European Commission

JRC TECHNICAL REPORT

Territorial Economic Data viewer: A data integration and visualization tool

Authors:
Marques Santos, A.
Conte, A.
Ojala, T.
Meyer, N.
Kostarakos, I.
Santoleri, P.
Shevtsova, Y.
De Quinto Notario, A.
Molica, F.
Lalanne, M.

2023

Joint
Research
Centre

Territorial Economic Data viewer: A data integration and visualization tool

Version: 15/05/2023

Anabela M. Santos* | corresponding author (Anabela.MARQUES-SANTOS@ec.europa.eu)

Andrea Conte* (Andrea.CONTE@ec.europa.eu)

Tauno Ojala* (Tauno.OJALA@ec.europa.eu)

Niels Meyer* (Niels.MEYER@ec.europa.eu)

Ilias Kostarakos* (Ilias.KOSTARAKOS@ec.europa.eu)

Pietro Santoleri* (Pietro.SANTOLERI@ec.europa.eu)

Yevgeniya Shevtsova* (Yevgeniya.SHEVTSOVA@ec.europa.eu)

Alicia de Quinto Notario* (Alicia.DE-QUINTO-NOTARIO@ec.europa.eu)

Francesco Molica* (Francesco.MOLICA@ec.europa.eu)

Marie Lalanne* (Marie.LALANNE@ec.europa.eu)

* Joint Research Centre, European Commission, Seville, Spain

JRC TECHNICAL REPORT

INDEX

Executive Summary

The [Territorial Economic Data viewer (TEDv)](#) is the first available tool that combines statistical territorial information from various European Union's Research and Innovation (R&I) funding programs within a single and coherent framework. It compiles information from multiple data sources (Figure 1) and combines beneficiary-level data (micro-level) with regional- or country-level data (macro-level) to produce comprehensive territorial statistics.

Figure 1. Data sources behind the Territorial Economic Data viewer (TEDv)
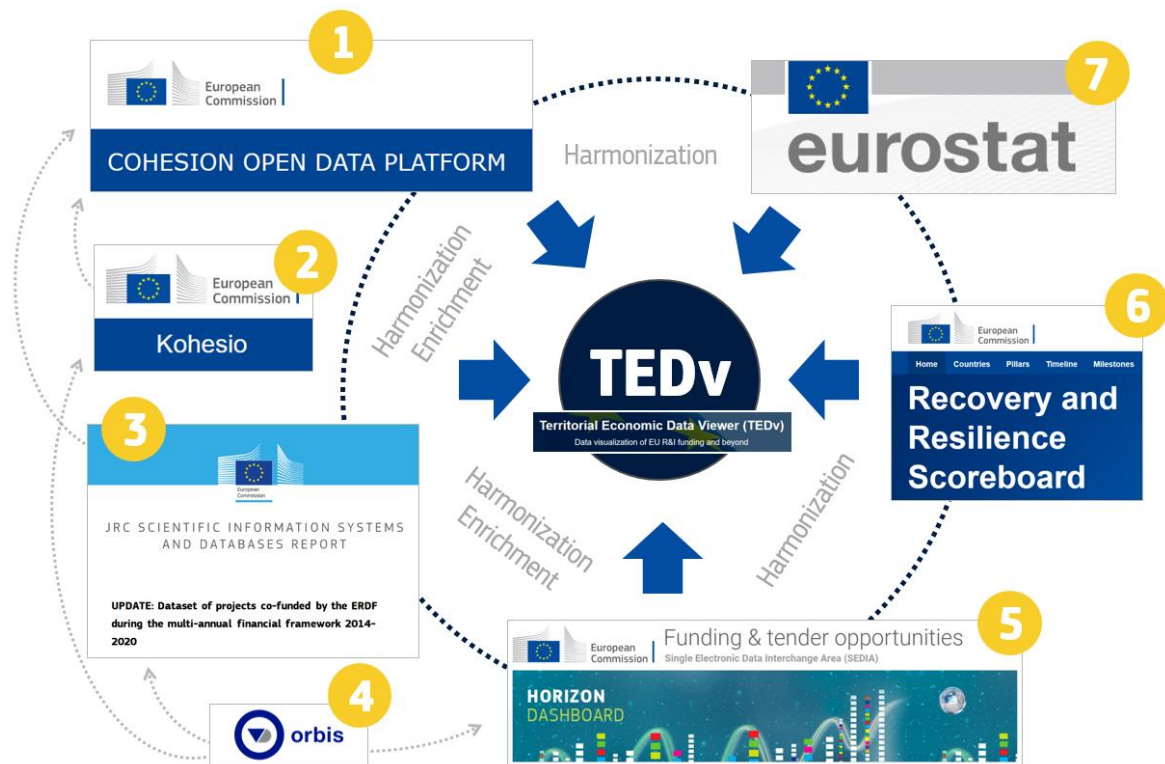
Source: Own elaboration.

The TEDv was developed to assist policymakers in monitoring the use of various Research and Innovation (R&I) funding programs. This task can be quite challenging as information about the final beneficiaries and territorial allocation of different funds is not available in a single data repository. Government bodies, managing authorities and stakeholders have to navigate through several data sources/webpages, sometimes with information expressed in different formats or taxonomies. The TEDv aims to address these challenges, by integrating data from multiple sources and presenting territorial statistics in a user-friendly format.

The TEDv includes territorial statistics (at the country-, NUTS 1- or NUTS 2-level) from three different R&I funding instruments: Cohesion Policy, supported by European Structural and Investment Funds (ESIF), Horizon Framework (Horizon 2020 and Horizon Europe) and Recovery and Resilience Facility (RRF) included in the Next Generation EU (Figure 2).

Figure 2. R&I funding programme included in the TEDv
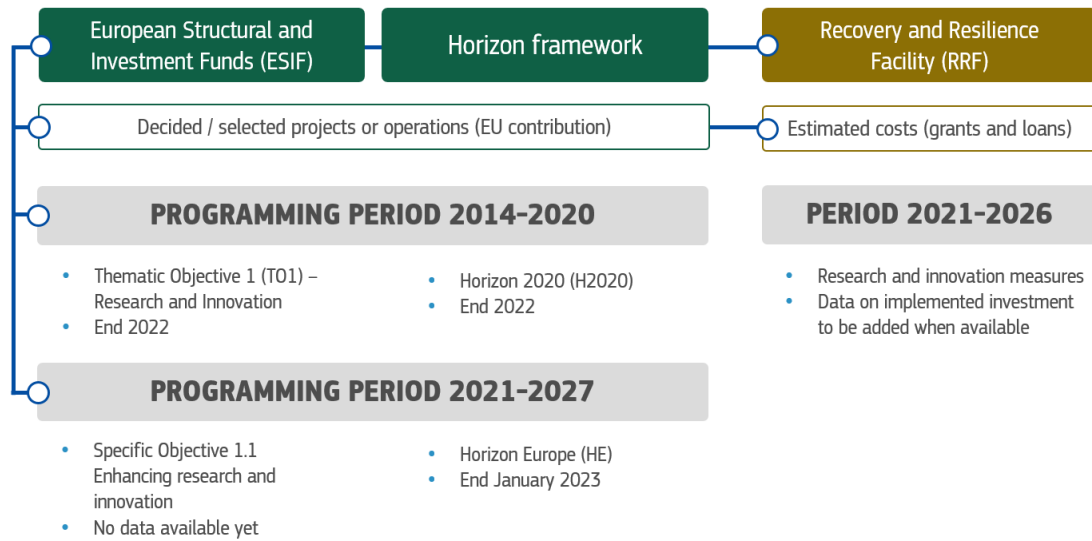
Source: Own elaboration.

Information regarding the territorial concentration of these funds can be found across three dashboards: (i) the Regional dashboard, (ii) the Sectorial dashboard and (iii) the Comparison dashboard. Beyond EU funding indicators, the TEDv also reports socio-economic and demographic statistics displayed in the 'Regional info-sheet' dashboard (Figure 3). The TEDv enables the comparison of different EU funds in a particular territory and displays the contribution of each funding source towards the total R&D expenditure of that territory. This statistical information can be particularly valuable for policy-makers as it allows them to compare their territory's relative position with respect to country and EU averages, as well as other regions within the EU.

Figure 3. TEDv dashboards' content

Territorial R&I funding statistics			Beyond EU funding indicators
Regional dashboard	Sectorial dashboard	Comparison dashboard	Regional info-sheet
Provides a territory snapshot of the allocation of different R&I funding (ESIF, Horizon and RRF)	Displays the sectorial concentration of different R&I funding (ESIF and Horizon)	Allows comparing territories in terms of the allocation of different R&I funding (ESIF and Horizon)	Compare regional performance over time with country and EU average, and provides information about the socio-economic and demographic position of the region (EUROSTAT)

Source: Own elaboration.

Territorial Economic Data viewer: A data integration and visualization tool

Anabela M. Santos* |corresponding author (Anabela.MARQUES-SANTOS@ec.europa.eu)

Andrea Conte* (Andrea.CONTE@ec.europa.eu)

Tauno Ojala* (Tauno.OJALA@ec.europa.eu)

Niels Meyer* (Niels.MEYER@ec.europa.eu)

Ilias Kostarakos* (Ilias.KOSTARAKOS@ec.europa.eu)

Pietro Santoleri* (Pietro.SANTOLERI@ec.europa.eu)

Yevgeniya Shevtsova* (Yevgeniya.SHEVTSOVA@ec.europa.eu)

Alicia de Quinto Notario* (Alicia.DE-QUINTO-NOTARIO@ec.europa.eu)

Francesco Molica* (Francesco.MOLICA@ec.europa.eu)

Marie Lalanne* (Marie.LALANNE@ec.europa.eu)

* Joint Research Centre, European Commission, Seville, Spain

Version: 15/05/2023

Abstract

The present working paper aims to describe the data sources and methods used to develop the Territorial Economic Data viewer (TEDv), as well as to explain the purpose and usefulness of the different dashboards available in the current version of the tool. Additionally, this paper includes practical examples with policy lessons that can be drawn from the available information, as well as a glossary of the indicators within the TEDv.

Disclaimer: The views expressed are purely those of the author(s) and may not in any circumstances be regarded as stating an official position of the European Commission.

Data behind the Working Paper: R&I funding statistics described in this Working Paper, as well as other original data are available for download in the Territorial Economic Data viewer (TEDv): https://web.jrc.ec.europa.eu/dashboard/TEDV/index.html

1. Context

Access to relevant and timely data is crucial for supporting policy-decision making and improving the effectiveness of policy interventions. However, policy analysts face several challenges in finding the right data and integrating them into an accurate reporting system. Some of the main bottlenecks that policy experts encounter include different data sources with varying taxonomies, unstructured data, and limited time and resources.

The TEDv was developed to support policymakers in monitoring the use of different Research and Innovation (R&I) funding programs. This task can be very daunting as information about the final beneficiaries and the territorial allocation of the different funds is not readily available in a single data repository. Government bodies, managing authorities, and/or stakeholders need to navigate through several data sources/webpages, sometimes with information translated into different formats and taxonomies.[1]

The TEDv is a tool that addresses these challenges by integrating multiple data sources and presenting the information in a user-friendly format, thanks to Qlik Sense Enterprise (QSE) server hosted by the Joint Research Centre (JRC).

The TEDv is the result of six main data-related activities developed within the Regional Economic Monitoring (REMO) pillar of the Territorial Data Analysis and Modelling (TEDAM) team, and developed to support the different phases of the policy decision-making process:

- Data research: identifying needs, following data trends;
- Data collection: data gathering from different sources, combining micro and macro-level data
- Database construction: cleaning, harmonisation, and enrichment
- Data analysis: producing derivate indicators to support monitoring and policy evaluation
- Data visualisation: generating maps, graphs, and tailor-made statistics
- Data sharing: data import for policy decision-making and research

The present working paper starts by explaining the purpose and usefulness of the TEDv (section 2). Then, it describes the data sources (section 3) and methods behind the development of the TEDv (section 4). It also includes practical examples with policy insights that can be drawn from the available information (section 5), as well as a glossary and definitions of the indicators existing in the TEDv.

---

[1] Such differences and the lack of a centralized data repository are mainly the result of the different governance model of the existing R&I funding programme. For instance, Cohesion policy operational programmes are managed at regional and/or national level, whereas Horizon Europe is centrally managed by the European Commission.

2. The Territorial Economic Data viewer (TEDv): An overview

The TEDv includes information from three different R&I funding programmes with different objectives and governance models (Table 1):

a) Cohesion Policy, supported by European Structural and Investment Funds (ESIF), under the thematic objective (or policy-specific objective) related to R&I;

b) Horizon Framework (Horizon 2020 for 2014-2020 and Horizon Europe for 2021-2027);

c) Recovery and Resilience Facility (RRF), included in the Next Generation EU, under the thematic area R&I.

Table 1. Objectives and governance model of main EU R&I funding programmes

Fund	Objective	Governance model
European Structural and Investment Funds (ESIF)	To correct territorial inequalities, enhance job creation and promote a sustainable and healthy European economy and environment. R&I is one of the main focus areas.	Depending on the EU countries, management of the funds could be regional, national or a combination of both
Horizon Framework	Supporting research and innovation with an emphasis on excellent science, industrial leadership, and tackling societal challenges	Management is centralised at European Commission-level
Recovery and Resilience Facility (RRF)	Temporary recovery instrument created to support the recovery of EU countries from the economic and social impact of the COVID-19 pandemic. R&I plays a key role in the recovery and twin (green and digital) transition	Mainly centralized at country-level

Source: Own elaboration based on Conte and Santos (2023).

The ESIF includes the European Regional Development Fund (ERDF), the European Social Fund (ESF), Cohesion Fund (CF), the European Agricultural Fund for Rural Development (EAFRD), and the European Maritime and Fisheries Fund (EMFF). For instance, in the programming period 2014-2020, the ESIF budget targeted strengthening research, technological development and innovation (under Thematic Objective 1: Research and Innovation – TO1) comes from ERDF (94%) and EAFRD (6%).[2] The values reported in the TEDv refer only to the ERDF part.

---

[2] https://cohesiondata.ec.europa.eu/themes/1/14-20 [accessed on 12/05/2023].

The TEDv covers two programming periods, 2014-2020 and 2021-2027. For 2014-2020, it includes data from ESIF under the thematic objective 1 (TO1) – R&I and Horizon 2020 (H2020) about the EU contribution of the total eligible costs for decided/selected projects. For 2021-2027, the current version of the TEDv only includes data on the on-going execution of Horizon Europe funds. Data about the RRF-R&I investments refer to the estimated costs included in the Recovery and Resilience Plans (RRP) for the all period 2021-2026 and financed by grants and loans. Therefore, data on implementation EU R&I funding programmes in 2021-2027 are going to be updated over time, as soon as the information becomes available.

Territorial statistics in the TEDv are aggregated in four dashboards:

a) **Regional dashboard**: provides a territory snapshot of the allocation of different R&I funding (ESIF, Horizon, and RRF) at the country-, NUTS 1- or NUTS 2-level;

b) **Sectorial dashboard**: displays the sectorial concentration of different R&I funding (ESIF and Horizon) at the country-, NUTS 1- or NUTS 2-level, as well as the total R&D expenditures by economic activity (EUROSTAT);

c) **Comparison dashboard**: allows comparing territories (countries and regions at NUTS 1- or NUTS 2-level) in terms of the allocation of different R&I funding (ESIF and Horizon); the selection of territories can be done manually by the user or using a pre-defined selection of peer regions (for more details about the methodological approach used to identify similar regions see section 4.2);

d) **Regional info sheet**: reports territorial statistics beyond funding indicators and allows to compare the regional performance (NUTS 2-level) with the country and EU average regarding a set of selected socio-economic and demographic indicators, to see the region evolution in the Regional Innovation Scoreboard (RIS), as well as its rank in the Regional Competitiveness index (RCI) and Cohesion criteria classification (2014-2020 programming period).

3. Data source behind TEDv statistics

To produce the main territorial statistics expressed at the country-, NUTS 1- or NUTS 2-level, the TEDv combines information from seven data sources:

- Kohesio (micro-level) database containing the list of projects and beneficiaries co-funded by EU Cohesion policy;

- JRC-WIFO (micro-level) database (Bachtrögler-Unger et al., 2021) comprising the list of projects funded by ERDF in the 2014-2020 programming period;

- Cohesion Open Data Platform (macro-level data) to estimate the cumulative amount of EU funding share of total eligible costs decided under the ESIF-R&I (TO1);

- eCORDA (COmmon Research DAta Warehouse) containing the list of Horizon programme participants (micro-level data), which is also the main data source behind the Horizon dashboard;

- Recovery and Resilience Scoreboard (macro-level data) to extract the cost estimated for R&I expenditures included in the Recovery and Resilience Plans of the EU Member States, except for Bulgaria, Hungary and Sweden where the values are from the Commission Staff Working Document (SWD) accompanying the proposal for a Council Implementing Decision. Data from Poland are taken from the European Parliament webpage.

- BvD ORBIS (micro-level data) to extract information about the location and economic activity of ESIF-R&I and Horizon beneficiaries.

- EUROSTAT (macro-level data) for socio-economic and demographic indicators, as well as, Research & Development (R&D) statistics;

Data included in the 'Regional info sheet' also comes from the European Commission's Regional Innovation Scoreboard (RIS) and Regional Competitiveness Index (RCI).

4. Methodological approach behind the TEDv

4.1. Estimation of territorial statistics

The main source of information for ESIF-R&I statistics is the "Cohesion Open Data Platform", namely the database containing the categorisation data from the ERDF/ESF/CF programmes. From this database, it is possible to extract data related to the location (NUTS 2- and, in some cases, NUTS 3-level) of total eligible costs decided (selected operations) and the respective EU funding share. Data reported at the NUTS 3-level are aggregated at the NUTS 2-, NUTS 1- or country-level to produce the respective territorial statistics. When the amount of ESIF-R&I funds for some operations is only available at the country or NUTS 1-level, the population share is used to regionalise the data at NUTS 2-level. Operational programmes associated with Territorial Cooperation (TC)[3] are excluded from the values generated. The costs classified in Multi-Thematic Objectives (TO) are split by TOs based on their priority axis classifications. Using data about ESIF 2014-2020 finances planned details, we estimated the share of EU budget for each TO over the total amount of each priority axis code by Operational Programme. Then, the estimated shares (using data from the planned amounts) is applied to Multi-TO to split their amounts by

---

[3] The main ESIF interregional collaboration projects refer to INTERREG projects. INTERREG is one of the key instruments of the European Union (EU) supporting cooperation across borders through project funding, with the aim to help their economic and social development and tackle the obstacle of borders (for more details, see Commission webpage).

TO. When the location of the operations refers to the country- or Nuts 1-level region, meaning that project/operation is supporting the whole country (or Nuts 1-level) population, we use the population share to regionalize the values. When needed, we use the NUTS converter (Joint Research Centre, 2022) to allow all the statistical information to be reported in the NUTS version 2021. To estimate the amount of ESIF R&I broken down by economic activity (NACE Rev.2), the sectorial shares - generated by combining the micro-level data on ERDF beneficiaries (from Kohesio and Bachtrögler et al. 2021) enriched by BvD ORBIS - are used.

The territorial statistics related to the Horizon funding programme are directly estimated using beneficiary-level data of Horizon 2020 Grants Database (or of the ongoing Horizon Europe). Information on the funds allocated to each project-participant (participant net EU contribution) is used to estimate the total funding allocated to each region (NUTS 2- or NUTS 1-level) or country. The amount of H2020 funds allocated by region/country and economic activity (NACE Rev.2) is based on the NACE classification of the participants extracted from BvD ORBIS.. The original dataset contains information on the sector of economic activity for around 42% of unique beneficiaries (for more details see Appendix A). To improve the coverage of sectorial classification we first match the original dataset with BvD ORBIS data that should provide a better NACE coverage. The matching is done using the VAT number available for around 90% of unique beneficiaries. The remaining 10% of beneficiaries with missing VAT information are matched with the BvD ORBIS data using string-match techniques. As a result, we are able to obtain information on the sector of economic activity for around 70% of beneficiaries (Table A1 in Appendix A). To further improve on this result we allocate some of the beneficiaries with missing NACE information using string-search techniques. For example, the beneficiaries that contain "Pharma" as part of their company name are allocated to NACE sector 21 (Pharmacompanies) and further to NACE sector "C" (Manufacturing). The beneficiaries with "UNIVERSIT" in their name are allocated to NACE sector 85 (Education) and further to sector "P" (Education). As a result the final dataset contains the information on the sector of economic activity for around 75% of beneficiaries. These data are further aggregated by region (NUTS 2-level, NUTS 1-level) or country and sector level. The funds where the sectorial classification is still missing are allocated to sectors using the shares of those sectors in total Horizon funds allocated to a specific country/region. Finally, the number of beneficiaries by sector and country are calculated by aggregating all the participants/projects into sector, region/country.

The RRF cost estimates for R&I expenditures for the period 2021-2026 include the amount financed by grants and loans. For the 22 EU Member States (MS) with their National Recovery and Resilience Plans (NRRP) approved until April 2022, the data are directly extracted from the 'Thematic Analyses' Research and Innovation performed by Commission's Directorate-General for Economic and Financial Affairs (ECFIN) available in the RRF scoreboard. For Bulgaria, Hungary and Sweden, we use the information available in the Commission SWD on the approval of the assessment of the RRPs:

- SWD(2022) 106 final (Bulgaria): total of the component 2 - Research and Innovation - reported in Table 4 (page 30);

- SWD(2022) 686 final (Hungary): total of the investment classified with the intervention field code 21 (*Technology transfer and cooperation between enterprises, research centres and higher education sector*) listed in the Annex – Climate and Digital tagging;

- SWD(2022) 102 final (Sweden): total of the investments classified with the intervention field codes 22 (*Research and innovation processes, technology transfer and cooperation between enterprises focusing on the low carbon economy, resilience and adaptation to climate change*) and 23 (*Research and innovation processes, technology transfer and cooperation between enterprises focusing on circular economy*) listed in the Annex - Climate and Digital Tagging.

For Poland, as the information in the Commission SWD on the approval of the assessment of the RRP - SWD(2022) 161 final - doesn't allow to clearly identify the innovation measures, we use the information reported by Sapala and Szczepański (2022) in the Table 1 of the briefing of the European Parliamentary Research Service (EPRS). From this table we use the total of investments of the subcomponents A2 and D3. Any discrepancies in the values of the RRF cost estimates are due to rounding and because the reporting unit in the cited documents is in million euros.

Statistical data about R&D expenditures (total and by sectors of performance – business enterprises, government, higher education, and private non-profit) are extracted from EUROSTAT (code: rd_e_gerdreg).

In addition to the cumulative amount of different R&I funding, the TEDv also includes some derivative indicators, for instance:

1) Average annual value of R&I funding amount = cumulative R&I funding amount divided by 7 years for ESIF and Horizon (or by 6 years for the RRF);

2) Cumulative R&I funding per capita = cumulative R&I funding over total population;

3) R&I funding as % of total annual R&D expenditure = average annual value of R&I funding amount over total R&D expenditure (GERD);

4) Relative size of two R&I funding = cumulative R&I funding amount of programme (a) over cumulative R&I funding amount of programme (b).

For more details, see also the "Glossary" and "Definitions" sections.

4.2. Territorial benchmarking: identifying similar regions

Comparing R&I funds allocation across regions represents a useful exercise that has the potential to inform policy learning and design. The key challenge when comparing regions is avoiding the comparison of dissimilar areas, leading to inaccurate assessments. To that end, we developed a data-driven approach based on statistical matching to identify similar regions at the NUTS2-level. The first step required to identify

similar regions is choosing which observable characteristics to take into account. As no universally accepted approach determines which regional factors should be considered, we selected a set of variables based on previous similar exercises (Navarro, et al. 2014) and on recent empirical studies using NUTS 2-level data (see, e.g., Cerqua and Pellegrini, 2022; Diemer, et al. 2022). Following these studies, we identified a set of variables intended to capture the structural conditions of regions, largely related to their innovation capabilities and economic development. These can be grouped into five main dimensions: geo-demography, economy, education, innovation, industrial structure, and employment. A detailed list of all variables included in these dimensions is outlined in Table 2.

Table 2. Variables used for matching

Dimension	Time-span	Source	Code
Geo-demography:			
Population	2011-2020	EUROSTAT	nama_10r_3popgdp
Population density	2011-2020	EUROSTAT	demo_r_d3dens
Life expectancy at birth	2011-2020	EUROSTAT	demo_r_mlifexp
Old-age dependency ratio	2011-2020	EUROSTAT	demo_r_pjangroup
Economy:			
Labour productivity per hour worked (PPP)	2011-2020	ARDECO	RHVGDH
GDP per capita (PPP)	2011-2020	ARDECO	SHVGDP
Education:			
Population with tertiary education (%)	2011-2020	EUROSTAT	edat_lfse_04
Innovation:			
EPO patent applications (1,000 inhabit.)	2011-2018	REGPAT	
GERD per inhabitant	2011-2019	EUROSTAT	rd_e_gerdreg
Industrial structure and employment:			
Employment in high-tech sectors (%)	2011-2020	EUROSTAT	htec_emp_reg2
Employment in science and technology (%)	2011-2020	EUROSTAT	hrst_st_rcat
Employment in NACE A (%)	2011-2019	ARDECO	SNETZ
Employment in NACE N-K (%)	2011-2019	ARDECO	SNETZ
Unemployment (%)	2011-2019	EUROSTAT	lfst_r_lfu3rt

Source: Own elaboration.

Note: All variables are measured at the NUTS2-level (2021 version). NACE A is "Agriculture, forestry and fishing". NACE N-K are "Financial and insurance activities; real estate activities; professional, scientific and technical activities; administrative and support service activities". In some cases, values for GERD are linearly interpolated (e.g. for Austria) or imputed using national trends when the series is not complete (e.g. for France and Belgium).

We opted to include variables at the NUTS-2 level that are available longitudinally and which provide a richer account of regional dynamics if compared with those that are only available cross-sectionally. In fact, using variables recorded at one point in time may arguably lead to less accurate comparisons due to possible

transitory shocks. One additional advantage of using longitudinal data is that they allow us to check whether the target region and its similar peers, experience similar dynamics in both levels and trends over time. To that end, we selected a 10-year interval spanning 2011-2020.[4] Before performing the matching procedure, we take averages over the entire period to attenuate any temporary shock, which may lead to misleading comparisons.

Having defined the set of variables, we use a statistical matching approach to identify similar regions called the Mahalanobis Distance Matching (MDM) algorithm (King and Nielsen, 2019), a technique that is widely used to build control groups in observational studies (see, e.g., Cerqua and Pellegrini, 2022). The MDM allows us to identify, for a given target region, the regions that are as similar as possible based on the set of variables outlined in Table 2. In principle, this approach allows the selection of several peer regions. However, having a large number of peers would arguably defy the main goal of the exercise, as it would lead to identify regions whose degree of similarity with the target region is very small. As a result, for any given region we restrict the maximum number of potential peers to three. Additionally, we allow the algorithm to identify the three closest peers in absolute terms, and the three closest peers outside the country of the target region. This allows the option to identify peers that do not belong to the same country (see Appendix B for an example).

Finally, it is important to stress that while this benchmarking tool makes it easier to conduct more appropriate comparisons across regions, it should not be regarded as a way to make causal statements or as a substitute for rigorous *ex-post* policy evaluation.

4.3. Data visualisation techniques

Once all the statistical data are expressed in a common taxonomy (NUTS version 2021 and NACE codes), the Qlik Sense Enterprise server is used for data integration and visualisation. Then, territorial statistics are displayed in four dashboards with different functionalities, as described in the next section.

5. Description of the applicability of TEDv

5.1. Regional Dashboard

The Regional Dashboard provides an overview of the allocation of different R&I funding (ESIF, Horizon, and RRF) for different territories. It combines data from EUROSTAT, Cohesion Open Data Platform (complemented by Kohesio data), and eCORDA, as explained in previous sections. The values of R&D statistics from EUROSTAT refer to the last year available for the programming period displayed in the

---

[4] Data availability is higher for this period compared to longer time spans. To test the sensitivity of our approach, we experimented with alternative time-spans (e.g. using the programming period 2014-2020) and obtained fairly similar results.

dashboard for ESIF-R&I or Horizon framework (i.e. 2014-2020 or 2021-2027). The amount of R&I funds from RRF refers to the period 2021-2026, and it appears in the dashboard for both programming periods (2014-2020 and 2021-2027) for comparison purposes.

In addition to reporting the cumulative value of ESIF R&I and H2020 funding programmes associated with selected projects or operations for a territory (country, NUTS-1 or NUTS-2 level regions), the dashboard also displays the average annual value of these funds and their estimated contribution to the total R&D expenditure in a territory. Furthermore, the dashboard also shows the relative size of the different R&D funds, where a value equal to one means that two funds are equivalent in terms of their size.

Figure 4 illustrates an example of TEDv visualisation for Portugal and Figure 5 for the Portuguese region of Alentejo (PT18). Looking at the information displayed in Figure 4, we can see that in the 2014-2020 programming period, the contribution of the Cohesion Policy to finance Portugal's R&D expenditure is higher than that of the Horizon framework. For instance, on average, ESIF-R&I funds contributed to 22% of R&D expenditure whereas H2020 contributed only 5.3% (graph in the middle and at the bottom of the dashboard). Consequently, the size of ESIF-R&I is about 4 times larger than that of the H2020 (graph on the right and at the bottom of the dashboard). The amount of R&I investment under the RRF is expected to be lower than that of ESIF-R&I for 2014-2020 but similar to that of H2020. Figure 5 reports the same information (except for RRF, since costs estimated are only available at the country-level) for the Portuguese NUTS 2-level region Alentejo (PT18). The region reveals to be less innovative than the country average, based on the R&D expenditure intensity expressed as % of GDP (0.78% versus 1.4%). Furthermore, the ESIF-R&I contributed, on average, to around 54% of the total R&D expenditures of the region, more than twice the value registered at the country-level. The annual average contribution of H2020 to R&D expenditure is around 4% and below the country average (5%). The size of the ESIF-R&I fund is about 13 times more than that of H2020.

Figure 4. Example of TEDv visualisation: Regional Dashboard - Portugal
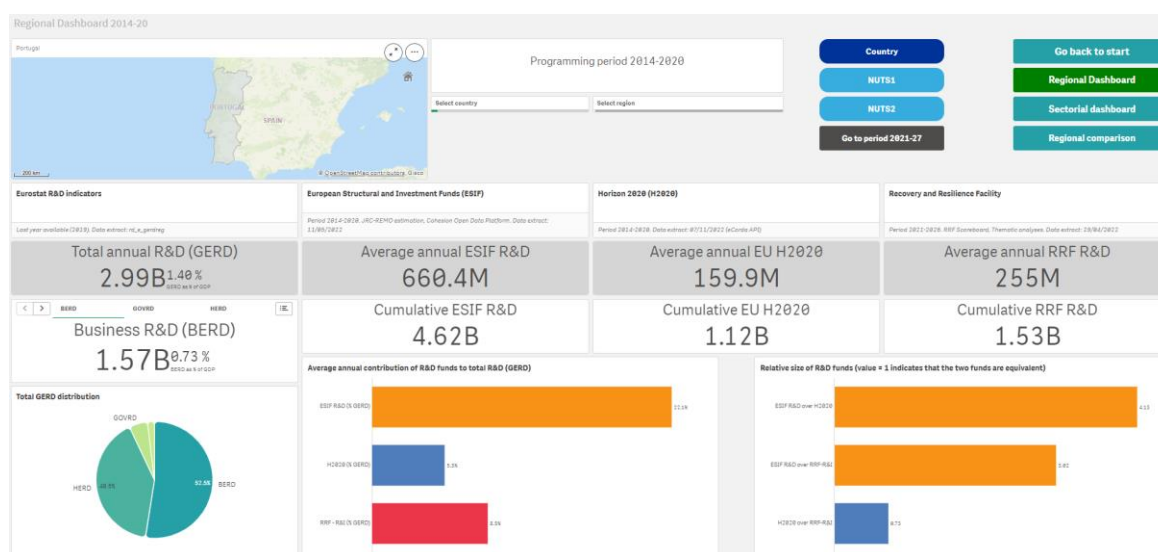
Source: TEDv (extracted on 31/03/2023).

Figure 5. Example of TEDv visualisation: Regional Dashboard – Portuguese region of Alentejo (PT18)
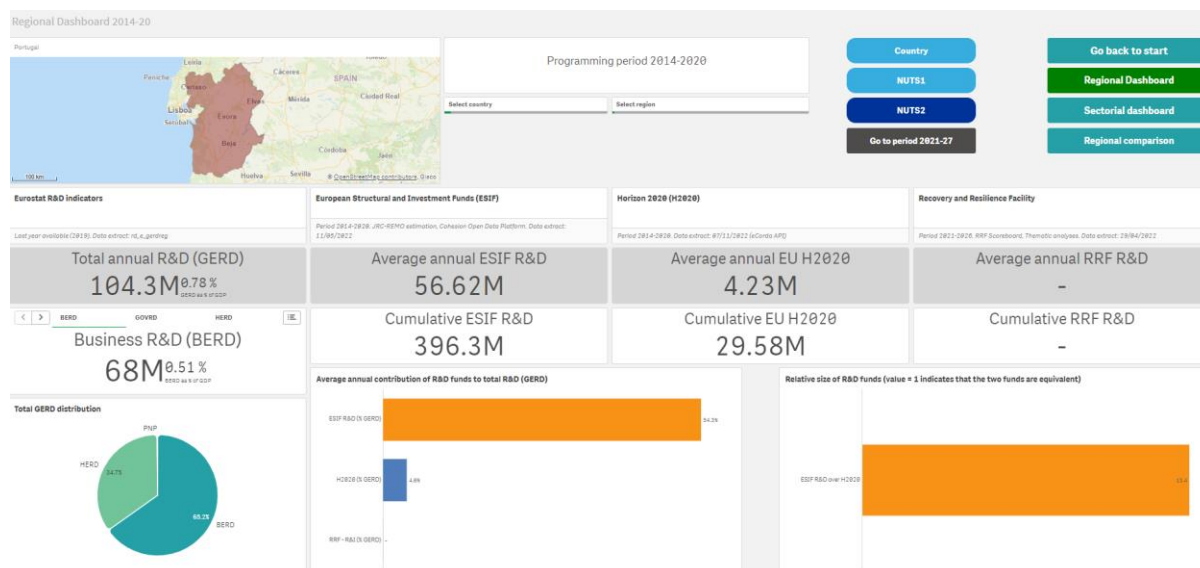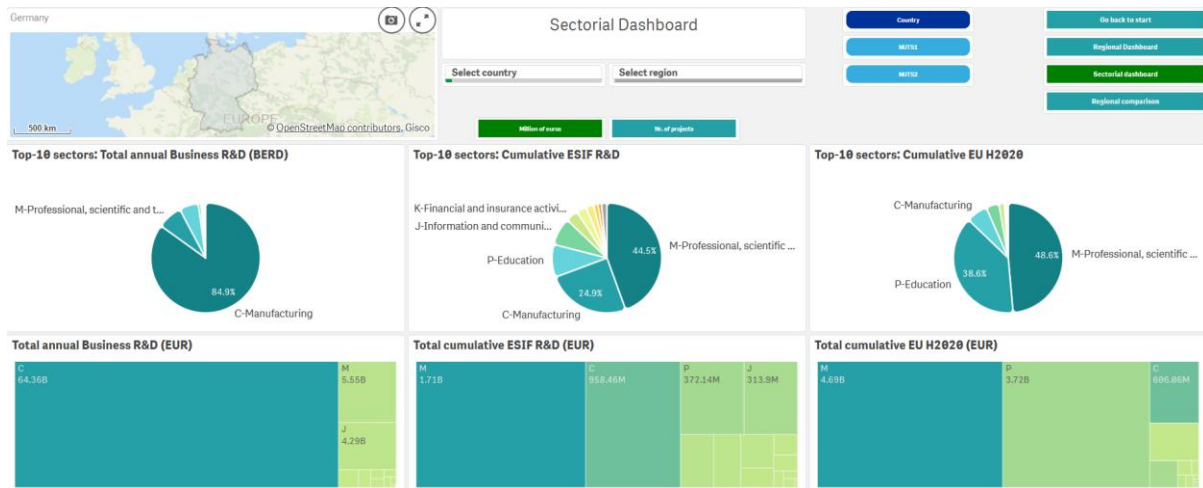

Source: TEDv (extracted on 31/03/2023).

5.2. Sectorial Dashboard

The Sectorial Dashboard displays the concentration of different R&I funds (ESIF and Horizon) by economic activity (NACE classification), expressed in relative (% total) and absolute terms (EUR or the number of projects/operations). It also reports the sectorial concentration of the total annual R&D expenditures (EUR) in a territory, extracted from EUROSTAT (country total). Those data include R&D statistics from Business R&D expenditures (BERD) by economic activities.

The territorial statistics displayed in this dashboard allow us to see if the different R&I funds show a similar sectorial concentration and whether they are similar to the territory trend in terms of R&D expenditures concentration.

Figure 6 illustrates an example of the TEDv visualisation for Germany. For instance, it shows that even if the manufacturing industry is responsible for around 85% of the total Business R&D expenditure and 58% of the total R&D of the country, this economic activity is mainly financed by EU funds coming from Cohesion Policy programmes. Indeed, H2020 is mostly concentrated on specialised services (NACE section M) and education (NACE section P). Indeed, these activities captured around 87% of total H2020 funds.

Figure 6. Example of TEDv visualisation: Sectorial Dashboard – Germany
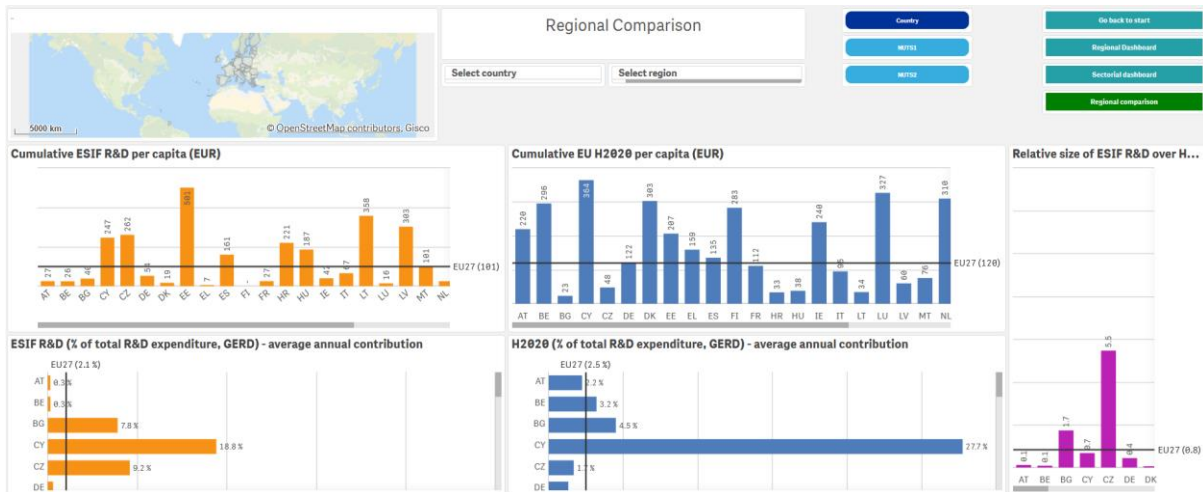

Source: TEDv (extracted on 31/03/2023).

5.3. Comparison Dashboard

The Regional Comparison allows comparing territories in terms of the allocation of different R&I funding (ESIF and Horizon). To this purpose, territorial statistics on R&I funding (ESIF and H2020) are expressed in per capita terms, % of total annual R&D expenditure, and in relative size ratio. The comparison between territories regarding the territorial statistics listed above is available under four options:

1) It shows the territorial statistics for all the EU countries (Figure 7)
2) It displays the territorial statistics for all the NUTS 2-level regions (see example in Figure 8) or NUTS 1-level regions in a selected country
3) It allows the user to own select the different territories to compare expressed in the same NUTS level (see example in Figure 9)
4) It allows the user to compare a NUTS 2-level region with the three most similar regions (irrespectively of the country) - predetermined in the TEDv (see example in Figure 10) or the three most similar regions outside the country.

Figure 7. Example of TEDv visualisation: Comparison dashboard – EU Member states
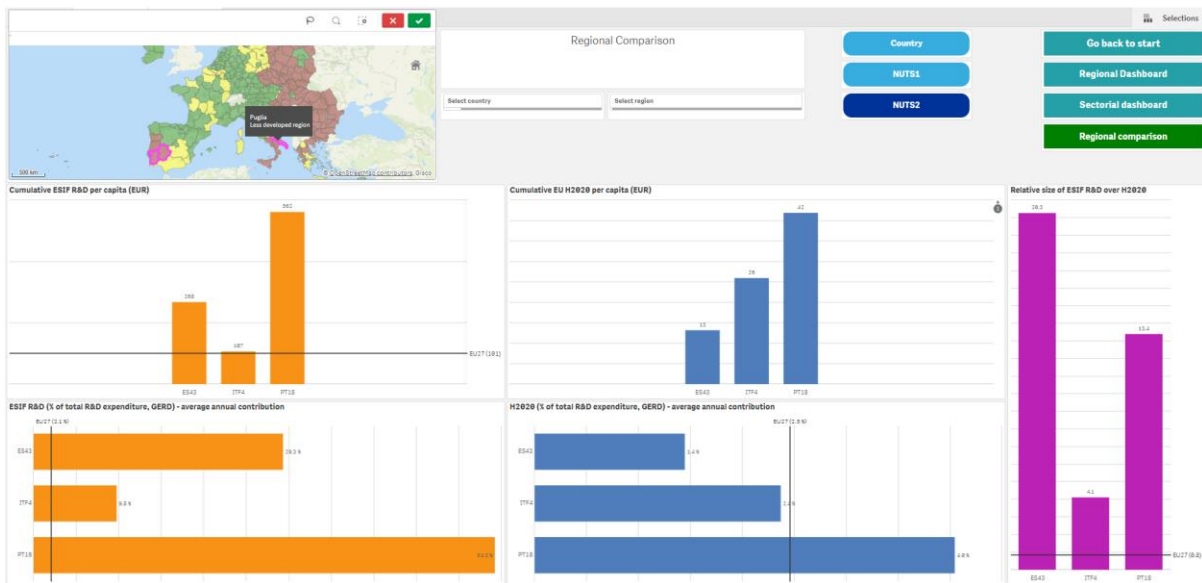

Source: TEDv (extracted on 31/03/2023).

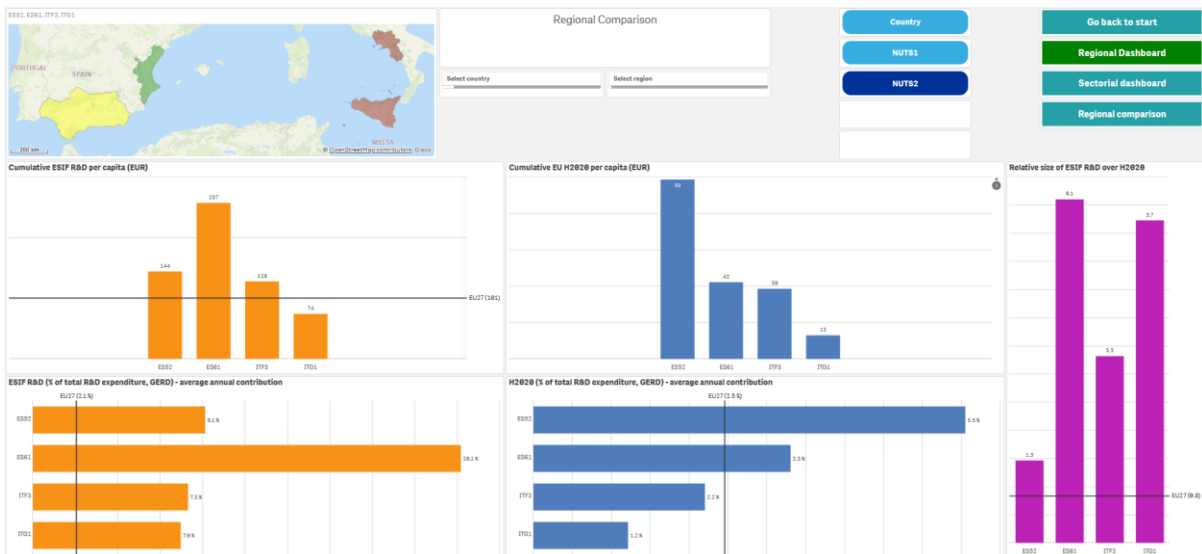Figure 8. Example of TEDv visualisation: Comparison dashboard – all Spanish regions
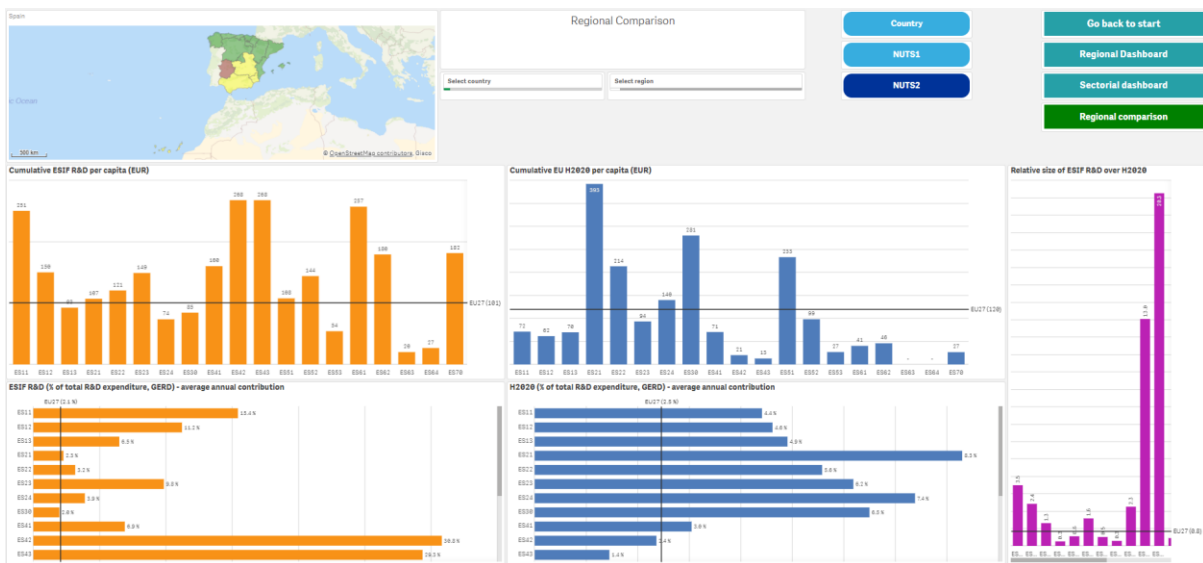
Source: TEDv (extracted on 31/03/2023).

Figure 9. Example of TEDv visualisation: Comparison dashboard – user selection

Source: TEDv (extracted on 31/03/2023).

Figure 10. Example of TEDv visualisation: Comparison dashboard, Andalucia (ES61) – peers regions selection

Source: TEDv (extracted on 31/03/2023).

The fourth options available in this dashboard can be particularly useful for policymakers, for instance, to see how much competitive R&I funds (e.g. through H2020 of HE) the region has managed to attract compared to regions with similar characteristics. The example illustrated in Figure 10 shows that the Spanish region of Andalucía (ES61) has attracted less H2020 funding per capita comparing with its national peer (Comunitat Valenciana – ES52), but more than its peer regions located outside Spain (Campania – ITF3 – and Sicilia – ITG1).

5.4. Regional info-sheet

The 'Regional info-sheet' dashboard reports regional statistics beyond R&I funding indicators, and aims to complement the information available in the other dashboards. It allows to compares the regional performance over time for some selected socioeconomic and demographic indicators with the country and EU average, as well as regarding well-know composite indicators like Regional Innovation Scoreboard (RIS) and Regional Competitiveness index (RCI). Those indicators are divided by four tabs, as described below:

- *Overview*:
  - Classification of the region according to the Cohesion criteria in the period 2014-2020 (more-developed, less-developed or transition region) according to the Commission Implementing Decision (2014/99/EU);
  - Regional Innovation Scoreboard[5]: reports the category and ranking of the region (last year available) and its evolution over time
  - Regional competitiveness Index[6]: displays the ranking of the region (last year available)
- *Economic*:
  - GDP per capita expressed in Purchasing Power Parity (PPP) and relative term face to EU27, where the EU27 is equal to 100 (EUROSTAT, nama_10r_2gdp)
  - Annual growth rates of the Total Factor Productivity: indicator measuring change in production's efficiency and extracted from Kostarakos (2023)
  - Investment rate: estimated by the ratio between the capital investment (Gross Fixed Capital Formation – GFCF – EUROSTAT, nama_10r_2gva) and output (Gross Value Added – GVA – EUROSTAT, nama_10r_2gfcf)
- *Employment and skills*:
  - Unemployment rate (EUROSTAT, lfst_r_lfu3rt): refers to the number of people unemployed (15 to 74 years of age) as a percentage of the labour force (active population);
  - Share of persons employed in science and technology (EUROSTAT, tgs00038): displays the total of human resources in science and technology (HRST) as a share of the active population in the age group 15-74. HRST refers to persons having successfully completed an education at the third level or being employed in science and technology.
  - Share of employment with tertiary education (EUROSTAT, lfst_r_lfe2emprtn): employment with an educational attainment level between 5 and 8 (tertiary education) according to the International Standard Classification of Education (ISCED 2011) over total employment from 15 to 64 years.

---

[5] https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/regional-innovation-scoreboard_en

[6] https://ec.europa.eu/regional_policy/information-sources/maps/regional-competitiveness_en

- *Demographic*:
  - Population density (EUROSTAT, demo_r_d3dens): displays the number of persons per square kilometer;
  - Old-age dependency ratio: reports the number of elderly people at an age when they are generally economically inactive (65 years old and over), compared to the number of people of working age (15-64 years old). Estimated using EUROSTAT - demo_r_d2jan;
  - Projected relative change of the population in 2030, 2040 and 2050 using as baseline scenario the population in 2019 (EUROSTAT, proj_19rp3) shows the estimated percentage change of population in comparison to 2019.

When past time series is displayed, the dashboard allows the users to select the year they want to visualize. Figure 11, Figure 12 and Figure 13 show examples of the visualisation available for each of the Regional info-sheet dashboards using the case of the Romanian's region Centru (RO12). For example, it shows that regions' per capita GDP is below the EU average but has improved over time (Figure 11 - bottom left graph). Furthermore, even though the job qualification in the region is below the EU, the regions' performance is above the country average (Figure 12 – bottom center and right graphs).

Figure 11. Example of TEDv visualisation: Regional info-sheet – Economic dashboard, Romanian's region Centru (RO12)
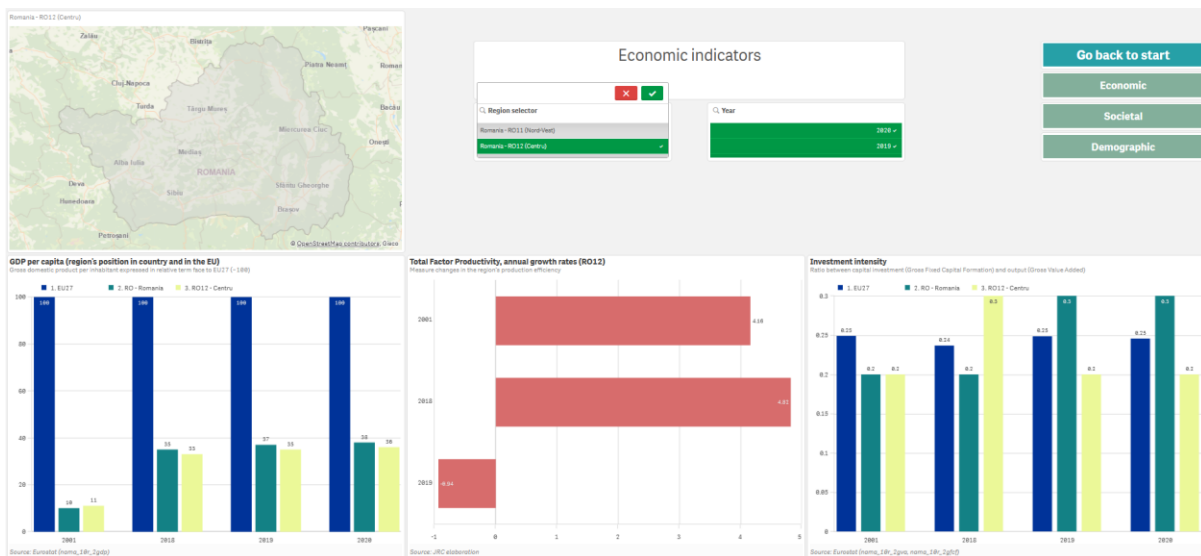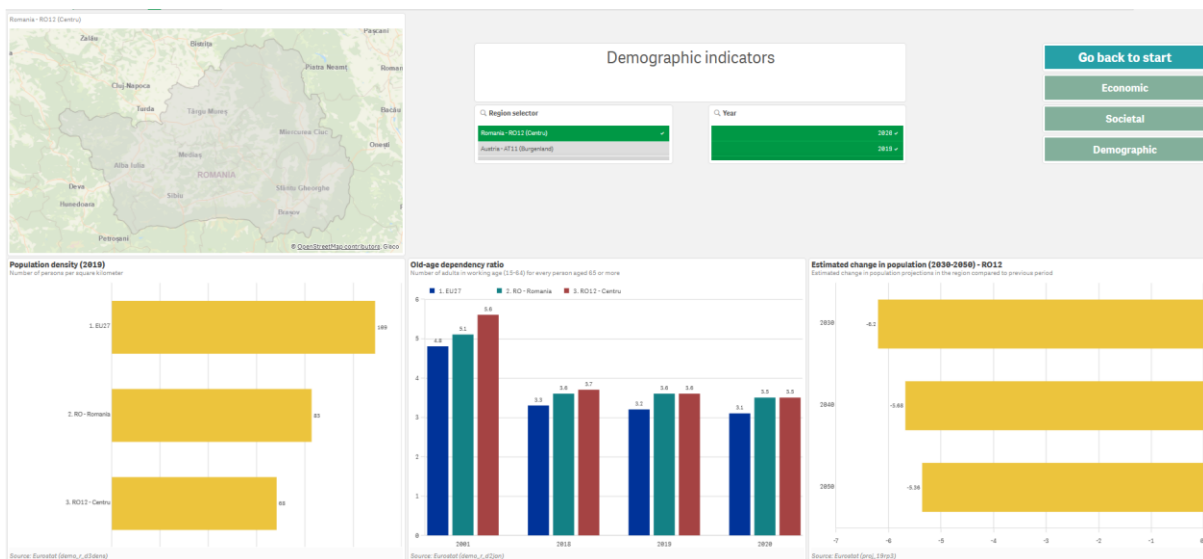

Source: TEDv (extracted on 31/03/2023).

Figure 12. Example of TEDv visualisation: Regional info-sheet – Employment and skills dashboard, Romanian's region Centru (RO12)
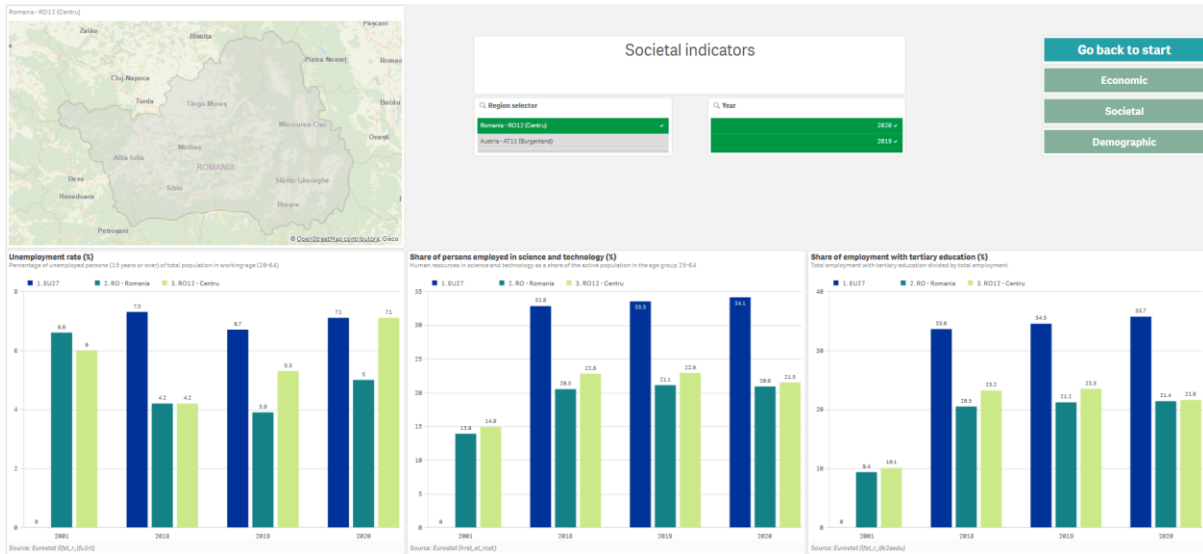
Source: TEDv (extracted on 31/03/2023).

Figure 13. Example of TEDv visualisation: Regional info-sheet – Demographic dashboard, Romanian's region Centru (RO12)

Source: TEDv (extracted on 31/03/2023).

6. Policy relevance of TEDv

The present working paper provides a detailed description of the data and methodological approach used to develop the Territorial Economic Data viewer (TEDv) launched by the Joint Research Centre. The TEDv is the first available tool that combines statistical territorial information from different EU funding programmes in a single and coherent framework, thanks to the methodological effort on territorial and

sectorial/thematic allocations (via taxonomy conversions). This allows users to compare the size of different EU funding programmes and their contribution to total R&D expenditures, making it a valuable resource for analysts and policy-makers. Furthermore, the Comparison dashboard includes a territorial benchmarking tool that allows appropriate comparison across regions, providing useful information to support policy cycle and design.

References

Bachtrögler, J., Arnold, E., Doussineau, M., and Reschenhofer, P. (2021). *UPDATE: Dataset of projects co-funded by the ERDF during the multi-annual financial framework 2014-2020*, JRC125008.

Cerqua, A., and Pellegrini, G. (2022). "I will survive! The impact of place-based policies when public transfers fade out", *Regional Studies*. DOI: 10.1080/00343404.2022.2136370

Conte, A. and Santos, A.M. (2023). "Territorial Mapping of EU Funding Programmes for Research and Innovation Activities". Edited by Capello, R. and Conte, A. in *Cities and Regions in Transition* (forthcominng)

Diemer, A., Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2022). "The regional development trap in Europe", *Economic Geography* 98(5), 487-509.

Joint Research Centre (2022). NUTS converter. https://urban.jrc.ec.europa.eu/nutsconverter

King, G., and Nielsen, R., (2019). "Why propensity scores should not be used for matching", *Political analysis* 27(4): 435-454.

Kostarakos, I. (2023). Regional productivity growth in the EU: An assessment of recent developments, *JRC Working Papers on Territorial Modelling and Analysis No. 04/2023*, European Commission, Seville, JRC133434.

Navarro, M., Gibaja, J. J., Franco, S., Murciego, A., Gianelle, C., Kleibrink, A., and Hegyi, F. B. (2014). Regional benchmarking in the smart specialisation process: Identification of reference regions based on structural similarity (No. JRC89819). Joint Research Centre.

Sapała, M. and Szczepański, M. (2022). "Poland's National Recovery and Resilience Plan", *Briefing Next Generation EU (NGEU) delivery – How are the Member States doing?*, European Parliamentary Research Service, Next Generation EU Monitoring Service, Members' Research Service, PE 733.665, September 2022.

Appendix

Appendix A. Horizon 2020

Table A1. Share of sector-classified H2020 funds, %

Country	H2020	ORBIS	Final
Austria	41.54	78.09	85.80
Belgium	40.08	73.54	77.47
Bulgaria	48.97	71.85	82.70
Cyprus	36.57	42.13	49.07
Czech Republic	50.64	65.89	68.86
Germany	42.24	74.09	82.34
Denmark	33.56	63.21	65.11
Estonia	30.50	82.63	83.40
Greece	44.24	46.00	54.49
Spain	45.96	95.49	95.82
Finland	37.46	96.53	96.99
France	40.15	71.43	75.82
Croatia	50.91	34.55	46.18
Hungary	42.83	62.40	67.05
Ireland	34.30	50.23	57.12
Italy	43.88	77.65	85.19
Lithuania	40.84	81.68	83.77
Luxembourg	33.78	75.00	78.38
Latvia	39.31	76.30	84.97
Malta	37.18	20.51	29.49
Netherlands	37.92	84.95	86.64
Poland	41.14	59.41	62.60
Portugal	49.89	89.80	91.34
Romania	48.07	44.01	64.27
Sweden	38.69	94.22	94.81
Slovenia	53.90	96.73	96.73
Slovakia	53.16	79.32	82.70
Average	42.14	69.91	75.15

Notes: H2020 stands for the percentage of NACE-classified H2020 funds in the original H2020 Grants database; ORBIS stands for the percentage of NACE-classified H2020 funds in the BvD ORBIS-enriched H2020 grants database; Final – stands for the percentage of NACE-classified H2020 funds in the final version of the database, after adding the data from BvD ORBIS and applying string-search technique to classify funds with missing NACE information.

Appendix B. An illustration of the territorial benchmarking tool included in the comparison dashboard

We illustrate the benchmarking tool by taking as a reference the Spanish region of Andalucía (ES61). As displayed in Figure B1, according to the matching procedure outlined in the previous section, the three most similar EU regions to Andalucía are i) Comunitat Valenciana, ii) Campania, and iii) Sicilia.

Figure B1. Andalucía and the three most similar regions
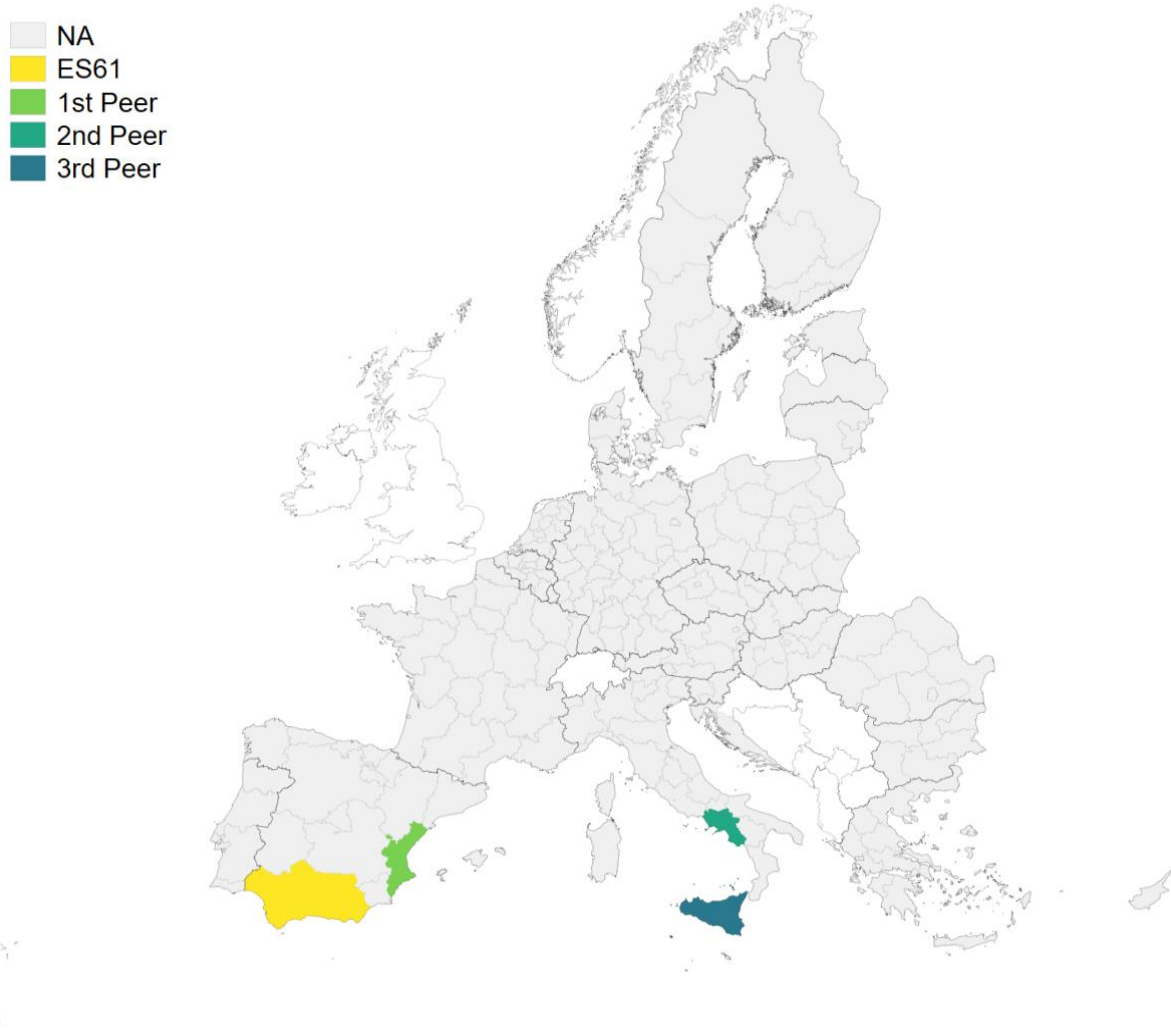
In Figure B2, we can inspect how close Andalucía is to its closest peer (Comunitat Valenciana – ES52) by plotting a selected number of variables used in the matching outlined in the previous section. For each variable considered, the blue lines represent, respectively, the target region (Andalucía on the right) and its closest first peer (Comunitat Valenciana on the left). The remaining gray lines represent all remaining NUTS2 regions that are not picked up by the matching algorithm as they are not as similar. We can see that the Comunitat Valenciana displays very similar values to Andalucía. This is true for both levels and trends over time.

Figure B2. Comparing variables across Andalucía (ES61) and Comunitat Valenciana (ES52) - 2011-2020.
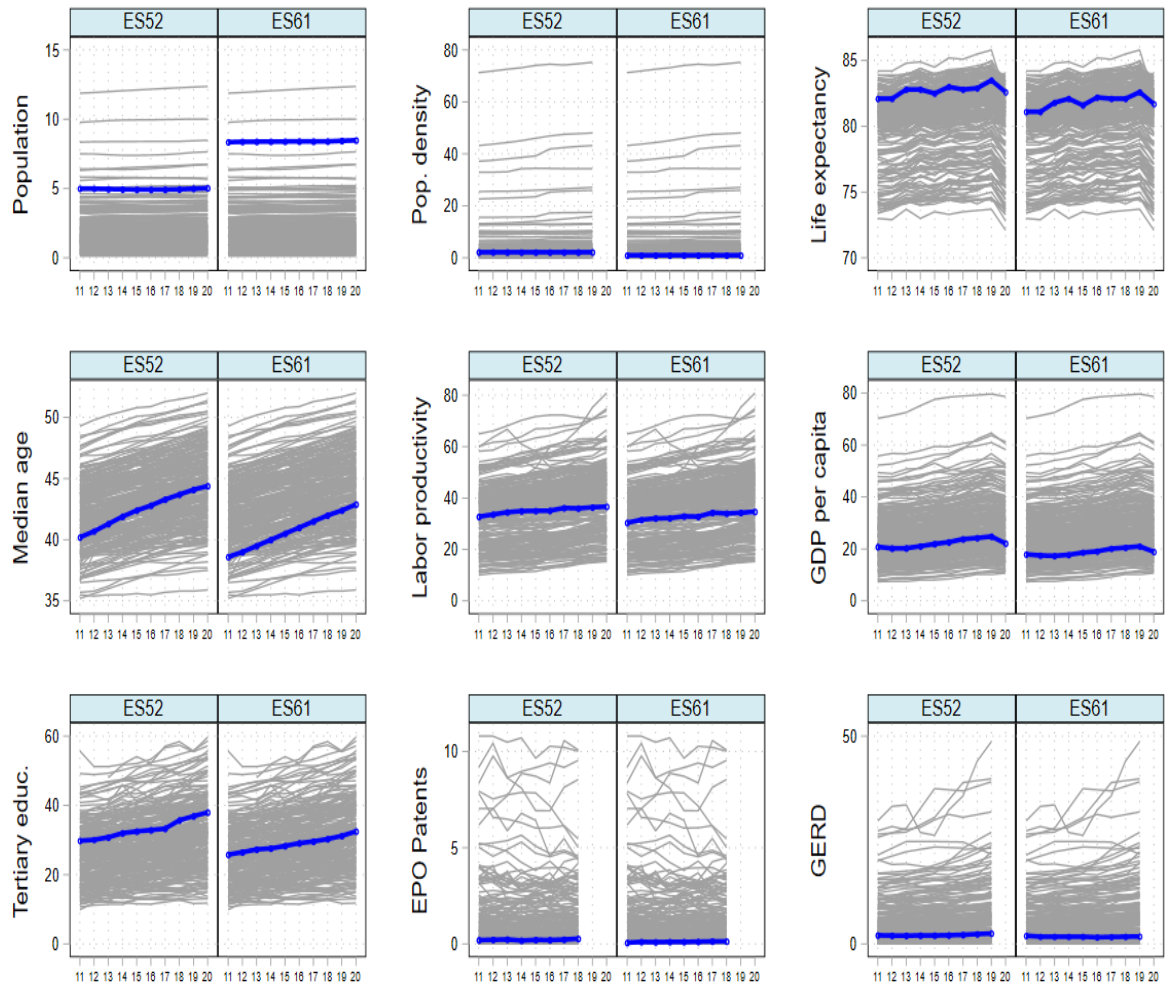
Definitions

- Average annual ESIF R&D: cumulative ESIF R&D divided by 7 years (2014-2020)

- Average Annual EU H2020: cumulative EU H2020 divided by 7 years (2014-2020)

- Average annual RRF R&I: cumulative RRF R&I divided by 6 years (2021-2026)

- Cumulative ESIF R&D per capita: cumulative ESIF R&D over total population

- Cumulative ESIF R&D: cumulative amount of EU funding share of total eligible costs decided until 2021 under the TO1

- Cumulative EU H2020 per capita: cumulative EU H2020 over total population

- Cumulative EU H2020: total H2020 funds allocated in programming period 2014-2020

- Cumulative RRF R&I: total cost estimated for R&I expenditures (2021-2026)

- ESIF R&D (% GERD): average annual ESIF R&D over total R&D expenditure (GERD)

- ESIF R&D over H2020: cumulative ESIF R&D over the cumulative Horizon 2020 (net EU contribution)

- ESIF R&D over RRF R&I: cumulative ESIF R&D over cumulative RRF R&D (only available at country level)

- ESIF R&D over RRF R&I: cumulative ESIF R&D over cumulative RRF R&D (only available at country level)

- GERD: Gross Domestic Expenditures on R&D, including the sum of R&D undertaken by resident businesses (BERD), universities and research institutes (HERD), the government (GOVRD), and private non-profit organizations (PNPRD).

- H2020 (% GERD): average annual EU H2020 over total R&D expenditure (GERD)

- H2020 over ESIF R&D: cumulative EU H2020 over cumulative ESIF R&D

- H2020 over RRF R&D: cumulative EU H2020 over cumulative RRF R&D

- H2020 over RRF R&I: cumulative EU H2020 over cumulative RRF R&I

Glossary

- BERD: Business enterprises expenditures on R&D.

- CF: Cohesion Fund

- EAFRD: European Agricultural Fund for Rural Development

- ECFIN: Commission's Directorate-General for Economic and Financial Affairs

- eCORDA: COmmon Research DAta Warehouse

- EMFF: European Maritime and Fisheries Fund

- EPRS: European Parliamentary Research Service

- ERDF: European Regional Development Fund

- ESF: European Social Fund

- ESIF: European structural and investment funds

- EU: European Union

- GDP: Gross Domestic Product

- GERD: Gross Domestic Expenditures on R&D

- GFCF: Gross Fixed Capital Formation

- GOVRD: Government expenditures on R&D

- GVA: Gross Value Added

- H2020: Horizon 2020

- HE: Horizon Europe

- HERD: Higher education expenditures on R&D

- HRST: human resources in science and technology

- ISCED 2011: International Standard Classification of Education

- MDM: Mahalanobis Distance Matching

- MS: Member States

- NACE: Statistical classification of economic activities in the European Community

- NRRP: National Recovery and Resilience Plan

- NUTS: Nomenclature of territorial units for statistics

- PNPRD: Private non-profit expenditures on R&D

- PPP: Purchasing Power Parity

- R&D: Research and Development

- R&I: Research and Innovation

- RCI: Regional Competitiveness Index

- RIS: Regional Innovation Scoreboard

- RRF: Recovery and Resilience Facility

- RRP: Recovery and Resilience Plan
- TC: Territorial Cooperation
- TEDv: Territorial Economic Data viewer
- TFP: Total Factor Productivity
- TO: Thematic Objective
- TO1: Thematic Objective 1 - Research and Innovation