

8 JUNE 2018

## Estimates of modalities of use of websites by enterprises

Istat releases the estimates of the modalities of use of websites by enterprises, produced using directly Internet data. These estimates show the rate of enterprises on the total reference population that own or use a website in which are present:

1. web ordering functions (e-commerce component) are available;
2. information on job vacancies;
3. links to social media (Facebook, Twitter, Instagram etc.).

The estimates, referring to 2017, concern a reference population of about 184,000 enterprises with at least 10 persons employed operating in different sectors of economic activity and are calculated, for each variable of interest, as well as at national level, also for the following domains:

- the 4 classes of persons employed;
- the 16 sub-domains derived from the intersection between classes of persons employed and the four macro-sectors of economic activity;
- the 26 classes of economic activity;
- the 21 regions and autonomous provinces.

These estimates are currently produced every year using traditional techniques (data collection through a web questionnaire in a sample of enterprises), by the survey common to all EU member countries, on the use of information and communication technologies (ICT)<sup>1</sup>.

---

<sup>1</sup> The annual survey on information and communication technologies (ICT) has been carried out since 2001 by Istat to produce indicators on the use of ICT in enterprises with at least 10 employees active in industry and in non-financial services. Since 2004, the survey is part of the Community statistics on the information society as set out in the Commission Regulation n. 808/2004 on the basis of which the contents of the questionnaire are harmonized and made common among all Member States. The survey provides information on the activities carried out by the company on the Internet (website, cloud computing, social media), types of connection used (fixed and mobile broadband), on the use of IT tools adopted to manage internal activities (for example, use of ERP software, CRM) or to share electronically supply chain management information (customers, suppliers, other companies), on e-commerce, the use of specialist ICT skills, e-invoicing. The ICT survey is one of the main sources of data for the Digital Agenda Scoreboard which is used by the European Commission to measure the progress of the European digital economy and contributes to define 11 indicators of the Digitization Index of the Economy and Society (DESI).

The experimental estimates reported here are obtained by an alternative procedure which is divided into the following phases:

1. the addresses of the related websites (about 100,000) were found in part through a procedure that made use of search engines and personal data contained in the Statistical Archive of Active Enterprises (ASIA), partly using information already available (from survey or from the CONSODATA archive);
2. through the addresses obtained, web scraping was carried out and the text content captured wherever possible;
3. the approximately 85,000 texts thus obtained have been elaborated with Natural Language Processing techniques in order to obtain terms (n-grams) relevant for the purpose of predicting the different target variables;
4. in the subset of units for which both the answers provided to the survey and the texts captured on the web were available (about 12,000) it was possible to develop the value prediction tools (presence / absence) of each of the three target variables based on the relevant terms identified in the previous step; in the case of web ordering and the online job offer, "random forest" models were used, while in the case of social media, "information retrieval" techniques have been applied;
5. these instruments were applied to the 85,000 enterprises for which it was possible to collect and process the texts contained within the corresponding sites;
6. alternative estimates compared to the current survey estimates (survey estimates) were calculated with two different estimators: one full model based (model estimates) and the other based on a combination of predicted values and survey data (combined estimates).

In the excel files made available (one for each variable of interest) are shown:

- the particular classification domain;
- the estimate of the variable of interest currently produced by the *survey*, together with the details of the relative confidence interval;
- the *model* estimate based on the count of the values predicted for each single site accessed;
- the *combined* estimate, i.e. using both the survey data and the predicted values.

Furthermore, a table is presented only with model estimates for the three target variables by economic activity at the two-digit level of Nace Rev. 2 (62 divisions). This detail represents new information compared to that produced by the survey which provides only a more aggregated sectoral detail.

With reference to the estimated phenomena, it is opportune to specify some differences between target variables asked in the questionnaire and defined by the [Methodological manual](#) and those predicted by model. In the year 2017, the survey asks enterprises with at least 10 persons employed who claim to have an Internet connection, if they have a website and, to the sub-group responding positively, are asked what features or services are available on the site itself. The question does not refer exclusively to the enterprise' website, but to the use of a website by the enterprise, including third party websites such as, for example, the website of the group of enterprises to which the respondent belongs, the website of a holding enterprise or affiliated with the respondent enterprise which in any case has Internet pages dedicated to the respondent<sup>2</sup>.

Regarding the web ordering function offered by the website or the possibility of making online orders or bookings, the respondents are advised not to consider this functionality as present if, to conclude the operation, there is a need for additional offline contacts or for e-mail, while payment may not be included in the procedure.

With regard to the presence on the website of a link to social media, reference is made to the presence of links to an account or profile of the respondent enterprise on a specific social media.

Finally, in 2017, only in the Italian questionnaire and not in the European one, it is asked if web site provided information on job vacancies including both the presence of simple information on job offers, and the presence of online tools for submitting job applications.

To the production of these experimental statistics have collaborated:

- Istat: G. Barcaroli, G. Bianchi, N. Golini, A. Nurra, P. Righi, S. Salamone, F. Scalfati, M. Scannapieco, D. Summa, D. Zardetto
- CINECA: M. Scarnò
- Università di Roma Sapienza: R. Bruni

---

<sup>2</sup> In the definition of enterprises having a web site are not included those present only in e-marketplace sites (such as Amazon, Booking, Expedia, Groupon, etc.) in which they have the opportunity to advertise and sell their prices or those that appear in third-party websites only with the name, address, contact information (e.g. Yellow Pages).