

The Web Intelligence Hub – A tool for integrating web data in Official Statistics

Fernando Reis

Eurostat, European Commission, Luxembourg.

Abstract

The use of the World Wide Web as a source of big data has become quite common. However, the use of web data in regular statistical production following the quality standards expected in official statistics has particular requirements and is not easy. The Web Intelligence Hub (WIH) is a European Statistical System (ESS) statistical infrastructure built to address those requirements.

The WIH is being developed in the context of the Trusted Smart Statistics, an ESS initiative launched to address the issues raised by the use of new types of data and sources. The initiative is organised in hubs, with each hub specialising on data sources with similar characteristics and processing similar types of data. The WIH is the pillar of the TSS that provides the fundamental building blocks for harvesting information from the Web to be used in statistics.

The WIH implements the principles adopted for the trusted smart statistics. It follows a modular architecture in order to be easily evolvable to the technological changes occurring on the Web. The processes running in the hub and the corresponding outputs are open, transparent and auditable. The WIH will provide several services to the ESS, such as commonly agreed partnership models with web portals, commonly agreed data gathering, processing and statistical methodologies, algorithms and ready to run scripts. The data and scripts running in the WIH will also be available to the partners as a service.
