v1





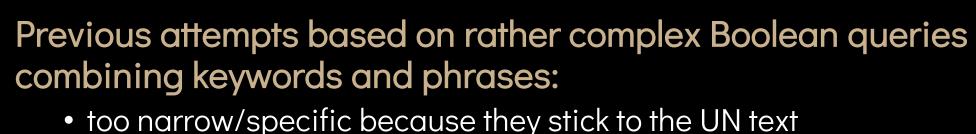
- too narrow/specific because they stick to the UN text
- too broad because of the polysemy of some words



SDG3: *AIDS* (disease) -> *aids* (noun or verb)

Physics-based modeling <u>aids</u> in designing efficient data center power and cooling systems

B Jayabalasingham, R Boverhof, K Agnew, and L Klein. 2019. Identifying research supporting the United Nations Sustainable Development Goals. *Mendeley Data* (2019). doi:10.17632/87TXKW7KHS.1



- too broad because of the polysemy of some words
- too complex, too long, or with a proximity operator not available in tools such as Tableau

Search query for SDG01 Bergen action approach	:h.														
Note that the query is spread over several colu	mns to fac	cilitate reading. A sir	mple version	on of copying	is available in	a separate file	e (02_AllSDG_	queries_simple	eformat)						
Target notes	Query														
	TS= (
1.1/1.2 Eradicate poverty		("	poverty"												
114 112 Endicate poverty		<u> </u>	, ,	NEAR/5	("decreas*")	OR "minimi*"	OR "reduc*"	OR "alleviat*" (OR "tackI*" OR "f	fight*" OR "comb	hat*" OR "end" O	R "ended" OR "e	nding" OR "elimi	nat*" OR "nreven	t*" OR "erad
		1		14271143	(accreas		Oit reduc	on uneride .	on won on	ngine on conn	out on the o	ii ciided oii e	maning on cann	nut on preven	
		- /													
1.3. Social proctections for all, inc. floors. Cove	OB	(I"cocial protect	tion¢" OP '	cocial floors	" OR "cosial cor	nuicoś" OB "w	olfaro custom	¢" OP "wolforo	convica¢"						
1.4. Access to basic services	OIL	(("social protection\$" OR "social floor\$" OR "social service\$" OR "welfare system\$" OR "welfare service\$") NEAR/10 ("implement*" OR "establish*" OR "coverage" OR "covered" OR "covering" OR "improv*" OR "strengthen*" OR "increas*" OR "enhanc*" OR "building")													
1.4. Access to basic services		,	LANTO	Inplement	OK establish	OK COVER	ige On tove	ieu Ok tovei	ilig OK illipiov	OK stiengthe	or increas	OK emilanc	OK build OK b	uliuling /	
	OR	1 0		:¢" OD	llandel flanckii.	OD !!i-!	uisséll on lluu	- 16 C1	" OD "	:¢"					
	UK		("social protection\$" OR "social floor\$" OR "social service\$" OR "welfare system\$" OR "welfare service\$"												
		OR "social welfare" OR "social security" OR "basic service\$" OR "essential service\$"													
)													
		N	IEAR/15						n poor" OR "the vi		vulnerable group\$	" OR "slum" OR	"slums"		
							usehold\$" OR	"people" OR "c	hildren" OR "com	imunit*"))					
					s" OR "women										
			OR (("person\$" OR "people" OR "adult\$") NEAR/3 ("disabled" OR "disabilities" OR "unemployed" OR "older" OR "elderly")) OR "disability"												
			OR (("work" OR "workeplace" OR "worker\$" OR "occupational") NEAR/3 ("injury" OR "injuries" OR "illness*")) OR "worker\$ compensation"												
			OR "bables" OR "infants" OR "newborn\$" OR "children" OR "child" OR "pregnant" OR "pregnancy"												
)											
)													
1.4. Equal rights to economic resources. Own	OR														
		((("income"	OR "wealth"	QR "money" C	OR "economic	" OR "natural	resource\$" OR	"banking")						
				NEAR/10	distribution" OR "redistribution" OR "equalit*" OR "inequalit*"										
				$\overline{}$	OR "right\$" (OR "ownership	p" OR "access	" OR "control" (OR "equitab*" OR	R "inequitab*" OI	R "benefit sharing	")			
)													
		A	ND	("poverty" (OR "the poor" C	R "the poore	st" OR "rural p	oor" OR "urbar	n poor" OR "the vi	ulnerable" OR "v	vulnerable group\$	"			
			AND ("poverty" OR "the poor" OR "the poorest" OR "rural poor" OR "urban poor" OR "the vulnerable" OR "vulnerable group\$" OR (("poor" OR "poorest") NEAR/3 ("household\$" OR "people" OR "children" OR "communit*"))												
			OR "workers" OR "women"												
			OR (("person5" OR "people" OR "adult5") NEAR/3 ("disabled" OR "disabilities" OR "unemployed" OR "older OR "elderly")) OR "disability"												
			OR (("work" OR "workplace" OR "worker5" OR "occupational") NEAR/3 ("injury" OR "injuries" OR "iliness")) OR "worker5 compensation"												
			OR "bables" OR "infants" OR "newborns" OR "children" OR "children" OR "pregnant OR "pregnancy"												
			OR "sustainable development"												
				ok sustain	able developm	ent									
		,)	-										
)													
	OR	('	("microfinanc*" OR "micro-financ*" OR "microinsurance" OR "micro-insurance" OR "wage theft")												
	OR		financial se												
		N	EAR/15	("right\$" OF	t "access*" OR	"availab*"									

C S Armitage, M Lorenz, and S Mikki.

Mapping scholarly publications related to the Sustainable

Development Goals: Do independent bibliometric approaches get the same results?

Quant. Sci. Stud. 1, 3 (2020), 1092—

1108 (2020).

doi:10.1162/qss a 00071



Previous attempts based on rather complex Boolean queries combining keywords and phrases:

- too narrow/specific because they stick to the UN text
- too broad because of the polysemy of some words
- too complex, too long, or with a proximity operator not available in tools such as Tableau

New approach:

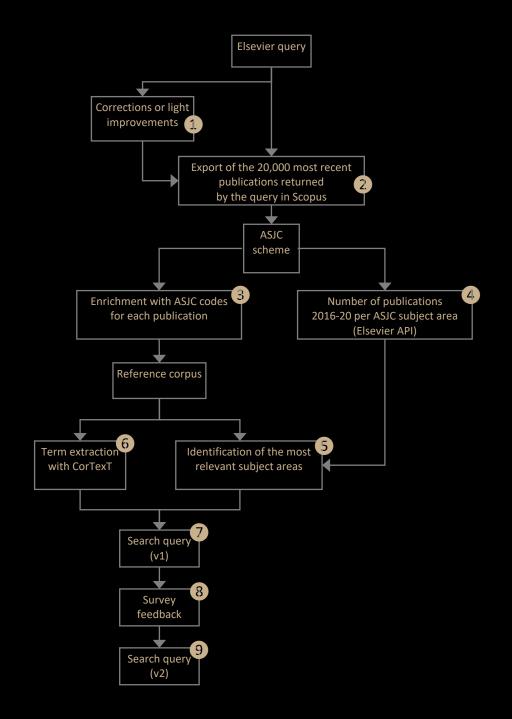
- based on Elsevier queries (Jayabalasingham et al., 2019)
- relying both on terms and journal fields (ASJC)
- to mitigate the polysemy of terms
- that can be used in Tableau and adapted or improved

ASJC scheme



Supergroups	Fields	No. of Subfields
(-)	Multidisciplinary	1
Health Sciences	Medicine	48
Health Sciences	Nursing	23
Health Sciences	Veterinary	4
Health Sciences	Dentistry	6
Health Sciences	Health Professions	16
Life Sciences	Agricultural and Biological Sciences	11
Life Sciences	Biochemistry, Genetics and Molecular Biology	15
Life Sciences	Immunology and Microbiology	6
Life Sciences	Neuroscience	9
Life Sciences	Pharmacology, Toxicology and Pharmaceutics	5
Physical Sciences	Chemical Engineering	8
Physical Sciences	Chemistry	7
Physical Sciences	Computer Science	12
Physical Sciences	Earth and Planetary Sciences	13
Physical Sciences	Energy	5
Physical Sciences	Engineering	16
Physical Sciences	Environmental Science	12
Physical Sciences	Materials Science	8
Physical Sciences	Mathematics	14
Physical Sciences	Physics and Astronomy	10
Social Sciences	Arts and Humanities	13
Social Sciences	Business, Management and Accounting	10
Social Sciences	Decision Sciences	4
Social Sciences	Economics, Econometrics and Finance	3
Social Sciences	Psychology	7
Social Sciences	Social Sciences	22
4 supergroups	26 fields	307 subfields

Bordignon, F. Tracking content updates in Scopus (2011-2018): a quantitative analysis of journals per subject category and subject categories per journal. *17th International Conference On Scientometrics & Informetrics*, ISSI, Sep 2019, Rome, Italy. pp.1630. (hal-02281351)



Altmetric DOAJ

TRA CTS OpenAlex

Crossref

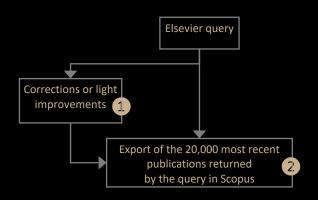
LENS.ORG

Clarivate
Web of Science

tableau

XII Excel

CSV

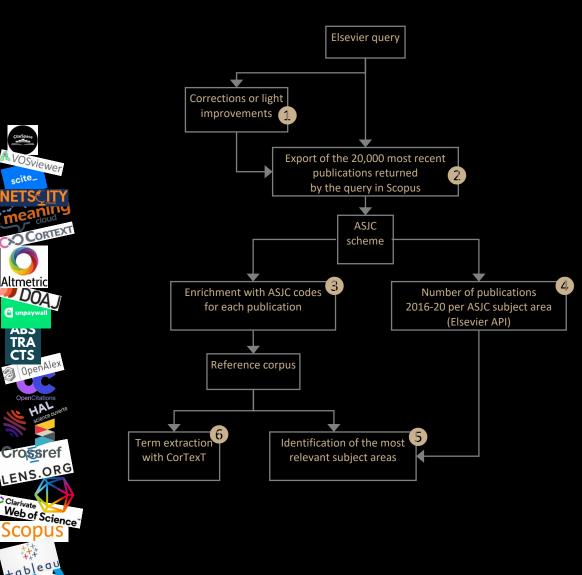


- review and improve each query proposed by the Elsevier team
 - SDG3: AIDS (disease) -> aids (noun)

Also a microscopic examination of fungal culture aids in the identification.

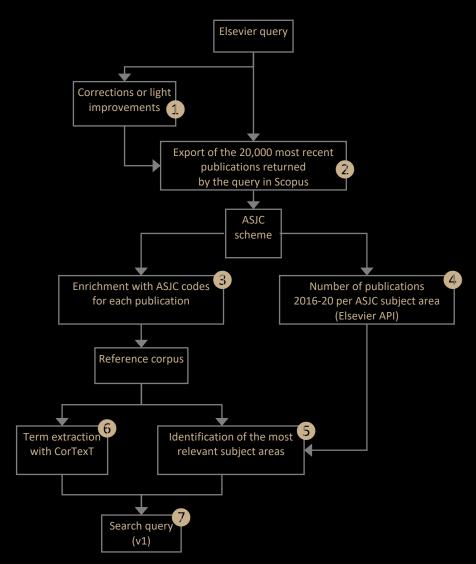


 retrieve the 20,000 most recent publications from Scopus to build a reference corpus



- identify the most relevant subfields (ASJC)
- extract the most relevant terms and phrases (with CorTexT)





(terms expressing poverty or referring to poor people (poor population, the rural poor, urban poor area...)

AND (Economics, Econometrics and Finance OR 8 selected subfields of Social sciences))

OR

terms expressing poverty or referring to poor people (poor population, the rural poor, urban poor area...)

AND specific phrases (food insecurity, income inequality, labor market...)

OR

unambiguous phrases (*reduction of the poverty*, *financial empowerment...*)



IF(

(([Code ASJC]="3300" OR [Code ASJC]="3301" OR [Code ASJC]="3302" OR [Code ASJC]="3303" OR [Code ASJC]="3304" OR [Code ASJC]="3305" OR [Code ASJC]="3306" OR [Code ASJC]="3307" OR [Code ASJC]="3307" OR [Code ASJC]="3311" OR [Code ASJC]="3311" OR [Code ASJC]="3311" OR [Code ASJC]="3315" OR [Code ASJC]="3315" OR [Code ASJC]="3315" OR [Code ASJC]="3316" OR [Code ASJC]="3317" OR [Code ASJC]="3317" OR [Code ASJC]="3319" OR [Code ASJC]="3319" OR [Code ASJC]="3320" O

AND

(CONTAINS([TI-ABS-KW], "poor population") OR CONTAINS([TI-ABS-KW], "poor households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poor communities") OR CONTAINS([TI-ABS-KW], "poorer communities") OR CONTAINS([TI-ABS-KW], "poor countries") OR CONTAINS([TI-ABS-KW], "poor countries") OR CONTAINS([TI-ABS-KW], "poor country") OR CONTAINS([TI-ABS-KW], "poor countries") OR CONTAINS([TI-ABS-KW], "poor nations") OR contains([TI-ABS-KW], "poor nations")

OR

METSCITY meaning COCORTEXT

Altmetric

TRA CTS

Crossref LENS.ORG

Web of Science

((CONTAINS([TI-ABS-KW], "social protection") OR CONTAINS([TI-ABS-KW], "food security") OR CONTAINS([TI-ABS-KW], "cash transfers") OR CONTAINS([TI-ABS-KW], "food insecurity") OR CONTAINS([TI-ABS-KW], "cash transfers") OR CONTAINS([TI-ABS-KW], "food insecurity") OR CONTAINS([TI-ABS-KW], "development assistance") OR CONTAINS([TI-ABS-KW], "income inequality") OR CONTAINS([TI-ABS-KW], "labour market") OR CONTAINS([TI-ABS-KW], "health insurance") OR CONTAINS([TI-ABS-KW], "protection system") OR CONTAINS([TI-ABS-KW], "financial development") OR CONTAINS([TI-ABS-KW], "north-south divide") OR CONTAINS([TI-ABS-KW], "financial inclusion") OR CONTAINS([TI-ABS-KW], "women empowerment"))

AND

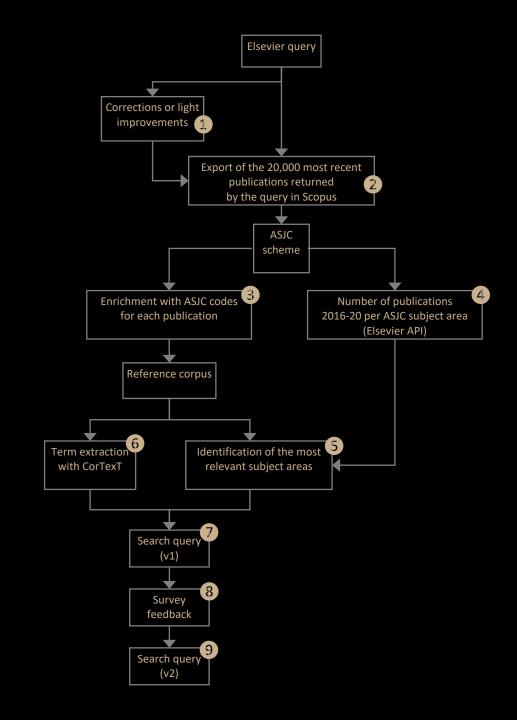
(CONTAINS([TI-ABS-KW], "poor population") OR CONTAINS([TI-ABS-KW], "poor households") OR CONTAINS([TI-ABS-KW], "poorer households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorest households") OR CONTAINS([TI-ABS-KW], "poorer communities") OR CONTAINS([TI-ABS-KW], "poor communities") OR CONTAINS([TI-ABS-KW], "poorer countries") OR CONTAINS([TI-ABS-KW], "poor countries") OR CONTAINS([TI-ABS-KW], "poor nations") OR CONTAINS([TI-ABS-KW], "poor nations") OR CONTAINS([TI-ABS-KW], "poorest nations") OR CONTAINS([TI-ABS-KW], "poor people") OR CONTAINS([TI-ABS-KW], "poorest people") OR CONTAINS([TI-ABS-KW], "poor people") OR CONTAINS([TI-ABS-KW], "poorest people") OR CONTAINS([TI-ABS-KW], "poor area") OR CONTAINS([TI-ABS-KW], "poor women") OR CONTAINS([TI-ABS-KW], "poor area") OR CONTAINS([TI-ABS-KW], "poor area")

OR

(CONTAINS([TI-ABS-KW], "no poverty") OR CONTAINS([TI-ABS-KW], "financial literacy") OR CONTAINS([TI-ABS-KW], "poverty") OR CONTAINS([TI-ABS-KW], "reduction of both poverty") OR CONTAINS([TI-ABS-KW], "reduction of the poverty") OR CONTAINS([TI-ABS-KW], "reduction of both poverty") OR CONTAINS([TI-ABS-KW], "reduction of the poverty") OR CONTAINS([TI-ABS-KW], "poverty reduction") OR CONTAINS([TI-ABS-KW], "poverty") OR CONTAINS([TI-ABS-KW], "distributional effect") OR CONTAINS([TI-ABS-KW], "distributional effects") OR CONTAINS([TI-ABS-KW], "distributional effects") OR CONTAINS([TI-ABS-KW], "distributional effects") OR CONTAINS([TI-ABS-KW], "aid the development") OR CONTAINS([TI-ABS-KW], "distributional effects") OR CONTAINS([TI-ABS-KW], "aid in the development") OR CONTAINS([TI-ABS-KW], "social assistance") OR CONTAINS([TI-ABS-KW], "system of social protection") OR CONTAINS([TI-ABS-KW], "social assistance") OR CONTAINS([TI-ABS-KW], "system of social protection") OR CONTAINS([TI-ABS-KW], "social assistance") OR CONTAINS([TI-ABS-KW], "system of social protection") OR CONTAINS([TI-ABS-KW], "flood banks"))

THEN "SDG1" ELSE "0" END





ADS

TRA

CTS

.

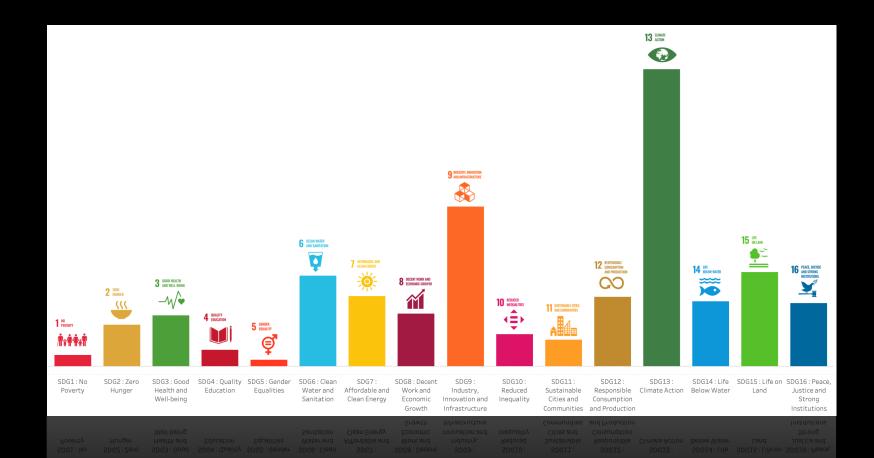
Do you think the following publication can be considered as a contribution to achieving the goals suggested below?

Paris, a Contested Construction of Metropolitan Space

Paris, a Contested Construction of Metropolitan Space This chapter tells the story of the difficult and contested efforts to construct Paris as a metropolitan space. Covering the period of 2000 to the present, it focuses on the search for a collective actor and the production of collective action through the quest for alliances between players, be they the central State, local governments or business associations. More specifically, the chapter concentrates on the relationships between core metropolitan stakeholders through the analysis of four political initiatives aiming at building collective action and territorial leadership at the metropolitan scale. In the Île-de-France region, many competing conceptions of metropolitan space are co-present and are conflicting. First the regional council, whatever its political leaning, has always considered the Metropolis to be the Île-de-France region. In that perspective the construction of metropolitan space is understood to mean the strengthening of the powers and resources of the regional authority. Second, the city of Paris considers the core area to be the most relevant space to address the important problems of the metropolis. Third, the State maintains an ambiguous position as regards its conception of the metropolitan space, identifying it as the city region in the period 2007–2012 when the State was controlled by the conservative parties, but restricting it to the core area in the period 2012–2017 by the new Socialist government. In this context, the most important business associations have been ready to accommodate any initiative, provided that the treatment of the most important issues of economic competitiveness, transport and housing, was taken care of. This story clearly indicates the importance of politics in the construction of metropolitan space and shows the critical role that politics plays in shaping scalar conflicts. In this context, decentralization and globalization can be understood to be both structuring forces and elements which are instrumentalized. First, there is the issue of territorial leadership which has consistently led to a stalemate, because of a governance system whose main feature, unregulated competitive decentralization, prevents any leader from being accepted as legitimate at a regional scale. Second, the building of a coalition at the metropolitan level, to ensure capacity to act at that scale, has proved impossible, first, because of a relative balance of powers between local governments and, second, because of the incapacity of the State to successfully act as the leader of a coalition because of its political instability and its lack of resources. Third, players also conflict about the vision they have of the future of the metropolitan area. Two clearly opposed visions coexist, one in which priorities are to fight against social and territorial inequalities and to initiate an alternative mode of development based on ecological transition, another one in which economic competitiveness is the top priority to which social, territorial and environmental issues must be subordinated. Since these two visions are supported by alliances of players which are roughly equal in strength, the not surprising result is conflict and the blockage of the governance system.

		OK, relevant	Irrelevant	I don't know
10 NEGACIES	REDUCED INEQUALITIES - Reduce inequality within and among countries			
11 SECONMENT OFFE	SUSTAINABLE CITIES AND COMMUNITIES - Make cities and human settlements inclusive, safe, resilient and sustainable			
16 PLACE STRING NOTIFICATION NO	PEACE, JUSTICE AND STRONG INSTITUTIONS - Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels			





Publications identified through this process cannot be considered as evidence of the commitment of authors and their institutions to actions towards the targets established by the UN. However, they can be an accurate indicator of which research is relevant to the issues addressed by the SDGs, whether or not it is a direct contribution.



Discussion



Concluding remarks

- Diversity of sources and formats, dependence on DOIs, fluctuating coverage of the tools -> degrading the data
- Successful combination of bibliometric indicators and text analysis to provide innovative studies

Concluding remarks



- Diversity of sources and formats, dependence on DOIs, fluctuating coverage of the tools -> degrading the data
- Successful combination of bibliometric indicators and text analysis to provide innovative studies -> with caution

Concluding questions

- Many tools, too many?
- What does this mean?
- What are the consequences?

Thank you!

Frédérique Bordignon Ecole des Ponts & Laboratoire LISIS, Paris, France





