

Ministrstvo za kohezijo in regionalni razvoj, Sektor za pametno specializacijo

1. odprti dan za gospodarstvo (SLAIF), 17. 11. 2025

Semantična analiza RRI dosežkov strategije S5

Matej Forte
Sektor za pametno specializacijo

Kdo smo?

Slovenska strategija trajnostne pametne specializacije (S5)



Kaj je S5?

- Nacionalna strategije pametne specializacije
- Tematski omogočitveni pogoj za Cilj politike 1 – Konkurenčnejša in pametnejša Evropa
- Podpira dvojni prehod (zeleni, digitalni)



10 prednostnih področij

Pavmetna mesta in skupnosti, Pametne zgradbe in dom z lesno verigo, Mreže za prehod v krožno gospodarstvo, Trajnostna pridelava hrane, Mobilnost, Tovarne prihodnosti, Zdravje – medicina, Informacijske in komunikacijske tehnologije (IKT), GoDigital, Materiali kot končni produkti, Trajnostni turizem



Proces podjetniškega odkrivanja

Stalni dialog med državo, akademsko sfero, podjetji in regijami o tem, kje so dejanske prebojne priložnosti.

Hipoteza

Prednostna področja Strategije S5 so bila pravilno/nepravilno zastavljena, na podlagi pomanjkljivih informacij.

Napoved „state of art“ tehnologij je bila pravilno/nepravilno zastavljena.

~107.000
dokumentov



Cilj

Izvesti **semantično analizo** dejavnosti in rezultatov raziskav, razvoja in inovacij (**RDI**) iz izbranih mednarodnih podatkovnih baz **z uporabo orodij** umetne inteligence.

Metoda: klasifikatorji, kategorizacija, modeliranje tem, hibridni model (2. faza)

Type of document	Data Source	Number of documents in Slovenia	Number of distinct actors in Slovenia
Scientific publications (only those with a DOI)	OpenAlex	~84,000	~7,400 disambiguated actors across the data sources
Patents (applications, to any patents office)	Patstat	~8,200 application ids, ~4,100 patent families	
Horizon projects	Cordis	~1,600	
ERDF projects	Kohesio	~8,400	
Interreg projects	Interreg	~930	

Kako učinkovito in verodostojno presejati veliko količin besedil iz različnih baz podatkov



Dostopnost podatkov

Prosto dostopne baze podatkov vs. plačljive baze podatkov

Primerljivost baz



Zanesljivost podatkov

Kakovost in struktura

Zmogljivost modelov (kalibracija)

Prekrivanje tem in horizontalnost tem
(višja zahtevnost klasifikacije)



Obremenitev kadra

Delo ekspertov v S5 ~100 ur na eksperta za vzpostavitev klasifikatorjev

0,8 FTE za obdobje 6 mesecev (vodja projekta)

Metodologija

Priprava podatkov: obsežen korpus dokumentov RRI iz več mednarodnih virov podatkov, iz odprtih podatkov, vključno z znanstvenimi publikacijami (OpenAlex), patenti (Patstat) in projekti (Horizon/CORDIS, Interreg/Keep.eu in ERDF/Kohesio). Ti podatki so bili standardizirani in integrirani v strukturirano bazo podatkov za analizo.

Razvoj taksonomije

začrtanje večnivojskega klasifikacijskega okvira
binarni klasifikatorji + tehnike modeliranja tem →
prepoznavanje tematskih skupin znotraj in med
prednostnimi področji

Postavitev semantičnega modela: LLM + fine-tuning
(hibridni pristop)

Klasifikacija vseh dokumentov: 50 ključnih tem na celotnem
področju pametne specializacije in 10 do 20 na
posameznem PP

Identifikacija in klasifikacija organizacij: Identificirali, razjasnili
in klasificirali smo organizacije, vključene v slovenski
ekosistem RDI, glede na njihov tip (akademski/raziskovalni,
industrija, javni sektor) in članstvo v SRIP (strateškem
partnerstvu za raziskave in inovacije).

AI-powered classification - Methodology overview

Challenge

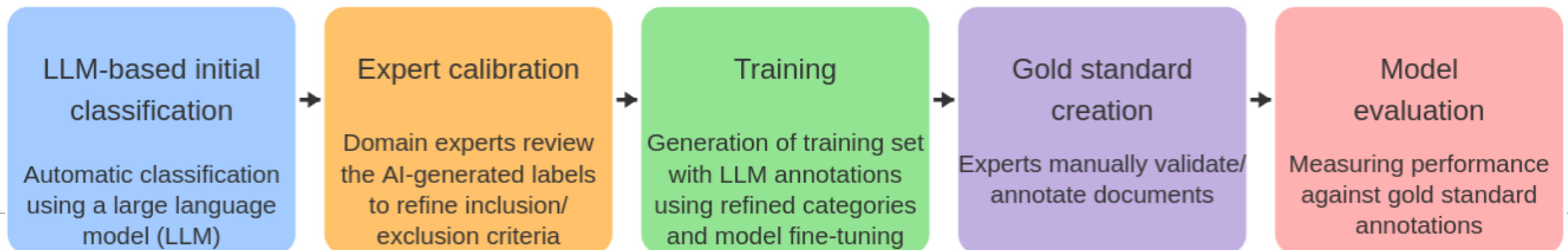
- ❖ Classifying 100,000+ documents into 10 Slovenian S5 priority areas require innovative AI-human expert collaboration.
- ❖ Large amount of training (pre-classified) data is needed for models to learn patterns in the data and make predictions.

Innovative approach

- ❖ Combining AI capabilities with **expert annotations** ensure **accurate and scalable classification** across diverse RDI documents.

Key contribution by six **thematic experts from the Cohesion Ministry's S3 Unit** for Calibration and Evaluation in January and February

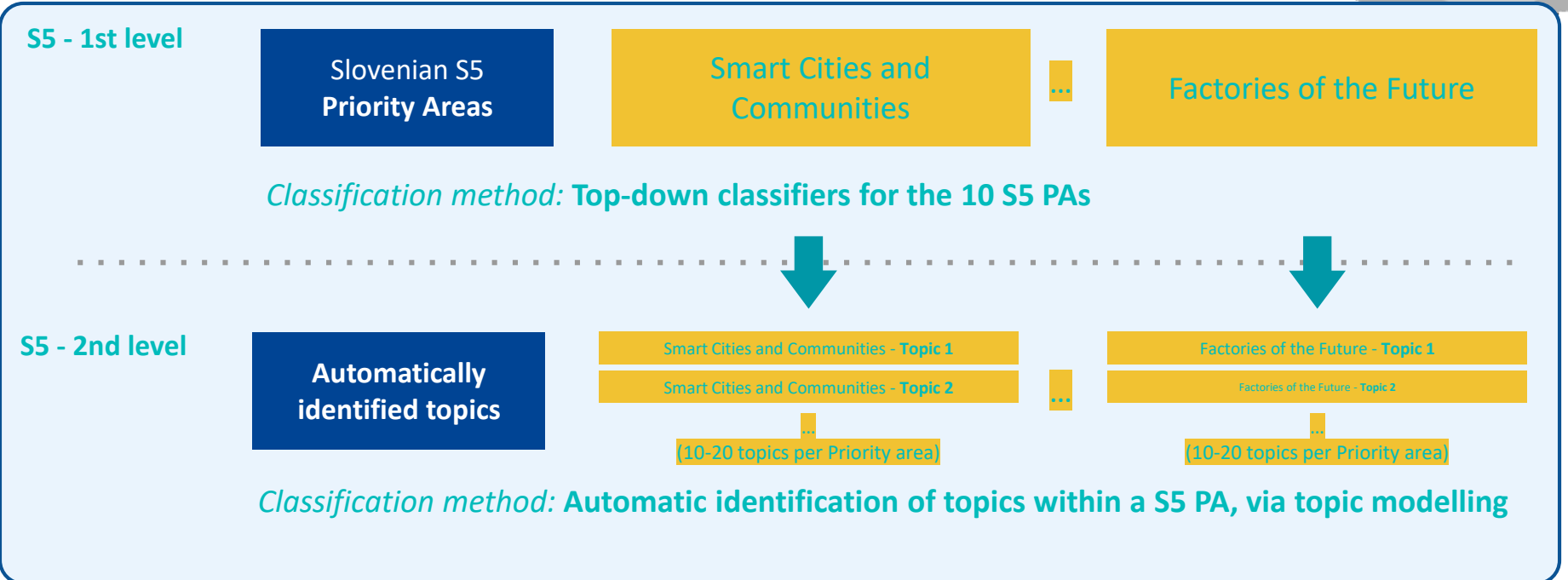
Methodology breakdown



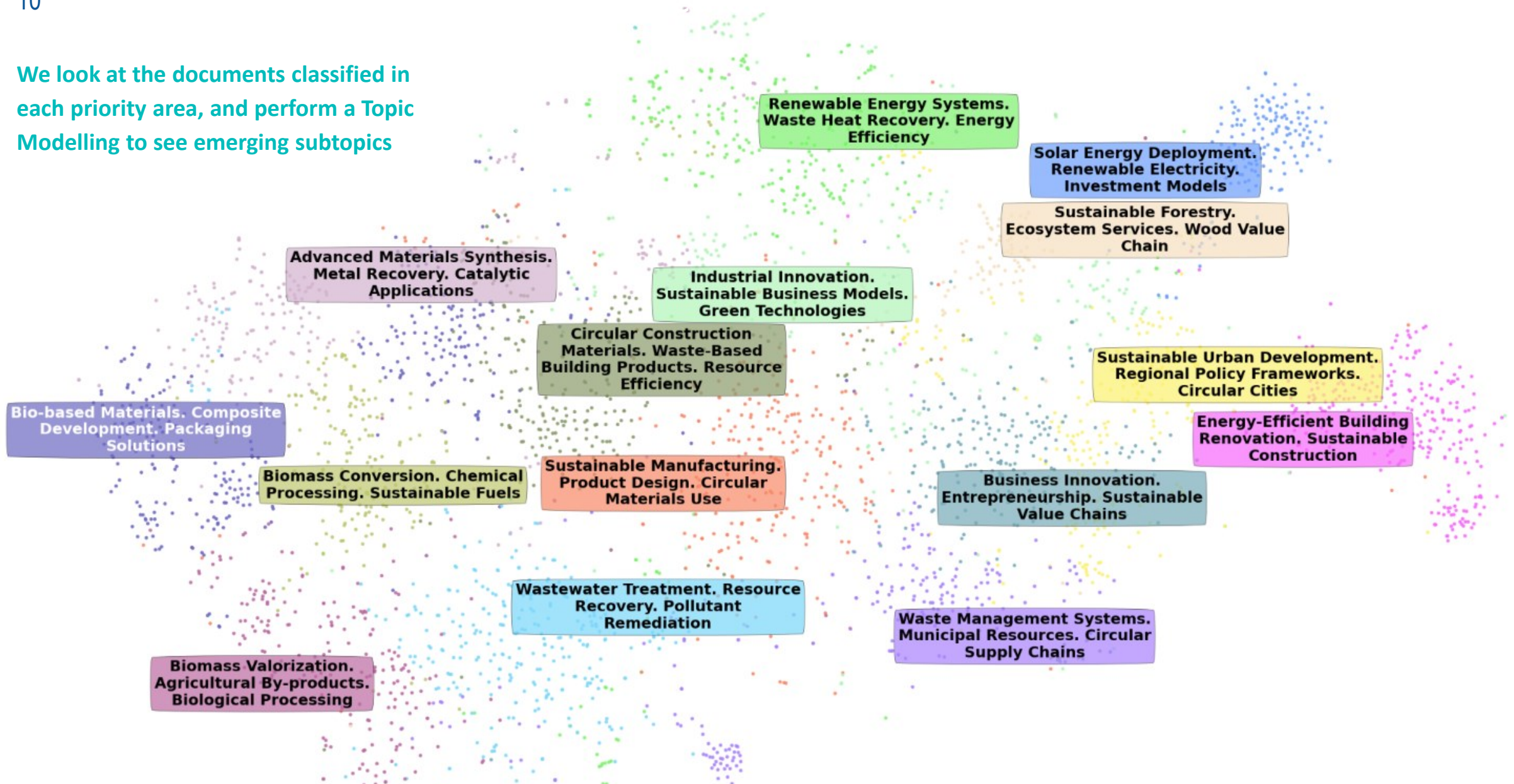
Target analytical taxonomy

Smart Cities and Communities	Information and Communications Technology (ICT) - Horizontal Network	Smart Buildings and
Health	Digital transformation	Smart, almost passive
Smart devices, sensors and tele-medicine	New business models and promotion of entrepreneurship related to digital transformation, including user experience	Modular and mobile devices
Monitoring the functional health parameters and the quality of living in smart living environments	Green & Digital - Digital transformation related to digital and green transition	Fast construction systems
Personalised long-term care for patients, the elderly and other target groups	Digital transformation of the economy and public administration	Wooden constructions
Smart integrated healthcare and patient care system	IoT (Internet of Things, embedded systems and sensors)	Massive constructions
Establishing a smart integrated healthcare and patient care system	Mobile, wireless and edge infrastructure and communications for IoT (including 5G and 6G)	Interior and architectural
Energy and utilities	IoT platforms and services	Construction
Energy transformation, distribution and management	Embedded sensor systems for IoT	Multifunctional element envelope
Comprehensive support for water services	IoT (Internet of Services, platforms)	Joinery and wooden cladding
Mobility, transport and logistics	Technologies for the development and establishment of a comprehensive service platform, i.e. the third platform	(Load-bearing) construction systems
Carbon neutral society	Open urban data platforms	Special, covering and internal systems
Using aggregate mobility data for better understanding of the migration dynamics within a municipality, as well as between municipalities	Innovative IoT services related to the Blockchain technology	Forest, timber and wood
Smart city traffic planning	Cyber security	Interior elements
Multimodal mobility platform	Development of security products and services	Light and lighting solutions

“the baseline table”
(S5 Priloga Tabela)

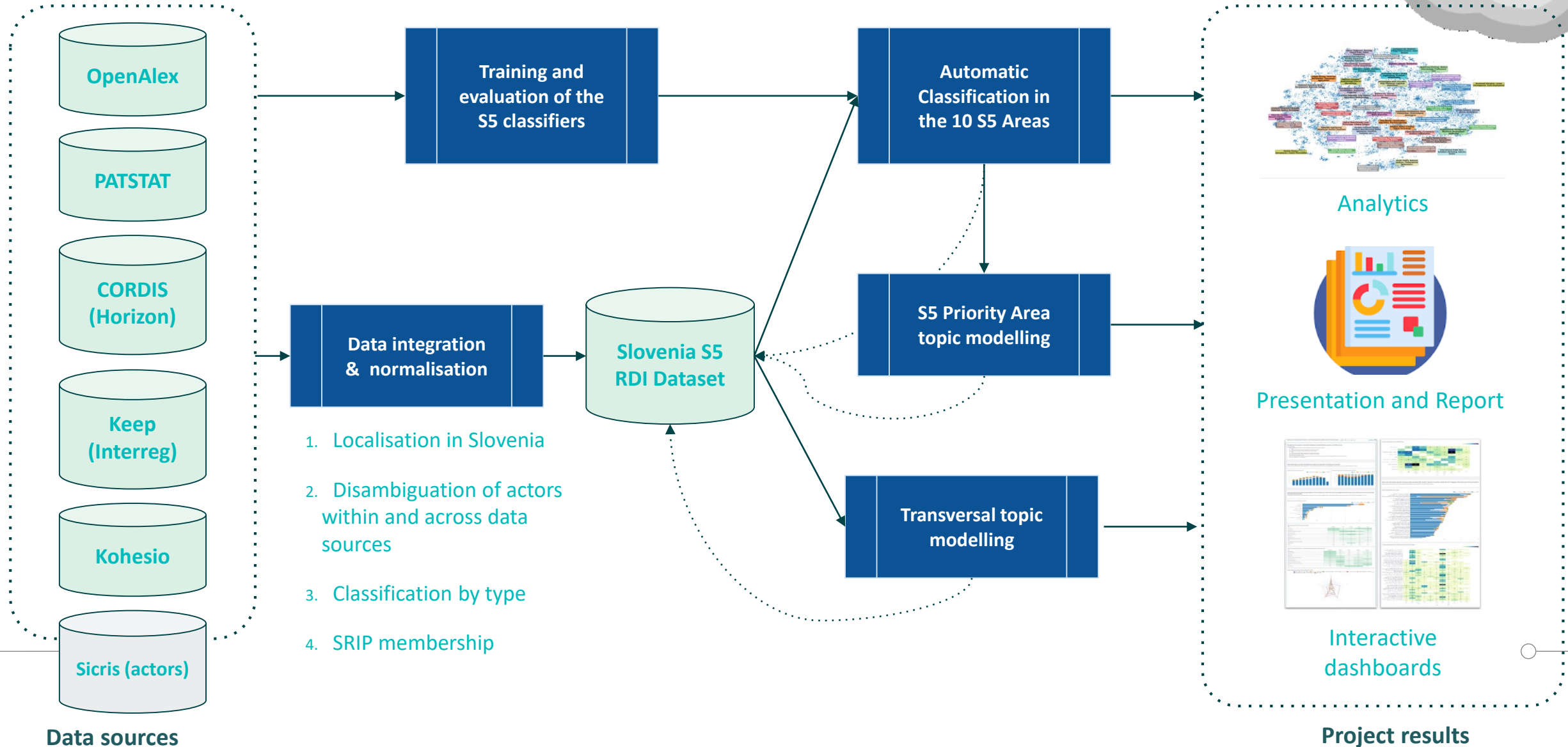


We look at the documents classified in each priority area, and perform a Topic Modelling to see emerging subtopics



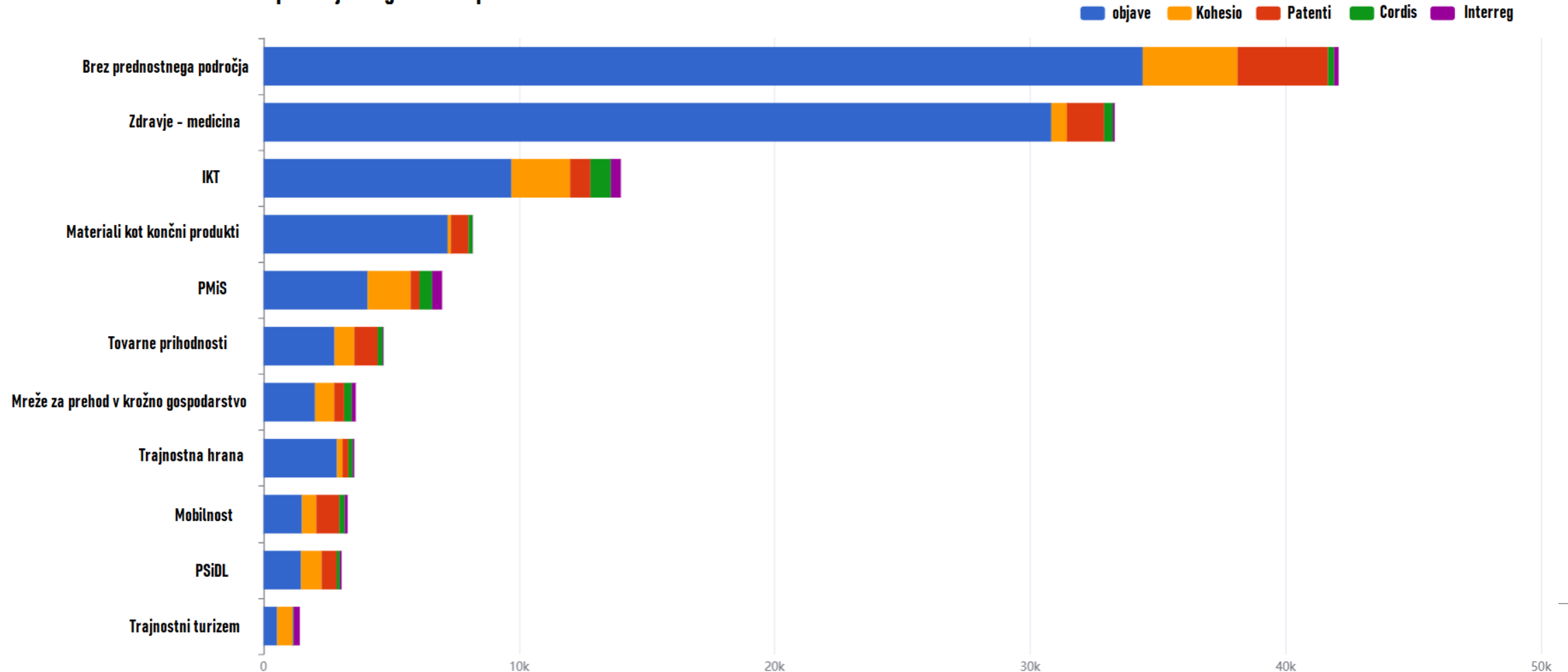
Identifying relevant RDI topics in “**Networks for the Transition to Circular Economy**”

Summary scheme of the full methodology

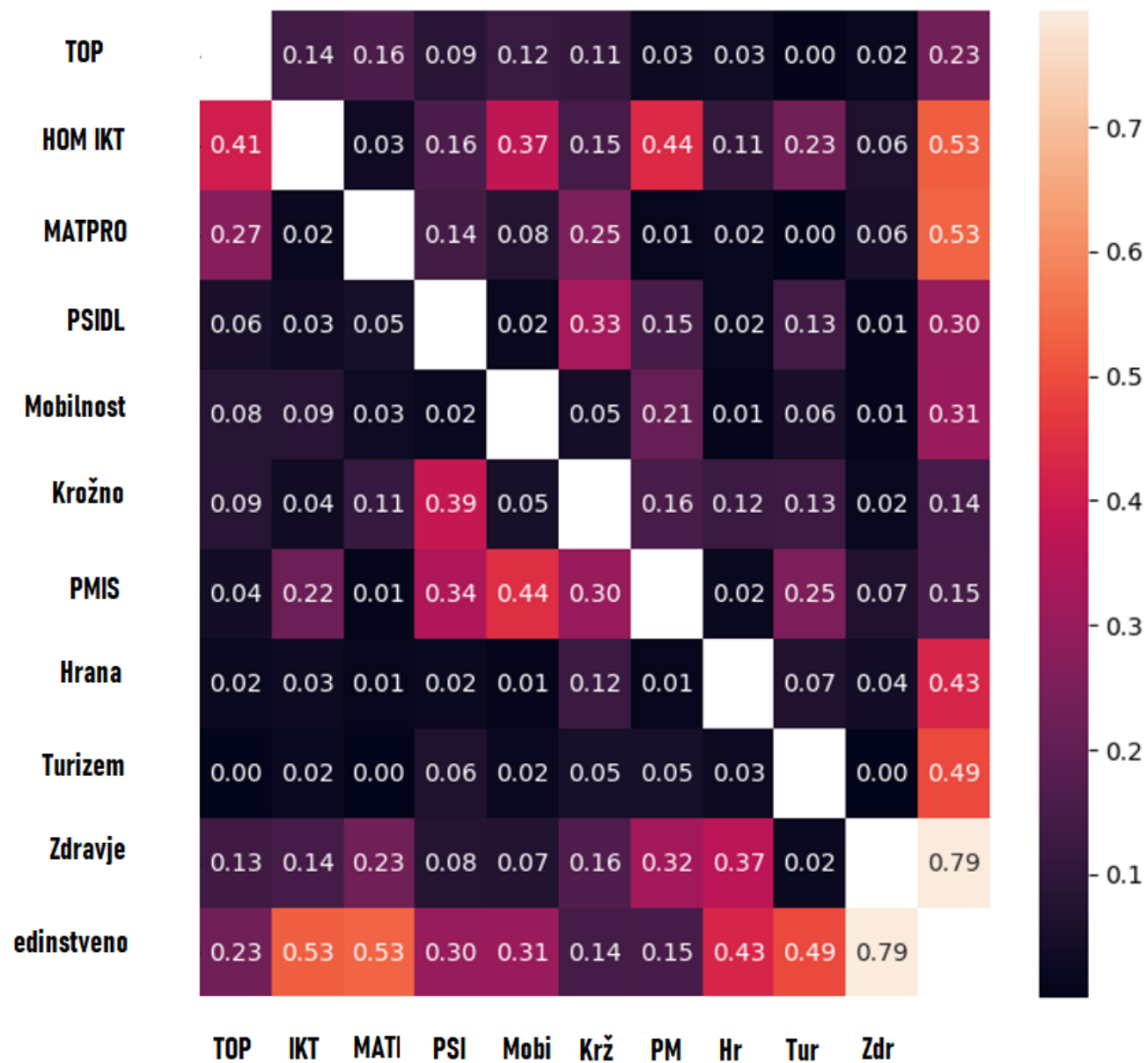


Razvrstitev slovenskih RRI dokumentov po prednostnih področjih S5 - od zgoraj navzdol

Število dokumentov v Prednostnem področju S5 glede na tip dokumenta



Stopnja (so)prekrivanja med prednostnimi področji S5



Understanding the overlap between S5 Priority Areas and the distribution of data sources per Areas



Documents classified in multiple S5 Priority Areas (co-occurrence)

	Factories of the future	Health and Medicine	ICT	Materials as End Products	Mobility	Circular Economy	Smart Buildings and Homes	Smart Cities	Sustainable Food	Sustainable Tourism
Factories of the future		626	1.91k	1.28k	391	405	263	181	102	7
Health and Medicine	626		2.02k	1.86k	243	594	239	2.27k	1.31k	27
ICT	1.91k	2.02k		271	1.22k	529	484	3.06k	383	324
Materials as End Products	1.28k	1.86k	271		255	911	422	70	86	5
Mobility	391	243	1.22k	255		168	57	1.46k	33	79
Circular Economy	405	594	529	911	168		1.18k	1.1k	444	189
Smart Buildings and Homes	263	239	484	422	57	1.18k		1.06k	64	193
Smart Cities and Communities	181	2.27k	3.06k	70	1.46k	1.1k	1.06k		81	361
Sustainable Food	102	1.31k	383	86	33	444	64	81		95
Sustainable Tourism	7	27	324	5	79	189	193	361	95	
	Factories of the future	Health and Medicine	ICT	Materials as End Products	Mobility	Circular Economy	Smart Buildings and Homes	Smart Cities	Sustainable Food	Sustainable Tourism

Color scale shows the level compared to the levels of all cells. Period: 2014 - mid 2024.

- Some S5 Priority Areas are very transversal, with significant output shared with several Priority Areas : this is the case of **Factories of the Future, Health and Medicine, ICT, Smart Cities and Circular Economy**.
- In contrast, other S5 Priority Areas have lower levels of co-occurrence of their scientific output with other Priority Areas, except with the most transversal ones : this is the case of Sustainable Food, Sustainable Tourism, Smart Buildings and Homes, and Mobility

Identifying relevant RDI topics automatically via Topic Modeling: How does it work?

Topic Modelling (TM) is a **machine learning technique** that serves to **automatically “discover” the topics from a collection of texts** (in this case, titles and abstracts of scientific publications and R&I projects).

Semantically-similar texts, identified by deep learning models, **are clustered together, forming the topics**.

Advantages:

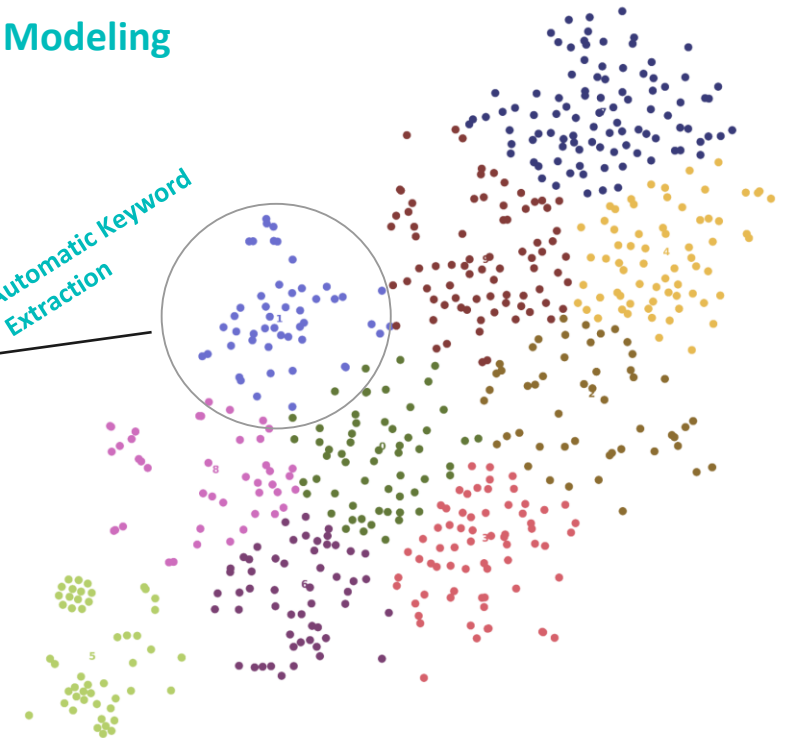
- Custom taxonomy to a specific perimeter
- Find interdisciplinary topics
- Find new topics
- Mix data from different sources (e.g. publications + projects + patents)

Photogrammetry and 3D Modeling

GenAI + human
review

photogrammetry,
photogrammetric, scan,
architectural, archaeological
excavation, model
archaeological,
reconstruction, 3d survey,
monument, camera

Automatic Keyword
Extraction



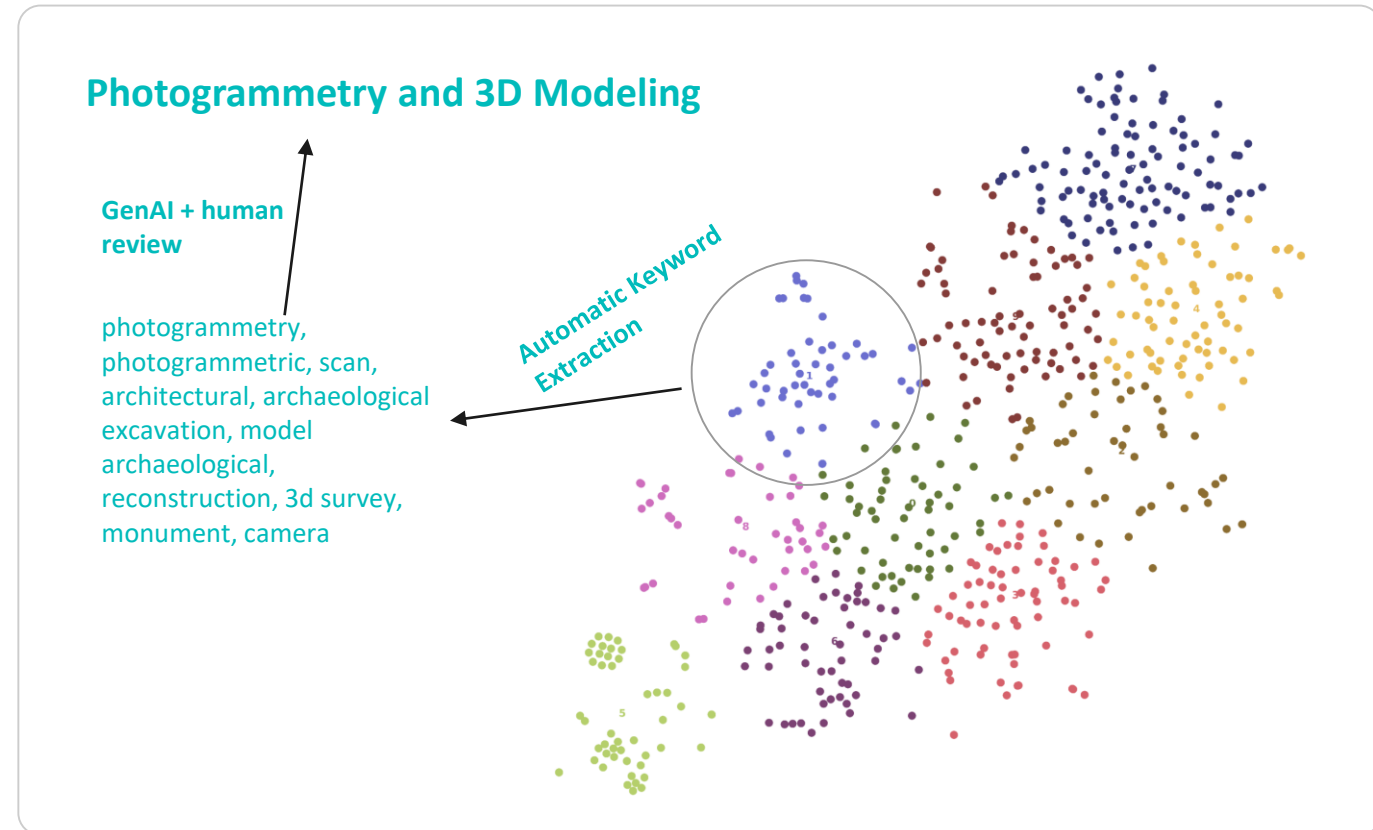
Identifying relevant RDI topics automatically via Topic Modeling: How does it work?

First, we use SPECTER*, a transformer-based model to convert texts into vectors.

Second, vectors are clustered into topics (using clustering method called K-Means).

Third, we use generative large language models and human expert curation to produce labels for each cluster.

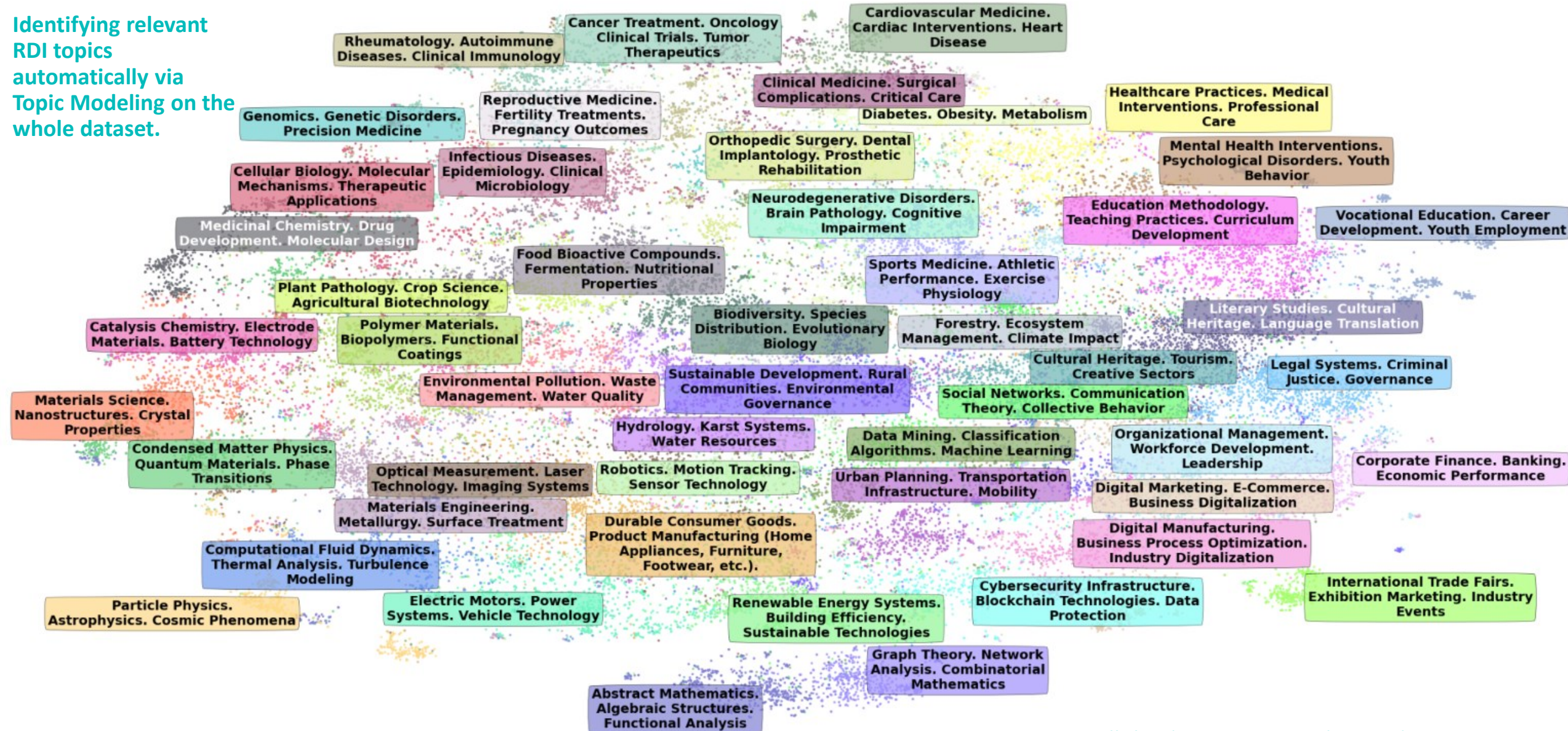
Example of a topic modelling, unrelated to Slovenia



* "SPECTER [is] a new method to generate document-level embedding of scientific documents based on pretraining a Transformer language model on a powerful signal of document-level relatedness: the citation graph."



Identifying relevant
RDI topics
automatically via
Topic Modeling on the
whole dataset.



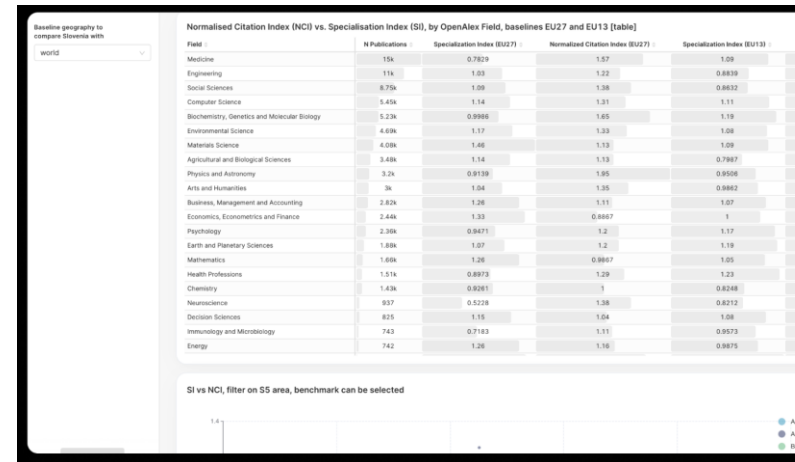
All the documents are clustered into 50 topics,
and the topics automatically labeled for human inspection

An interactive dashboard is built using this methodology: this presentation showcases the main results.

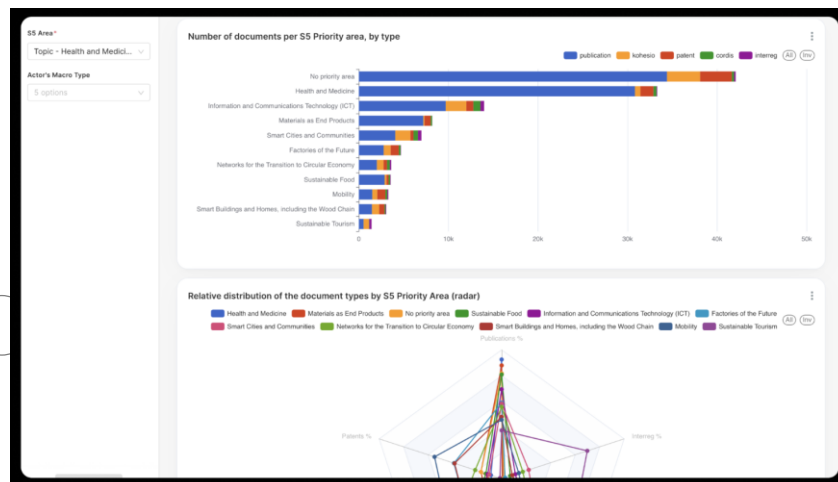
1. Overview of the Slovenian RDI activities in the S5 Priority Areas, their evolution, and their thematic content



3. RDI International benchmarking in publications and patents, using the original classifications systems of their data sources



2. Focus on individual S5 Priority Areas



4. Indicators on Slovenian R&I actors in the S5 Priority Areas and the SRIPs

5. Indicators on Slovenian R&I actors in the transversal topics

Actors summary indicators per S5 Priority Area

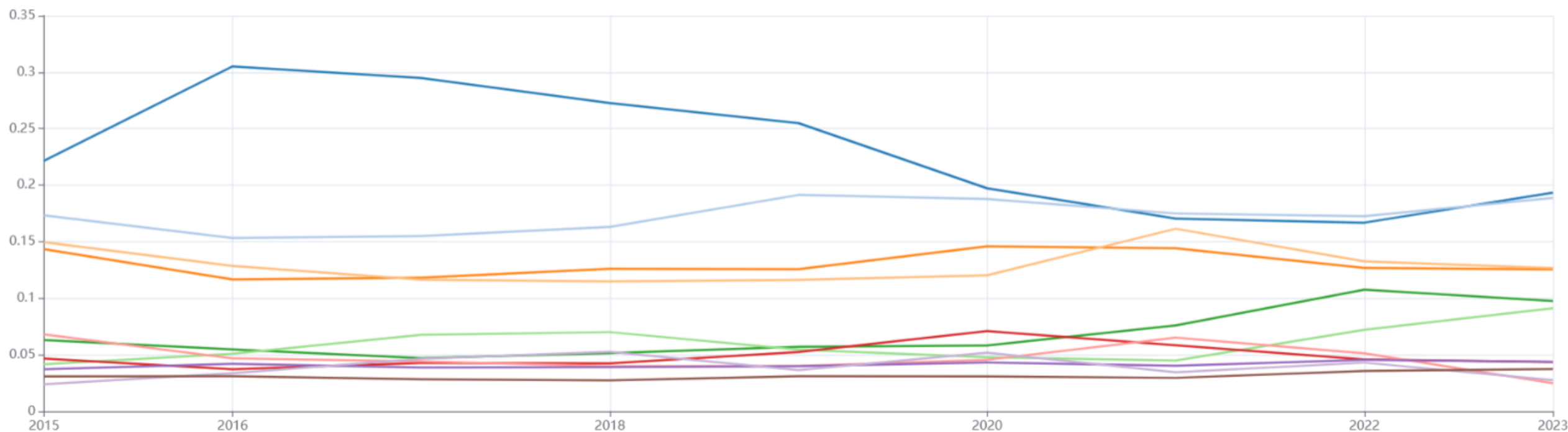
Show: 200 entries

ID	Name	Macro Type	S5 Area	SRIP Member	Score	Publications	Patents	SRIP Patents	Horizon Projects	Horizon Funding	Interreg Projects	Horizon Projects
607	Elaphe Population Technologies	Industry	Health	✓	0.0057	15	22	9	17	8M	0	4
4446	PIPISTREL VERTICAL SOLUTIONS D.O.O. PODJETJE ZA NAPREDNE LETALNIKE	Industry	Health		0.0336	0	0	0	14	10.5M	0	0
2762	KOLPORATOR GROUP D.O.O.	Industry	Health	✓	0.0319	9	11	6	6	4.54M	0	2
4196	ADAMOVIC D.O.O.	Industry	Health		0.0195	0	126	21	0	0	0	0
4873	SLGO D.O.O.	Industry	Health		0.019	0	0	0	0	0	0	4
3222	HOLDING SLOVENSKIE ELEKTARNE D.O.O.	Industry	Health		0.0185	0	0	0	1	6.23M	0	0
1348	SEVA D.O.O.	Industry	Health	✓	0.0182	0	10	3	0	0	0	3
2982	GEM MOTORS D.O.O.	Industry	Health		0.0154	1	5	2	0	0	0	3
4813	COS LABS D.O.O.	Industry	Health		0.0143	0	0	0	0	0	0	3
513	BERN-VEHOVAR KOMPOTI D.O.O.	Industry	Health		0.0143	0	0	0	0	0	0	3
8021	HORSA D.O.O.	Industry	Health	✓	0.0131	2	6	6	1	393k	0	2
4503	DOMEL D.O.O.	Industry	Health	✓	0.013	0	4	0	2	914k	0	2
8099	POLYCOM SPOJNA LOKA D.O.O.	Industry	Health	✓	0.0112	0	0	0	1	507k	0	2
5154	MAKNET D.O.O.	Industry	Health	✓	0.0102	1	0	0	9	2.87M	0	0
6311	TPV Group Corporation	Industry	Health	✓	0.0098	1	1	0	0	0	0	2
7045	OPTIMAX D.O.O.	Industry	Health		0.0095	0	0	0	0	0	0	1
4600	FURJAN TECH D.O.O.	Industry	Health		0.0095	0	0	0	0	0	0	1
6225	INTERCOM DEL D.O.O.	Industry	Health		0.0095	0	0	0	0	0	0	2
7412	RTC D.O.O.	Industry	Health		0.0095	0	0	0	0	0	0	2

Razumevanje porazdelitve in razvoja slovenskih RRI dejavnosti v prednostnih področjih S5

Časovni razvoj koncentracije dejavnosti RRI na prednostno področje S5 (v % skupaj, normalizirano glede na velikost vira podatkov)

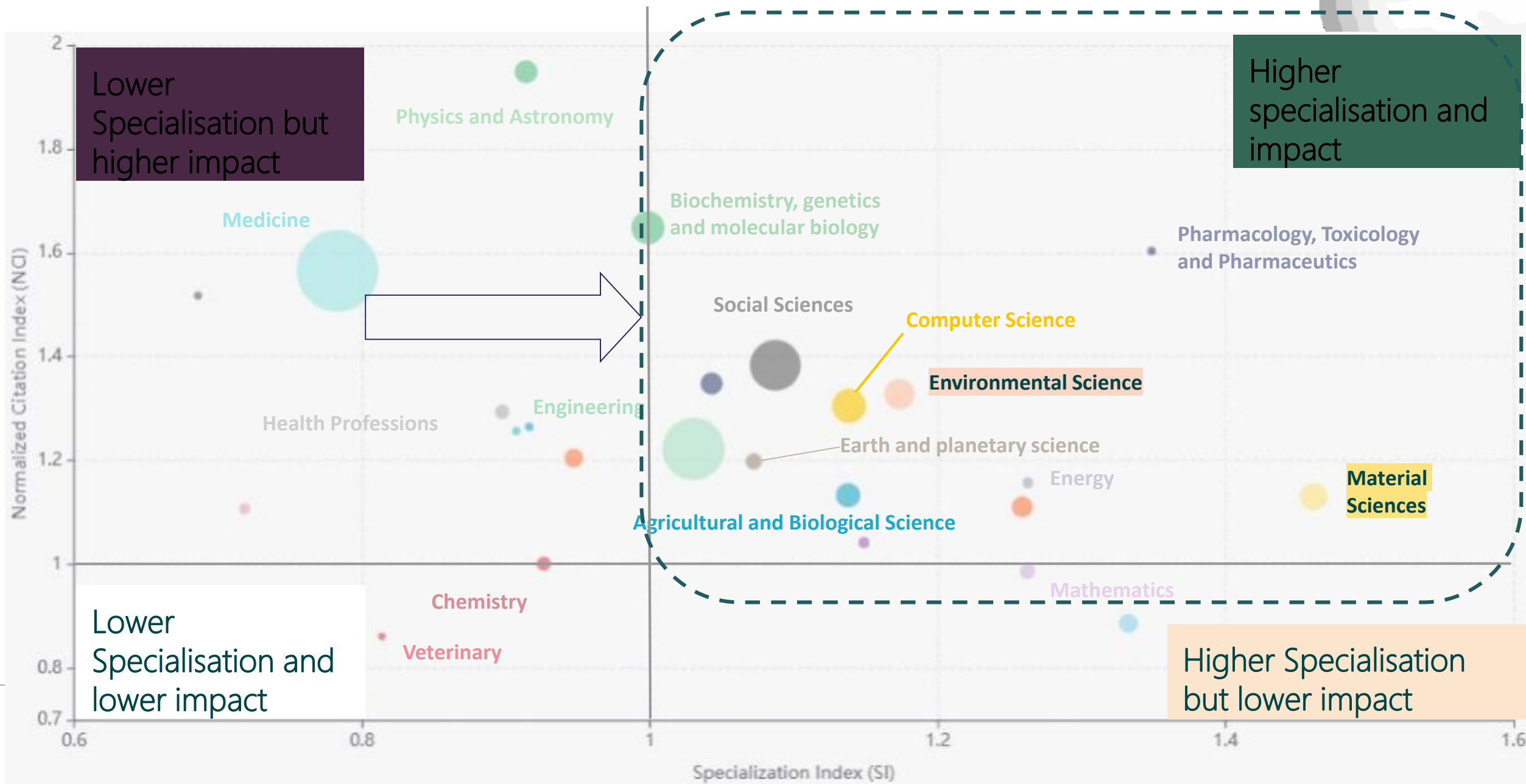
brez Predn. P. IKT Zdravje-medicina PMiS Krožno TOP
PSiDL Materiali kot končni produkti Trajnostni turizem Trajnostna hrana Mobilnost



- Ugotovitve:
 - zmanjšujejo se dejavnosti, ki se ne nanašajo na nobeno prednostno področje, kar izraža pozitiven trend specializacije in
 - nedavna relativna rast SRIP Krožno gospodarstvo in SRIP Mobilnost.

Benchmarking of publications to the EU 27

Normalised Citation Index (NCI) vs. Specialisation Index (SI), by OpenAlex Field.



Razumevanje prekrivanja med prednostnimi področji S5 in porazdelitev virov podatkov po področjih

Ključni vpogled

Prečna prednostna področja

Tri od prednostnih področij S5 so zelo medsektorska.

Prispevajo k številnim drugim, ne da bi sami imeli izstopajočo edinstveno dejavnost:

- Tovarne prihodnosti
- Pametna mesta
- Krožno gospodarstvo

Manjši delež edinstveno razvrščenih dokumentov na prednostnem področju S5, večji delež soobstoja z drugimi področji

Hibridna področja z edinstvenimi jedrom in prečnimi značilnostmi

Tri od prednostnih področij S5 so zelo medsektorska, obenem pa imajo tudi močno edinstveno dejavnost na svojem področju:

- Zdravje - medicina
- IKT
- Materiali kot končni izdelki

Večji delež edinstveno razvrščenih dokumentov na prednostnem področju S5, večji delež soobstoja z drugimi področji

PP s pretežno edinstvenim jedrom

Preostala štiri prednostna območja imajo veliko bolj edinstveno jedro:

- Mobilnost
- Trajnostna hrana
- Trajnostni turizem
- Pametne stavbe

Večji delež edinstveno razvrščenih dokumentov na prednostnem področju S5, večji delež soobstoja z drugimi področji

Gonilne sile preobrazbe

- Digitalizacija
- Trajnost in krožnost

Sektorski inovacijski grozdi

- Agroživilstvo
- Grajeno okolje
- Mobilnost
- Napredni industrijski sektorji (*vključno s stroji, materiali, **obrambo***)
- IKT / digitalni sektor
- Zdravstvena industrija in zdravstvene storitve
- Turizem

*Pametna mesta se
"razpusti" v 2, 4 in 5*

Visokotehnološki temelji – Znanost in tehnologija

- Napredni materiali in napredna proizvodnja, vključno z robotiko in optiko/laserji.
- Zelene tehnologije in okoljske znanosti
- Biomedicinske raziskave in biotehnologija

Rešitev izziva

- Prednostna področja Strategije S5 so bila pravilno zastavljena. Načrtovanje PP in „bottom-up“ proces v okviru procesa podjetniškega odkrivanja (EDP) je bil ustrezno upravljan.
- Napoved „state of art“/“beyond state of art“ tehnologij je bila ustrezna.
- Transformacija prednostnih področij po letu 2027.
- Semantična analiza ni nadomestilo za EDP, kvečjemu je lahko podpora temu procesu podjetniškega odkrivanja.
- Transparentnost metodološkega pristopa je nujna.
- RRI stičišče kot logični naslednik (odprto vprašanje).
- Rezultati analize predstavljajo analitično podlago in ne uradnega stališča države.

»Lessons learned«

Brez dobrih podatkov ne gre.

Izkušeni izvajalci so plus.

Interpretacija izsledkov raziskav na podlagi modelov umetne inteligence je ključnega pomena.

Motivirani sodelavci so zlata vredni.

Semantična analiza raziskovalno-razvojnih in inovacijskih (RRI) dosežkov v okviru S5

Matej Forte
v imenu ekipe S5

za več info:
matej.forte@gov.si
041-363-234

Hvala za pozornost

