STI Microdata lab

Purpose

The STI Micro-data Lab is a data infrastructure project within STI Directorate, collects and links large-scale administrative and commercial micro-level datasets. These mainly relate to administrative data such as intellectual property (IP) assets, including patents, trademarks and registered designs; scientific publications; and information on companies from private providers. IP data are collected in the framework of the OECD-led IP Statistics Task Force, which gathers representatives from IP offices worldwide.

These micro-data, which complement and enhance official statistics like macro-aggregated or survey-based data, have the advantage of being granular in nature and comprehensive in time and geographical coverage.

By providing detailed information about the behaviour of economic agents and the way science and technology develop, these data help address policy-relevant questions, such as those related to the generation and diffusion of new technologies, the different ways in which firms innovate, science-industry links, researchers' mobility patterns or the role of knowledge-based assets in firms' economic performance.

Objectives and outputs

The STI Micro-data Lab serves as a platform for the development of new metrics and methodologies, and feeds into a large range of analyses. Experimental indicators built using this infrastructure are regularly published by the OECD, notably in the OECD Science, Technology and Industry Scoreboard. Details about the location of authors, inventor and designers or about applicants' names and addresses shed light on scientific, inventive, branding and design activities occurring in different countries, and on how different actors interact. The technology, design and product domains that IP rights protect can be inferred from the technology or product classification codes provided in IP documents.

Non-member countries involved in the activity:

Albania, Argentina, Armenia, Asia, Azerbaijan, Barbados, Belarus, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, China, Chinese Taipei, Colombia, Costa Rica, Croatia, Cyprus, Dominican Republic, Ecuador, Egypt, El Salvador, Europe, G20, Georgia, Guatemala, Honduras, Hong Kong, India, Indonesia, Jamaica, Kazakhstan, Kyrgyzstan, Liechtenstein, Lithuania, Macedonia, Malaysia, Malta, Moldova, Mongolia, Morocco, Nicaragua, Other, Panama, Paraguay, Peru, Philippines, Republic of Montenegro, Republic of Serbia, Romania, Russian Federation, Saudi Arabia, Serbia and Montenegro, Singapore, Slovenia Former, South Africa, Tajikistan, Thailand, Turkmenistan, Ukraine, Uruguay, Uzbekistan, Venezuela, World.

Main Developments for 2018

General aspects:

The different micro datasets of the STI Microdata Lab are being used in an independent fashion, e.g. to develop indicators related to specific analytical questions, or combined in such a way as to generate new information related to a broader array of issues or to more complex dynamics.

As IP data do not contain firm identification codes, IP rights are assigned to firms using harmonization and linking techniques to connect firms' and IP assignees' names using word-matching algorithms.

To improve the granularity of analysis, addresses are linked to about 5 500 intra-country regions in more than 40 countries.

Experimental data mining techniques have been developed to exploit information from IP and bibliographic data and identify accelerations in science and technology-related activities ("bursts").