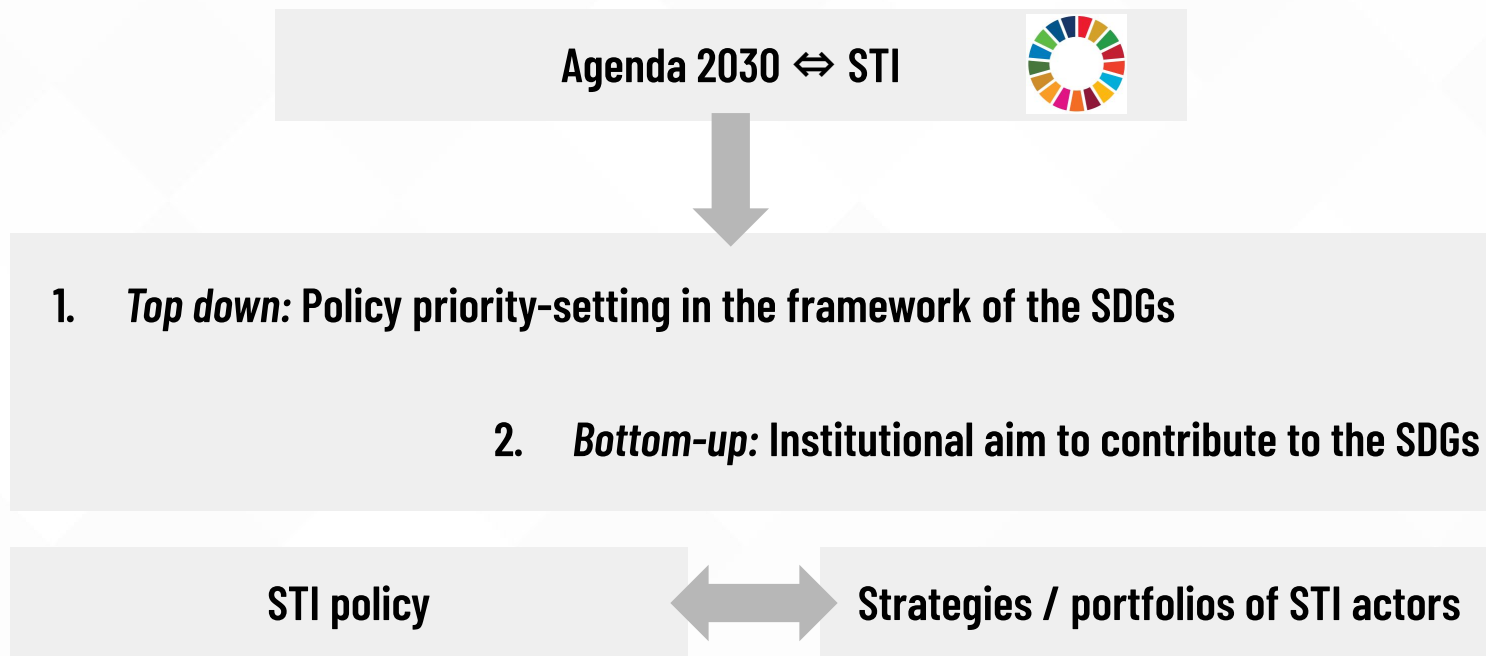


Mapping Research related to the SDGs - Workshop

Francesco A. Massucci,
SIRIS Academic

04 September 2020





A pressing need to map and analyse new Science, Technology and Innovation (STI) domains of interest

Governments and Agencies



Who are the actors researching on SDG-related themes in my territory and on what topics are they working on?
Is the SDG-related portfolio in my territory aligned with the societal / environmental needs?

Higher Education and Research Organisations



What is my research potential on SDGs? Who are the main SDG-related researchers in my institution?

Philanthropic Foundations



What is my research portfolio? Are there underfunded SDG-related topics?

Identifying and mobilising SDG-oriented STI actors

Classification of activities & actors
Collaboration networks
Emerging trends

Comparing SDG-oriented STI portfolios

Classification of activities & actors
Collaboration networks
Emerging trends
Specialisation / Complementarity
Benchmarking

Aligning STI policies and ecosystems with the SDGs

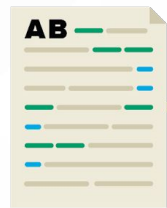
Classification of activities & actors
Collaboration networks
Emerging trends
Specialisation / Complementarity
Benchmarking
Policy priorities within the SDGs
SDG indicator statistical assessment

Semantic mapping techniques vs. taxonomies

- It is based on the **actual language and textual content** used and written by domain **stakeholders**
- Adaptable to the **current questions and to emerging domains**
 - Emerging domains may be poorly formalised (eg Industry 4.0)
 - Taxonomies do not incorporate emerging domains, move slowly
- **Each record is analysed individually** (no statistical aggregates)
 - Avoids misallocations and thematic simplifications
 - Facilitates the identification of niches and transversality

Building a controlled vocabulary for SDGs

STEP 1.
Manual extraction
of key terms



**INITIAL SET
OF TERMS**

STEP 2.

Download a reference corpus
of the texts indexed by these
seed keywords and train a
Word2Vec model



STEP 3.

Enriching the initial
set of terms

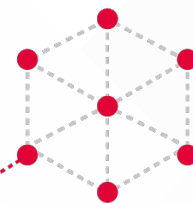


3a

Use the Word2Vec model to
extract co-occurrent keywords
“semantically similar” to the
seed keywords

3b

Use the Word2Vec model to
extract terms with a similar word
context - emerging synonyms



*The reference textual corpus is used
to train a Word2Vec word embedding
model (a machine learning model
based on neural networks)*

3c

Use ontologies to extract
homonyms, hypernyms and
other terms linked with the
initial eyword set



WikiData, DBpedia, Yago...

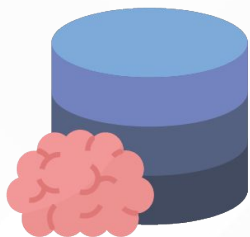
STEP 4.

Compilation and revision
of the final list of terms

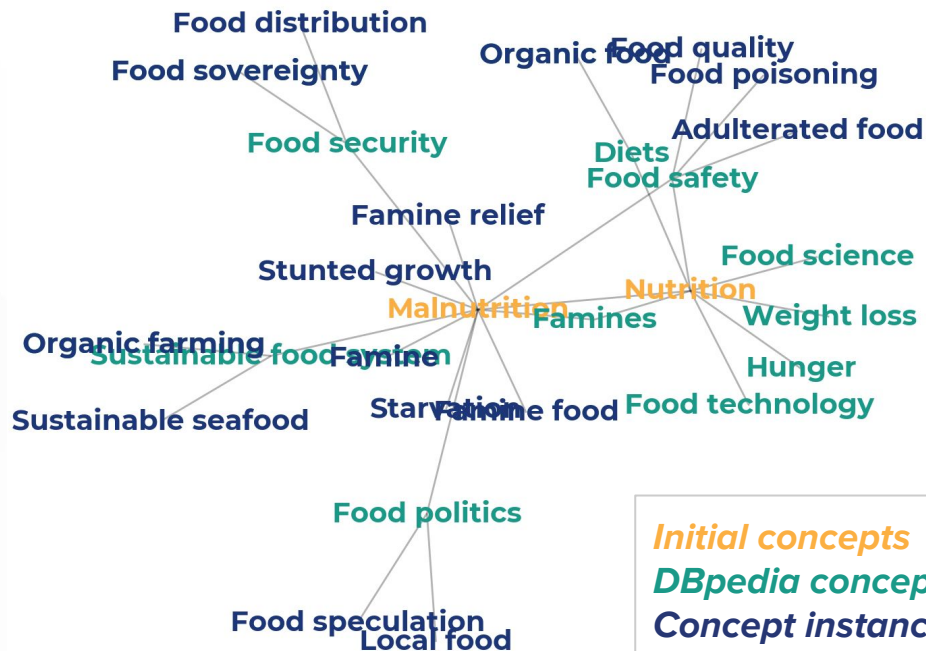


3c

Use ontologies to extract homonyms, hypernyms and other terms linked with the initial eyword set



WikiData, DBpedia, Yago...



Vocabulary snapshot

SDG	N. Terms	Terms with contextualisation
SDG 1	102	40
SDG 2	121	17
SDG 3	628	151
SDG 4	114	59
SDG 5	133	33
SDG 6	291	133
SDG 7	206	22
SDG 8	146	58
SDG 9	268	99
SDG 10	110	65
SDG 11	275	95
SDG 12	153	44
SDG 13	301	20
SDG 14	163	49
SDG 15	186	31
SDG 16	297	33
Grand Total	3494	949

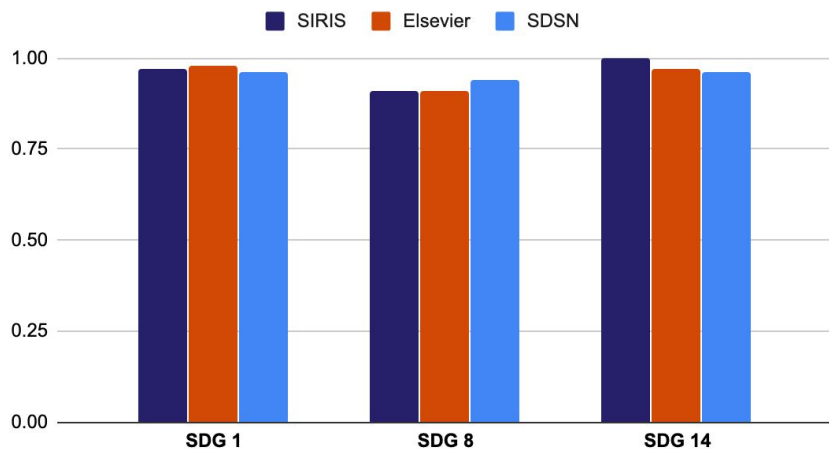
SDG 12 - Responsible consumption and production

keyword	extra
biodegradable bag	
biotic material	sustainable affordable reliable
biowaste	
circular economy	
closing the loop	Sustainable ecological equitative
compost	Sustainable ecological equitative
compost pile	

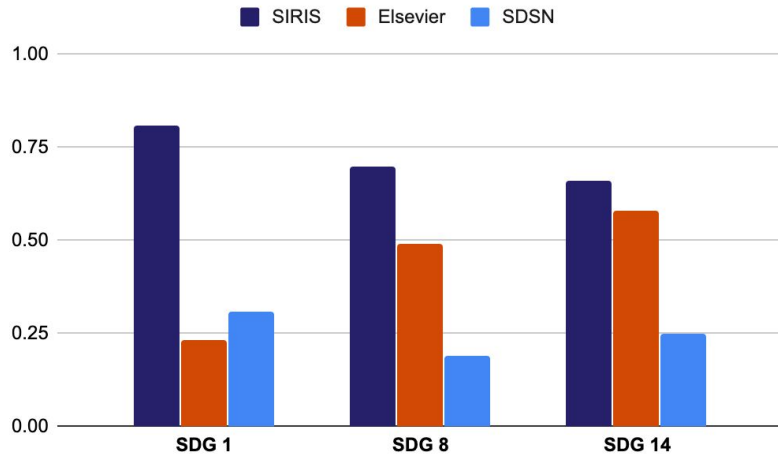
...

On a textual corpus manually annotated by **3 different persons**.
Texts are linked to the SDG if there is **consensus**.

Precision



Recall



The challenge is to be extensive enough, yet precise enough!

- This is **a** possible view of the SDGs: needs and challenges **may differ geographically**

The methodology allows to efficiently build extensive vocabularies, so it could be applied **to extend/change the vocabulary** for specific needs

- Which textual datasets can fairly cover STI contributions to SDGs?

It is essential to look at **different sources** to effectively capture different STI efforts towards SDGs

- All in all, in most of the cases the **linkage text-SDG is subjective**

This should be seen as a means and not as a result - i.e. the vocabulary is a tool to elicit discussions and analyses, not

- **Issues** with transversal **Goal 17** and “meta-STI” **Goals 8 and 9**

- Goal 17 not included in the vocabulary

- Goals 8 and 9 not applied (yet) to project descriptions due to false positives. Good results with scientific publications.

- **Limitations** of the controlled vocabulary in **highly technical texts** (e.g. patents) and their applications in **non-technological goals**

- More reliable for goals 2, 3, 6, 7, 11, 12, 13

- Combination with other semantic techniques (seed corpus)

Applications

1. **Pilot methodology for mapping Sustainable Development Goals in the context of Smart Specialisation Strategies**



European Commission

STI Roadmaps for SDGs

Pilot activity in Serbia

Joint Research Centre

European Commission | SUSTAINABLE DEVELOPMENT GOALS

The Joint Research Centre (JRC) supports the Sustainable Development Goals

Context and more information: <https://s3platform.jrc.ec.europa.eu/pilot-methodology>

Aligning STI policies and ecosystems with the SDGs

Policy priorities within the SDGs

SDG indicator statistical assessment

Classification of activities & actors

Collaboration networks

Emerging trends

Specialisation / Complementarity

Benchmarking

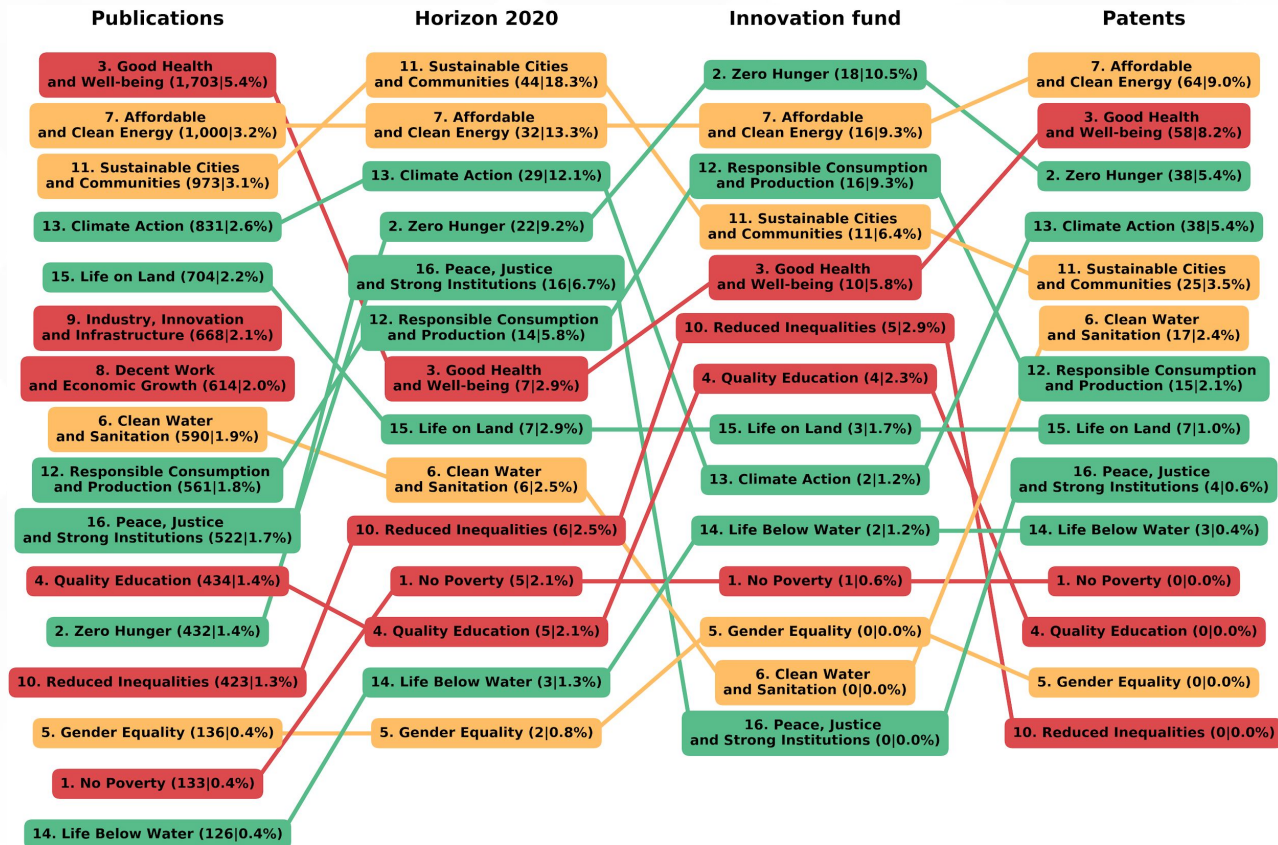
Higher priority challenges in Serbia



Goals with higher potential in the context of smart specialisation

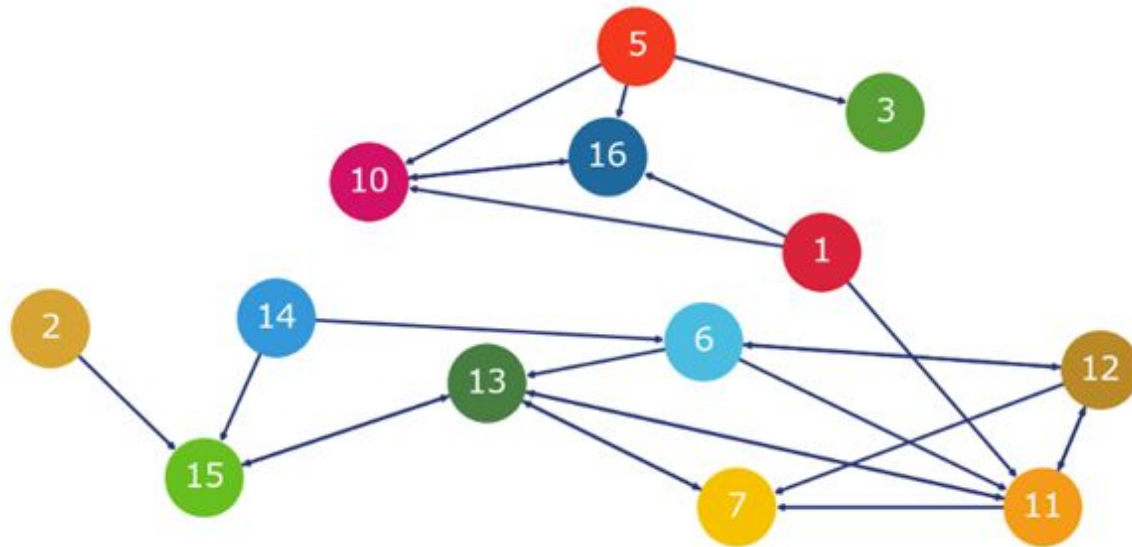


SDG-related STI activities in Serbia (by priority level)



Relationship between the goals

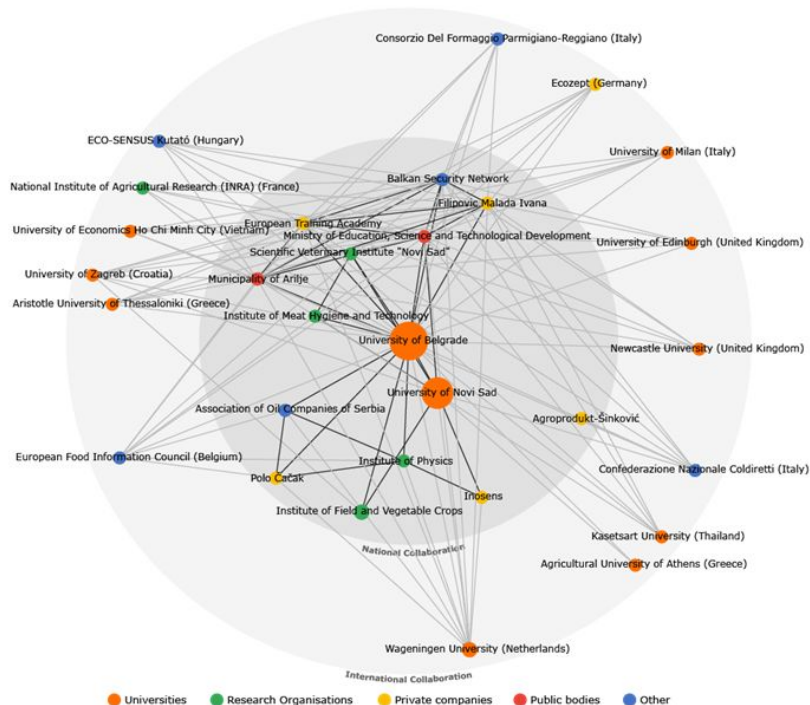
Goal similarity network
Goals which share more than 14% of STI records in common



2



Goal 2-related STI collaboration network presenting the TOP 15 Serbian institutions and TOP 15 international partners



TOP 10 COMPANIES PARTICIPATING IN R&I PROJECTS

Inosens doo novi sad

Biounik

Prometno drustvo sa ogranicenom
odgovornoscu polo cacak

Poslovni sistem global seed doo curug

Institute for forage crops krusevac ltd

Filipovic malada ivana

Dnet labs doo novi sad

Agroprodukt-sinkovic doo

White lemur

Urbigo

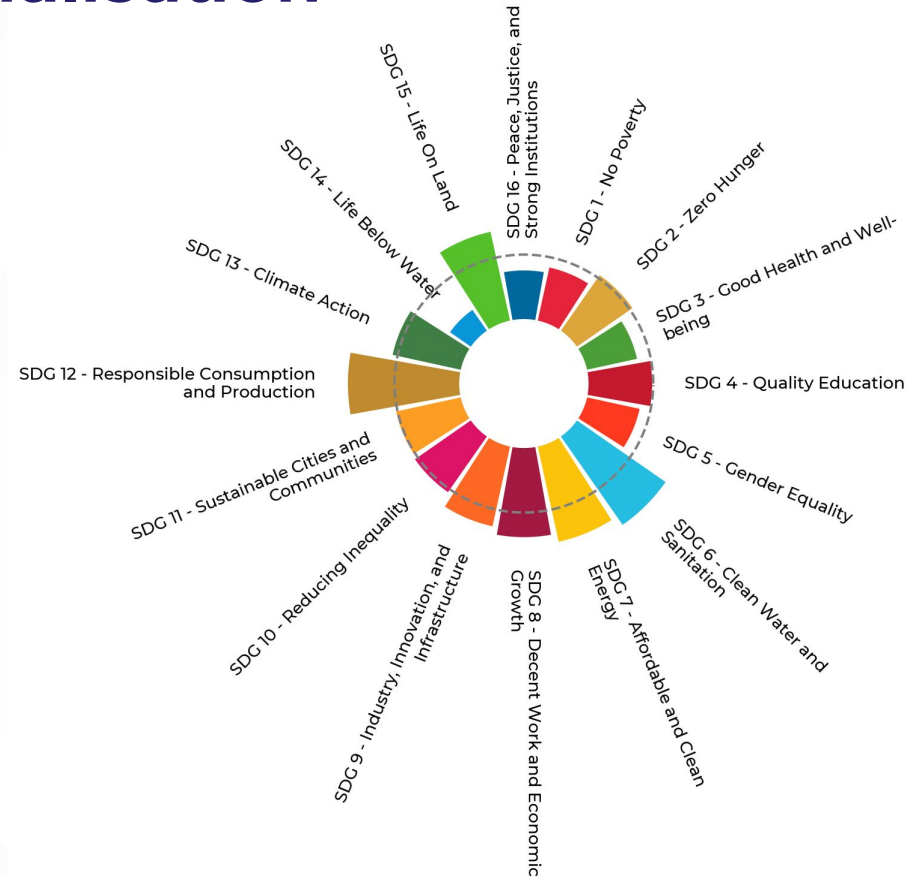
Applications

2. Identifying niches of specialisation for committed universities

Research portfolio specialisation

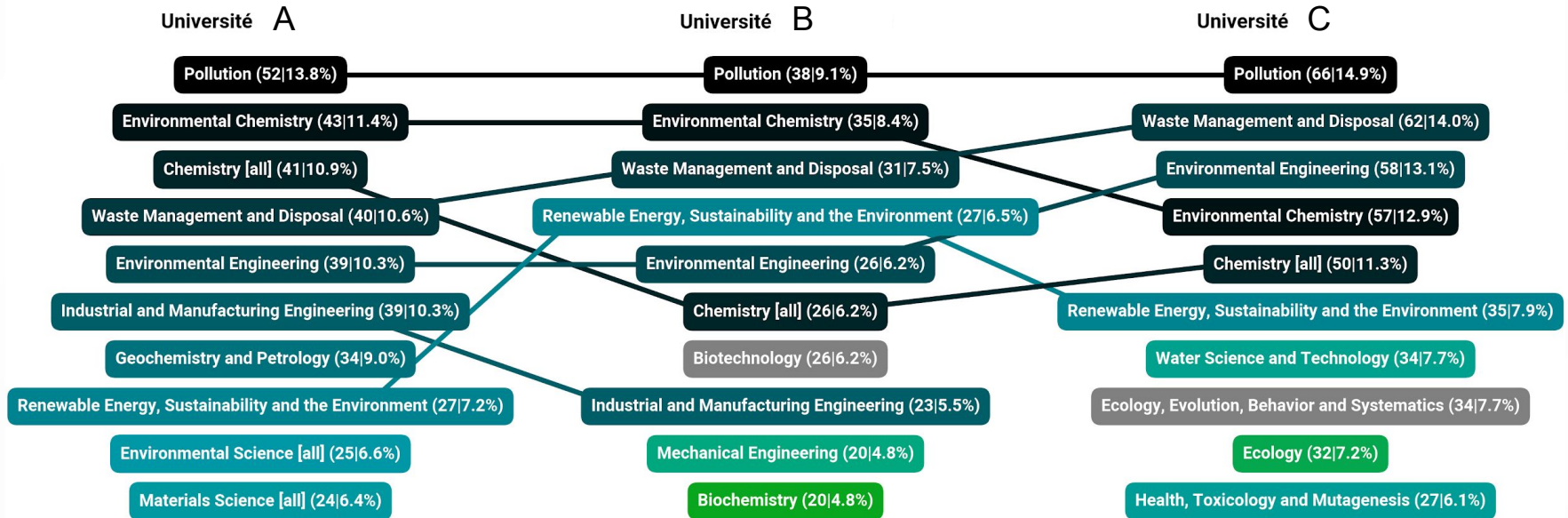
Specialisation is computed over classified publications, with respect to the French Higher Education Institutions' production.

	Specialization Index*
SDG 1	0.84
SDG 2	1.05
SDG 3	0.79
SDG 4	0.98
SDG 5	0.83
SDG 6	1.67
SDG 7	1.51
SDG 8	1.38
SDG 9	1.27
SDG 10	1.05
SDG 11	1.02
SDG 12	1.72
SDG 13	1.08
SDG 14	0.39
SDG 15	1.41
SDG 16	0.75



Focus on SDG 12

Responsible Consumption and production



Wrap-up and next steps

A tool-kit to identify, map and characterise SDG-related texts

Still early for conclusions...

Challenges / lessons:

Localisation → adapt and extend the vocabulary, translate texts, include local data sources

Social / non-technological innovation → Requires a special focus (harder to find)

Interaction with stakeholders → use the process to build a common understanding of a topic / challenge, and to strengthen mutual knowledge

Upscaling

- **The SDG vocabulary is open**, ready to be used anywhere: [A controlled vocabulary defining the semantic perimeter of Sustainable Development Goals](#)
- **Can be used** for any institution or perimeter (municipality, region, country...)
- Useful **methodologies for priority-setting and stakeholder mobilisation** towards [responsible research and innovation](#), [smart specialisation](#), [STI roadmaps for the SDGs](#), etc.

