OECD Regional Innovation and Education indicators dataset - *Metadata*

Dataset access	http://stats.oecd.org/Index.aspx?DataSetCode=REGION_SOCIAL
Contact	GovRegionStat@oecd.org
Dataset last update	15-July-2015
Next dataset major update	March 2016
Metadata content	**1. OECD Regional Database -** *General Overview (common to all regional datasets)* 1.1 Objective 1.2 Datasets 1.3 Geography covered a) Countries b) Territorial levels c) Regional Typology 1.4 Dimensions common to all datasets **2. Regional Innovation and Education indicators dataset -** *Specifics* 2.1 Innovation and Education Indicators 2.2 Innovation and Education Data specificities - Data definitions

1. OECD Regional Database: General Overview (common to all regional datasets)

1.1 Objective

The Regional Database contains annual data from 1995 to the most recent available year (e.g. in July 2015, data are generally available up to 2014 for demographic and labour market data, 2013 for regional accounts, innovation and social statistics).

The data collection is undertaken by the Directorate of Public Governance and Territorial Development, within the Regional Development Policy division (GOV/RDP). Statistics are collected through an annual questionnaire sent to the delegates of the Working Party on Territorial Indicators (WPTI), and through access to the web-sites of National Statistical Offices and Eurostat.

The WPTI is responsible for developing regional (subnational) and urban statistics and providing analysis to support policy evaluations. The Regional Database includes statistics on the regional distribution of resources, regional disparities, and how regions contribute to national growth and the well-being of society. Under this framework, the Regional Database is one of the pillars for providing indicators to the publication *OECD Regions at a Glance* (link).

1.2 Datasets

The OECD Regional database is composed by five datasets:

- **Regional demography** (population and number of deaths by age and gender; population density; life expectancy; infant mortality; inter-regional migration; demographics indicators)
- **Regional economy** (regional GDP; GVA by industry; employment by industry; labour productivity; labour utilisation; primary and disposable income; growth index)
- **Regional labour** (labour force at place of residence; employment and unemployment by gender; part-time employment; long term and youth unemployment)
- **Regional innovation** (patent and co patent by technology - fractional count, by inventor and priority year ; R&D expenses and R&D employees; Labour Force education attainment and Student enrolment by ISCED level)
- **Regional social** (*Environment:* air pollution, municipality waste; *Health:* physician and hospital beds density; *Safety:* homicides, car theft, mortality due to transport; *Social inclusion*: rate of young people neither in employment nor in education and training (NEET), rate of early leavers from education and training, *Housing*: number of rooms per capita, housing cost as a share of household income).

1.3 Geography covered

a) Countries

The Regional Database covers subnational statistics for the 34 OECD countries, plus 10 non-OECD member countries (Brazil, Colombia, Costa Rica, China, India, Indonesia, Latvia, Lithuania, Russia and South Africa).

b) Territorial Levels (TLs)

In order to facilitate greater comparability of regions, the OECD has classified two levels of subnational units. This classification is officially established, relatively stable in all member countries, and is used by many countries as framework for implementing regional policies:

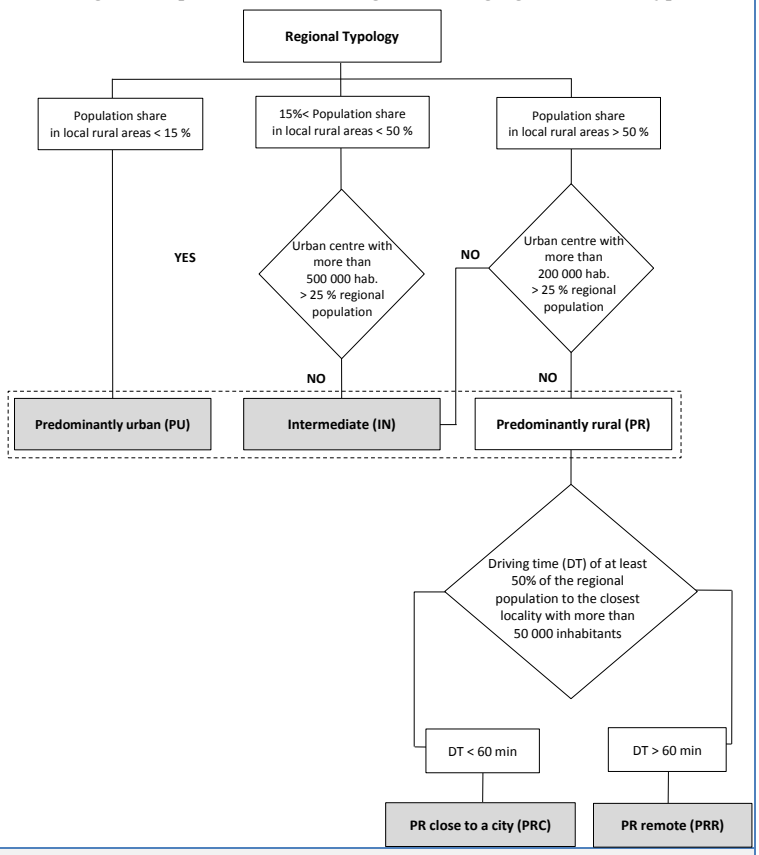
- Territorial Level 2 (TL2): covers the first administrative tier of subnational government, consisting in 362 large regions for the OECD zone. No regions have been defined for Luxembourg.
- Territorial Level 3 (TL3): is composed by 1 802 small regions under the OECD zone. TL3 regions are available for OECD countries plus Latvia and Lithuania. During 2015, Portugal is in the process of implementing new TL3 regions, therefore both classifications are reported in the Regional Database during this process.
- Non Official Grid (NOG): labour-market indicators in Canada are presented for groups of TL3 regions, labelled as non-official grids.

Country	Territorial level 2	Territorial level 3
Australia	States/territories (8)	Statistical divisions (60)
Austria	Bundesländer (9)	Gruppen von Politischen Bezirken (35)
Belgium	Régions (3)	Provinces (11)
Canada	Provinces and territories (13)	Census divisions (288)
		For Labour statistics, Non Official Grid (NOG): Economic areas census 2011, group of TL3 (72)
Chile	Regions (15)	Provincias (54)
Czech Republic	Oblasti (8)	Kraje (14)
Denmark	Regioner (5)	Landsdeler (11)
Estonia	Region (1)	Groups of maakond (5)
Finland	Suuralueet (5)	Maakunnat (19)
France	Régions (22)	Départements (96)
Germany	Länder (16)	Spatial planning regions (96)
Greece	Groups of development regions (4)	Development regions (13)
Hungary	Planning statistical regions (7)	Counties + Budapest (20)
Iceland	Regions (2)	Landsvaedi (8)
Ireland	Groups regional authority regions (2)	Regional authority regions (8)
Israel	Districts (7)	-
Italy	Regioni (21)	Province (110)
Japan	Groups of prefectures (10)	Prefectures (47)
Korea	Regions (7)	Special city, metrop. area and province (16)
Luxembourg	State (1)	State (1)
Mexico	Estados (32)	Grupos de municipios (209)
Netherlands	Landsdelen (4)	Provinces (12)
New Zealand	Groups of regional councils (2)	Regional councils (14)
Norway	Landsdeler (7)	Fylker (19)
Poland	Vojewodztwa (16)	Podregiony (66)
Portugal	Comissaoes de coordenaçao e desenvolvimento regional & regioes autonomas (7)	Grupos de municipios (30)
Portugal (new TL3 regions)		Grupos de municipios (25)
Slovak Republic	Zoskupenia krajov (4)	Kraj (8)
Slovenia	Kohezijske regije (2)	Statistične regije (12)
Spain	Comunidades autonomas (19)	Provincias (59)
Sweden	Riksomraden (8)	Län (21)
Switzerland	Grandes regions (7)	Cantons (26)
Turkey	Regions (26)	Provinces (81)
United Kingdom	Regions and countries (12)	Upper tier authorities or groups of lower tier authorities or groups of unitary authorities or LECs or groups of districts (139)
United States	States and the District of Columbia (51)	Economic areas (179)
Brazil	Estados + districto federal (27)	-
China	Provinces (31)	-
Colombia	Departamentos (32) and Capital District	-
Costa Rica	Provinces (7)	-
India	States and union territories (35)	-
Indonesia	Provinces (33)	-
Latvia	Region (1)	Statistical regions (6)
Lithuania	Region (1)	Counties (10)
Russian Federation	Oblast or okrug (83)	-
South Africa	Provinces (9)	-

c) Regional Typology

The OECD has established a regional typology for small regions (TL3 level), to better take in account the different "geography" of each geographic unit. This typology, based on settlement patterns calculated on the percentage of population living in rural communities, enables meaningful comparisons between regions belonging to the same type and level.

The OECD regional typology is primarily based on a criterion which identifies rural communities according to population density. A community is defined as rural if its population density is below 150 inhabitants per km2 (500 inhabitants for Japan and Korea to account for the fact that the national population density exceeds 300 inhabitants per km2). The next steps of the methodology follows the scheme presented on the right. For further information, the detailed methodology is available here.



Thus, each TL3 regions have been classified as:

- Predominantly Urban (PU)
- Intermediate (IN)
- Predominantly Rural (PR)

Predominantly Rural regions are further distinguished between rural regions that are:

- Close to larger urban centres (PRC)
- Remote (PRR)

This extended typology is currently not available for Australia and Chile, and for which rural regions are therefore presented under predominantly rural (PR).

1.4 Dimensions common to all datasets

The data selection in the datasets can be done through the following common dimensions:

- **Territorial Level and Typology**: data can be selected at different levels:
  - country level, code=[1]
  - large regions, code=[2]
  - small regions: predominantly urban [3_PU], intermediate [3_IN], predominantly rural close to a city [3_PRC], predominantly rural remote [3_PRR] (for Australia and Chile, only [3_PR] currently available for rural regions).
  - aggregation of typology at country level: predominantly urban [1_PU], intermediate [1_IN], predominantly rural [1_PR], predominantly rural close to a city [1_PRC], predominantly rural remote [1_PRR]
- **Region**: regions are presented as a tree structure, in other words, TL2 large regions are contained in their respective countries and TL3 small regions are contained under their respective TL2. There are two exceptions to this rule:
  - The United States, for which not all TL3 are contained in a single TL2, therefore the list of TL3 regions are presented below the list of TL2 regions.
  - The TL3 region 'Ost-Friesland' in Germany [code=DE12] which is contained in two TL2 regions, the 'Lower Saxony' [DE9] and 'Bremen' [DE5]. In order to ease the selection, this region has been put under 'Lower Saxony' [DE9], since more than 80% of its population is located in this TL2, and therefore, the values for these two TL2 don't correspond to the sum of the TL3 shown under their hierarchy.
- **Indicator**: the list of indicators is specific to the datasets (cf. next box).
- **Position**: this dimension allows to show the regional disparities the selection of extreme regional values within a country, by territorial level and by year. All regions can be selected through the position 'All regions' [code=ALL], which is the selection by default, or only the selection of regions that have the highest and the lowest values can be selected [codes=MAX,MIN]. This can be done only for a selection of relevant indicators, which is specific to each dataset. The highest or lowest values can return several values if more than one region have the same value. Highest and lowest values are retuned without taking in account if the indicator is a 'positive indicator' like life expectancy, or a 'negative indicator' like unemployment rate.
- **Time:** reference year

In addition to these common dimensions, some datasets have specific dimensions:

- **Gender** (for Regional Demography and Labour datasets)
- **Series** and **Measure** (for the Regional Economy dataset)

2. Regional Innovation and Education indicators dataset

2.1 Innovation and Education Indicators

- **Student Enrolment by Level of Education**
  - Student Enrolment Total
  - Student Enrolment at Elementary Level (ISCED 0-2)
  - Student Enrolment at Secondary Level (ISCED 3-4)
  - Student Enrolment at Tertiary Level (ISCED 5-6)
  - Student Enrolment Total Rate (in % of population)
  - Student Enrolment at Elementary Level Rate (in % of population)
  - Student Enrolment at Secondary Level Rate (in % of population)
  - Student Enrolment at Tertiary Level Rate (in % of population)

- **Educational Attainments of the Labour Force**
  - Labour Force Total
  - Labour Force with Elementary Education (ISCED 0-2)
  - Labour Force with Secondary education (ISCED 3-4)
  - Labour Force with Tertiary education (ISCED 5-8)
  - Labour Force with Unknown Educational Level
  - Share of Labour Force with Elementary Education (in % of labour force)
  - Share of Labour Force with Secondary Education (in % of labour force)
  - Share of Labour Force with Tertiary Education (in % of labour force)

- **R&D Personnel by Sector**
  - R&D Total Personnel
  - R&D Personnel Employed by the Business Sector
  - R&D Personnel Employed by the Government Sector
  - R&D Personnel Employed by the Higher Education Sector
  - R&D Personnel Employed by the Private and Non-Profit Sector
  - R&D Total Personnel Rate (in % of total employment)
  - R&D Personnel Employed by the Business Sector Rate (in % of total employment)
  - R&D Personnel Employed by the Government Sector Rate (in % of total employment)
  - R&D Personnel Employed by the Higher Education Sector Rate (in % of total employment)
  - R&D Personnel Employed by the Private and Non-Profit Sector Rate (in % of total employment)

- **R&D Expenditures by performing sector**
  - R&D Total Expenditure (millions of current national currency)
  - R&D Expenditures Performed by the Business Sector (millions of current national currency)
  - R&D Expenditures Performed by the Government Sector (millions of current national currency)
  - R&D Expenditures Performed by the Higher Education Sector (millions of current national currency)
  - R&D Expenditures Performed by the Private and Non-Profit Sector (millions of current national currency)

  - R&D Total Expenditure (millions USD, constant prices, constant PPP, base year 2010)
  - R&D Expenditures Performed by the Business Sector (millions USD, constant prices, constant PPP)
  - R&D Expenditures Performed by the Government Sector (millions USD, constant prices, constant PPP)
  - R&D Expenditures Performed by the Higher Education Sector (millions USD, constant prices, constant PPP)
  - R&D Expenditures Performed by the Private and Non-Profit Sector (millions USD, constant prices & PPP)

  Share of R&D Total Expenditure (in % of GDP)
  - Share of R&D Expenditures Performed by the Business Sector (in % of GDP)
  - Share of R&D Expenditures Performed by the Government Sector (in % of GDP)
  - Share of R&D Expenditures Performed by the Higher Education Sector (in % of GDP)
  - Share of R&D Expenditures Performed by the Private and Non-Profit Sector (in % of GDP)

- **Patent applications - PCT applications (fractional count; by inventor place of residence and priority year)**
  - PCT Patent Applications, count (fractional count; by inventor and priority year)
  - PCT patent applications per million inhabitants (fractional count; by inventor and priority year) - level

  Patents collaboration by regional location - PCT applications (fractional count; by inventor and priority year):
  - PCT co-patent applications, count (fractional count; by inventor and priority year)
  - Percent of PCT co-patent applications that are done within the region
  - Percent of PCT co-patent applications that are done within the country
  - Percent of PCT co-patent applications that are done with foreign regions

  Patents international collaboration, ownership
  - Domestic ownership of foreign patents
  - Foreign ownership of domestic patents

2.3 Innovation and Education Data specificities

.

Data definitions

- *Student enrolment* is defined as the total number of students enrolled in all types of schools and education institutions, including public, private and all other institutions that provide organised educational programmes regardless of age. A student is a person of either sex, not classified as usually economically active, who attends any regular educational institution, public or private, for systematic instruction at any level of education. Levels of education are classified according to International Standard Classification for Education ISCED-97.
- *Educational attainment of the active population aged 15 years old or more:* Number of persons with primary, secondary and tertiary education. Educational attainment is defined as the highest grade completed within the most advanced level attended in the educational system of the country where the education was received Educational attainments are internationally standardised trough the ISCED 2011 (International Standard Classification for Education) used to define the levels and fields of education (1 - basic school 8-10 grade;2 - general upper secondary school; 2 - vocational upper secondary school; 4 - short-cycle higher education; 5 - medium-cycle higher education; 6 - bachelor; 7 - phd-degree). The ISCED classification is grouped due to data availability constraints as follows: ISCED 0-2, ISCED 3-4, ISCED 5-8 and Total.

- *R&D expenditures and R&D employed personnel* can be divided in different sectors:
    a) the business sector (see Frascati Manual section 3.4)
    b) the government sector (see Frascati Manual section 3.5)
    c) the higher education sector (see Frascati Manual section 3.7).
    d) the private and non-profit sector (see Frascati Manual section 3.6).
    e) The total R&D expenditures and total R&D personnel are the sum of the four performing sectors (business, government, higher education and private non-profit).

- *Patent indicators:* a patent is an exclusive right granted for an invention, which is a product or a process with industrial applicability that provides, in general, a new way of doing something, or offers a new technical solution to a problem ("inventive step"). A patent provides protection for the invention to the owner of the patent. The protection is granted for a limited period, generally 20 years. Data refer overall patent applications to Patent Cooperation Treaty (PCT) applications. Patent documents report the inventors (where the invention takes place), as well as the applicants (owners), along with their addresses and country of residence. Patent counts are based on the inventor's region of residence and fractional counts. If on the patent document are registered two or more inventors, the patent is classified as a co-patent.
    - Patent intensity is defined as the number of patent applications per million population in a region.
    - Domestic ownership of foreign patents is the percent of PCT patents that have one or more foreign inventors and one or more domestic applicants in the total number of patents owned domestically (i.e. with one or more domestic applicants)
    - Foreign ownership of domestic patents is the percent of PCT patents that have one or more domestic inventors and one or more foreign applicants in the total number of patents invented domestically (i.e. with one or more domestic inventors)