



Policymakers' sessions: RISIS-KNOWMAK – A tool for exploring knowledge production in Europe

Benedetto Lepori* and Thomas Scherngell

*Università della Svizzera italiana, Lugano and
University of Paris Est, Paris, blepori@usi.ch

**Austrian Institute of Technology, Vienna,
Thomas.scherngell@ait.at

Brussels, 19 December 2019

Programme

- 14.00 14.30 The RISIS-KNOWMAK tool as an interface to RISIS integrated data
- 14.30 15.15 Demonstration of the KNOWMAK tool
- 15.15 15.45 Break
- 15.45 16.15 The spatial distribution of knowledge production in Europe. A latent class analysis
- 16.15 17.00 Future development of RISIS-KNOWMAK and open discussion

The RISIS ambition

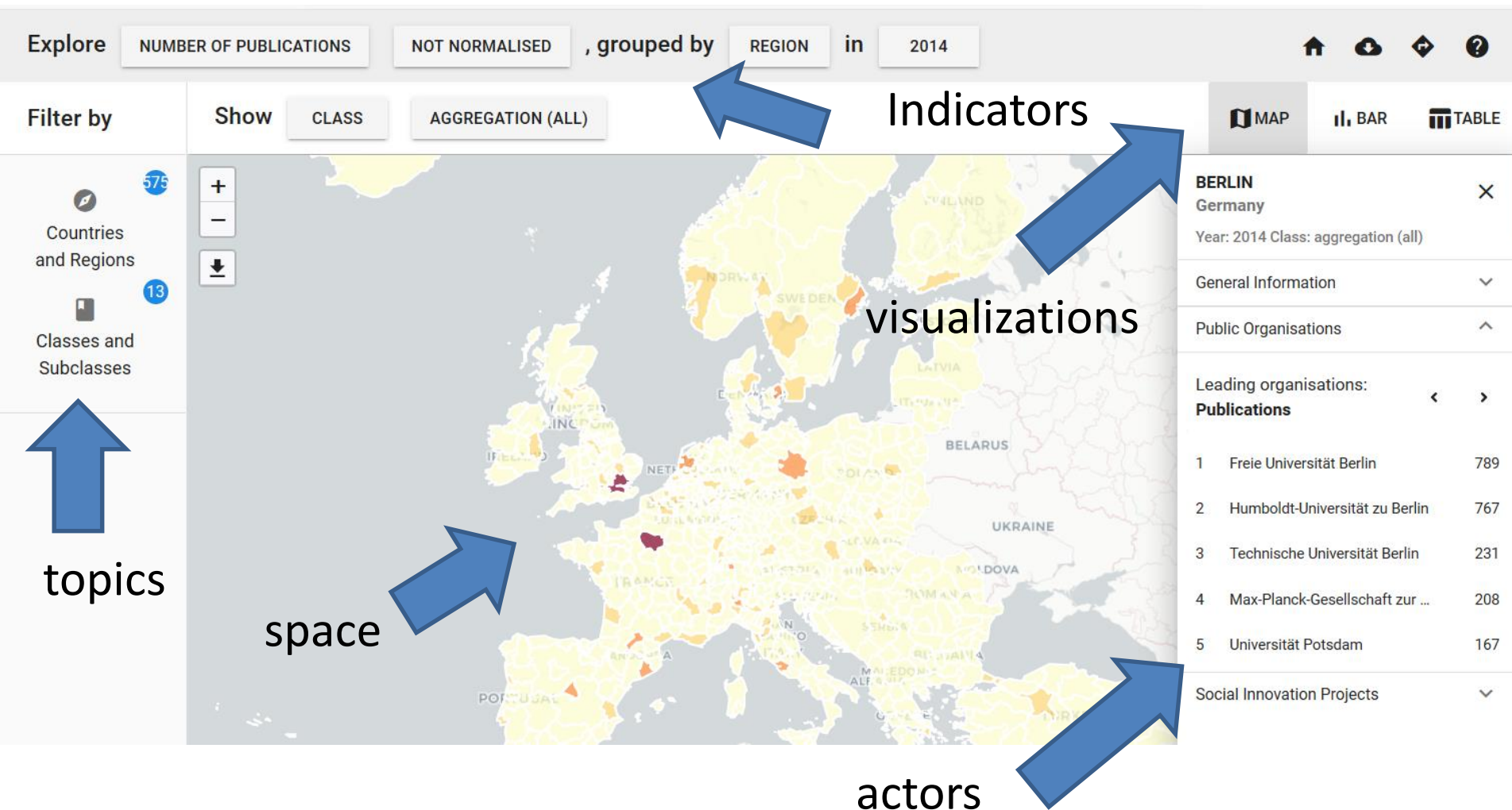
- RISIS2: H2020 research infrastructure project (risis2.eu)
- Creating a data infrastructure for the study of science and technology
 - By integrating and opening existing datasets (publications, patents, etc.)
 - By developing additional dataset (social innovation, research funding programs, etc.)
- Three core integrative dimensions
 - Space (regions, countries)
 - Topics (ontology)
 - Research organizations (public and private)
- Beyond access to micro-data
 - Provide ready-made indicators to policy users for further analysis

The RISIS-KNOWMAK tool

- An on-line tool to display and download selected indicators on knowledge production in the European Research Areas
- Exploits the core datasets in RISIS publications, patents, projects, social innovation projects,...)
- Integrated through a common ontology, space (countries/regions), actors (public and private)
- Developed within an H2020 project (www.knowmak.eu)
 - To be further extended and maintained within RISIS

The tool:

<http://knowmak.eu/dashboard>



The tool

- To explore geographical spaces of knowledge production
 - Countries and regions
 - Looking into regional actors (HEIs, PROs, firms)
- To explore topics of knowledge production
 - KET/SGC and their subclasses
- To combine different types of data
 - Established: publications, patents, projects.
 - New: social innovation projects and citizens' attention based on social media
- To visualize knowledge production
 - Through an interactive tool
 - And the download of ready-made indicators for further analysis

A focus on indicators

- The KNOWMAK tool provides indicators, not data
- Relevant and condensed information
- No direct access to micro-data (also for legal reasons)
- To answer to relevant questions like:
 - Which are top-European research actors on a certain research topic?
 - In which topics is a given European region specialized ?
- Indicators deal mostly with
 - Volume of activities and outputs (raw or normalized)
 - Linkages between topics, actors, geographical spaces
 - Combinations of dimensions (composite indicators)

Who could use it?

- Policy analysts at European, national and regional level
 - Overview of knowledge production
 - Fine-grained views by space *and* topics
 - Ready-made maps and indicators
- Researchers
 - Fine-grained ready-made datasets with multiple indicators
- Research managers
 - Identifying partner regions and actors
 - Particularly at the topic/subtopic level

What is new

- Linking policy topics to data by using ontologies
 - as a flexible approach to combine different types of sources in a single topical view
- Integrating heterogeneous data sources
 - To provide single views based on space/actors
 - To develop composite indicators and alternative views of knowledge production
- Providing data on social innovation and user attention
 - Covering the realm of 'other' types of innovation, driven mostly by civil society organisations
 - Inquiring about the broader reach of science in society

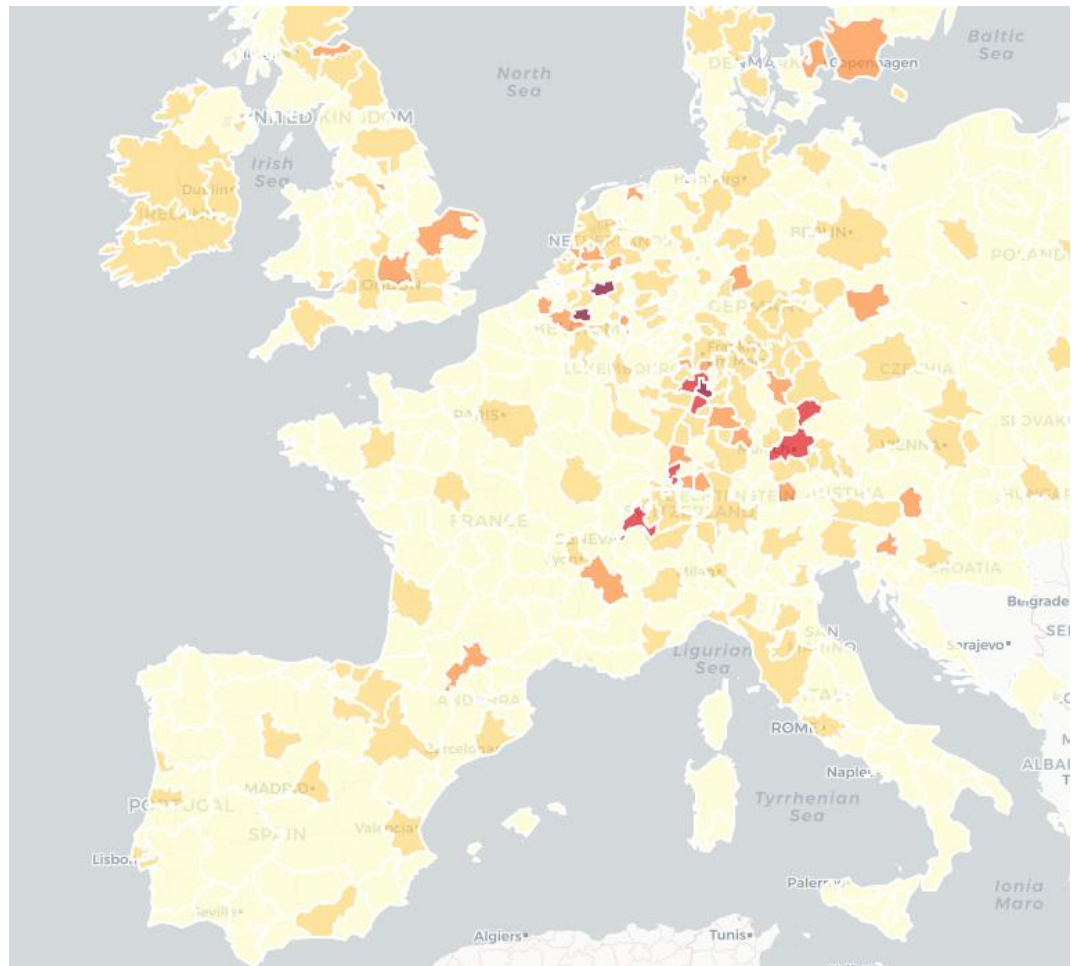
Space

- Countries
- Regions combining metropolitan areas and NUTS2/3 regions

You will be able to

- Display and combine data and indicators
- View regional profiles including topical specialization and actors

A spin-out of RISIS work on space and geolocalisation



Topics

- A fine-grained ontological structure (13 KET/SGC and 135 subclasses)
- Keywords used to annotate data and to attribute them to classes to build indicators
- A very flexible approach in terms of policy questions *and* of structure of the data

KEY ENABLING TECHNOLOGY advanced manufacturing technology advanced materials industrial biotechnology micro- and nano-electronics nanoscience and technology optics and photonics SOCIETAL GRAND CHALLENGE bioeconomy climate change	MNE manufacturing advanced materials for manufacturing manufacturing biotechnology nanotechnologies for manufacturing photonics in manufacturing software for manufacturing	advanced design software for manufacturing logistics supporting software for manufacturing production supporting software
---	--	---

Selected Class: http://www.gate.ac.uk/ns/ontologies/knowmak/amt_software

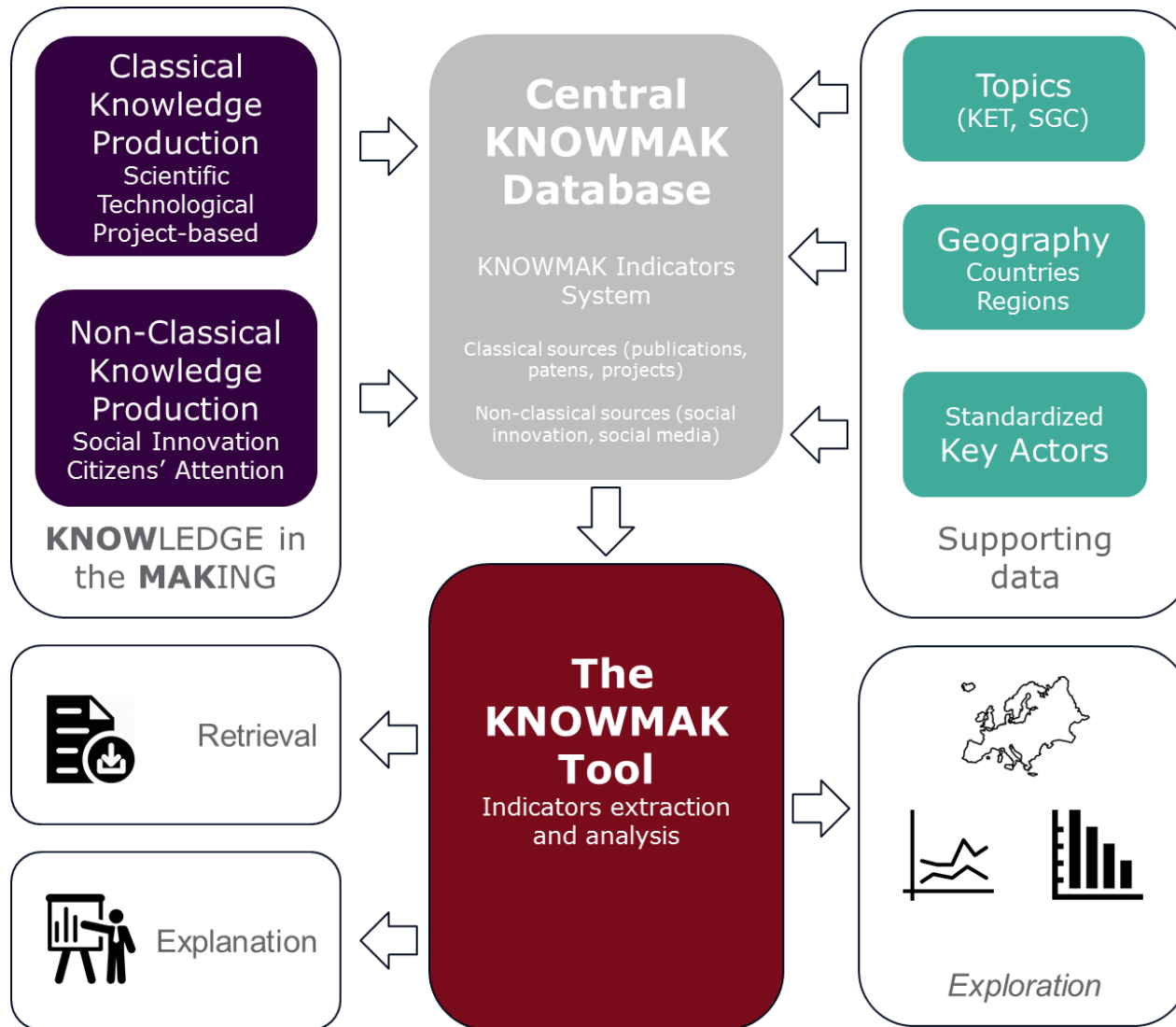
Software for manufacturing. AMT Software.

Related Keywords: amt software, cloud-based, custom, customised, customization, hardware, matlab, middleware, multimedum, runtime, software for manufacturing, vendor, workflow

Actors

- Standardization of public-sector actors
 - Building on RISIS Public Research Organizations Register (orgreg.joanneum.at) and Firms Register (firmreg.joanneum.at)
 - Covering HEIs (ETER+), PROs and firms
- Extensive matching of actors with data sources
- KNOWMAK currently allows to
 - Identify top-actors by data source and topic for each country/region
 - Actors-level data will be made available within RISIS2

Data integration in practice

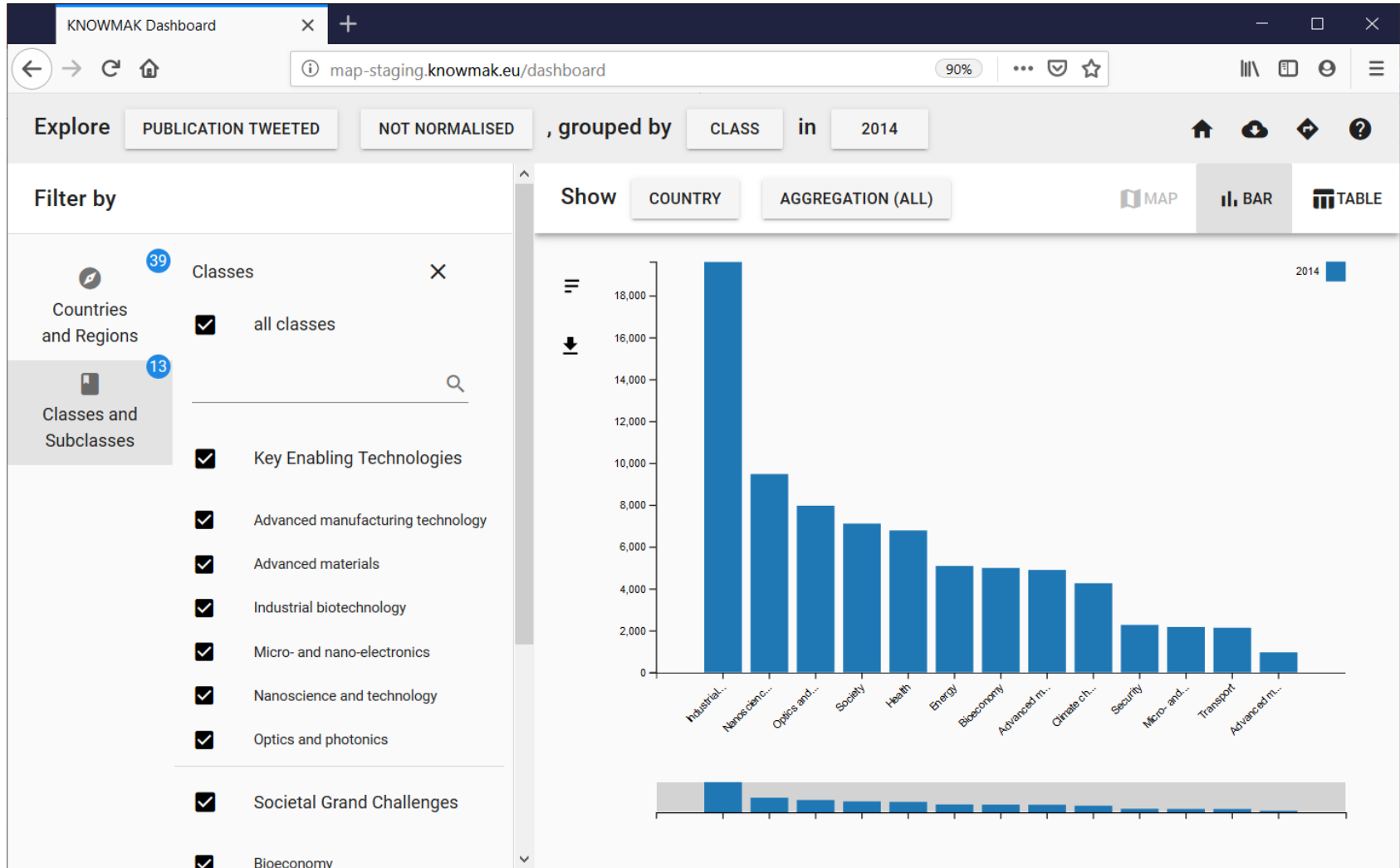


- Central database for the data integration
- Design of workflows for primary data annotation, data transfer, integration.
- On-line interface built on the central database (API-based)
- Allows for reproducibility of the process

Usages

- On-line visualization of patterns of knowledge production
 - To explore and navigate accross spaces and indicators
- Simple analyses based on the data
 - Combining dimensions and using descriptive statistics
 - The KNOWMAK stories
- More far-reaching analyses
 - Downloading the indicators for further analyses

On-line visualizations



Simple analyses. Open access and user attention

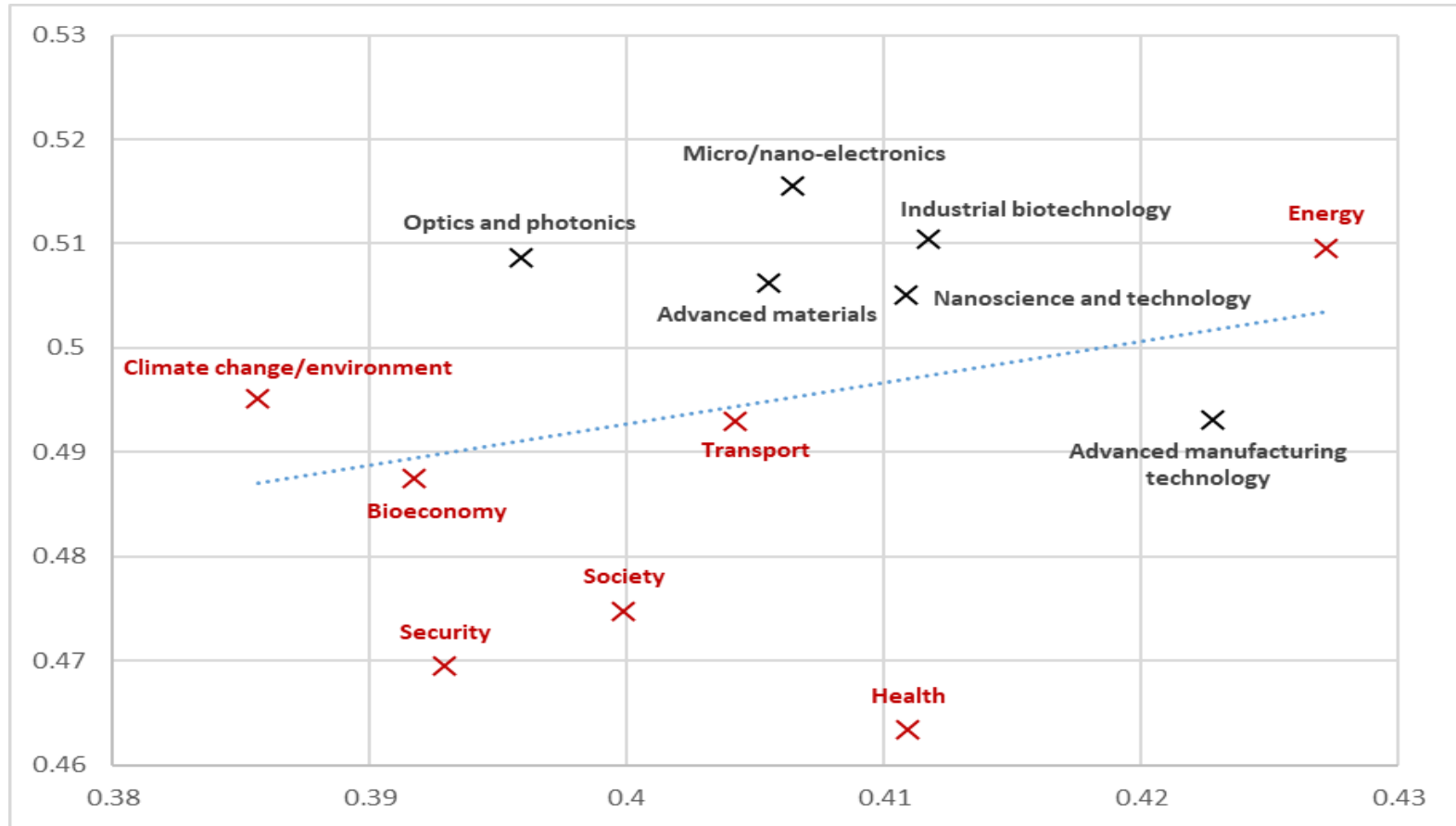
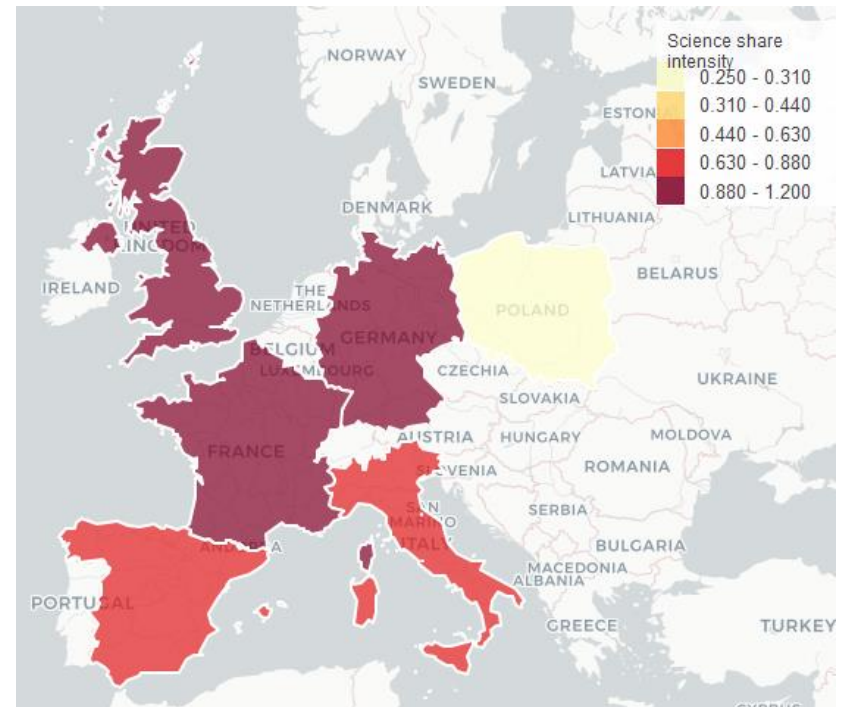
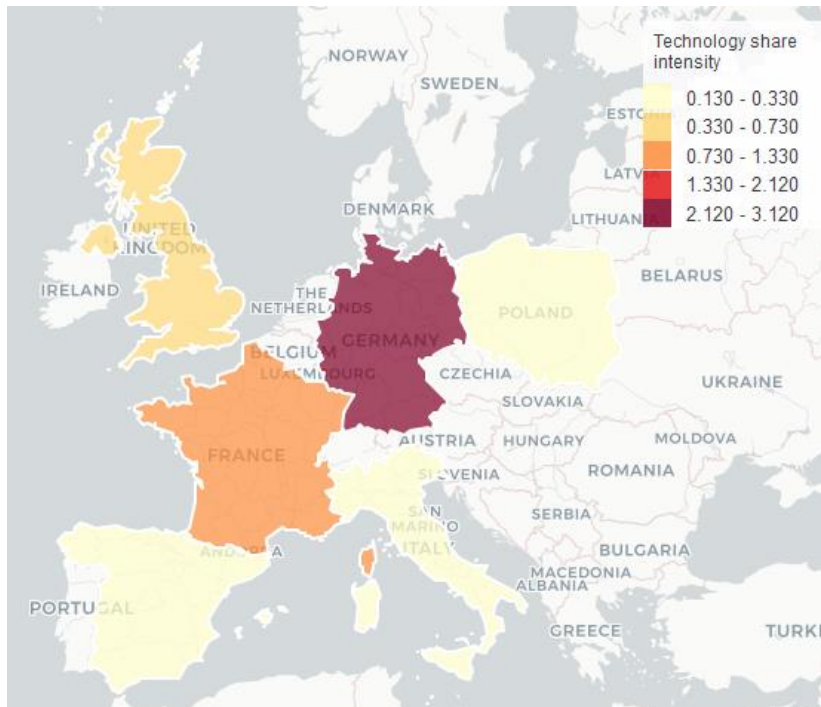


Figure. Percentage of OA (vertical) and tweeted (horizontal) publications by topic

Composite indicators



Science production intensity: based on publications and EU-FP indicators normalized by the population.

Technology production intensity: based on patent indicators

Outlook

- Integration must be grounded on conceptual dimensions and the identification of suitable indicators
 - Beyond the 'big data' hype
- Need for tools to make indicators available for exploration and further analysis
 - Reduce usage barriers
 - Allows for reproducibility and standardization of processes
- Wide usage of a tool helpful also for testing and improvement
- Some clear areas of further development (see final discussion)
 - More fine-grained ontology
 - Extending actors' coverage to firms



Demonstration of the RISIS-KNOWMAK tool

Thomas Scherngell

Brussels, 19 December 2019

Outline


1) Exploring the interactive KNOWMAK dashboard and its functionalities

- a. Selecting and filtering indicators
- b. Visualising and extracting indicators

2) Some selected usage scenarios

- a. Scenario A: Spatial distribution of knowledge creation in Genomics research (different types, combined and networking)
- b. Scenario B: Open Access and Social Media for disseminating new knowledge (total and by topics)
- c. Scenario C: Identify Top knowledge producing Actors and Social Innovation Projects in Energy in Germany, Poland and France

The KNOWMAK user handbook

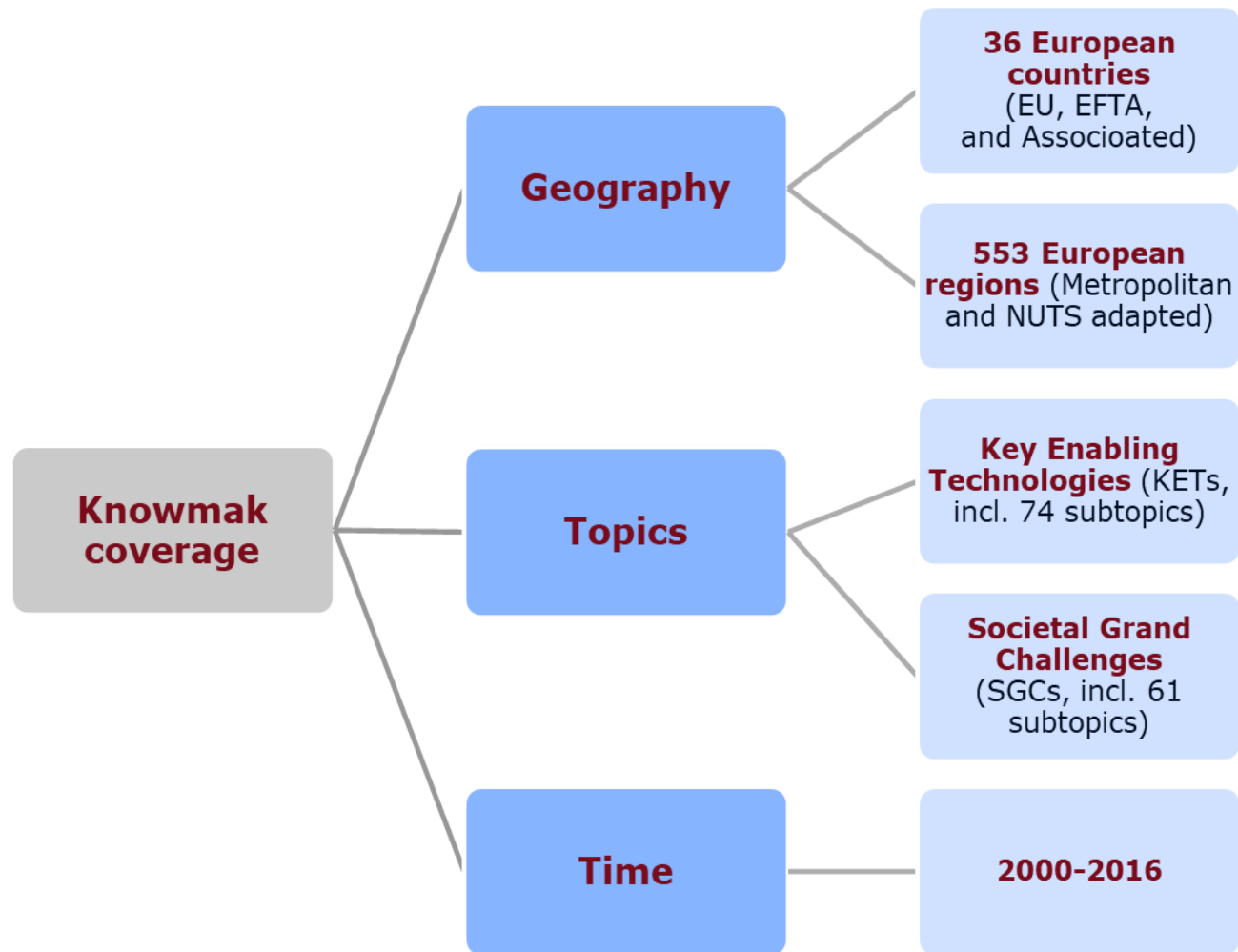
- Available for download directly in the tool under the help button. 
- Provides in a compact manner a
 - guide through the tool and definition of all indicators and functionalities
 - glossary of main terms and their definitions (e.g. actor, region, indicator, etc.)
 - Pointers to limitations specifically relevant for interpretation of results

Main functionalities





The tool allows the user for three main analytical possibilities

- **Exploration:** Allows the user to explore indicators in the tool in form of tables and with different visualisation forms (maps, bar charts and line diagrams)
- **Retrieval:** Allows the user to download selected indicators in a suitable format for further statistical analysis or combination with other data
- **Explanation:** Supports the user in the navigation through the tool with online help features, user handbook, as well as some exemplary analysis (data stories)

KNOWMAK coverage



Indicators available in the tool

Category	Indicator
	Number of publications Number of publications in the Top10% cited Number of intercontinental scientific collaborations` Number of Open Access publications Number of tweeted publications (user attention)
	Number of patent applications Number of transnational patent applications
	Number of EU-FP participations Number of EU-FP coordination
	Lists of social innovation projects per spatial entity and topic, with information on project title, website and actors (available via factsheets, see Section 4.3)

Publications

Patents

EU-FP projects

Social Innovation

Note: To be updated in further releases

Composite indicators: Composite knowledge production



Knowledge production share

Average of the shares of projects, publications and patents; gives an overall impression of knowledge production activities, in particular when comparing a larger set of regions/countries or whole Europe



Knowledge production intensity

Total production share normalised by population



Science share intensity

Total scientific knowledge production share normalised by population



Technology share intensity

Total technological knowledge production share normalised by population

The dashboard (I)

The dashboard view is the user space for **data exploration and retrieval in the tool**. The basic steps to display or retrieve data are

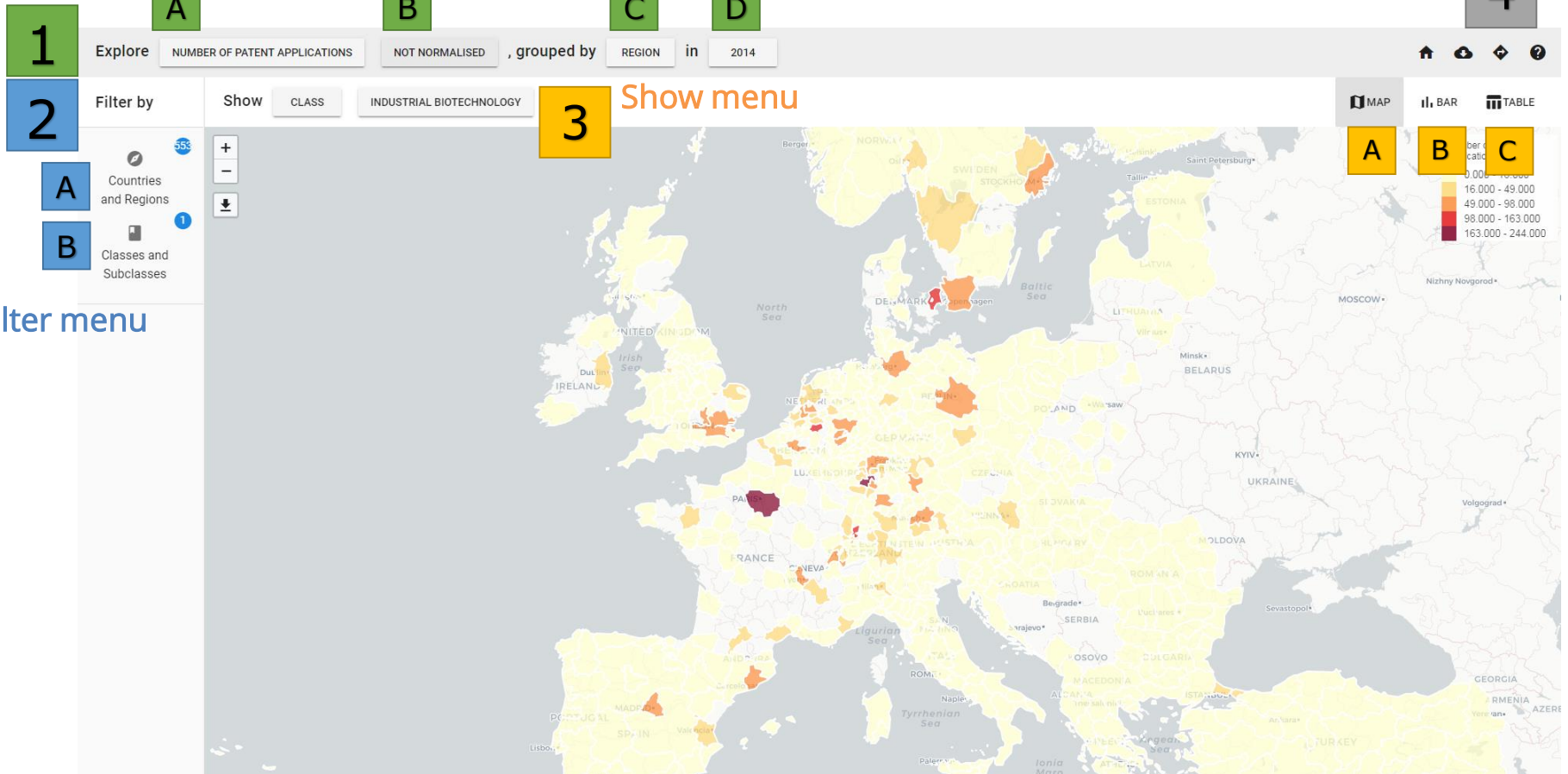
1. The **explore menu** for (A) selecting the indicator; (B) normalisation (not normalized or normalized by population); (C) dimension (location or topics) and (D) time. These three choices are mandatory.
2. The optional filter menu for further filtering data in terms of (A) geographical and (B) topical coverage according to user's needs.
3. The **show menu** for **analytical choices** to investigate selected data with the different (A) (B) (C) visualization (exploration mode) or (C) export data (retrieval mode)
4. The **help and download menu** featuring one functionality to access the handbook, a guided tour, and a functionality for downloading selected indicators

The dashboard (II)

Explore Menu

Help and
download menu

4



Filter menu

How to start?

Explore NUMBER OF PUBLICATIONS by COUNTRY in 2014 ←

1. **Selection** of an **indicator**, and specify the exploration mode ("*explore by*") in terms of **geography** and/or **topics** and *in time*. The selection of the indicators to be displayed requires three choices by the user:

A Selecting the indicator to be explored

B Normalisation (not normalized or normalized by population);

C Explore selected indicator by

Geography **Country or Region**

OR

Topics **Class:** (6 KET and 7 SGC) or **Subclass:** (135 subareas of KETs and SGCs)

D In Year

a. In year **Years** 2010 – 2016

b. default year 2014 (no patents for 15/16)

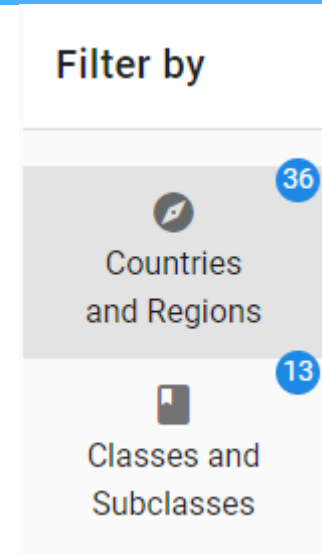
How to filter?

2 Filtering: the pre-selection from **1**, i.e. the selected indicator can be filtered by

A Countries and regions

B Classes and subclasses

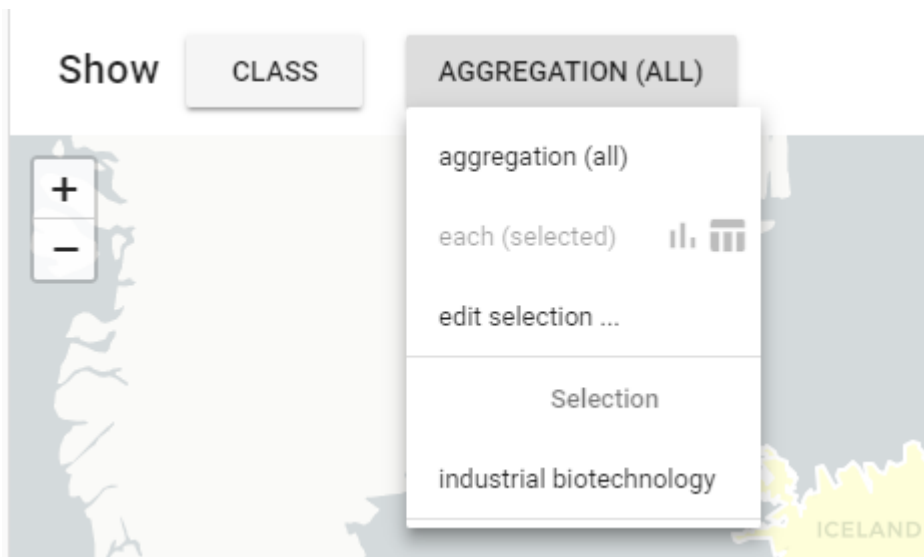
- By default, all countries and topics are selected. Ticking the all countries/all regions and the all topics/all subtopics box selects/de-selects all entities
- Specific countries/regions and classes/subclasses can be selected by ticking single boxes



How to explore? (I)

The analytical possibilities available are **Map View**, **Bar Chart** and **Table View** (incl. data download)

The **Show** option is needed to specify in the three analytical views the class or subclass to be displayed in case more than one classes/subclasses have been selected in the **filter**



Aggregation all: Total of the indicator (not of the selected topics)

Each (selected): all active classes/subclasses; switches therefore to bar chart (map view not possible)

Edit Selection: Switch between classes when only want to show one selected in the bar chart

How to explore? (II)

The **Map View** contains the following elements/functions (see handbook for screenshots)

- a. Legend (right top): contains the colours and the range for the five classes (natural breaks) produced for the indicator values
- b. Zoom into map by using + and – button or mouse wheel
- c. To display value mouse over country/region
- d. Display country/regional profile on right side of dashboard by

How to explore? (III)

The **Bar Chart** displays the selected indicators by means of a classical bar chart, with each bar representing an indicator value for a country/region and/or class/subclass (see handbook for screenshots)

- a. Sort the data in ascending/decending alphabetical order (A-Z/Z-A), or by increasing/decreasing numerical values (0-9/9-0)
- b. Zoom into the bar chart by narrowing down (mouse over changes to an arrow) and/or moving the grey area of the small chart on the bottom of the page
- c. To display value mouse over bar
- d. Display country/regional profile on right side of dashboard by

How to display data and extract?

The **table view** displays raw values of the selected indicator and contains the download function (for download in .csv format);

Unique to the table view is the possibility to display data for multiple years (data from year to year), which is in particular useful for data retrieval. Moreover you are able to

- a. Select rows per page (default = 10) and click to go to next row
- b. Sort function (for each column right hand to the columns header)
- c. Display country/regional profile on right side of dashboard by
- d. Download as selected in csv (data retrieval in .csv format)

Advanced download dialogue (multiple indicators and/or years) can be reached in the help menu by clicking the cloud symbol

Some usage scenarios

Scenario A: Spatial distribution of knowledge creation in Genomics research (different types, combined and networking)

Scenario B: Open Access and Social Media for disseminating new knowledge (total and by topics)

Scenario C: Identify Top knowledge producing Actors and Social Innovation Projects in Energy in Germany, Poland and France

Scenario A: Spatial distribution of knowledge creation in Genomics research

1. Select

- base indicator: three types of knowledge creation, i.e. number of publications, number of patents, number of EU-FP participations (select only one first, can be switched during show)
- normalised or non-normalised, group by: spatial setting (country or region), time

2. Filter

- Tick „Classes and Subclasses“ menu and de-select all classes
- Then tick under the class button **Show** for selecting subclasses;
- Tick the arrow in the right next to the KET „Industrial biotechnology“
- Select „Genomics“, and close the filter menu by ticking „x“

3. Show

- Switch between indicators, normalisations and spatial settings in the explore menu
- Add networking indicators to show network centrality of countries/regions
- Switch between different show views
- Download the indicators (from table view or download dialogue)

Scenario B: Open Access and Social Media for disseminating new knowledge (by topic)

1. Select

- base indicator: Number of publications Open Access, number of publications tweeted (select only one first, can be switched during show)
- normalised or non-normalised, group by: class/subclass, time

2. Filter (optional)

- Tick „Countries and Regions“ menu if you want only specific spatial entities to be included

3. Show

- By grouping along classes/subclasses the tool switches to the bar chart view
- Sort the indicator using the sort function
- Switch between classes and subclasses; use the zoom function to detect most prominent subclasses for the two indicators
- Show the data in table; sort and extract/download
- Filter out countries/regions and/or classes/subclasses with the filter function

Scenario C: Top knowledge producing Actors and Social Innovation Projects in Energy

1. Select

- Select any base indicator (or leave the default, i.e. number of publications)
- normalised or non-normalised, group by: spatial setting (country or region), time

2. Filter

- Tick „Classes and Subclasses“ menu and de-select all classes (subclasses not necessary in this scenario)
- Select ,Energy‘, and close the filter menu by ticking ,x‘

3. Show

- To show top knowledge producing actors and Social Innovation projects (SIPs) in a specific country/region, click on it the map view (use the mouse overlay for displaying the name of the spatial entity); opens region/country factsheet;
- Show top-5 organisations by ticking the header and switch between patents, publications and projects; Show SIPs by ticking the header
- Click on an actor and/or SIP to get additional information



The spatial distribution of knowledge production in Europe. A latent class analysis

Benedetto Lepori, Massimiliano Guerini, Thomas Scherngell and Philippe Larédo

Brussels, 19 December 2019

Background

- The creation of new knowledge is the essential basis for successfully generating innovation, and thus
 - a major determinant of the overall socio-economic development of regions
- Increasing interest in the investigation of the distribution of knowledge at the regional level of analysis,
- and in the classification of regions into different types in terms of their knowledge creation activities
- How KNOWMAK can contribute to address these questions

Motivation

- Most empirical works so far rely mainly on technology based indicators (mainly patents) (see e.g. Verspagen et al. 2010),
 - and combine these indicators in a linear-additive manner to a synthetic regional innovation index (e.g. Hollanders et al. 2009)
- these works usually
- underestimate or neglect the importance of different dimensions of knowledge creation (e.g. academic or project-based),
 - and rely on spatial breakdowns (mostly NUTS2 for the European case) with pure administrative borders, often intersecting agglomerations of knowledge creation

Objective

→ address this gap by using the RISIS-KNOWMAK dataset, and the application of different methodological and empirical advancements

Main goal is to analyze the spatial distribution of knowledge production across European regions, focusing on different modes of knowledge production, including

- Science-based' knowledge production
- R&D Project-based knowledge production
- Technological knowledge production

Intersecting these dimensions, we develop a new classification of the regions in terms of knowledge production

Data

Derived from the KNOWMAK dataset complemented with ETER data on higher education.

- Using the NUTS 3 adapted regional structure from KNOWMAK

In the initial analysis, a set of indicators are included:

- N. of publications, N. of international publications, N. of publications in the top 10%
- N. of participations and coordinations to EU-FP projects
- N. of priority patent applications and of transnational priority patent applications
- N. of students ISCED5-7 (ETER) and ISCED 8 (ETER)
- Area and Population Density

The advantage of KNOWMAK : these indicators available ready-made for download.

Methods and empirics

- Principle Component Analysis (PCA), is mobilized to identify relationships between indicators and to single out the main dimensions differentiating regions
- Latent Class Analysis (LCA) is used to classify regions according to different types in terms of knowledge production
- Empirical setting:
 - Indicators in the initial analyses are measured for a cross-section; year: 2013
 - Regional breakdown comprising 553 metropolitan and other European regions

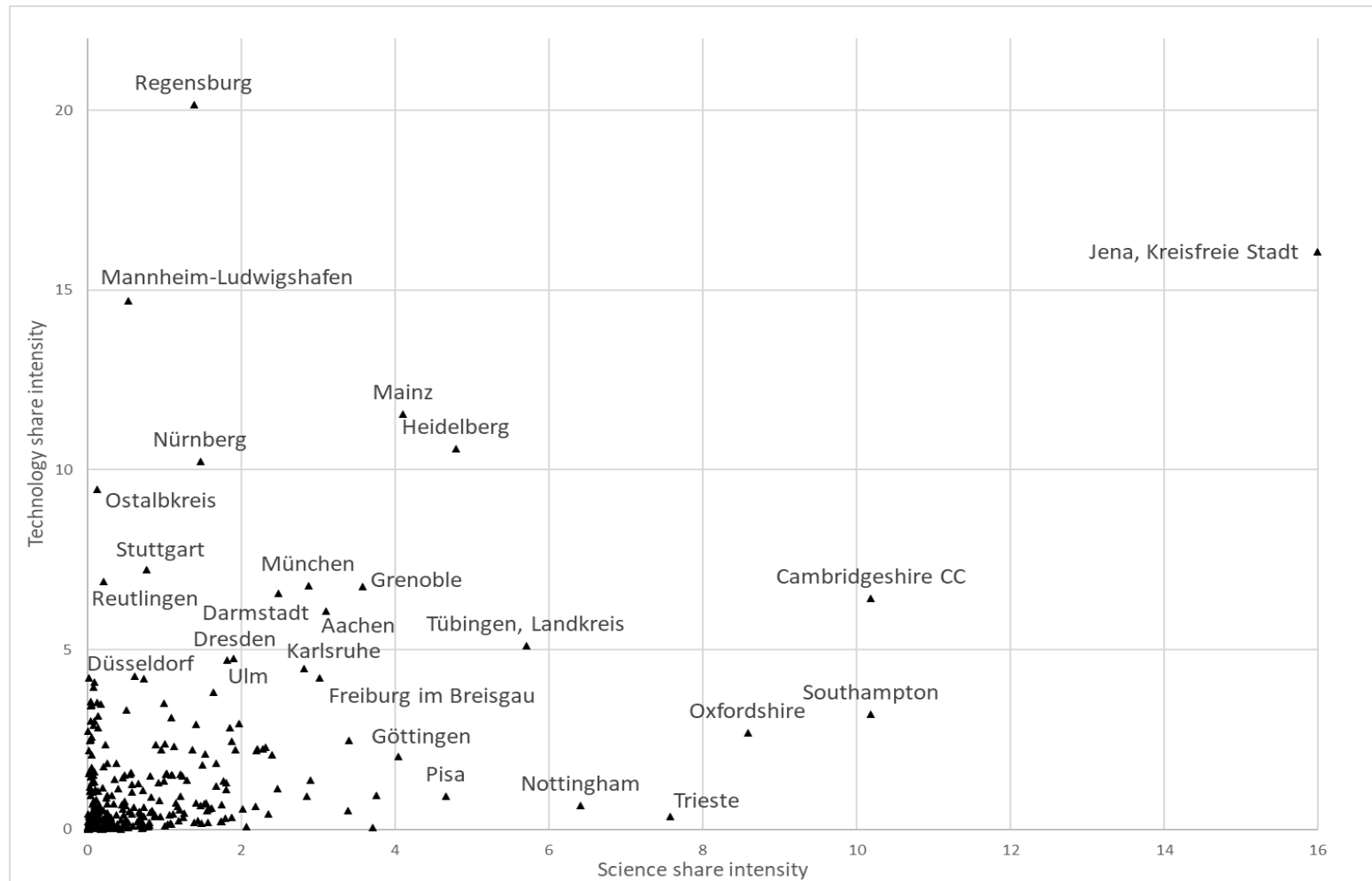
Results PCA

Variable	Science	Technology	Education	Rural	Unexplained
publications	0.412	0.000	-0.041	-0.015	0.033
top10% publications	0.418	-0.008	-0.081	-0.021	0.047
international co-publications	0.416	0.002	-0.076	-0.021	0.037
patents	0.033	0.676	0.004	0.006	0.009
transnational patents	-0.023	0.731	0.021	0.002	0.009
eu-fp participations	0.382	0.060	-0.070	-0.018	0.095
eu-fp coordinations	0.401	0.014	-0.077	-0.031	0.087
students ISCED5-7	-0.035	0.017	0.888	-0.048	0.090
students ISCED8	0.328	-0.065	0.240	0.010	0.218
area	-0.005	0.003	-0.029	0.983	0.012
population	0.257	-0.035	0.358	0.170	0.258

Four main components that account for 88% of the variance and can be labeled as

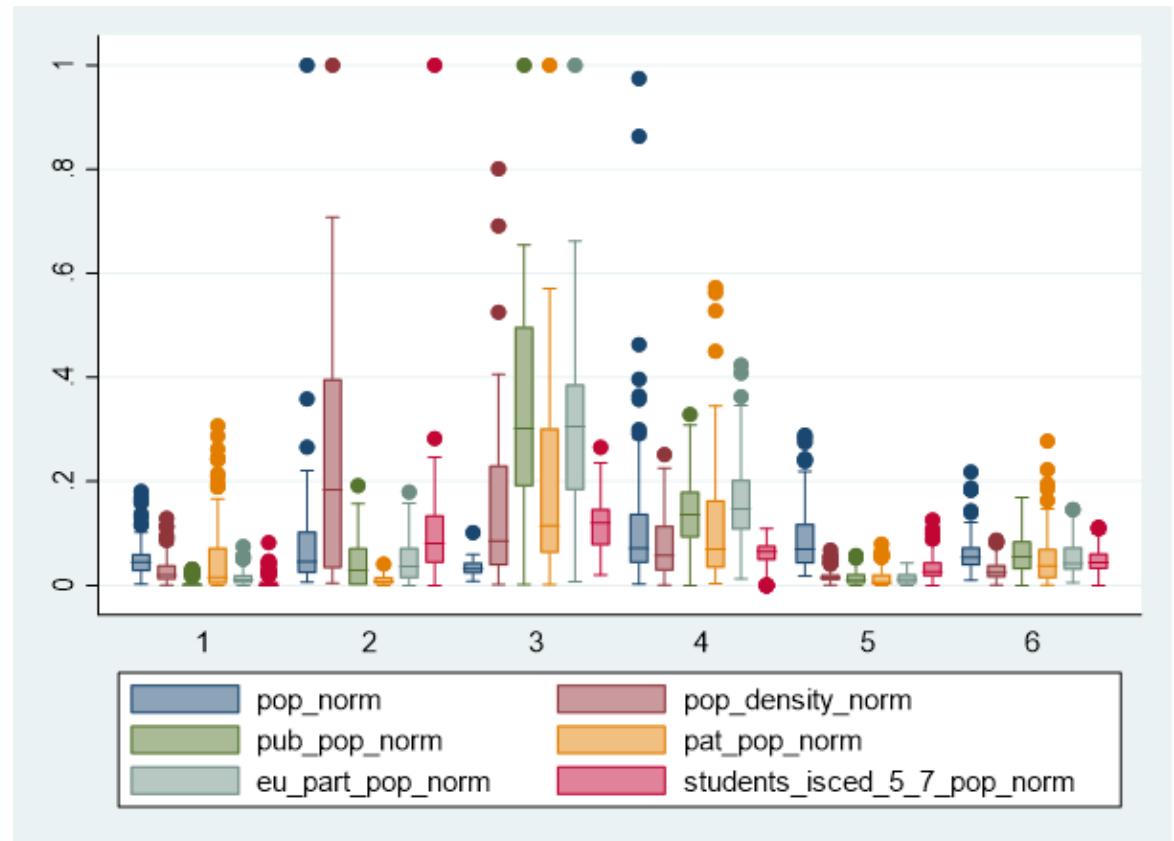
- science-based knowledge production (with high levels in publications and projects indicators)
- technological knowledge production (with high levels in patent indicators)
- education-based regions (with high number of students)
- rural regions (low levels for all indicators in large areas)

Regions: Technological or Academic?



Results Latent Class Analysis

- Class 1. Urban and rural, comparatively small, low intensity for science and education, but quite some patent activity (e.g. Prato, rural Veneto, Taranto with some industrial basis)
- Class 2. Urban high-density areas with strong education, some science and very low technology (mostly Eastern and Southern European cities, some UK, Istanbul, Naples)
- Class3: high intensity regions, such as Zurich, Cambridge, Oslo, Stockholm.
- Class 4. Large cities with high science basis (including London, Paris, Vienna, Rome, Milano, Turin, Bologna)
- Class 5. Urban and rural regions, with some education, but low science and technology (Italy: Palermo, Abruzzo, Molise, Basilicata)
- Class 6. Middle-science cities with some science basis and education (Italy: Verona, Parma, Bergamo)



Variables normalized by the population and rescaled [0,1]

Class characterization

Class	Name	Area	Density	Science	Technology	Education
1	Industrial regions	Large	low	very low	low	very low
2	Low-science metropolises	small	very high	low	very low	high
3	Knowledge hubs	small	high	very high	very high	high
4	Large metropolises	Large	high	high	high	high
5	Rural areas	Very large	low	low	very low	low
6	Middle-size cities	Large	low	low	low	low

Discussion

- Identify science, technology and education as distinct components of knowledge production of regions
 - Regions positioned differently in these dimensions
- Identify distinct groups of regions characterized by:
 - Size (urban vs. rural)
 - Intensity of knowledge production
 - Orientation towards education, science and technology
- Major advances
 - A broader understanding of knowledge production
 - Data availability thanks to data integration within the RISIS project
 - A regional classification tailored to knowledge production

Further extensions

- Use the whole knowledge output (beyond KET and SGCs)
- Include additional variables
 - Network centrality
 - GDP
 - R&D expenditures (based on imputation)
- Refine classification
 - And interpret in a more fine-grained way results

This work can also contribute to extent the KNOWMAK tool.



Future development of RISIS-KNOWMAK and open discussion

Benedetto Lepori and Thomas Scherngell

Brussels, 19 December 2019

From KNOWMAK to RISIS-KNOWMAK

- From 01.01.2020 the KNOWMAK tool will be maintained and further extended within RISIS
 - As a service providing indicators based on the core RISIS datasets
- Shows the core function of RISIS to integrate datasets, tools and services
 - Beyond short-term contracts
- Data on S&T are a public resources
 - Cannot be sustained without a public infrastructure
 - Does not exclude some 'premium' analytical services, but data should not be privatized

Integrated access

- In the future, RISIS will offer different types of access to data and services
- Fully open public data (for research)
 - Ready-made indicators through the KNOWMAK tool
 - Data on research organizations through OrgReg, FirmReg and the RISIS mirror of ETER
- Controlled access to micro-data
 - Depending on usage and licensing options

Planned developments (tbd)

- Refinement of the ontology
 - And extending to FP9 missions
- Covering the whole of S&T production (beyond KET and SGCs)
- Including data on higher education (from ETER)
- Updating the data yearly

These tasks can be integrated within RISIS with minor changes to the contract and budget:

- Strongly expanding usage of RISIS dataset and functioning as gateway to micro-data.

Further usage will hint to future extensions to be planned more long-term.

DISCUSSION