UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Economia Marco Biagi

DEMB Working Paper Series

N. 148

R&I smart specialisation strategies: classification of EU regions' priorities. Results from automatic text analysis

Pasquale Pavone[1], Francesco Pagliacci[2],
Margherita Russo[3], Anna Giorgi[4]

June 2019

[1] Sant'Anna School of Advanced Studies, Department of Excellence, Economics and Management in the Era of Data Science (EM-beDS), and CAPP
E-mail: pasquale.pavone@santannapisa.it
[2] University of Padua, Dipartimento Territorio e Sistemi Agro-Forestali and CAPP
E-mail: francesco.pagliacci@unipd.it
[3] University of Modena and Reggio Emilia, Department of Economics Marco Biagi and CAPP, Center for the Analysis of Public Policies
Address: Viale Berengario 51, 41121, Modena, Italy
E-mail: margherita.russo@unimore.it
[4] University of Milan, Leader AG1 EUSALP Lombardy Region representative, and Gesdimont Research Centre
E-mail: anna.giorgi@unimi.it

ISSN: 2281-440X online

R&I smart specialisation strategies: classification of EU regions' priorities. Results from automatic text analysis

Pasquale Pavone[1], Francesco Pagliacci[2], Margherita Russo[3], Anna Giorgi[4]

[1] Istituto di Economia - Department of Excellence, Economics and Management in the Era of Data Science (EMbeDS), Scuola Universitaria Superiore S.Anna di Pisa and CAPP (Centro Analisi Politiche Pubbliche), Università degli Studi di Modena e Reggio Emilia. Viale Berengario 51, 41121 Modena, Italy

[2] Dipartimento di Territorio e sistemi agro-forestali – TESAF and CAPP (Centro Analisi Politiche Pubbliche), Università degli Studi di Modena e Reggio Emilia. Viale Berengario 51, 41121 Modena, Italy

[3] Dipartimento di Economia Marco Biagi and CAPP (Centro Analisi Politiche Pubbliche), Università degli Studi di Modena e Reggio Emilia. Viale Berengario 51, 41121 Modena, Italy

[4] Leader AG1 EUSALP Lombardy Region representative, and Gesdimont research centre, University of Milan, Milano, Italy

e-mail addresses:
pasquale.pavone@santannapisa.it (corresponding author)
francesco.pagliacci@unipd.it
margherita.russo@unimore.it
anna.giorgi@unimi.it

Abstract

Building on automatic text analysis, this paper proposes an original categorization of Research and Innovation Smart Specialisation Strategy (RIS3) priorities and provides a common language (with detailed dictionaries) to classify priorities and then to associate EU regions to multiclass categories of priorities. This result is a powerful tool to interpret the current state of diversification across regions, with its potential of complementarities and synergies that might support territorial integrated development paths. It would also support regions in their future strategic programmes on RIS3. A case study on the Alpine macro-region shows innovation development paths to outline macroregion strategic planning.

Keywords: Research and Innovation Strategy for Smart Specialisation (RIS3); Future Cohesion Policy; Data classification with non-supervised techniques; automatic text analysis; EU macro-regional strategies

JEL codes: R58 regional development policies; O31 innovation and inventions: processes and incentives, C8 Data Collection and Data Estimation Methodology, Computer Programs

Funding

This work is part of the Work Package Nr: T-3 "Enhancing shared Alpine Governance project" of the Project "Implementing Alpine Governance Mechanism of the European Strategy for the Alpine Region" (AlpGov) of the Interreg Alpine Space Programme - Priority 4 (Well-Governed Alpine Space), SO4.1 (Increase the application of multilevel and transnational governance in the Alpine Space).

Acknowledgment

For the discussion on the topics presented in the paper, the authors wish to thank all the EUSALP's Action Group 1 members.

1. Introduction

Started almost ten years ago as an academic debate (Foray et al., 2012; Foray 2015 and 2018), Research and Innovation Strategies for Smart Specialisation (RIS3) has become in the 2014-20 programming period a pillar in the ex-ante conditionality for accessing the European Structural Investment Fund with programmes aiming at pooling, focusing and concentrating resources on research and innovation to maximize the impacts of the structural funds themselves (European Commission, 2014a; 2014b; 2014c; 2015; Radosevic et al. 2017). Nowadays, capitalizing on the results of the RIS3s implemented by almost all regions in Europe may contribute to the discussion of the future of EU Cohesion Policy, to outline its future goals and the development paths to achieve them. In order to appreciate such results, it would be necessary to provide a systematic and comparative analysis of RIS3 across EU regions (as stressed by McCann and Ortega-Argilés, 2016), but,so far, no systematic categorization of RIS3 priorities has been proposed, nor a classification of all EU regions according to their strategies. A recent contribution (Gianelle et al. 2018) proposes a comparative analysis of Italy and Poland, suggesting priorities as a *distinctive combination* of four main dimensions[1] and assigning them according to their occurrence in the RIS3s documents (21 regional strategies and one national strategy in Italy; 16 regional strategies and one national strategy in Poland). That approach returns a partial and non-systematic analytical tool that relies exclusively on expert judgment of policy documents.

A different strategy could be followed by using the database Eye@RIS3, implemented by the European Commission in the S3 Platform. Built by EU Joint Research Center as a source of information on RIS3, it aims to support regions in learning from other regions' practices (Kuznetsov & Sabel, 2017), as learning may help them improve in implementing ongoing projects and future RIS3 strategies. Eye@RIS3 provides a unique opportunity to have broader information on RIS3 across EU regions. The underlying database is populated by regions/nations and stakeholders on a volunteer basis. Indeed, information refers to different levels of territorial units, according to the level of governance of the territorial entity that elaborated the smart specialisation strategy (in most cases, Nuts 2 level regions, but also information at Nuts 0, Nuts 1 or even Nuts3 level can be found). The tool can be used for specific queries on priorities by entering a free text or selecting one or more codes (at one- or two-digit level) in each of the three domains categorizing the RIS3s of EU territorial entities: economic, scientific and policy objectives. In this way, the query returns the set of regions showing exactly those features.

The wealth of information available in the database Eye@RIS3 deserves attention for testing its use in a more effective search for regions presenting similar characteristics and in implementing a systematic comparative analysis on RIS3s across EU regions, both at country level and for those intermediate soft spaces represented by the macro-regional strategies[2].

[1] (A) Sectors or value chains of primary interest for the intervention; (B) Transformative processes to be activated (technology applications); (C) Societal challenges to be addressed; (D) Natural and/or cultural resources to be used (e.g. maritime ecosystem, alpine ecosystem, cultural heritage).

[2] See Stead (2014) and the references to the relevant literature surveyed by him.

The issue to be addressed in both search and comparative analysis is a grounded definition of similarity. Being multidimensional by nature, a classification of RIS3s needs to take into account what makes a priority specific, i.e. is it the context of economic domains in which it was elaborated, the scientific domains that the regions aim to leverage in their implementation, or the policy objectives they intend to reach? Since Eye@RIS3 offers the description of priorities both in free text - all translated in English - and in coded format, the information can be treated with automatic techniques of text analysis, an essential tool for topic classification (Oecd-WPTIP, 2018). By applying a text analysis and a non-supervised automatic classification of the priorities in free text format and of their codes (of economic domains, scientific domains and policy objectives), this paper provides both a complementary tool that could be implemented by Eye@RIS3 for online querying and a new multidimensional perspective on RIS3 that could be applied for comparative analysis.

To contextualise the RIS3 priorities' classification and analysis of regions, proposed in this paper, Section 2 starts with a brief summary of the background issues shaping the RIS3 in the EU regions and of the state-of-art in the overall process of implementation, monitoring and assessment of this policy, in the broad discussion of the future EU Cohesion Policy. Section 3 describes the dataset drawn from Eye@RIS3 that has been used in the empirical analysis: the types of information available on RIS3s' priorities (free and coded texts), the territorial entities selected for the analysis, the sources of data and date of records' update. Section 4 presents a methodology for automatic text analysis and the multidimensional analysis adopted in the paper. Section 5 returns the main results on the two main corpora created for the analysis: the one with free text description of priorities and the one with codes of economic domains, scientific domains and policy objectives associated to each of the priorities entered in the online platform. Section 6 builds on the results of both analyses and discusses their implications for policy and possible future strands of this research. A case study on the Alpine macro-region shows innovation development paths to outline macroregion strategic planning. Annexes present some detailed results.

2. Literature review and analytical framework

In November 2009, the European Commission (EC) published the report "Knowledge for Growth", the results of an expert advisory group to the European Union (EU). Tasked with finding an alternative to public policies, the expert group proposed that national and, especially, regional governments should encourage investments in domains that would complement the country's productive assets, to create future domestic capability and interregional comparative advantage. Aiming at building development policies as the result of a bottom-up process and at creating opportunities for place–based policies, the 2014-20 programming period of EU structural funds has oriented regional policies in designing smart specialisation strategies (Foray, 2018; Foray et al., 2015). This strategic proposal was adopted in the EU 2020 agenda (along with its objectives of smart, sustainable and inclusive growth). Today, smart specialisation strategy is an ongoing EU programme, begun in 2014 and expected to go beyond the current programming period. The basic ideas surrounding the Research and Innovation Strategies for smart specialisation agenda have been well articulated by Foray (2015), Foray et al. (2012; 2015), McCann (2015) and McCann & Ortega-Argilés (2016). In the ongoing debate for future EU Structural Fund Programmes, RIS3 seems to maintain the pivotal role in maximizing the impacts of the structural funds, to concentrate resources on research and innovation. To some extent, RIS3 represents the realisation of the rationale that lies behind the smart specialisation concept, namely, the concentration of the knowledge resources and the identification of a limited number of socioeconomic activity priorities to be linked with. Eventually, this will permit countries and regions to become and remain competitive in the current global economy (European Commission 2015).

In fact, concentration of knowledge resources for economic specialisation will allow each EU region to benefit from both scale and scope economies. In addition, spill overs in knowledge production and knowledge use are expected to occur[3]. Furthermore, smart specialisation adopts the so-called ‘entrepreneurial process of discovery’ (EDP), which is used to address the difficult problem of prioritisation and resource allocation decisions. EDP is an inclusive and interactive bottom-up process in which different participants explore and open-up new domains of opportunities (e.g. technological and market), potentially rich in numerous innovations that emerge as feasible and attractive (Foray, 2015). This entrepreneurial stand would throw light on what the most promising areas for future regional development are and would be a way to reveal what a country or region does best in terms of R&D and innovation.

In line with these general aims, a smart specialisation strategy becomes a tool for regional policy aiming at following a place-based approach to economic development, which has been promoted by both the European Commission and the Organisation for Economic Co-operation and Development (OECD). This is true, even though RIS3 makes no specific recommendations regarding which particular policy approaches to adopt in which places (McCann, 2015). However, it requires structured and explicit processes of analysis, reflection and prioritization (Kroll, 2015), which are based on key principles, allied with monitoring and evaluation activities, all of which are to be tailored to the context (McCann & Ortega-Argilés, 2016; Kuznetsov & Sabel, 2017). The place-based fundamentals of the RIS3 approach are intended to be articulated and developed by the local actors on the basis of the analysis, consultation and engagement activities, and cannot be imposed top-down by authorities, exactly as the Barca report argued (Barca, 2009; Barca et al., 2012). Yet, these requirements also make significant demands on governance capabilities, and especially so in regions with more limited institutional capacity, most of which are also economically weaker regions (McCann & Ortega-Argilés, 2016).

Within the overall design and implementation of RIS3s, consensus has emerged on the need for comparing the RIS3 of the EU regions, supporting in this way improvements

[3] All of them are key drivers of productivity. Accordingly, smart specialisation is about generating unique assets and capabilities based on the region’s distinctive industry structures and knowledge bases. In this regard, RIS3 aims to overcome some of the "weaknesses that have affected previous regional innovation strategies, such as: i) a lack of an international and trans-regional perspective (in many cases, regional innovation was considered in isolation); ii) discordance with the industrial and economic fabric of the region, public involvement in R&D being not sufficiently business driven; iii) the missing (lack) of a sound analysis of the region’s assets; iv) copying the best performing regions without considering the local context."

in more effective implementation and revamping of the strategy. The web platform created under the JRC coordination, Eye@RIS3[4], "developed as a tool to help strategy development rather than a source of statistical data" (McCann & Ortega-Argilés, 2016: 1409), might support such a comparative perspective on RIS3. The purpose of the database is to give each region an overview of the EU regions' priorities, and in this way regions would become able to position themselves, to find their unique niches and to seek out potential partners for collaboration.

A critical aspect is how to compare EU regions. JRC proposes a benchmarking grounded in the socio-economic characteristics of each of them. This tool, developed by Orkestra (the Basque Institute of Competitiveness), in association with the S3 Platform, is of utmost interest. It allows regions to identify reference regions across Europe based on a regional benchmarking logic (Navarro et al., 2014).

With regard to RIS3, the online tool Eye@RIS3 includes a range of benchmarking facilities and tools which aid regions in their RIS3 profiling and self-analysis. In the Eye@RIS3, it is possible to query regions by selecting one or more characteristics out of three features of the regions' specialisations: the economic domain (82 two-digit codes of the Nace Rev.2[5]), the scientific domain (109 NABS two-digit codes)[6] and the policy objective (a list of 72 items, elaborated by JRC from EU strategic documents). Although it is not intended to be used as a source of statistical data, this paper uses information available in Eye@RIS3 for a systematic analysis of RIS3 across the EU[7]. As will be seen in detail in the following section, the large coverage of information across regions and the level of detail of information entered by the regions suggest exploring the database, which so far has only been populated by regions, with no systematic analysis of its contents and with no development of effective tools for querying. The potential of this database calls for a more focused perspective in looking for similarities among regions, not just as a combination of a few selected characteristics. Using a tool supporting a multidimensional perspective would be able to return a categorization of priorities to be applied in a classification of regions. The next sections will present how to build such a tool and how to apply it.

3. Data

The documents about RIS3, usually written in national languages, can be accessed in English in the single dataset published by JRC on the webpage "Eye@RIS3: Innovation Priorities in Europe"[8]. Data are updated, based on the inputs from EU regional and national authorities and their stakeholders (McCann & Ortega-Argilés, 2016).

For each territorial entity, classified according to the NUTS (Nomenclature of Territorial Units for Statistics) 2013 classification, the database contains the Region/Country

[4] http://s3platform.jrc.ec.europa.eu/eye-ris3

[5] Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008)

[6] Nomenclature for the Analysis and Comparison of Scientific Programmes and Budgets, Eurostat (2007)

[7] McCann and Ortega-Argilés (2016) have already pointed out some of the major EU priorities and sub-priorities for EU regions and member states, but only selecting some of them – and not in a systematic way. Gnamus (2017) refers to the prospective use of the tool for comparative analysis, but so far there is no evidence of such implementation.

[8] http://s3platform.jrc.ec.europa.eu/map (access on 1st October, 2018).

Name; the Priority Name (a short title of the priority promoted in the RIS3) and the Priority Description (a longer description of the same priority); Source date and type (in particular: Final RIS3 Document, Draft RIS3 Document, Peer Review, Presentation at public event, other study or source), Date Encoded. To allow for easy comparisons of information on priorities across regions, the tool presents, for each of the priorities, classifications of their economic domains (NACE Rev.2 two-digit classification), scientific domains (two-digit NABS 2007 classification) and policy objectives. On 1st October 2018, the dataset downloaded for the analysis presented in this paper, included a total number of 1,295 records that refer to 214 territorial entities, across the EU-28 Member States.

Territorial entities. The dataset provides information on both regions and whole countries. In fact, given the diverse structure of governance across EU Member States, mandates for RIS3 processes and strategy implementation were assigned to a widely varying set of spatial levels of governance (Kroll, 2015; Capello and Kroll, 2016). Although the EC has worked towards achieving a certain level of uniformity in the RIS3 process, any final decisions lay with the individual Member States. Likewise, many Member States delegated the RIS3 process to certain regional levels, either too high or too low, simply because these had previously been responsible for EU Funding (Kroll et al., 2014). As a result, too many RIS3 processes were either conducted at rather high levels of governance or assigned to rather small regions[9] (Capello and Kroll, 2016). The dataset Eye@RIS3 points out such a situation perfectly, and the choice of the territorial statistical observations represents an empirical issue: while in most cases RIS3s have been implemented at NUTS 2 level, in some other cases, they have been implemented either at country (i.e. NUTS 0), or at NUTS 1, or at NUTS 3 level (see Annex 1). Given this heterogeneous mix of governance features of the RIS3s, in the rest of the paper, the analysis refers to 1225 records covering 206 territorial entities[10], after the exclusion of those country-level RIS3s that belong to countries where all the regions have also implemented their own RIS3.

Source of data. In the Eye@RIS3 platform, the selected 1225 records refer to five main sources: Final RIS3 Document, Draft RIS3 Document, Peer Review, Presentation at public event, other study or source. The majority of records refer only to one source and the most quoted (in seven out of ten records) is the Final RIS3 document. As expected, the most frequent years are 2013 and 2014, at the beginning of the present cycle of EU policy, in which RIS3 was enforced as an ex-ante conditionality (Annex 1 Table 2 returns details on source types and year of the documents).

Corpus: descriptions, codes. From the original data base two aligned sets of information – Descriptions and Codes - have been extracted. The Corpus “Descriptions” is the combined set of texts in the original fields “Priority Name” and “Priority Description”.

[9] Capello and Kroll (2016: 1397) point out that reverse consequences occur. In the former case, the whole notion of specialisation becomes questionable as, on a a national level, social and regional cohesion must play a central role. In the latter case, counties or local agencies with very limited administrative means could do little more than channel further funding into those one or two fields that are evident anyway.

[10] The number of records entered by the regions with information on their priorities ranges from 2 to 15 (details upon request).

The Corpus "Codes" comprises the three sets of economic domains (NACE Rev.2), scientific domains (NABS 2007) and policy objectives, all of which are taken at the two-digit classification to avoid replication of the same one-digit level across different records of the same region. Alignment of the two corpora means that, after their independent analysis and classification of topics, results are associated to each of the records entered by the regions, creating new information on which to elaborate querying and statistical analysis.

The database over time. A first pilot analysis was performed on the database downloaded on 31/05/2018. Composed of 1215 records, referring to 207 regions, its Corpus Description (29,012 occurrences) is smaller than the one extracted on 01/10/2018 (34,046 occurrences), elaborated in this paper. The comparison of the two downloads highlights that in some cases regions have changed the number of records (11 regions have deleted 47 records; 27 regions have added 75 more records), and in other cases, the textual description of priorities has changed. By focusing on the 73 regions that have not changed the number of records, it emerges that 47 regions present their priority descriptions with a larger text, while in 26 regions the descriptions have a smaller-sized text. Annex 2 Figure 1 shows the difference in the number of occurrences in the Corpus Description for those 73 cases. The changes have improved the overall quality of the information: some regions with only the name of the priority and no description have added it[11], and regions with a very long description have reduced it by dropping general text[12].

4. Methodologies

This paper adopts automatic lexical textual analysis (Bolasco 2013) and cluster analysis to identify the main specialisations characterising the RIS3 of the 206 selected territorial entities. Automatic text analyses, supported by expert selection of topics, are generally used when the domain is largely known and topic modelling is appropriate (Blei et al., 2003; Griffiths & Steyvers, 2004). In the case of the Eye@RIS3 database, it would be better to leave the pre-definition of topics to be selected as open as possible, and then to adopt multidimensional techniques to explore the Corpus bag-of-words, allowing the emerging of topics on the basis of the characteristics of the texts under analysis (Bolasco, 1999).

The analysis of the two corpora, Descriptions and Codes, is carried out independently: the former set consists of information in free text format, while the latter has a structure of recurrent graphic form.

The identification of the main specialisations contained in the corpus is carried out through correspondence analysis and cluster analysis (Benzecri, 1973, 1992; Greenacre, 1984, 2017). Multidimensional analysis allows us to observe the similarity of records

[11] For example, on 31/05/2018, ITI1 had three records only returning the priority names, of one or two words. On 01/10/2018, they present, in a synthetic form (overall, 132 words), an effective level of detail of the priorities presented.

[12] For example, PL11 had six records with 779 words, which now have been reduced to 291, with no abbreviations and a more focused description.

based on their lexical content. The word-based clustering phase is a non-supervised classification[13] that reflects the semantic similarity among records. This conceptual homogeneity makes explicit the theme or semantic trait prevalent in that group of records, which can be summarized in a category not defined *a priori* but obtained through the analysis. Categories are labelled through the expert reading of the dictionaries associated to each group of records.

The analysis of each corpus was carried out using two software. Taltac2 (Bolasco, 2010) was used for: the lexical-textual analysis; the identification of the lexical analysis units (intended both as simple and multiword expressions occurring in the corpus); and the definition of the lexical and textual matrices (to be explored through the multidimensional analysis). Spad® was used for multidimensional analysis and cluster analysis.

5. Results

The main results are presented, separately, with regard to the Corpus Description and to the Corpus Codes, and combined.

5.1 Classification of regions according to their priorities' descriptions

The corpus under analysis consists of 1220 records with descriptions of priorities[14]. By using Taltac2 software, the text information (which is unstructured by its nature) is structured within a Document Warehouse, consisting of the Vocabulary DB (lexical units of analysis) and the Documents DB (textual units of analysis). After this preliminary step, the Corpus Description ends up being constituted by 3,878 different words for a total of 34,046 occurrences. The corpus under analysis is characterized by a great variability in terms of the size of the records. In fact, 209 texts have descriptions with less than five occurrences (22 records only contain a single word) while the largest record contains 337 occurrences. In describing the regions' priorities, records present different structures: from a simple list of activities to a more detailed form of discourse.

Furthermore, in order to avoid the fragmentation of words according to the number of their flexions (e.g., singular/plural), the analysis is carried out on the lemmatised corpus, thus obtaining 3,521 different graphic forms, instead of 3,878 forms belonging to the non-lemmatised Corpus Description.

<u>Identification of lexical analysis units</u>

Through the grammatical annotation of the graphical forms of the Vocabulary and the application of a lexical-textual model (Bolasco & Pavone, 2010), it is possible to identify multiword expressions (MWEs) through the search for syntactic structures (Pavone, 2010, 2018). In order to avoid excessive fragmentation of the graphic forms, given the very small size of the records, it was considered appropriate to lexicalize only

[13] Here a hierarchical clustering is adopted using 10 coordinates of the Correspondence analysis for the aggregation. This classification was performed with the Spad® software, which follows the Ward aggregation criterion, with 10 iterations of consolidation with mobile centers.

[14] Compared to the total dataset, 5 records are missing: they refer to Galicia (ES11), for which there are only the priority codes but not the description.

MWEs with more than 10 occurrences[15]. Thanks to the grammatical classification of the graphic forms it has also been possible to distinguish the Active forms (words of content: nouns, verbs, adjectives and adverbs) and the Supplementary forms (in the literature also defined as 'stop words' or 'words of speech structure': conjunctions, prepositions, articles, etc.). Aiming at identifying the themes that characterized the various records, all graphic forms (lemmas and MWEs) classified as nouns and adjectives at the threshold of five occurrences were selected as active units of lexical analysis.

Accordingly, 773 lemmas and MWEs were selected and used to define the textual matrix *records* × *active lexical units* (1220×773), to be processed through a factor analysis.

Identification of RIS3s's categories of priorities based on Description

Through the factor analysis and the cluster analysis carried out on the textual matrix it is possible to classify in a univocal way the records under analysis[16]. To avoid a distortion of the factorial representation, 295 records with less than five active lexical units have been excluded as active elements of factorial analysis. These records have been considered as additional elements, and therefore they do not affect the definition of the factorial subspace but they can still be classified according to their similarity to the groups defined through the active elements[17].

By observing the resulting dendrogram, four main groups with 23 groups of related records emerge. For each of them, the reference topic of the priority is defined, based on the expert reading of their characteristic dictionaries of terms[18].

The classification of the textual matrix *records* × *active lexical units* (1220×773) disentangles 23 clusters of records, covering 21 priorities. They are listed in Table 1, grouped in the four macro-categories. Clusters' labels are assigned by expert reading of the dictionaries of terms encompassed in each category: by ordering terms in decreasing order of their test-value and by considering only terms with p-value less than 0.001, labelling is quite easy[19].

[15] The lexical-textual model has allowed the identification of 85 MWEs. The main MWEs identified are: new technology, renewable energy, energy efficiency, life science, creative industry, advanced material, value chain, smart grid, quality of life, advanced manufacturing.

[16] Seven out of 1220 records were not classified because the terms of their description are below the threshold value of 5 occurrences within the corpus, and in particular those cases present only hapax: Bio-Agro-Food; Hydraulics; IT; Mathematics and physics; Multilingualism; Railroads; Tobacco.

[17] The classes assigned automatically to the 295 additional records have been individually checked. For 256 records, the assigned classes were confirmed. For 39 cases, the automatic classification was not considered appropriate: in 13 cases, the text description allowed an effective assignment to a different class; in the remaining 26 cases, a manual classification was not possible because of a too generic description.

[18] For each category, the characteristic dictionary of the priorities descriptions is available in Annex 2. The dendrogram associated to this analysis is available upon request.

[19] Groups c-9 and c-10 both refer to the Digital & ICT category, and groups c-15 and c-18 both refer to the Mechatronics category. Different groups were formed because of the specific textual expressions adopted by regions.

Table 1 - List of detailed priorities and macro-categories, obtained from automatic classification of Description (free texts)

cl-ID	Cluster label	Macro-category
c-1	Health	c-1 Health & Life science
c-2	Life Science	
c-3	Agrofood	c-2 Agrofood
c-4	Healthy Food	
c-5	Creative industry	
c-6	Fashion	
c-7	ICT & Tourism	
c-8	Tourism	c-3 New economy & Leisure industry
c-9	Digital &ICT	
c-10	Digital & ICT	
c-11	Bioeconomy	
c-12	Growth & Welfare	
c-13	Optics	
c-14	Photonics	
c-15	Mechatronics	
c-16	Automotive	
c-17	Manufacturing	c-4 Prod. & Transp. Manufact. & Energy
c-18	Mechatronics	
c-19	Transport & Logistics	
c-20	Marine & Maritime	
c-21	Water jet cutting	
c-22	Energy Production	
c-23	Sustainable Energy	

Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018

This result of the automatic text analysis allows the multi classification of each of the 206 territorial entities according to 21 categories of priorities, thus reducing the multitude of 3,878 different non-lemmatized words describing the RIS3s' priorities, which occur in the Corpus. This result paves the way to a comparative analysis across EU regions that will be discussed in section 6. Table 2 reports the number of regions by category of priority description, listed in decreasing order of occurrences.

Table 2 - Number of regions by category of priorities' descriptions

macrogroup	category of priority description	# regions	share out of 206
New economy and Leisure industry	Digital &ICT	117	0,57
Prod. & Transp. Manufact. & Energy	Sustainable Energy	116	0,56
Agrofood	Agrofood	97	0,47
Health & Life Science	Health	77	0,37
Prod. & Transp. Manufact. & Energy	Automotive	72	0,35
Health & Life Science	Life Science	64	0,31
New economy and Leisure industry	Tourism	63	0,31
Prod. & Transp. Manufact. & Energy	Manufacturing	58	0,28
Prod. & Transp. Manufact. & Energy	Transport & Logistics	54	0,26
New Economy and Leisure industry	Bioeconomy	52	0,25
Prod. & Transp. Manufact. & Energy	Mechatronics	41	0,20
Prod. & Transp. Manufact. & Energy	Energy Production	38	0,18
Prod. & Transp. Manufact. & Energy	Marine & Maritime	35	0,17
Prod. & Transp. Manufact. & Energy	Photonics	35	0,17
New economy and Leisure industry	Growth & Welfare	28	0,14
New economy and Leisure industry	ICT & Tourism	28	0,14
Agrofood	Healthy Food	19	0,09
New economy and Leisure industry	Creative industry	17	0,08
New economy and Leisure industry	Fashion	9	0,04
Prod. & Transp. Manufact. & Energy	Optics	5	0,02
Prod. & Transp. Manufact. & Energy	Water jet cutting	1	0,00

Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018

5.2 Classification of regions according to the codes of their priorities

In the Eye@RIS3 database, regions can categorize their priorities through a series of codes belonging to three different domains: the economic domain; the scientific domain; the policy objective. For each domain, each priority may be categorized by referring to more than one code, thus generating a multitude of different combinations of codes across the 1225 records. In order to group the records according to the similarity of the codes associated to the description of the RIS3 priorities, a Boolean type matrix *Records* × *Codes* is defined (1225×263[20]), where the internal elements of the matrix are the presence/absence (1/0) of the codes in the record. Also in this case, with a factor analysis and a cluster analysis, it is possible to obtain groups of records characterized by dictionaries of priorities codes (for each cluster, the complete list of codes is reported in Annex 3). In particular, through the analysis of the dendrogram, five macro-groups of records and 11 related groups emerge: the specific combination of economic domains, scientific domains and policy objectives in each category provides a key to interpreting the main trait of the RIS3 priority they characterize.[21]. Table 3 lists the labels assigned to the clusters of priorities emerging from codes. Also in this case, clusters' labels are assigned by expert reading of the dictionaries made up of the codes encompassed in each cluster[22].

Table 3 - List of detailed priorities and macro-categories, obtained from automatic classification of Codes

cl-ID	Cluster label	Macro-category
c-1	Creative industry & cultural and recreative services	c-1 New economy and Leisure industry
c-2	Traditional Tourism	
c-3	Social innovation & education	
c-4	ICT & digital transformation	
c-5	Sc.domains enhancing social activities	
c-6	Health & Life Science	c-2 Health & Life Science
c-7	New technologies for health	
c-8	Agrofood, forestry and tobacco	c-3 Agrofood, forestry and tobacco
c-9	Blue economy_transport, marine resources	
c-10	Blue economy_fishing and aquaculture	
c-11	Bioeconomy & Waste collection, treatment etc	c-4 Blue Economy & Energy
c-12	Energy efficiency & Sustainability	
c-13	Environment Protection (water, sewerage, waste)	
c-14	Energy Production	
c-15	Smart Mobility System & Transport equipment	
c-16	Transport & logistics	
c-17	Automotive & Aeronautics industry	
c-18	Aerospace	c-5 Logistic & Manufacturing
c-19	Advanced manufacturing systems & Mechatronics	
c-20	Sc.Domains enhancing traditional manufacturing	
c-21	Basic metals, traditional & machinery manuf.	

Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018

[20] This is the sum of 82 Nace codes, 109 Nabs codes, 72 Policy objective codes. As already mentioned, to avoid redundancy, only second-level codes were considered for cluster analysis. Data on the relative frequency of codes in the database is available on request.

[21] A hierarchical clustering has been implemented as in the analysis of the corpus Descriptions. The dendrogram associated to this analysis is available upon request.

[22] For each category, the characteristic dictionary of the codes' clustering is available upon request.

This result of the automatic classification of the records allows the multi classification of each of the 206 territorial entities according to 11 categories of priorities defined through codes. Table 4 shows the list of priorities sorted by decreasing number of regions.

Table 4 – Clusters of codes (economic domains, scientific domains, policy objectives), by macro-group

macrogoup	category of codes of priority	# regions	share out of 206
Health & Life Science	Health & Life Science	145	0,70
Logistic & Manufacturing	Manufacturing	142	0,69
Agrofood, forestry and tobacco	Agrofood, forestry and tobacco	122	0,59
Bioeconomy, Blue Economy & Energy	Bioeconomy & Waste collection, treatment etc	115	0,56
New Economy & Leisure industry	ICT & digital transformation	111	0,54
Bioeconomy, Blue Economy & Energy	Energy Production, Efficiency & Sustainability	96	0,47
New Economy & Leisure Industry	Creative industry, Tourism & cultural and recreative services	91	0,44
Logistic & Manufacturing	Transport & logistics	56	0,27
New Economy & Leisure Industry	Social innovation & education	41	0,20
Logistic & Manufacturing	Aeronautics, Aerospace & Automotive industry	28	0,14
Bioeconomy, Blue Economy & Energy	Blue Economy	23	0,11

Source: authors' elaboration of Eye@RIS3 database, download date: 1st October 2018

5.3 Classification of regions' priorities: descriptions and codes

The two classifications can now be combined to provide, for each region, the set of categories in which their priorities have been categorized. The complete list is available upon request.

The cross tabulation of the two classifications reveals a coherent attribution of codes to descriptions. In particular, categories of codes in the cluster "Agrofood, forestry and tobacco" elaborate descriptions also in other related domains (such as: bioeconomy, tourism, leisure, sustainable energy), while categories in the macro-groups of codes referring to "Health & Life Science", "New economy & Leisure industry", "Logistic & Manufacturing" largely elaborate within the same domain of descriptions. In the case of the macro-category "Blue Economy & Energy", these groups of records cross many diverse descriptions, with a significant overlapping with descriptions in the macro-group of "Production & Transport, Manufacturing & Energy"[23] (Table 5). In general, the results of cross tabulation provide indications of the specific priorities emerging both within and outside the overlapping of the same categories of descriptions and codes[24].

[23] This result is due to the higher cut-offs in clustering the two classifications, the one referring to descriptions and the other one referring to codes: a similar set of macro-groups emerge, but in the case of codes a better cut-off is with five macro-groups, instead of four (as in the case of descriptions), with a split of "Bio Economy & Energy" from "Logistic & Manufacturing".

[24] For instance, in the case of NL2-Eastern Netherlands, the text description mentioning "development of robotics for. transcranial Magnetic Stimulation" is classified as "Mechatronics" in Description Classification and as "Health & Life Science" in Codes Classification.

Table 5 - Eye@RIS 3 records by category of RIS3' priorities: descriptions and codes

		categories of codes											
		Agrofood, forestry and to-bacco	Health & Life Science	New Economy & Leisure industry			Bioeconomy, Blue Economy & Energy			Logistic & Manufacturing			
		Agro-food, fo-restry and to-bacco	**Health & Life Science**	**Creative industry, Tourism & cul-tural and recrea-tive ser-vices**	**ICT & digital transforma-tion**	**Social in-novation & educa-tion**	**Bioecon-omy & Waste col-lection, treatment etc**	**Blue Eco-nomy**	**Energy Pro-duction, Ef-ficiency & Sustainabi-lity**	**Aeronau-tics, Aero-space & Automotive industry**	**Manufactu-ring**	**Tran-sport & logistics**	
categories of descriptions: Agrofood	**Agrofood**	7,92	0,16				0,57		0,08	0,16			8,90
	Healthy Food	1,22	0,16										1,39
Health & Life science	**Health**	0,16	5,63	0,08	0,16	0,16	0,33			0,24			6,78
	Life Science	0,08	5,71										5,80
New Economy & Leisure industry	**Bioeconomy**	0,57	0,33	0,41	0,33	0,98	1,14	0,41	0,08	1,22	0,24		5,71
	Creative industry			0,16	1,22								1,39
	Digital &ICT	0,08	0,65	8,08	0,33	0,57	0,57	0,08		0,41	0,08	0,16	11,02
	Fashion				0,08					0,73			0,82
	Growth & Welfare	0,08	0,65	0,41	0,33	1,22	0,24	0,08		0,49	0,08	0,08	3,67
	ICT & Tourism			0,08	1,96	0,08	0,08			0,08			2,29
	Tourism	0,41	0,33	0,33	4,49	0,16	0,16						5,88
Prod. & Transp. Manufact. & Energy	**Automotive & Aerospace**	0,08		0,16		0,16	0,73	0,16	0,08	3,59	1,06	1,06	7,10
	Energy Production						0,49	2,69	0,08	0,16			3,43
	Manufacturing	0,33	0,08				0,65	0,08		4,24	0,33	0,08	5,80
	Marine & Maritime						0,41	0,65	1,39	0,41	0,41		3,27
	Mechatronics		0,08	0,08			0,16	0,08		2,78	0,16	0,08	3,43
	Optics		0,08				0,08			0,24			0,41
	Photonics		0,08	0,41			0,16		0,08	1,88		0,16	2,78
	Sustainable Energy	0,49					6,69	4,33		0,57	0,24		12,33
	Transport & Logistics			0,16			0,16	0,08		0,73	2,78	0,73	4,65
	Water jet cutting									0,08			0,08
No Description		0,33	0,08	0,08		0,57	0,16	0,24	0,16	1,31	0,16		3,10
		11,76	14,04	10,45	8,90	3,92	12,82	8,90	1,96	19,35	5,55	2,37	100,00

Source: authors' elaboration of Eye@RIS3 database, download date: 1st October 2018

6. Discussion and policy implications

The broad coverage in terms of territorial entities that entered information and the large homogeneity of information at NUTS-2 level suggest that information in the Eye@RIS3 platform can be treated as a collective effort to support a robust comparative analysis of RIS3s' priorities across EU-28. By taking seriously what regions say when entering information in the Eye@RIS3 platform, the paper provides a classification of RIS3s' priorities as they are described in free text format and in the series of related codes of economic domains, scientific domains and policy objectives. Such classification allows to associate each EU region to one or more categories of priorities (according to the definition of their strategy): food-for-thought when the goal is learning from other regions' experience (Kuznetsov & Sabel, 2017). Which regions? It depends on the specific needs in the learning process.

The paper suggests a cross tabulation for a first systematic view of different patterns that could be explored both when cross combinations have plenty of cases and in the combinations with rarer cases. If implemented as an online tool, this perspective would support a more focused search in querying Eye@RIS3 for similar regions[25].

In the analysis proposed in this paper, each cell of the cross tabulation of categorization of priorities descriptions and codes returns either no region or one or more regions. Regions are characterized also by other features, but the ones summarized in that specific combination may be considered as the common features, i.e. the ones that could guide an exploration of similarities on research and innovation paths of development, which are the ones that are scrutinised when regions intend to learn from other practices, within the same priorities. With regard to RIS3s' priorities, the results of the analysis presented in this paper support such a search for similarities in a multidimensional perspective; they are not identified by the exact combination of codes nor by browsing the words in the descriptions entered in the database: each category of codes embraces a statistically significant combination of the different sets of codes, and each category of descriptions refers to a statistically significant semantic domain. The dictionaries associated to each category help in checking for nuances (but also in controlling for ambiguity and misinterpretation).

In general, if it were implemented in the EC-JRC Eye@RIS3 platform, the perspective on RIS3 proposed in the paper might incentive regions to use information available in the platform to learn from other regions' experience, and, in turn, it could support them in re-elaborating strategies that are more effective.

In particular, it might help in capitalizing on RIS3, when searching for complementarities and synergies on research and innovation in the EU macro-regions (already active

[25] It would be also more effective than the present query available on the S3 Platform that allows a simple browsing for exact free text, to select a priority, where necessary filtered by codes for each of the three groups available to encode priorities (economic domains, scientific domains, and policy objectives). On 01/10/2018, when searching, for instance, for the priority "agro-food", Eye@RIS3 returned 22 records, but that priority might have also been expressed in other graphic forms, such as "agri-food", "agrofood", "agrifood", "agro food", "agri food", "agricultural and food" or "agriculture & food", occurring respectively in 62, 5, 15, 0, 2, 9 and 1 regions, for a total number of 94 more records. Some of them might include more than one expression.

in the Adriatic-Ionian, Alpine, Baltic, Danube regions). The paper addresses those soft spaces in terms of the specific common features that emerge from the RIS3s' classification, a perspective that may enhance the design and implementation of innovation programs within the areas.

An application of cross tabulation with regard to macro-regions may support a comparative analysis of specific policy measures and projects implemented by regions within the same domain of priorities. Let us consider the regions in the Alpine area that are included in the macroregional strategy EUSALP, and let us focus on the regions with at least one priority classified under the category "Bio economy, Blue Economy & Energy", which is advocated as the core of the 2019 Presidency of the EUSALP. Indeed, the analysis confirms that almost all the territorial entities (19 out of 23) under EUSALP have that priority. What is relevant for enhancing an effective integrated territorial strategy is that the 19 regions have elaborated that priority with regard to "Bio economy & Waste collection and treatment", applied to Agrofood (AT12), Digital &ICT (DE1 and ITH1), Growth & Welfare (AT33), Automotive & Aerospace (AT32, ITC3, SI), Energy Production (FR72), Photonics (FR72), Sustainable Energy (AT12, AT22, FR43, ITC1, ITC2, ITH2, SI), Transport & Logistics (FR43). Another specific set of application refers to "Energy Production, Efficiency & Sustainability", which only marginally overlaps the same set of regions[26].

If the classification proposed by this paper were implemented online, in the JRC platform or in the Platform of Knowledge (recently created under the Alpine strategy EUSALP)[27], regions could start elaborating more focused analyses and leverage on more effective learning from other regions. This could turn into a more fruitful dialogue on potential synergies or complementarities across regions that might support integrated territorial developments within macro-regions.

References

Barca, F. (2009). *An agenda for a reformed cohesion policy: A place-based approach to meeting European Union challenges and expectations*. An Independent report prepared at the request of the European Commissioner for Regional Policy. Danuta Hübner. Brussels: European Commission.

Barca, F., McCann, P., & Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. *Journal of Regional Science, 52(1)*, 134–152.

Benzecri, J.P. (1973) L'analyse des données. L'analyse des correspondances. Paris: Dunod.

Benzecri, J.P. (1992). *Correspondence Analysis Handbook*. New York: Marcel Dekker.

Blei, D.M., Ng, A.Y., & Jordan M.I., (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, Vol. 3*, 993-1022.

Bolasco, S. (1999). *Analisi multidimensionale dei dati* [multidimensional analysis of data]. Roma: Carocci.

[26] The list of regions and priorities classified under the main category "Bioeconomy, Blue Economy & Energy" is available upon request.

[27] Details on this platform are presented at https://www.alpine-region.eu/results/platform-knowledge, accessed on 27/10/2018.

Bolasco, S. (2010). *Taltac2.10. Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi* [Taltac2.10. Developments, experiences and essential elements of automatic test analysis]. Milano: LED.

Bolasco, 2013, *L'analisi automatica dei testi* [Automatic text analysis]. Roma: Carocci.

Bolasco, S., & Pavone, P. (2010). Automatic Dictionary- and Rule-Based Systems for Extracting Information from Text. In F. Palumbo, C. N. Lauro, & M.J. Greenacre (Eds) *Data Analysis and Classification* (pp. 189–98). Berlin: Springer.

Capello, R., & Kroll, H. (2016). From theory to practice in smart specialisation strategy: emerging limits and possible future trajectories. *European Planning Studies 24(8)*, 1393-1406.

European Commission (2014a). Horizon 2020 in Brief. The EU Framework Programme for Research & Innovation. Luxembourg: Publications Office of the European Union.

European Commission (2014b). 'RIS3 and Cohesion Policy'. Available at: https://bit.ly/1SIK8M5.

European Commission, Directorate-General, Regional and Urban Policy (2014c). Policy and Legislation Guidance on Ex ante Conditionalities for the European Structural and Investment Funds. Available at: http://ec.europa.eu/regional_policy/sources/ docgener/informat/2014/eac_guidance_esif_part2_en.pdf

European Commission (2015). 'European Structural and Investment Funds 2014-2020'. Publications Office of the European Union. Available at: https://bit.ly/ 2HaRHMT.

Foray, D. (2015). Smart specialisation: Opportunities and challenges for regional innovation policy. London: Routledge.

Foray, D. (2018). Smart Specialisation Strategies and Industrial Modernisation in European Regions—Theory and Practice'. Cambridge Journal of Economics, October. https://doi.org/10.1093/cje/bey022.

Foray, D., Goddard, J., Morgan, K., Goenaga Beldarrain, X., Landabaso, M., Neuwelaars, C., & Ortega-Argilés, R. (2012). Guide to research and innovation strategies for smart specialisation (RIS3), S3 Smart Specialisation Platform. Seville: IPTS Institute for Prospective Technological Studies, Joint Research Centre of the European Commission. Available at: http://s3platform.jrc.ec.europa. eu/en/c/document_library/get_file?uuid=e50397e3-f2b1-4086-8608-7b86e69e8553&groupId =10157

Foray, D., McCann, P., & Ortega-Argilés, R. (2015). Smart specialisation and European regional development policy. In D.B. Audretsch, A. Link, & M. Walshok (Eds.), *Oxford handbook of local competitiveness*, (pp. 458–480). Oxford: Oxford University Press.

Gianelle, C., Guzzo F., Mieszkowski K. (2018). Smart Specialisation at work: Assessing investment priorities. JRC Working Papers. S3 Working Paper Series No. 14/2018.

Gnamus, A. (2017). Smart Specialisation: Thematic S3 Platforms (TSSP) and Macro-Regional Strategies. Slides presented at the "Knowledge and Efforts for Sustainable Growth in the Marine and Maritime Sectors in the Northern Adriatic", Piran, March 21. Available at: http://s3platform.jrc.ec.europa.eu/documents/ 20182/206164 /S3P_A.Gnamus-Piran-3.17.pdf/97ca255d-5e7d-4930-99c5-bd 51409ff6d4.

Greenacre, M. (1984). Theory and application of Correspondence Analysis. London: Academic Press.

Greenacre, M. (2017). *Correspondence Analysis in practice*. New York: Chapman and Hall/CRC.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National academy of Sciences101.suppl 1: 5228-5235.

Kroll, H. (2015). Efforts to implement smart specialisation in practice – leading unlike horses to water. *European Planning Studies 23*: 2079-2098.

Kroll, H., Muller, E., Schnabl, E., & Zenker, A. (2014). From smart concept to challenging practice – How European regions deal with the commission's request for novel innovation strategies (Working paper Policy and Region No. R2/2014). Karlsruhe: Fraunhofer ISI.

Kuznetsov, Y. and C. Sabel. 2017. Managing Self-Discovery: Diagnostic Monitoring of a Portfolio of Projects and Programs. In S. Radosevic, A. Curaj, R. Gheorghiu, L. Andreescu, & I. Wade (Eds.) *Advances in the Theory and Practice of Smart Specialization* (pp. 51–72). London: Academic Press.

McCann, P. (2015). The regional and urban policy of the European Union: Cohesion, results-orientation and smart specialisation. Cheltenham: Edward Elgar.

McCann, P., & Ortega-Argilés, R. (2016) The early experience of smart specialisation implementation in EU cohesion policy. *European Planning Studies, 24:8,* 1407-1427.

Navarro, M., Gibaja, J.J., Franco, S., Murciego, A., Gianelle, C., Hegyi, F.B., Kleibrink, A. (2014). Regional benchmarking in the smart specialization process: Identification of reference regions based on structural similarity. S3 Working Paper Series no. 03/2014 – February 2014. Seville: IPTS Institute for Prospective Technological Studies. Available at: http://s3platform.jrc.ec.europa.eu/documents/20182/114903/JRC89819_RegionalBenchmarking.pdf/699a6115-f685-4567-969d-921d116a304e [latest access on: July, 27, 2018]

OECD-WPTIP (2018). Summary of the Workshop: "Semantic Analysis for Innovation Policy". Paris, 12-13 March 2018. Innovation Policy Platform. Available at: https://www.innovationpolicyplatform.org/system/files/imce/OECD%20TIP_SemanticAnalysis_summary_FINAL_0.pdf.

Pagliacci, F., Pavone P., Russo M., (2018). Socioeconomic benchmarking for regions in the EU-28, Deliverable prepared for the EU project "AlpGov", Work Package T-3.1 "Mapping the governance in the research and innovation field of the Eusalp regions", 31th August 2018.

Pavone, P. (2010). Sintagmazione del testo: una scelta per disambiguare la terminologia e ridurre le variabili di un'analisi del contenuto di un corpus [syntagmatization: a choice to disambiguate the terminology and reduce the variables in a content analysis of a corpus]. In S. Bolasco, I. Chiari and L. Giuliano (Eds.). *Statistical Analysis of Textual Data: Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles*, Sapienza University of Rome, (pp. 131-140). Roma: LED.

Pavone P. (2018) Automatic Multiword Identification in a Specialist Corpus. In: Tuzzi A. (eds) *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences*. Quantitative Methods in the Humanities and Social Sciences. Springer, Cham.

Radosevic, S., Curaj, A., Gheorghiu, R., Andreescu, L. & Wade, I. (2017). *Advances in the Theory and Practice of Smart Specialisation*. London: Academic Press.

Stead, D. 2014. European Integration and Spatial Rescaling in the Baltic Region: Soft Spaces, Soft Planning and Soft Security. *European Planning Studies 22 (4)*, 680–93.

Supplemental materials (Annexes)

Annex 1 - Data

With regard to the territorial observations that are considered in the paper, the following Table 1 lists the territorial entities, by country and NUTS level, in the JRC Eye@RIS3 database. Country-level RIS3s occur in 21 out of 214 territorial entities. However, eight of these cases (for a total number of 70 records) have been excluded from the analysis, due to the fact that – for the same countries – all the regions have also implemented their own RIS3 as well (Austria, Germany, Greece, Denmark, Poland, Portugal, Sweden and Romania): in these cases, nation-level RIS3 seems to just replicate region-level RIS3s. On the opposite, in one additional cases (namely, the Czech Republic), RIS3 at the country level has been kept into the analysis, due to the fact that only one region of this country has elaborated its own RIS3 strategy. This country-level RIS3 includes 10 records in total. In twelve more cases, only information at country level are available (with 84 records in total), so they have been kept into the analysis. In six cases, information refers to small countries (namely, Cyprus, Estonia, Lithuania, Luxembourg, Latvia and Malta), for which the country level is equal to the NUTS 2 level. Thus, these countries have implemented a nation-level RIS3, which – to some extent – may actually represent a NUTS-2 level RIS3. In the remaining six cases (i.e. Bulgaria, Croatia, Hungary, Ireland, Slovakia and Slovenia), RIS3 could have been elaborated at NUTS 2 level, but they have not. So, this latter is an example of the too-high level of governance mentioned by Capello and Kroll (2016). The opposite case (hence, a too-low level of governance) refers to two countries (Finland and Sweden), which have implemented RIS3 at NUTS 3 level. Moreover, it is worth noticing that three countries (Belgium, Germany and the Nether-lands) have implemented RIS3 at NUTS 1 level. The remaining EU Member States implement RIS3 at NUTS 2 level.

Annex 1 Table 1 – Territorial entities, by country and NUTS level, in the JRC Eye@RIS3 database (download, 1st October 2018)

	Country		RIS3 programming level				Total
			NUTS 0	NUTS 1	NUTS 2	NUTS 3	
***	AT	Austria	1		9		10
	BE	Belgium		3			3
§	BG	Bulgaria	1				1
▪	CY	Cyprus	1				1
**	CZ	Czech Rep.	1		1		2
***	DE	Germany	1	16	1		18
***	DK	Denmark	1		5		6
▪	EE	Estonia	1				1
***	EL	Greece	1		13		14
	ES	Spain			17		17
	FI	Finland				18	18
	FR	France			26		26
§	HR	Croatia	1				1
§	HU	Hungary	1				1
§	IE	Ireland	1				1
	IT	Italy			21		21
▪	LT	Lithuania	1				1
▪	LU	Luxembourg	1				1
▪	LV	Latvia	1				1
▪	MT	Malta	1				1
	NL	Netherlands		4	1		5
***	PL	Poland	1		16		17
***	PT	Portugal	1		7		8
***	RO	Romania	1		7		8
***	SE	Sweden	1		1	19	21
§	SI	Slovenia	1				1
§	SK	Slovakia	1				1

UK	Utd Kingdom		3	4	1	8
Total		21	26	129	38	214
After exclusion		13	26	129	38	206
Total in the EU-28 (2016)		28	104	281	1348	

*** country-level RIS3 excluded because all region-level RIS3 are available

** country-level RIS3 not-excluded, because only some regions have their own RIS3

§ country-level RIS3 not-excluded, but the country adopts it as a policy decision (in fact, it could have implemented it at NUTS-2 level, as well)

▪ country-level RIS3 not-excluded, because these are small countries and the NUTS-0 level coincides with the NUTS-2 level

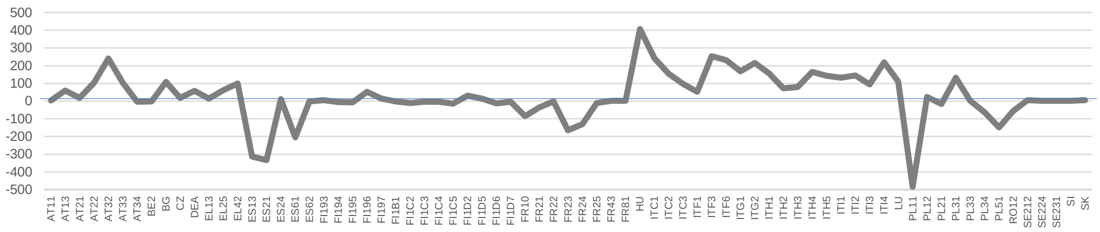
Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018

Annex 1 Table 2 – Number of records in the database Eye@RIS3, by sources and year

Etichette di riga	**2007**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**Total**
Final RIS3 Document	7	6	9	157	353	123	165	70	80	970
Draft RIS3 Document			11	101	41	3		13	6	175
Peer Review			2	13	24					39
Other Study				1	7	10		12	2	32
Presentation at public event					8			1		9
Total # of records	**7**	**6**	**22**	**272**	**433**	**136**	**165**	**96**	**88**	**1225**

Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018

Annex 2 Figure 1 – Difference in the number of occurrences describing RIS3' priorities, by 73 regions with the same number of records in the two downloads, on 31/05/2018 and on 01/10/2018



Source: authors' elaboration on Eye@RIS3 database, download dates: 31/05/2018 and 01/10/ 2018

Annex 3 – Corpus Description: characteristic dictionaries of the priorities' categories

Terms with p-value less than 0.001 are listed in decreasing order of their test-value

Clusters' label assigned by expert reading

Cluster 1 – 82 records, **Health**

health, medicine, medical, care, diagnosis, public health, healthcare, molecular, preventive, personalized, ageing, therapy, pharmaceutical, prevention, biotechnology, life science, therapeutic, biology, research, disease, quality of life, population, rehabilitation, biomedicine, drug, diagnostic, well-being, early, human, clinical, surgery, patient, assistance, diagnostics, tissue.

Cluster 2 – 71 records, **life science**

therapy, ageing, life science, patient care, disease, medical device, health, diagnostics, medical technology, medicine, e-health, chronic, regenerative, living care, personalised, formulation, diagnosis, diagnostic, treatment, pharmaceuticals, pharmaceutical, biomedical, medical, delivery, pathology, well-being, healthy, complex, biotechnology, imaging, oncology, infectious.

Cluster 3 – 110 records, **Agrofood**

food, agrofood, agricultural, packaging, animal, livestock, agriculture, product, food quality, agrofood sector, crop, processing, chain, farming, production, aquaculture, plant, cultivation, quality, functional, sustainable food, meat, food production, gastronomy, certification, aspect, vegetable, food industry, nutritional, breeding, processed, food safety, fishery, biotechnology, nutrition, precision, value chain, genetic.

Cluster 4 – 19 records, **Healty food**

traceability, producing, safe food, healthy, adequate societal value, minimised environmental impact, enhanced ecosystem, agrofood, waste, agrofood industry, food quality, functional, supply chain, food, food safety.

Cluster 5 – 17 records, **Creative industry**

media, art, creative industry, music, fashion, game, creative, film, visual, architecture.

Cluster 6 – 10 records, **Fashion**

textile, clothing, footwear, smart textile, fashion, leather, industry.

Cluster 7 – 29 records, **ICT & Tourism**

cultural industry, cultural, new technology, creative industry, valorisation, tourism, ict, experiential, cognitive, cross-sector innovation area, cultural heritage, social innovation, enhancement, experience, recovery, platform, digital, creative, diagnosis.

Cluster 8 – 72 records, **Tourism**

tourism, cultural, experience, culture, eco-tourism, nature, sustainable tourism, tourist, leisure, creative, wellness, sport, commercialization, ecotourism, natural, targeted, local, destination, event, links.

Cluster 9 – 72 records, **Digital & ICT**

ict, digital, smart city, cloud computing, virtual, community, language, information, digital economy, software, applications, digital service, media, sector, e-health, reality, big data, service.

Cluster 10 – 65 records, **Digital & ICT**

digital, data, software, internet, information, security, internet of things, applications, ict, service, content, visualisation, cloud computing, mobile, intelligence, game, big data, cyber-security, telecommunication, satellite, simulation, multimedia, cyber, artificial, solution, financial, computer, object.

Cluster 11 – 74 records, **Bioeconomy**

innovation, offer, promoting, company, knowledge, opportunity, world, new, bioeconomy, enabling, increased, regional.

Cluster 12 – 45 records, **Growth & Welfare**

cluster, knowledge, growth, effort, work, international, entrepreneurship, welfare, cooperation, benefit, partnership, initiative, job, innovation, company, business, research, venture, skill, area, collaboration, enterprise, creation, order, professional, more, private, greater, export, ensure, regional, university, central, result, centre, global, investment, take, world, climate, challenge, access, public.

Cluster 13 – 6 records, **Optics**

lighting, optics, laser, sensor, system, fiber, optical, photonics, communication, outdoor, dynamic.

Cluster 14 – 36 records, **Photonics**

photonics, electronics, sensor, industrial, material, technology, robotics, automation, system, network, micro, metrology, machine, advanced manufacturing.

Cluster 15 – 37 records, **Mechatronics**

manufacturing, mechatronics, equipment, production technology, housing, material, additive, interface, machine, automation, process, engineering.

Cluster 16 – 88 records, **Automotive & Aerospace**

automotive, developing, system, manufacturing, component, vehicle, specialized, projection, material, electric, factory, mobility, aerospace, production process, industry, machine, advanced manufacturing, implementing, installation.

Cluster 17 – 73 records, **Manufacturing**

wood, furniture, chemical, metal, material, plastic, composite, polymer, ceramic, product, manufacture, organic, chemistry, textile, paper, mineral, processing, forest, surface, coatings, machinery.

Cluster 18 – 5 records, **Mechatronics**

mechatronics, automatisation, production technology.

Cluster 19 – 66 records, **Transport & Logistics**

transport, vehicle, mobility, logistics, traffic, air, railway, aeronautics, system, road, aircraft, car, reliability, motor, engine, transport system, driving, safety, automotive, intelligent, aviation, space, automated, embedded system, e-mobility, mechatronics, component.

Cluster 20 – 42 records, **Marine & Maritime**

marine, maritime, shipbuilding, offshore, port, shipping, ship, coastal, blue economy, boat, naval, economy, marine energy, repair, construction, resource, blue, wind, fishery.

Cluster 21 – 1 records, **Water jet cutting**

jet, cutting, water,

Cluster 22 – 42 records, **Energy Production**

integration, energy storage, grid, power plant, fuel, wind energy, hydrogen, storage, energy network, solar energy, smart grid, assembly, heat, renewable energy, building, distribution, energy, turbine, gas, engineering, connection, maintenance, e-Mobility, photovoltaics, control.

Cluster 23 – 152 records, **Sustainable Energy**

energy, water, building, renewable energy, energy efficiency, sustainable, waste, construction, management, renewable, natural resource, waste management, recycling, water management, green, smart grid, cycle, resource, sustainable energy, energy production, construction material, efficient, source, treatment, technique, storage, circular economy, biomass, environmental, bio, monitoring, integrated, green economy

Source: authors' elaboration on Eye@RIS3 database, download date: 1st October 2018