RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

EXPLORING KNOWLEDGE PRODUCTION IN EUROPE. THE RISIS-KNOWMAK TOOL

RISIS Policy Brief Series

Issue # 2 | December 2019

risis.communication@ircres.cnr.it

CONTENTS

PRESENTATION

The KNOWMAK on-line tool provides a synthetic view on knowledge production in Europe, approximated through indicators on scientific knowledge (captured by publications), technological knowledge (patents), and project-based knowledge (European R&D projects). The tool allows to visualize and download about fifteen indicators on knowledge production constructed around three integrative dimensions, i.e. topics, geographical spaces and research organizations.

A core element of KNOWMAK is an ontology of knowledge production structured around Key Enabling Technologies (KET) and Societal Grand Challenges (SGC). The ontology allows providing indicators to answer policy questions, such as which are the most active regions/countries and research organizations in selected topics, both for what concerns scientific development (based on publications) and technology (based on patents).

By selecting their topics of interest and a geographical space (country or region), users can visualize on-line maps of Europe that display hot spots of knowledge production. Indicators cover quantity and quality of 'classical' knowledge production (scientific publication output and impact, patenting activities, participation to European Framework Programs). Further, users can identify the most active research organizations (public and private) and access a selection of social innovation projects. Finally, indicators measuring the broader societal outreach of knowledge production are provided, as captured by the number of publications tweeted and in open access.

The KNOWMAK tool will be further maintained and extended within the RISIS infrastructure project; a core development will be the introduction in the ontology of the 'mission' areas in the Horizon Europe program, such as cancer, smart cities and healthy ocean. This will allow RISIS-KNOWMAK to support with empirical evidence the development of the current European research policy.

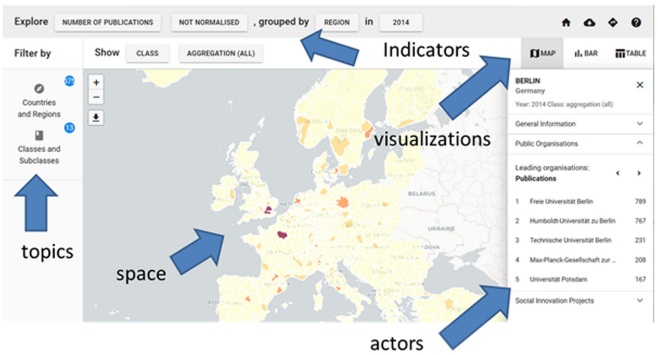
1. INTRODUCTION TO KNOWMAK

KNOWMAK is a free service that can be accessed on-line at www.knowmak.eu. The KNOWMAK dashboard (see Figure 1) allows the interactive selection of the knowledge production indicators, of their aggregation level (topics and geographical space), as well as of the year and of the type of visualization (maps, charts, table with the numbers). Further, it is possible to perform a more fine-grained selection of the topics of interest at two levels (6 KETs and 7 SGCs with 135 subclasses).

Additionally, for each country or region, a factsheet displays the most important public and research organizations for the topics selected (by number of scientific publications, patents and European projects), as well as a selection of social innovation projects pertaining to that region and topic.

Finally, users can download the indicators for their selection of spaces and topics in a csv format for further analysis.

Figure 1. The KNOWMAK dashboard



2. METHODOLOGY. THE KNOWMAK ONTOLOGY

An important issue in science and policy studies is to be able to analyze knowledge production by subject domains and, especially, to associate it with topics of political interest. However, established classification systems, such as the Frascati Manual fields of Science (OECD, 2015) and IPC codes for patents (Debackere and Luwel, 2005) are too rigid and incommensurable across data, while data-driven methods are too fine-grained to provide broader representations of knowledge production and are difficult to connect to political priorities (Cassi et al., 2017).

KNOWMAK focuses on a subset of scientific and technological domains, as identified by the Key Enabling Technologies (KET) and Societal Grand Challenges (SGC)

identified by the European Commission. They correspond to a sizeable share of knowledge production in Europe (for example about half of all scientific publications). Based on documentary analysis and expert knowledge, the 13 KET and SGCs have been further decomposed in 135 subtopics.

For example, the KET nanoscience and technology, which is defined as the branch of science that studies systems on the nanometer scale, has been disaggregated in ten subclasses, i.e. computational nanotechnology, DNA nanotechnology, food nanotechnology, graphene, nanobiotechnology, nanomedicine, nanoscale devices, nanoscale materials, nanoscience techniques and nanotoxicology (Maynard et al., 2019).

Each class and subclass was then populated with keywords, partly from documentary analysis, partly enriched through automated methods such as word embeddings. The current version of the ontology includes more than 6,000 unique keywords. Finally, documents, such as patents and publications, have been attributed to classes and subclasses based on the (relative) frequency of keywords in their textual content.

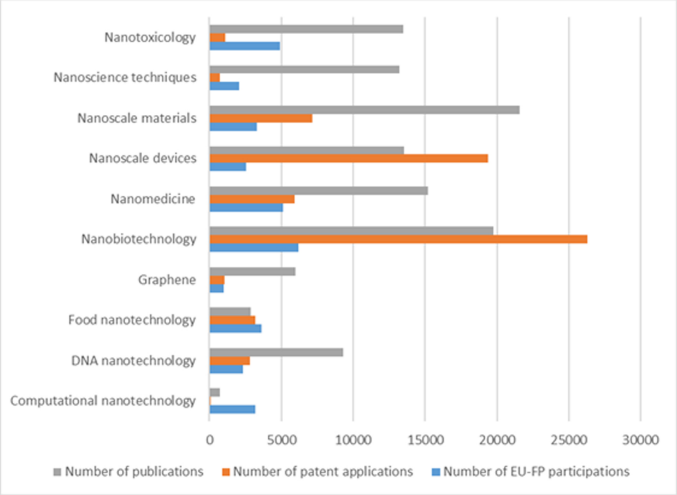
Moreover, the indicators on knowledge production were attributed to regions based on the addresses of their authors (publications), participating organizations (EU-FP projects), and inventors (patents) respectively. In this respect, KNOWMAK developed an adapted regional classification that is based on the EUROSTAT nomenclature of statistical territorial units (NUTS) at NUTS3-level. The adapted classification treats in a uniform way large metropolitan cities, such as London and Paris, and allows the identification of non-metropolitan regions by considering regional units in which metropolitan cities are excluded from the original NUTS regions.

This allows for a fine-grained view of knowledge production, which is connected to topics of policy interest and covers both scientific and technology production. As shown in Figure 3, out of the ten subclasses of nanoscience and technology, scientific output is concentrated in five classes, while technology production, as measured by patents, is heavily concentrated in three classes only, i.e. materials, devices and nanomedicine. On the contrary, European projects are more distributed across classes.

This analysis can lead to important insights on the complementarities between science and technology (Bonaccorsi, 2008), but also on the specific role of European funding to fill in gaps in the scientific and technological basis.

Further, by using the KNOWMAK tool, the analysis can be refined by looking to specialization by topics of regions and to the emergence of hot spots, i.e. regions and/or research organizations strongly specialized in a specific area.

Figure 2. Knowledge production in nanoscience and technology, 2010-2014



3. ANALYZING REGIONAL KNOWLEDGE PRODUCTION

The geography of knowledge production is highly complex. Regions are very heterogeneous in terms of their size, population and wealth; moreover, there are strong differences in the intensity of knowledge production, in the orientation towards science vs. technology and, finally, in the specialization in different topics. The KNOWMAK on-line tool allows for the first time a systematic analysis of such differences (see Lepori et al., 2019; Figure 2).

To this aim, KNOWMAK has developed two composite indicators of knowledge production. The composite indicator of knowledge production volume is the average of the share of projects, publications and patents associated to a region. The composite indicator of knowledge production intensity is defined as the ratio between the composite indicator of knowledge production volume and the population of the region. The reference year for the data presented in this analysis is 2013.

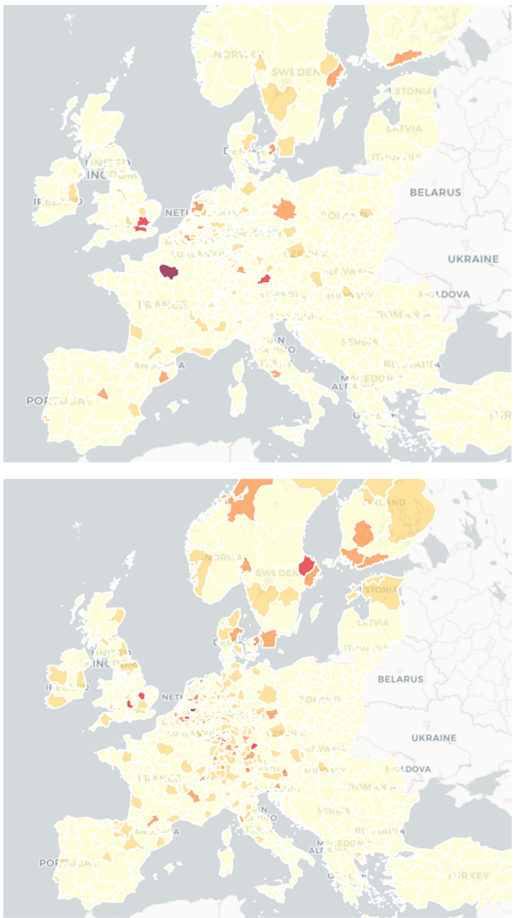
Using these indicators, we show that the regions with higher knowledge production volumes are mostly concentrated in large metropolitan regions, with Paris (in France), London (in the UK) and Munich (in Germany) that rank in the first three positions. The distribution of the volume of knowledge production appears highly skewed, with the first 10 regions (mostly large metropolitan regions with a population higher than 2M inhabitants) that account for more than 20% of the European total. Among these regions, only Munich ranks in the top

10 regions as to intensity. Paris and London, which are by far the most important regions in terms of volume, are only average in terms of knowledge production intensity. Their position is therefore largely accounted for by their sheer demographic and economic size.

On the other hand, some small- to mid-sized regions, such as Eindhoven, Leuven and Uppsala, produce more knowledge per capita than their mega-city counterparts London, Paris and Berlin, while still having rather large volumes of knowledge production. This emphasizes the important role of medium-size regions that is likely to emerge even more clearly when analyzing specific research domains.

Non-metropolitan areas typically exhibit lower levels of knowledge production, except when they include university cities, like Cambridge, Oxford, and Zuid-Holland (Leiden and Delft). Such areas are also characterized by a high level of knowledge production intensity. Finally, Eastern European countries are characterized by low volumes of knowledge production, with the exception of large capital cities like Prague, Warsaw, and Budapest. Production intensity of regions in Eastern Europe is generally lower.

Figure 2. Regional indicators of knowledge production volume (top) and intensity (bottom)



4. TOWARDS A TYPOLOGY OF KNOWLEDGE PRODUCTION REGIONS

RISIS-KNOWMAK allows overcoming two relevant shortcoming of the current literature on regional innovation, i.e. a) the focus on patent data and innovation indicators and the lack of information on the public knowledge base and on human capital formation and b) the use of regional breakdowns (such as NUTS2 for the European case) that often intersect agglomerations of knowledge creation, leading to problematic interpretations in a spatial context.

Based on the different indicators available in KNOWMAK, characteristic groups of regions in terms of their knowledge production activities can be derived. The analysis has been performed by considering indicators of publication output, patents, EU-FP participation and education at the higher education level, as well as regional statistics on area and population from EUROSTAT.

The NUTS-adapted classification introduced in KNOWMAK was adopted. Using different classification methods (principle component analysis, latent class models, see Lepori et al., 2019) allowed identifying six classes of regions based on the level and orientation of knowledge production (see Table 1):

- Class 1 are urban and rural regions, comparatively small, with low intensity for science and education, but some patent activity as associated with a strong industrial basis (e.g. Prato, rural Veneto, Lorraine region).

- Class 2 are high-density urban areas, mostly Eastern and Southern European cities such as Istanbul and Napoli, with strong educational activities, some science and low level of technology production.

- Class 3 high intensity regions in all dimensions of knowledge production, such as Zurich, Cambridge, Oslo, Stockholm.

- Class 4 are large cities with large science basis, including London, Paris, Vienna and Rome.

- Class 5 includes regions with some educational activities, but low science and technology, such as Castilla-La-Mancha in Spain and Calabria

- Finally, class 6 includes middle-size cities with both a science basis and education activities, such as Bremen, Bordeaux or Brno.

Table 1. Regional classification

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

Despite the preliminary nature of these results, this analysis shows how the rich availability of ready-made indicators in KNOWMAK allows for advancing the studies of regional knowledge production. An important future extension of these studies will be a more fine-grained analysis of the regional specialization in different topics of knowledge production by using the KNOWMAK ontology.

5. BEYOND CLASSICAL KNOWLEDGE PRODUCTION

A limitation of many available indicators in S&T is that they are focused on the knowledge being codified and recognized by the scientific community. While they are relevant to map the existing science and technology basis, most users require on-time indicators on emergent topics and actors (knowledge in the making). Furthermore, a focus on knowledge developed within structured R&D contexts does not allow for the tracking of broader processes of social innovation, constituting a core dimension of the new modes of knowledge co-creation. There is therefore a need for indicators capturing also the broader societal outreach of knowledge production (Zahedi et al., 2014).

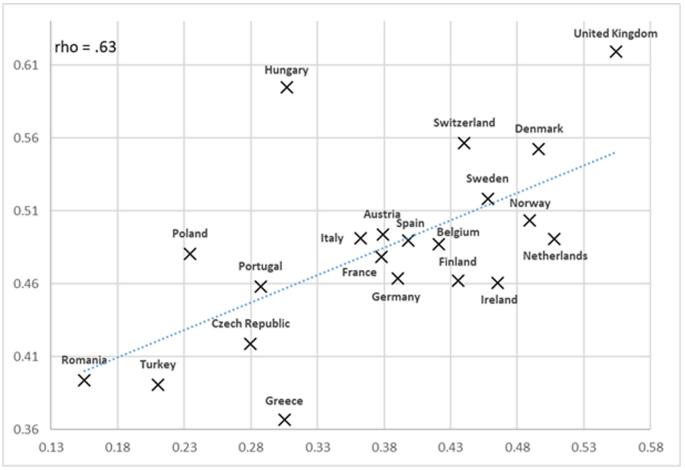
Therefore, the KNOWMAK tool provides some experimental data on three other dimensions:

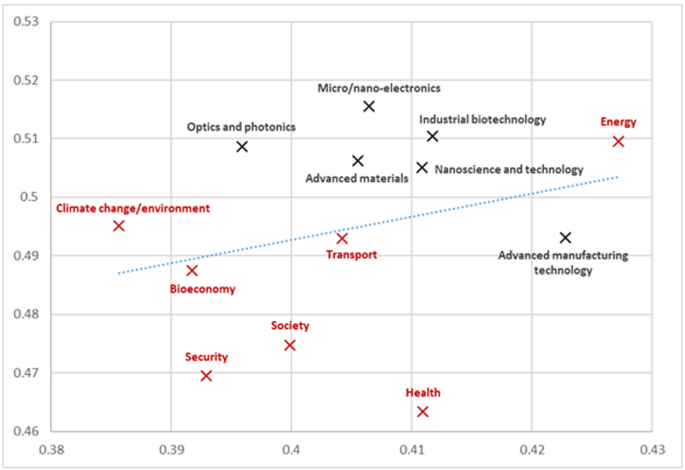
• social innovation (Edwards-Schachter and Wallace, 2017), broadly defined as innovations that satisfy societal needs including the needs of particular social groups and are primarily created by social actors such as non-governmental organizations or grassroots movements. To this aim, KNOWMAK provides information on about 1,000 social innovation projects derived from the RISIS European Social Innovation Database (ESID), such as the title, a short summary, website and location; these projects can be sorted by region and by subtopic, mostly in the societal grand challenge, in order to look to good examples of social innovation in specific areas.

• the number and share of publications that are either published on open access journals or are available on a public repository, as an indicator of the broader availability of scientific publication beyond the scholarly audience (Martín-Martín et al., 2018; see also www.cwts.nl/blog?article=n-r2w2a4&title=indicators-of-open-access-publishing-in-the-cwts-leiden-ranking-2019).

• user attention to scientific production, as measured by the number and share of publications tweeted (Haustein and Costas, 2015). Within KNOWMAK, it becomes therefore possible to analyze variations in the extent of user attention and open access by country and by topic to understand how these are related to different characteristics of scientific domains, but also to institutional cultures across Europe (see Figure 4).

Figure 4. Percentage of open access (vertical axis) and tweeted (horizontal axis) publications by country (top) and topic (bottom)





REFERENCES

Bonaccorsi, A., 2008. Search Regimes and the Industrial Dynamics of Science. Minerva 46, 285-315

Cassi, L., Lahatte, A., Rafols, I., Sautier, P., De Turckheim, E., 2017. Improving fitness: Mapping research priorities against societal needs on obesity. Journal of Informetrics 11, 1095-1113

Debackere, K., Luwel, M., 2005. Patent data for monitoring S&T portfolios, in Glänzel, W., Moed, H. (Eds.), Handbook of Quantitative Science and Technology Research. Springer, Dordrecht, pp. 569-585

Edwards-Schachter, M., Wallace, M.L., 2017. 'Shaken, but not stirred': Sixty years of defining social innovation. Technological Forecasting and Social Change 119, 64-79

Haustein, S., Costas, R., 2015. Identifying Twitter audiences: who is tweeting about scientific papers. Proceedings of the ASIS&T SIG/MET Metrics 2015 workshop , 9-11

Lepori, B., Guerini, M., Scherngell, T., Larédo, P., 2019. The spatial distribution of knowledge production in Europe. Evidence from KET and SGC. Proceedings of the 17th International Conference on Scientometrics & Informetrics , 685-690

Martín-Martín, A., Costas, R., van Leeuwen, T., López-Cózar, E.D., 2018. Evidence of open access of scientific publications in Google Scholar: A large-scale analysis. Journal of Informetrics 12, 819-841

Maynard, D., Lepori, B., Laredo, P., 2019. Using ontologies to map between research and policy data: opportunities and challenges.

OECD, 2015. Frascati Manual 2015. Guidelines for Collecting and Reporting Data on Research and Experimental Development. OECD, Paris

Zahedi, Z., Costas, R., Wouters, P., 2014. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. Scientometrics 101, 1491-1513

RISIS2 - European Research Infrastructure for Science, technology and Innovation policy Studies aims at building a data and services infrastructure supporting the development of a new generation of analyses and indicators on STI fields.

To develop a deeper understanding of knowledge dynamics and policy relevant evidence, RISIS goes beyond established quantitative indicators, developing positioning indicators, in order to reduce asymmetries in actors producing new knowledge, in places where knowledge is generated, and in themes addressed.

RISIS community is dealing with sensitive issues as social innovation, non-technological innovation, the role of PhDs in society, and portfolios of public funding instruments, studying both universities and firms.

RISIS - KNOWMAK

The development of the KNOWMAK tool has been supported by the European Commission through the H2020 grant 726992 (2017-2019). From 2020, the tool will be further maintained and updated within the RISIS2 H2020 infrastructure project (grant 824091).

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

RISIS Policy Brief Series *aim at disseminating key results coming from RISIS2 to improve the use of data for evidence-based policy making. The outcomes are presented through short documents pointing out the main policy issues at stake, demonstrating the contribution provided by RISIS, and what new avenues for research are now open.*

AUTHORS OF THE CURRENT ISSUE:

Benedetto Lepori | USI - UPEM
Philippe Laredo | UPEM
Thomas Scherngell | AIT
Massimiliano Guerini | POLIMI

www.risis2.eu

www.knowmak.eu