

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

## Regional knowledge production across Europe: Insights from combined indicators and an innovative regional delineation

*Thomas Scherngell, Benedetto Lepori, Massimiliano Guerini, Lionel Villard, Michael Barber*

**RISIS Research Seminars, December 15, 2021 (online)**



This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n° 824091

# Background and Motivation



- Creation of new knowledge essential basis for innovation, and thus, for the overall socio-economic development of regions or countries (Audretsch and Feldman, 2004)
  - Investigation of the geography of knowledge has become one of the main research domains in Economic Geography, but also in STI studies
  - Up to now, most studies concentrate on technological innovation activities while other facets are often underemphasized (see e.g. Moreno et al., 2006; Navarro et al., 2014; Verspagen, 2010)
  - Usually, regional breakdowns are used that intersect agglomerations of knowledge creation, leading to problematic interpretations in a spatial context
  - Alternatives proposed, such as Functional Urban Areas (Brezzi et al., 2012), provide a better delineation, but at the price of strongly reducing the scope of indicators available
- ***need to provide a richer, conceptually more in-depth empirical basis accounting for the multifaceted nature of knowledge production, and to come to a more meaningful and fine-grained regional classification system***

# Objective

Our main goal is to analyze the spatial distribution of knowledge production across European regions by focusing on

- (i) the volume of knowledge production (absolute and relative to the population), and
- (ii) the balance between a more ‘academic’ and a ‘technological’ component
- (iii) introducing a regional classification reasonable for analyzing knowledge production, but staying compatible for linking with external data

# Contents

1. State-of-the-art and regional classification
2. Data and indicators
3. Towards a new regional classification: NUTS adapted
4. Spatial distribution of regional knowledge production
5. Dimensions and regional types
6. Conclusions and Outlook



- High interest in empirically observing regional knowledge production, triggered by *Geography of Innovation* literature (Feldman 1994); by and large, two main challenges for advancing empirical research:
    - Multi-faceted, heterogenous and **complex nature of knowledge production** processes (Nowotny et al. 2003, Asheim et al, 2011)
    - Tracing information on knowledge production at finer and more **reasonable geographical breakdowns** than e.g. NUTS2 challenges data quality and geocoding algorithms
  - First research endeavours addressing both challenges at the same time over the past two decades (e.g. Innovation Scoreboard, Hollanders et al. 2009, or RIS3 monitor); mostly relying on aggregated data provided by statistical authorities (mostly NUTS2)
- Most recently (e.g. **RISIS**) new focus on **geocoding raw data** on knowledge production
- based on addresses of e.g. authors (publications), inventors (patents), research project participants, firms and research organisations, etc.,
  - to provide a richer picture of the geography of knowledge production at much more fine-grained and meaningful spatial levels (e.g. urban areas)

# Dimensions of knowledge production

- Vast amount of literature providing frameworks to categorize the heterogeneity of the knowledge production process (e.g. Asheim et al. 2005; Moodysson et al. 2008)
  - One basic distinction discussed extensively in the literature concerns
    - **explorative** (exploring new grounds in research; scientifically driven, academic settings)
    - **exploitative** (economic commercialization of new knowledge, technologically driven and application-oriented)
  - Knowledge production modes due to differences in underlying knowledge bases, research strategies, etc.
  - Also, strong heterogeneity of outputs, e.g. **quality** of new knowledge (so far strong focus on quantity)
- This study proposes an indicators framework that
- **accounts for these dimensions of heterogeneity** in grasping regional knowledge production (explorative, exploitative, quality / impact), and
  - applies these indicators to an **innovative set of European regions** by geocoding raw information on diverse kinds of knowledge production activities

# NUTS adapted: Towards a new regional classification for observing knowledge production



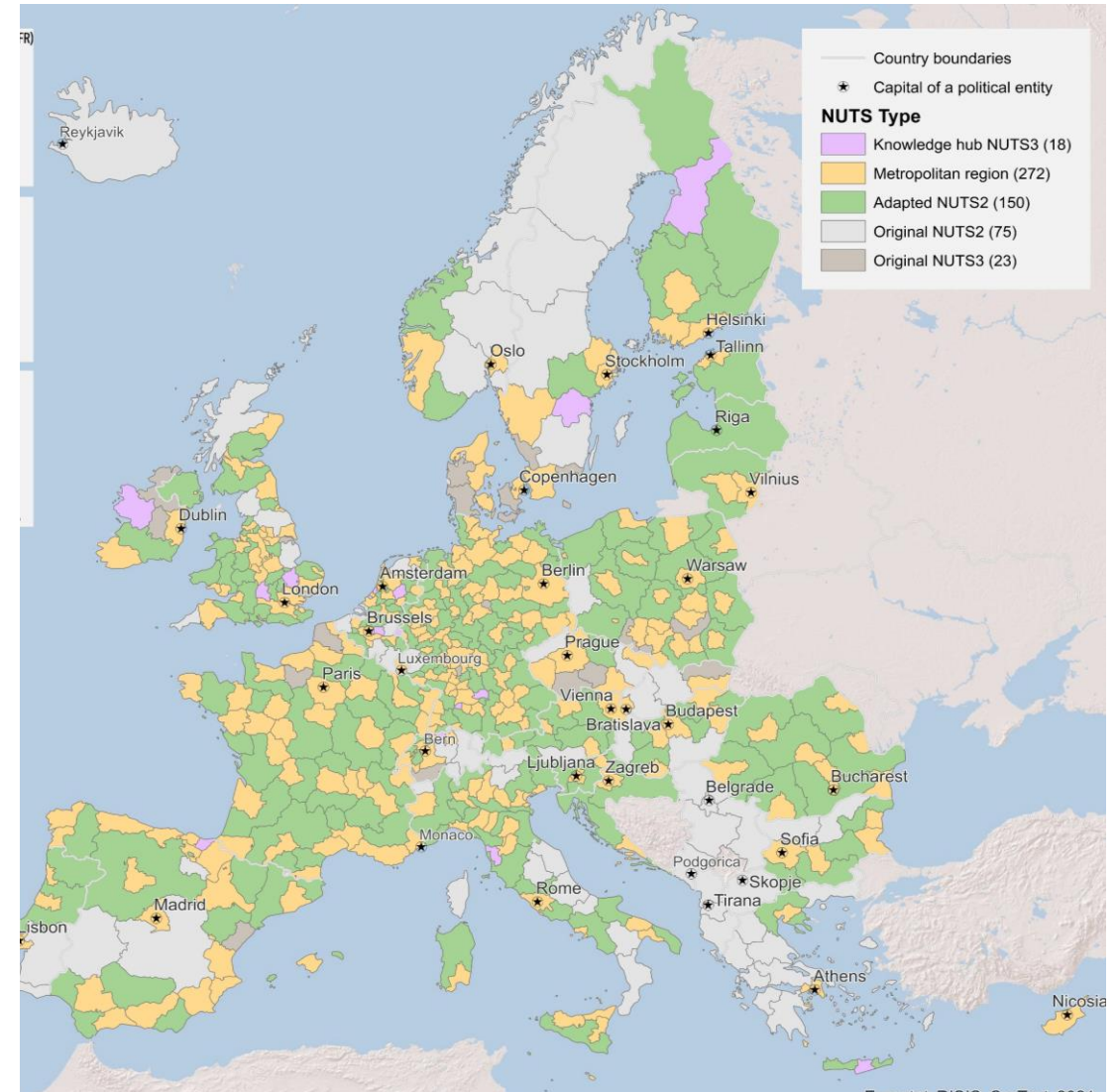
- Eurostat **NUTS classification as *de facto* standard** based on administrative boundaries; many well-known problems (only NUTS2 for main variables, very different levels of granularity between countries or different treatment of large cities)
  - Recent alternatives: identifying **functional regions** based on the characteristics of the phenomenon studied; example Functional Urban Areas (FUAs) for population; however, not compatible with NUTS
- We propose a regional classification that builds on **Eurostat metropolitan regions (MR)**
- Providing a **better delineation than original NUTS3**, remaining compatible with it, and still representing an approximation of functional urban areas
  - As of non-urban regions we **regroup all non-urban NUTS3 regions** within the same NUTS2 region in a single region
  - In addition, we single out a number of NUTS3 regions with **sizeable knowledge production** not included in the MR list as of their population
- Only possible by attributing information on different dimensions of knowledge production to the NUTS3 level using **RISIS geocoding facilities** (so far mostly just available at NUTS1 or NUTS2)

# Going for NUTS adapted: Comparing regional classifications

- NUTS adapted classification: 547 regions (in contrast to 1,441 NUTS3)
- NUTS3 regions have been grouped into 274 metropolitan regions as defined by EUROSTAT
- Nearly 1000 outside MRs have been grouped into 251 adapted regions (**drastic reduction of very small regions**)
- Effective for countries like Germany, with very many NUTS3 regions
- Knowledge hubs are smaller regions in terms of population and typically are prominent university centers (e.g., Jena, Pisa, Leuven, Cambridge, and Oxford)

→ **a far more uniform and consistent regional classification in terms of size**

**Metro areas quite similar to FUAs, i.e. keep functional approach but at the same time compatible with NUTS**





# The RISIS datasets used

## Explorative knowledge production

- **CWTS publication database** Full copy of Web of Science (WoS) with additional information on standardised organisation names and other enhancements
- **EUPRO**: Unique dataset providing systematic and standardized information on R&D projects of different European R&D policy programmes

## Exploitative knowledge production

- **RISIS Patent**: Enriched and cleaned version of the PATSTAT database, with a focus on standardised organisation names and geolocalisation

## Firm innovation

- **VICO**: A database comprising geographical, industry and accounting information on start-ups that received at least one venture capital investment in the period 1998-2014

## Inputs to knowledge production

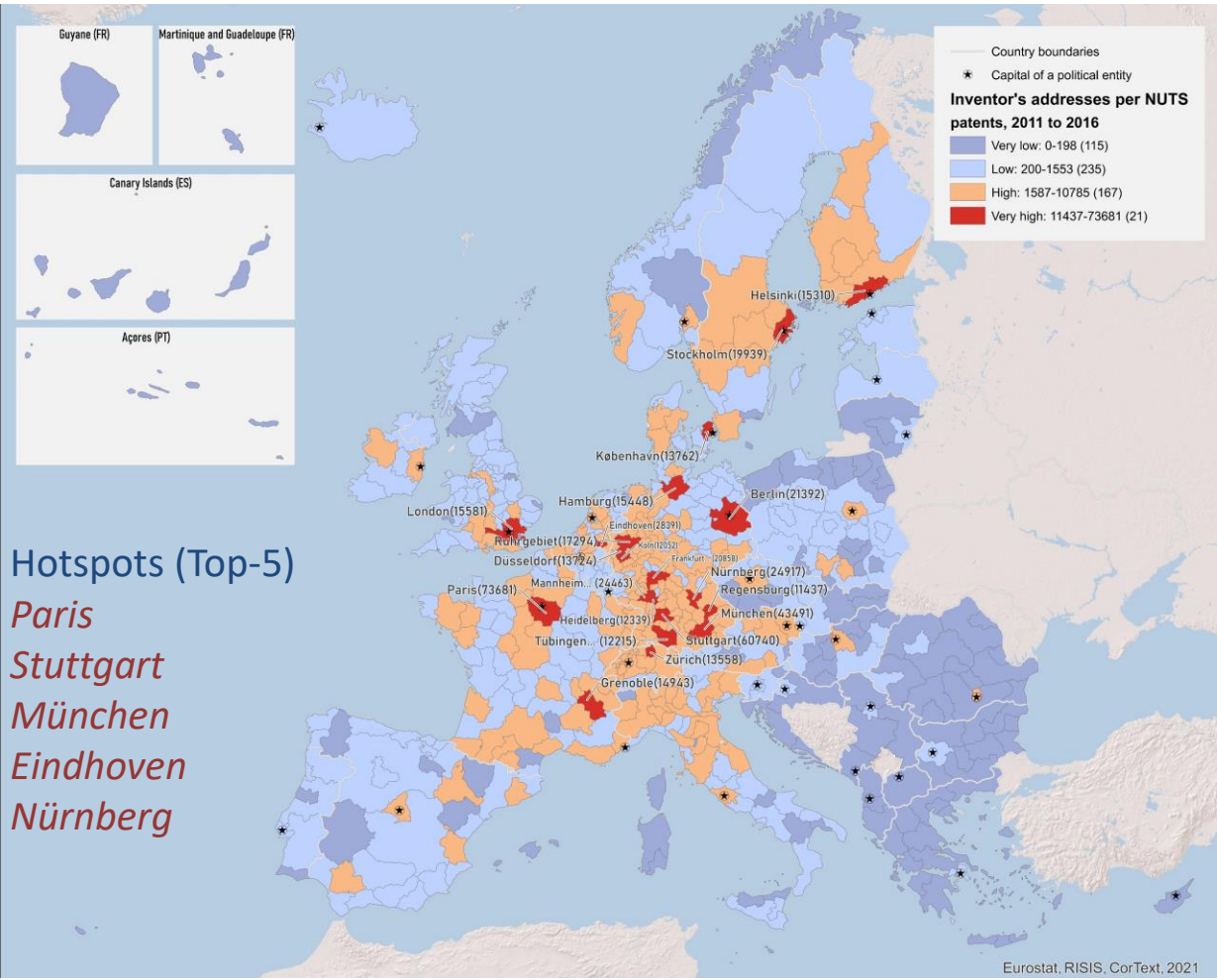
- **ETER**: Extension with indicators in terms of research activities of the European Tertiary Education Register database

**Other RISIS datasets for potential additions:** [rcf.risis2.eu](http://rcf.risis2.eu)

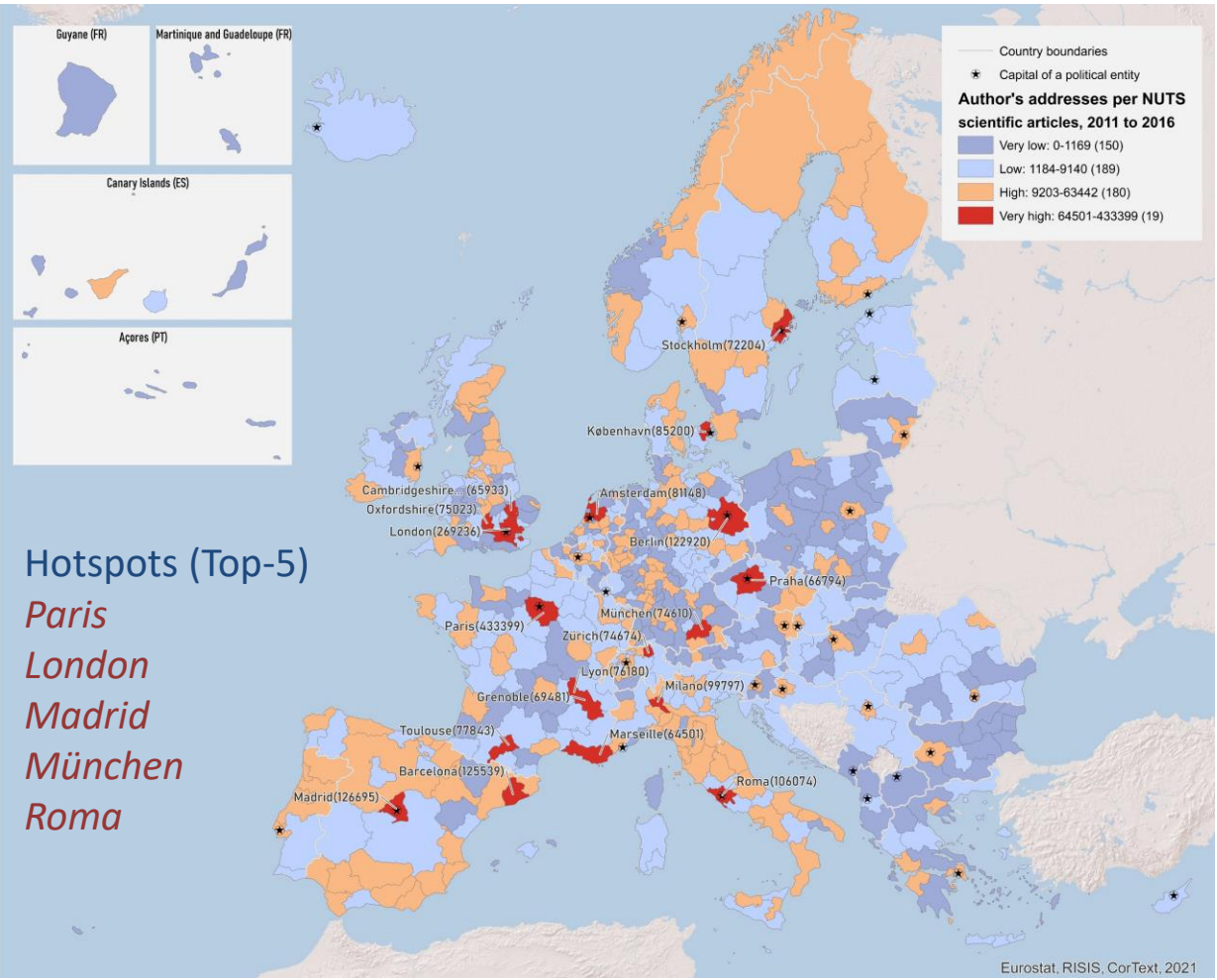
<i>Indicator</i>	<i>Definition</i>	<i>Source</i>
<b>N. of publications</b>	Count of publications from with at least one address in the region	<b>CWTS WoS database</b>
<b>N. of international publications</b>	Count of publications with at least one address in the region and at least one address outside Europe	
<b>N. of publications in the top 10%</b>	Count of publications in top10% cited with at least one address in the region	
<b>N. of participations to EU-FP projects</b>	Count of EU-FP projects (active) with at least one organization located in the region	<b>EUPRO</b>
<b>N. of coordinations to EU-FP projects</b>	Count of EU-FP projects (active) with the coordinating organization located in the region	
<b>N. of priority patent applications</b>	Count of priority patents with at least one inventors' address in the region	<b>RISIS Patent</b>
<b>N. of transnational priority patent app.</b>	Count of priority patents that have been extended in at least another patent office	
<b>N. of students ISCED5-7 (ETER)</b>	Students enrolled at level 5 (diploma), 6 (bachelor) and 7 (master) of the ISCED	<b>ETER</b>
<b>N. of students ISCED 8 (ETER)</b>	Students enrolled at level 8 (PhD) of the ISCED	<b>ETER</b>
<b>N. of start-ups</b>	Start-ups that received at least one venture capital investment in the period 1998-2014	<b>VICO database</b>
<b>Total population</b>	Number of inhabitants established in the region.	EUROSTAT
<b>Gross Domestic Product</b>	Regional Gross Domestic Product in Purchasing Power Parities	EUROSTAT
<b>Area</b>	Regional surface in square kilometers	EUROSTAT

\*time period: aggregated counts from annual observations 2011-2016, startups 2014-2016

# Geography of knowledge production: number of publications and patents



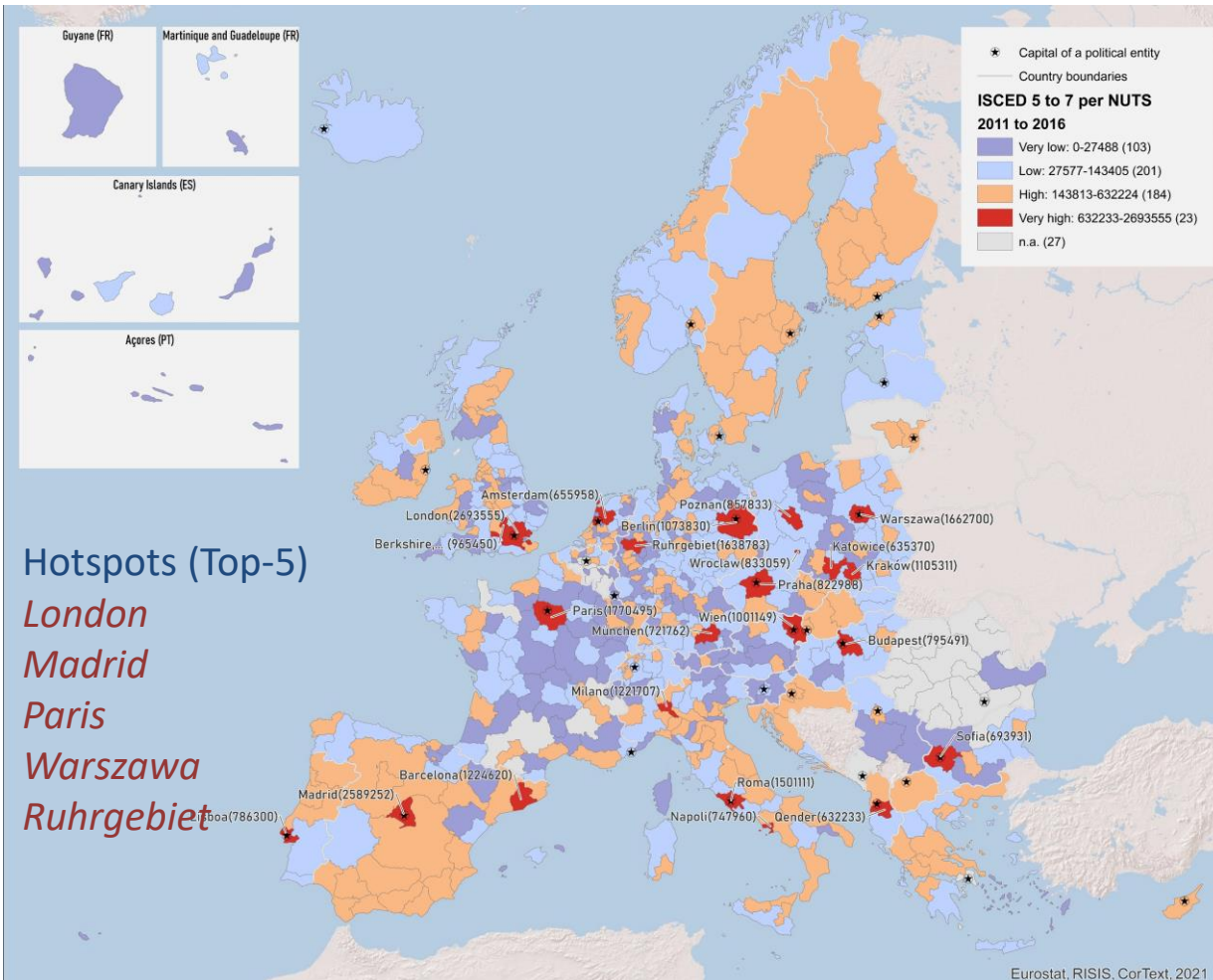
Patents



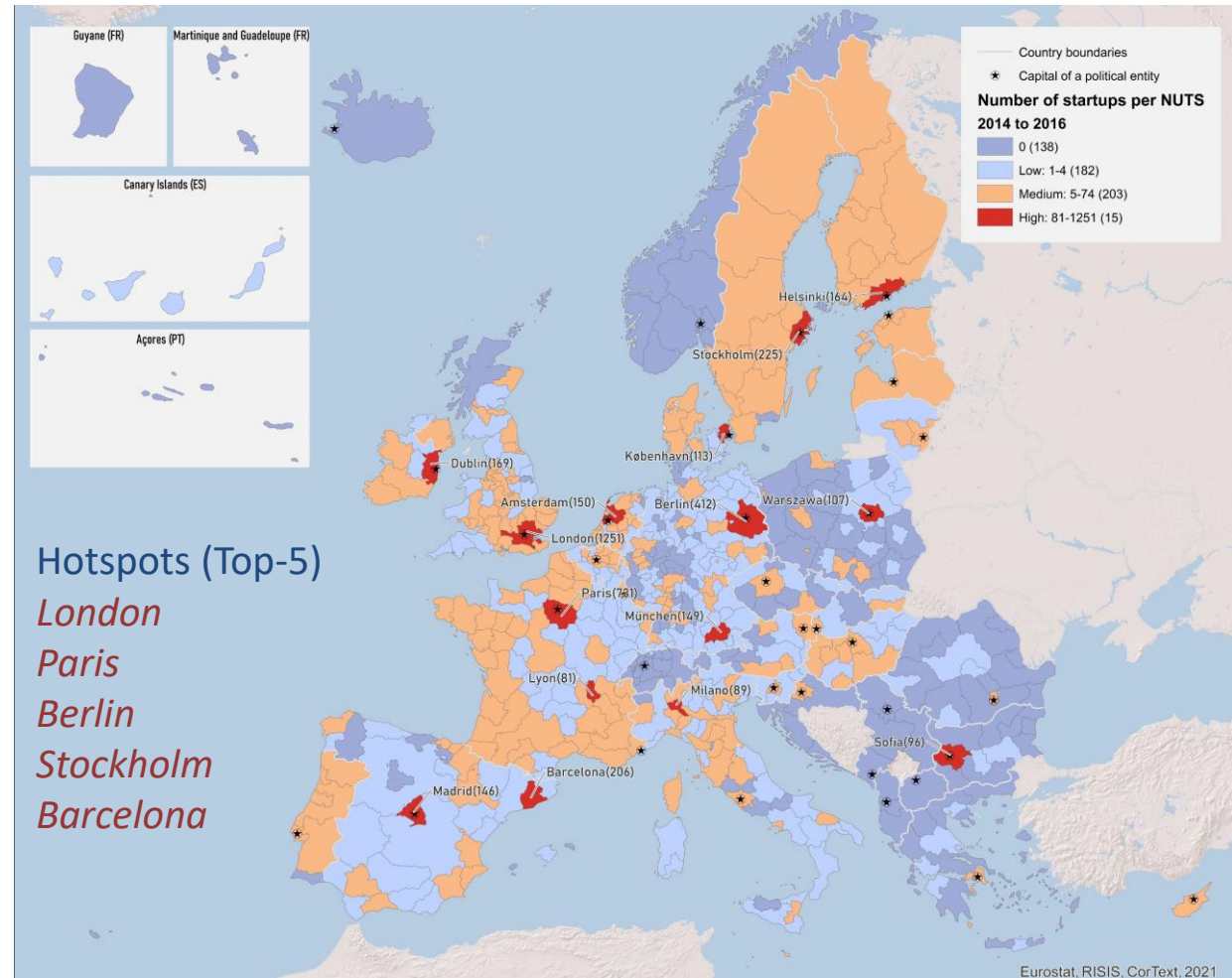
Publications

# Geography of knowledge production: number of students and startups

# RISIS



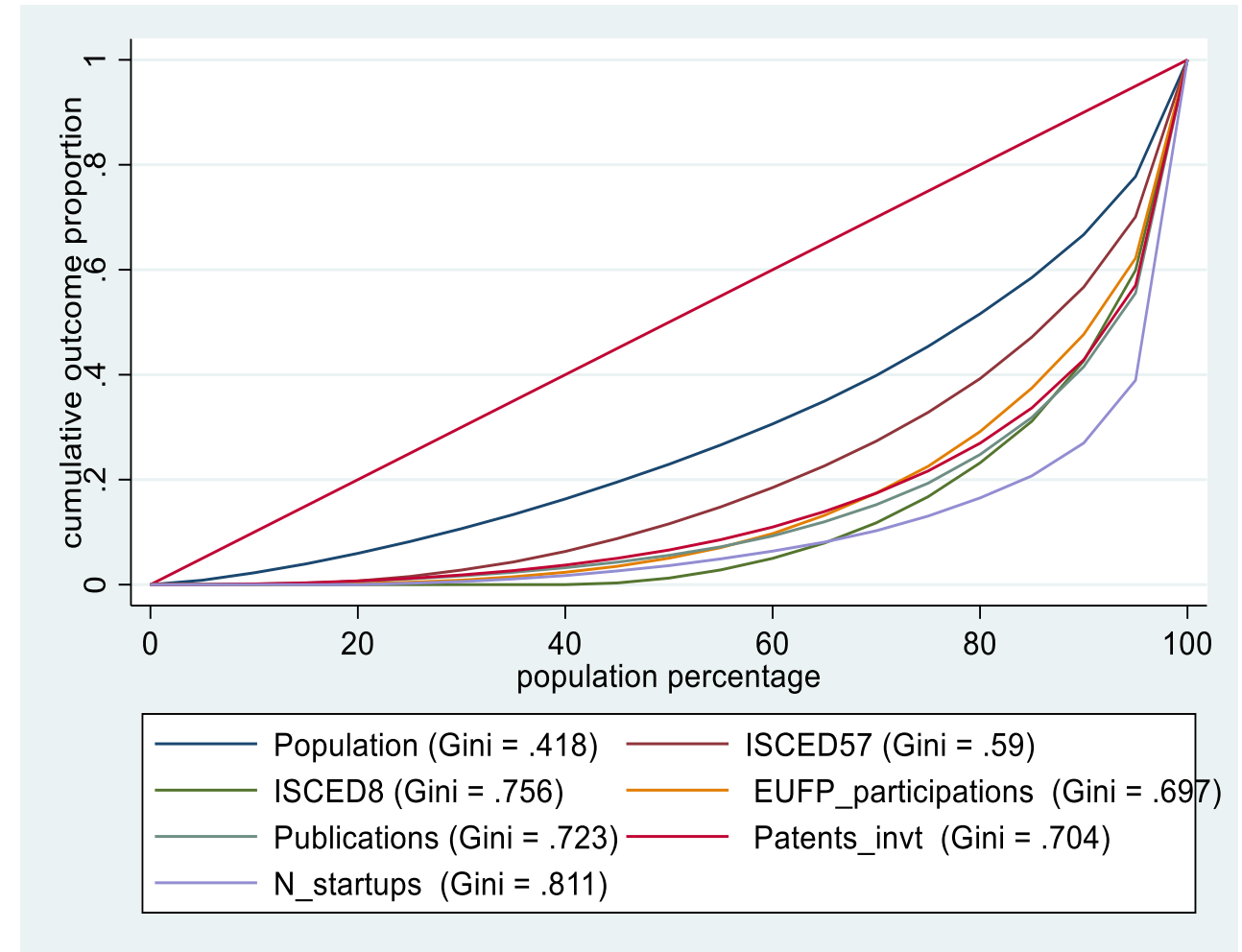
Students (ISCED5 to 7)



Startups

# Regional concentration

- High concentration of knowledge production
- More than 80% of knowledge produced by top 20% regions
- Increases when moving from inputs (HR) to innovation (patents, startups)
- Higher concentration of publications than of EU projects
- VC backed Startups: a dozen of regions accounting for 50%
- Knowledge hubs: 2% of population, but more than 6% in explorative production (fine grained breakdown necessary!)



# What about spatial spillovers?

- An old discussion in the Geography of innovation literature: Can we identify spatial clusters of innovation as an expression of spatial spillover and/or agglomeration effects?
- How to identify them statistically? Instruments from spatial statistics:
  - Moran's I statistic measuring the global spatial dependence (i.e. degree of spatial autocorrelation or spatial clustering of a specific variable  $x$  in a system of regions  $i, j = 1, \dots, n$ , defined as :

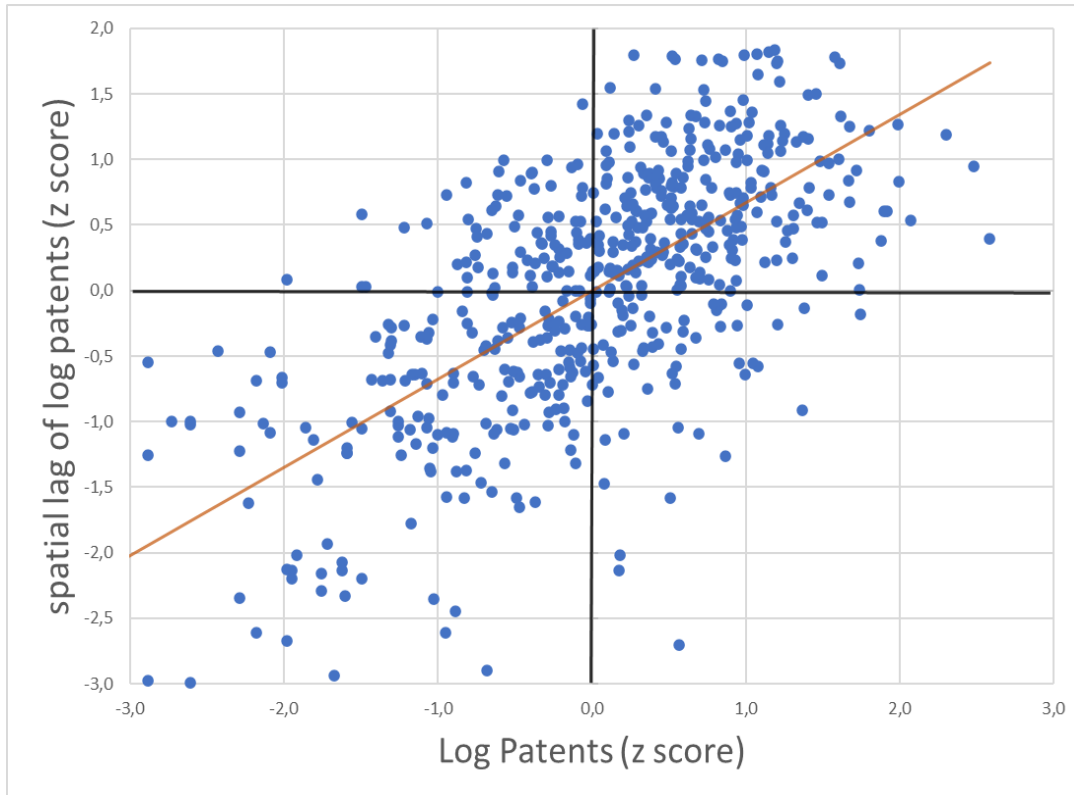
$$I = \frac{n}{\sum_{i=1}^n \sum_{j \neq i}^n W_{ij}} \frac{\sum_{i=1}^n \sum_{j \neq i}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{with} \quad W(ij) = \begin{cases} 1 & \text{if } i, j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases}$$

- Range [-1, 1]: Values significantly above 0 indicate positive spatial autocorrelation
- Spatial lag ( $Wx$ ) is the weighted sum or, when  $W$  is row-standardized, the weighted average of the neighboring values for that variable

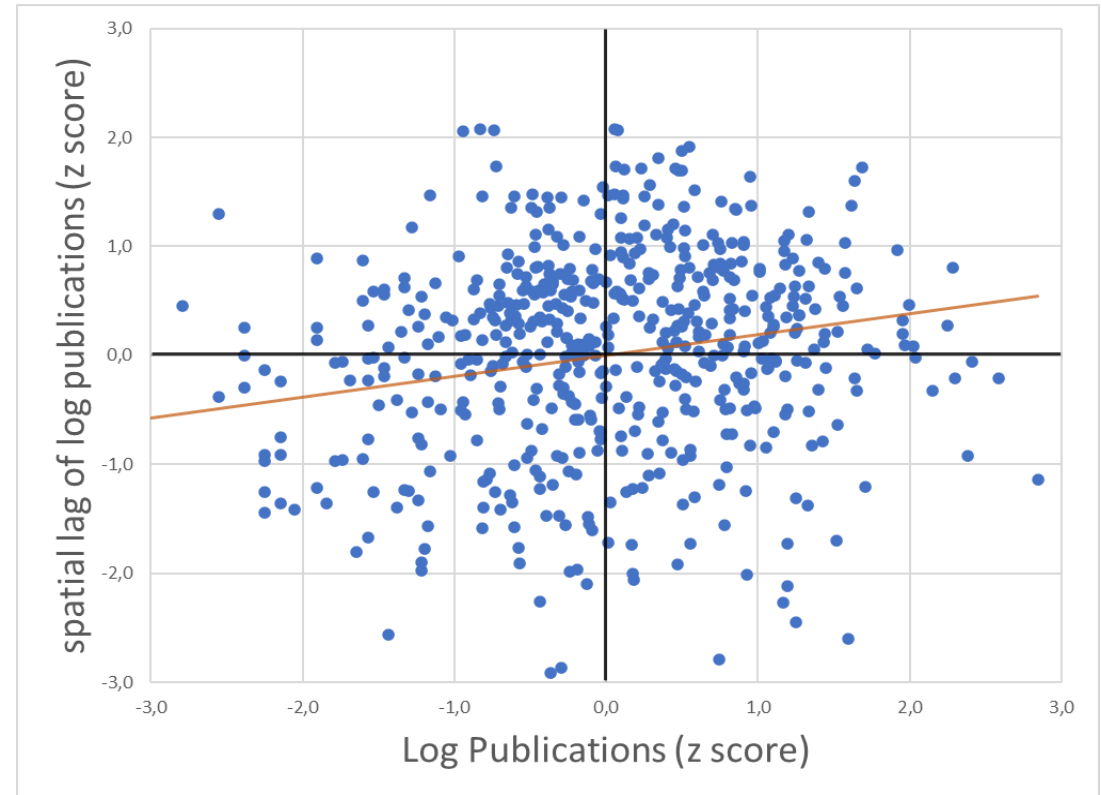
# A spatial dependence structure?



Technological Knowledge Production



Scientific Knowledge Production



Spatial autocorrelation analysis with Moran's I (using k 5-nearest neighbours and contiguity (Queen 1))

	Moran (K5)	p-value	Moran (Queen)	p-value
<b>Patents</b>	<b>0.58</b>	0.001	<b>0.63</b>	0.001
<b>Publications</b>	<b>0.17</b>	0.001	<b>0.24</b>	0.001
<b>Startups</b>	0.07	0.018	0.11	0.006
<b>Students</b>	0.02	0.229	0.05	0.034

# Spatial clustering tendencies

<p><b>Technological knowledge production (patents)</b> Highest clustering tendencies; neighbouring regions of regions with a high number of patents also tend to have more patents than the average</p>	High
<p><b>Scientific knowledge production (publications)</b> Lower but still significant spatial clustering; in comparison to patenting, knowledge production is more bounded within regions (but more related to other scientific knowledge producers located further away)</p>	Moderate
<p><b>Tertiary education</b> Very low spatial clustering, i.e. the number of students in a region is not depending on the number of students in neighbouring ones</p>	Low
<p><b>Startups</b> No spatial clustering; the number of startups is unrelated to the number of startups in neighbouring regions; strongly bounded in very specific regions, no spatial spillovers</p>	None



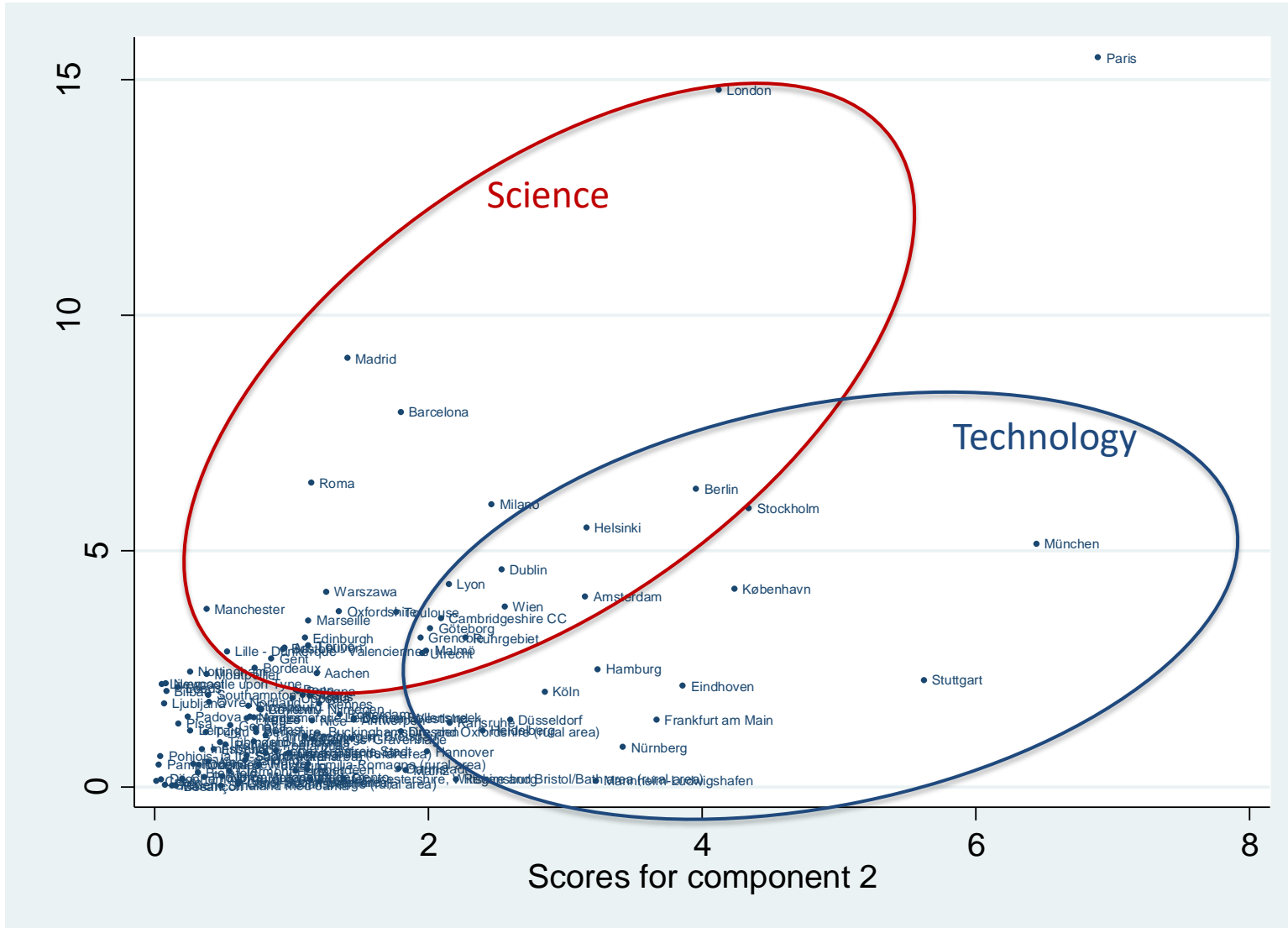
# Different region types?

- Principal component analysis (square root transformations): 3 components retained (82% of variance explained)
- Rotated components (varimax):
  - Comp1: related to scientific knowledge production and entrepreneurship (start-ups)
  - Comp2: related to the local level of innovativeness and competitiveness
  - Comp3: large less populated areas
- Comp1 and Comp2 positively correlated (interplay between science and technology production); Comp3 negatively correlated with the others

Variable	Comp1	Comp2	Comp3
N. of start-ups	<b>0.39</b>	0.15	0.09
N. of participations to EU-FP projects	<b>0.45</b>	0.07	0.02
N. of priority patent applications	0.09	<b>0.62</b>	0.11
N. of publications	<b>0.40</b>	0.19	0.02
N. of students ISCED5-7	<b>0.48</b>	-0.09	0.01
N. of students ISCED 8	<b>0.47</b>	-0.24	-0.11
GDP / Total population	-0.08	<b>0.70</b>	-0.12
Total population / Area	0.12	-0.03	<b>-0.66</b>
Area	0.09	-0.06	<b>0.72</b>

	Comp1	Comp2	Comp3
Comp1	1.00		
Comp2	0.50	1.00	
Comp3	-0.13	-0.20	1.00

# Concentration of knowledge production but regional types diverge



- Positive correlation between science (comp1) and technology (comp2) production,
- but heteroskedastic relation!
- Paris and London, with Paris in particular being top in both
- German and Netherlands regions with high technology but moderate science production (Munich, Stuttgart, Frankfurt, Eindhoven)
- Large capital cities (Rome, Madrid) with high science production but moderate regional innovativeness

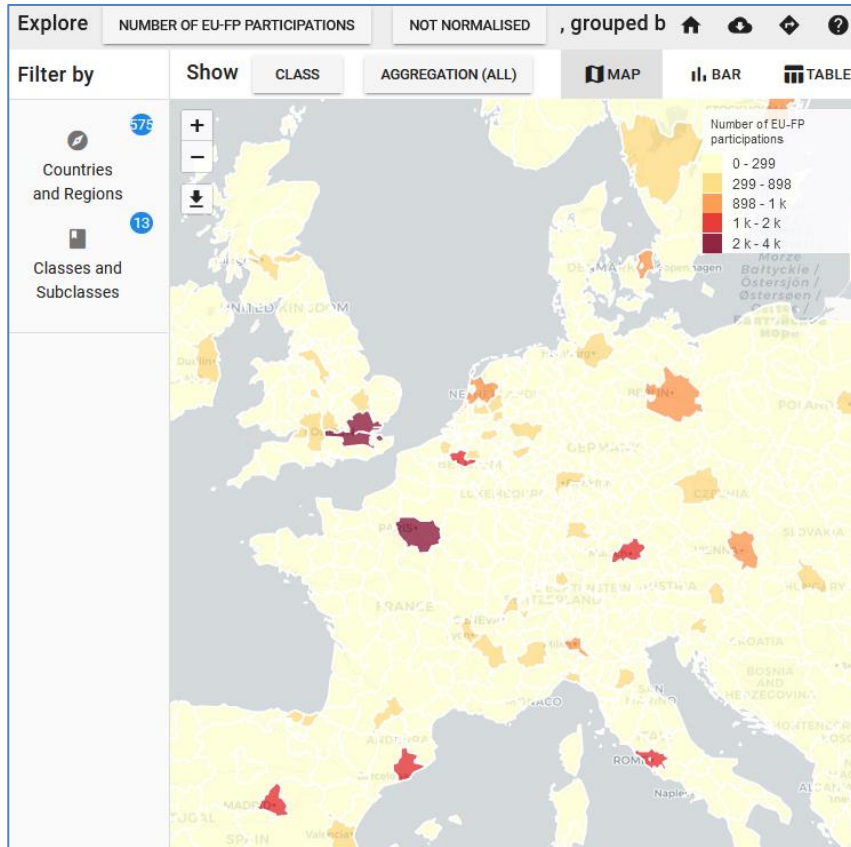
- Two major challenges for empirical investigation of the geography of innovation
  - multi-facted nature of knowledge creation
  - observation at meaningful and sufficiently fine-grained spatial levels
- The study presents a potential pathway to address both challenges by mobilizing and integrating new datasets and services available developed via the RISIS infrastructure
  - New indicators (e.g. students, startups) as well as traditional ones (e.g. patents, publications, projects) but geocoded from raw, organisation-level information on knowledge production
  - Applied to a new, more meaningful geographical breakdown, combining metropolitan regions with adapted NUTS regions, reaching a more balanced system of spatial entities
- Preliminary analytical exercises started, illustrating the differing spatial distribution of different indicators for knowledge production
- Strong spatial concentration, in particular for exploitation oriented knowledge production and innovation (startups); different hotspots for different knowledge production types
- Spatial clustering tendencies more robustly identified than in previous research, showing high spatial dependence for patents, while moderate for science, and none for startups and students
- Identification of three main regional types: 1) Knowledge exploring regions with high volume of education; 2) Knowledge exploiting regions (patents) 3) Rural and lagging regions

# Concluding remarks and outlook

- This work may provide an important entry point for a **new era of research** on the *geography of innovation*,
  - mobilizing **geocoded micro data** of different aspects of knowledge production,
  - applied to a **regional breakdown** suitable for knowledge production while staying **compatible** with regional statistics provided by Eurostat
- Future research needs to be targeted towards more specific questions and needs to shift from descriptive to an **explanatory** and **dynamic** perspective
- For a more direct connection to **policy debates** (e.g. smart specialisation), the indicator framework needs to be expanded, e.g. in terms of **regional knowledge complexity** or **technological relatedness** measures

# Directions

# RISIS



**Public access to Indicators on knowledge production**

**Knowmak.eu**

Access to RISIS datasets listed below, is free of charge and is offered through a two-step process:

- Accreditation: researchers need to register and agree on the conditions of use (good use of data, authorship, mentions to the RISIS project, agreement of posting results and aggregated datasets produced on RISIS website) via signing the [RISIS Code of Conduct](#).
- Selection: researchers need to propose a project (200 words) based upon the mobilisation of one or more datasets. Projects are reviewed both by the dataset producers and by the RISIS project review board that will give the final agreement for access.

All datasets can be accessed by [Registering](#) or [Login in](#), selecting the needed Dataset(s) then the **Continue Access** option, intending distant access to the selected dataset, or a physical visit to the location of the dataset owner, with costs covered by RISIS ([reimbursement form](#)).

**CHEETAH**

Cheetah is a database featuring geographical, industry and accounting information on three cohorts of mid-sized firms that experienced fast growth during the periods 2008-2011, 2009-2012 and 2010-2013

**Open access to underlying raw micro-data**

**risis2.eu;  
<https://rcf.risis2.eu/datasets>**

**Public access to organisation level indicators production**

**register.orgreg.joanneum.at  
eter-project.com**

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

# THANK YOU !

[CONTACT@RISIS2.EU](mailto:CONTACT@RISIS2.EU)



[@RISIS\\_EU](https://twitter.com/RISIS_EU)

[FACEBOOK.COM/RISIS.EU](https://FACEBOOK.COM/RISIS.EU)



[RISIS2 EU PROJECT](https://RISIS2.EU/PROJECT)

