



# .eu Insights

The EURid Insights series aims to analyse specific aspects of the domain-name environment. The reports are based on surveys, studies and research conducted by EURid in cooperation with industry experts and sector leaders.

## What's in a domain-name extension?

30-30-20-20: The secret code of the Internet

EURid analyses domain usage by categorising websites with .eu, .com, .net, .org, .info, .biz, .mobi and .pro extensions to illustrate how different domains reflect different uses.

## What's in a domain-name extension?

One often wonders what kind of websites are hidden behind a domain name. Is it mainly business? What is the market share of “pay-per-click” (PPC) sites? Is a specific usage linked to a specific extension?

These questions are often asked, but answers are difficult to find. Attempts have been made to determine particular aspects of website usage<sup>1</sup>. Research has been done to analyse and categorise websites found under the domain names that are registered in the major generic top-level domains (gTLDs). This report offers some insights into the use of domain names on the Internet.

# Pay-Per-Click Business Community Error Holding page Password Protected Institutional

## Method

Websites are rather difficult to categorise automatically. While it is theoretically possible to classify websites based on keywords found on the sites themselves, different languages and the way websites are used require human interpretation in order to obtain a reliable result. To this effect, EURid set up a test lab where evaluators visually assessed a random sample of websites found under each extension. Statistical methods were applied to estimate the error margin and automated scanning was used to verify some of the numbers.

The population of domain names used to determine the distribution among the specific usage categories was taken from the zone files of the eight studied TLDs. From these zone files, a random sample was generated which was then visually assessed by a team of testers. In total, around 5 000 domain names were assessed for each of the selected TLDs.

<sup>1</sup>One example of such undertakings is “Measuring the Perpetrators and Funders of Typosquatting” by Tyler Moore and Benjamin Edelman.

For the statistical analysis, an overall confidence level of 95% was used. The evolution of the distribution as the number of tested sites grew was carefully monitored and the convergence of the results was confirmed (see Exhibit 2).

This error margin does not take into account the subjective effect of the evaluators. Although all testers were trained to guarantee the same results as far as possible when presented with the same website, an additional control was added to incorporate the effect of subjectivity. A set of 200 URLs was assembled including a minimum of five sites from each category. The remaining sites were taken randomly from all categories. All evaluators were offered the same list to categorise and, for each website, a consensus was established based on what the majority of evaluators had chosen. Each evaluator was scored against the consensus. Subsequently, the distribution of the domain names over the different categories was statistically verified based on the number of divergences from the consensus in that category.

## Limits of the method

Zone files are readily available for most of the (g)TLDs and are therefore an ideal baseline for comparing different TLDs. However, this method does not take into account domain names that have been registered without configuring name servers (the so-called 'parked' domain names), which is typical of defensive registrations.

## Categories

The selection of categories is totally arbitrary and depends on what one wants to report. For this report, the choice was influenced by the questions EURid wanted to answer.

The (subjective) criteria used to classify the websites are summarised in Exhibit 1. The definitions of the categories clearly show that human intervention is needed to obtain reasonable results.

### ***Exhibit 1: Category descriptions***

**Business:** A site that clearly presents a commercial activity and is designed for customer interaction. These sites can be informative (presenting the services or products of a company) or allow visitors to shop online directly.

**Community:** A site that contains information for or about a community. Typical examples are sites for clubs, churches, families, etc. Personal websites, such as blogs, are another example. The key element of this category is that the sites' main objective is not commercial, although some merchandising, such as promoting the community, is acceptable.

**Pay-per-click:** A site that contains mainly advertising links. These come in many forms, but are generally easily recognisable by the fact that the same layout appears regularly in completely different contexts.

**Holding page:** This is typically a one-page website that can have several different formats:

- **A company holding page:** A page with no information other than the company's contact details – with or without the message that more information will be published later.
- **Under construction:** A page with just a few words telling the visitor they are viewing a holding page.
- **A (business) page of the hosting company:** some hosting companies set up a default webpage (even a full website) when their customer does not (yet) have a website available.

Remark: When the default page of the hosting company is a pay-per-click site, it is categorised as such and not as a holding page.

**Institutional:** A website that belongs to a government or government-related organisation. Examples are: city websites, public schools, government (sponsored) websites, etc.

**Password protected:** Websites that require the entry of a password before enough content is visible for the website to be categorised.

**Error:** Wrongly configured websites. Although the domain name was found in the zone file, it does not resolve to a 'www' website. This can be because the sub-domain 'www' was not configured or because it is only used for email.

**Pornography:** This was chosen as a separate category to evaluate the impact of pornography on the Internet.

## Results

The overall result across all considered TLDs is given in Table 1. Based on a sample of more than 41 000 sites, we found that around 27% of the websites of the investigated TLDs are business sites, followed by 24% of pay-per-click sites. Holding pages and domain names with no related website (or with a badly-configured site) each represent around 20%. In depth statistical analysis shows that a significant difference exists between the most important categories.

When a website was automatically redirected to a website with the same or a different extension, the domain name of the original extension was categorised according to the category of the website to which it was directed. Additional information on redirections can be found in Exhibit 3.

Quality-control checks indicated that some categories were very dynamic. For example, a site might fall into the Error category because of a temporary problem in the network (time out on the response) or an error in the configuration that was being fixed at the time of the evaluation, or the content of a website may have been temporarily changed due to hacking. Indications in that direction were found, as websites first categorised as pornography were recategorised as business websites upon re-investigation some days later.

Category	No. of sites	Pct
Business	11 259	27.3%
Error	8 695	21.1%
Holding page	7 809	19.0%
Institutional	1 194	2.9%
Password	534	1.3%
Community	1 477	3.5%
Pornography	408	1.0%
PPC	9 841	23.9%
<b>Sample Size</b>	<b>41 217</b>	<b>100.0%</b>

Table 1 – Average distribution of categories across all TLDs investigated

However, any research into website distribution will always be a snapshot of a given point in time. We would expect similar distributions (be it with different websites using the same sample) if the categorisation were to be redone at another time.

The details per TLD can be found in Table 2. Of all visually inspected .eu sites, 36.3% were categorised as business websites, which is significantly higher than the overall percentage and the highest among the TLDs. .mobi scores considerably lower than the others on business usage.

The number of websites falling into the Error category is very similar in all TLDs, but highest in .mobi. The latter can be explained by the high number of Chinese internationalised domain names (IDNs) in the sample, where firewalling issues were the cause of the non-responding website.

Holding pages are seen more frequently (in terms of percentage) in the more recent TLDs. While .com, .net and .org each have a distribution of around 12%, the younger TLDs have higher scores. .biz and .info are slightly higher and .eu, .pro and .mobi are above 20%.

Pay-per-click sites are notably present in all TLDs, ranging from a 22% share of the TLD portfolio in .mobi, to 29% in .info. The .eu TLD is clearly an exception, with only 14.5%.

Other categories are very limited in size. The prominent exception is .org, with about 9% of analysed sites falling into the Community category and a similar percentage in the Institutional category. Our research shows that, in terms of domain names, pornography websites are rather limited, although their presence is significantly higher in .com (more than 4% compared with less than 1% in all other TLDs).

	.eu	.com	.net	.org	.info	.biz	.mobi	.pro
Business	36.3%	30.5%	29.5%	23.9%	25.4%	31.5%	14.2%	26.7%
Error	19.1%	17.6%	22.9%	16.7%	22.8%	23.5%	27.3%	19.3%
Holding Page	23.8%	12.1%	11.8%	12.0%	14.9%	16.0%	34.3%	24.6%
Institutional	1.5%	2.3%	1.7%	8.2%	1.1%	3.1%	0.6%	5.4%
Password	1.4%	1.4%	1.6%	1.3%	1.1%	1.7%	0.7%	1.3%
Community	3.0%	4.2%	3.9%	8.7%	4.5%	1.6%	0.7%	1.6%
Pornography	0.3%	4.2%	0.7%	0.3%	1.0%	0.5%	0.4%	0.8%
PPC	14.6%	27.7%	27.9%	28.9%	29.2%	22.1%	21.8%	20.3%
<b>Sample Size</b>	<b>6 571</b>	<b>4 805</b>	<b>4 130</b>	<b>5 211</b>	<b>6 798</b>	<b>4 634</b>	<b>5 122</b>	<b>4 911</b>

Table 2 – Distribution of categories, by percentage, of all TLDs investigated during the study

The evolution of the category distribution was monitored and shows a stable situation as early as a sample size of 1 500 (see Exhibit 2).

### Exhibit 2: Convergence of the categorisation

During the process of categorisation, we kept track of the evolution of the distribution of each category in relation to the sample size. Figure 1, below, is an example of the convergence of categories of a TLD. It can be observed that the distribution stabilises within its error margins very rapidly.

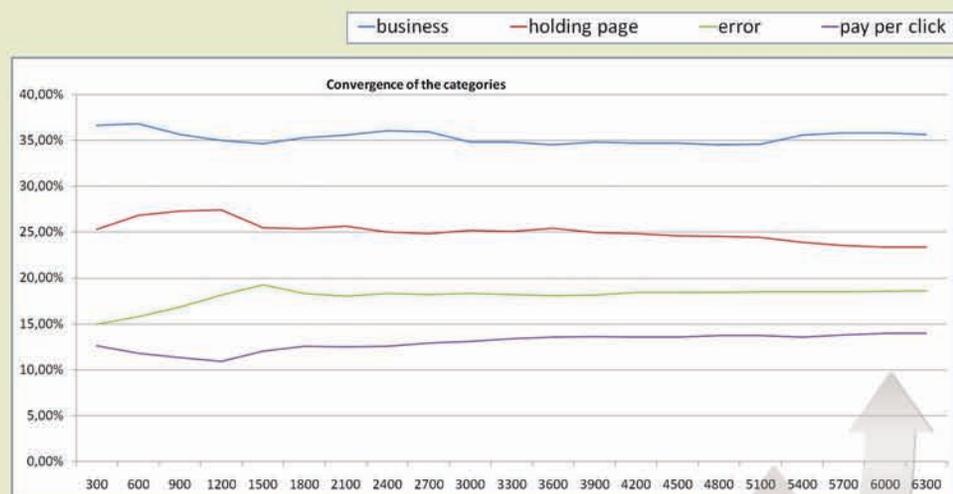


Figure 1 – Evolution of the distribution for one TLD as a function of the sample size

The significance of the differences in proportion of each category between .eu and the other TLDs was examined. The results, based on a confidence interval of 95%, are summarised in Table 3.

The indication 'significant' in Table 3 means that the measured difference between .eu and the TLD in column one is statistically significant for that category (in the top row).

	Business	Holding page	Pay-per-click	Error
.com	Significant	Significant	Significant	Not significant
.net	Significant	Significant	Significant	Significant
.org	Significant	Significant	Significant	Significant
.info	Significant	Significant	Significant	Significant
.biz	Significant	Significant	Significant	Significant
.mobi	Significant	Significant	Significant	Significant
.pro	Significant	Not significant	Significant	Not significant
overall	Significant	Significant	Significant	Significant

Table 3 – Summary of the statistical analysis of the difference in proportion between .eu and the other TLDs for the most important categories

## Conclusions

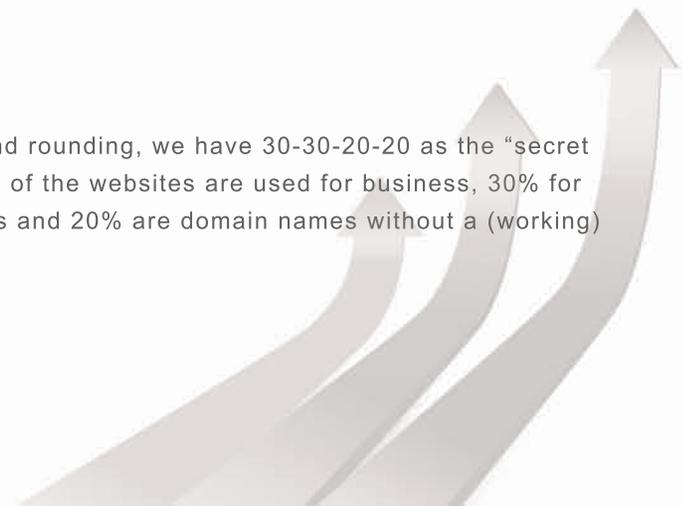
The profile graphs (Figures 2 and 3) show that the various categories are distributed similarly among the bigger gTLDs. These also happen to be the older gTLDs that were the first in the market. The graphs show the profile of each TLD against the average market profile, based on an equal sample from each TLD. Further comparisons among the TLDs are presented in Figures 4 and 5.

### *Uniformisation*

Although the different extensions were initially intended to serve different communities (.com: commercial organisations, .net: information-network companies, .org: non-profit, etc.), in general we see a very similar distribution in most of the extensions, with slightly different accents for some. The biggest deviation from the average distribution pattern is clearly .org, with its significantly higher population of community users, still reflecting its original population.

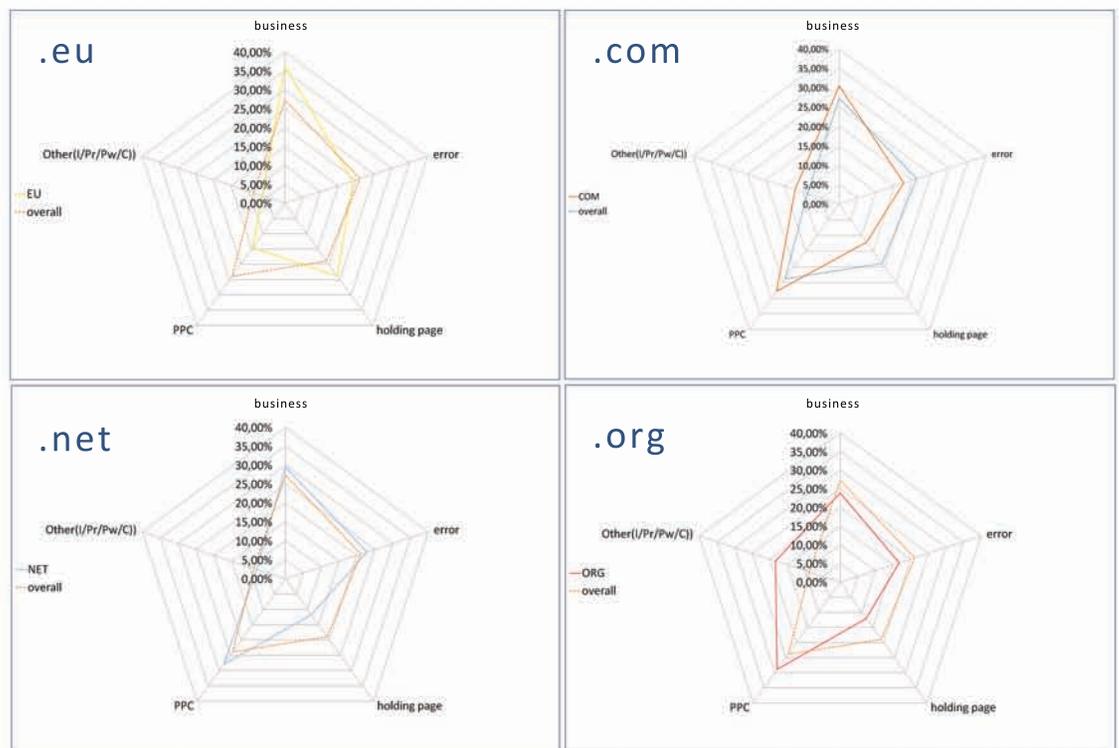
### *30-30-20-20*

Taking into account error margins and rounding, we have 30-30-20-20 as the “secret code of the Internet” for gTLDs. 30% of the websites are used for business, 30% for pay-per-click, 20% are holding pages and 20% are domain names without a (working) website.



## High percentage of pay-per-click

All studied TLDs have similar levels of pay-per-click websites. Although it is a relatively recent phenomenon, which has caused lively debates regarding its validity as a business model, the high numbers show that this online activity has won its place on the Internet. Whether this is good or bad is a completely different discussion. This high percentage is partly due to the habit of hosting companies to automatically set up a PPC site as the default page for a newly registered domain name.



*The percentage of pay-per-click sites is significantly lower for .eu.*

Figure 2 – Profile graphs depicting the category distribution of .eu, .com, .net, and .org compared with the average distribution documented in Table 1

## Holding Pages

The percentage of holding pages in a TLD's portfolio is typically higher in the more recent TLDs (except for .biz). However, they are also present in long-standing TLDs and have to be considered as a non-negligible subset. As with PPC, many hosting companies automatically set up a standard holding page after the registration of a new domain name.

## Exhibit 3: Auto-scan

To obtain additional information, an automated scan was run on a much larger sample of the domain names in each zone file. Information about redirects in particular was gathered by this means. Table 4, below, shows the percentage of redirected websites in each TLD and the most important extensions to which they are redirected.

The percentage represents the part of an extension's portfolio (in the top row) that is redirected to the corresponding extension in the first column.

This is an indication of domain names that have been registered but for which the existing (first) choice remained the main website. Nevertheless, the domain name is actively used as an addressable site.

	.eu	.com	.net	.org	.info	.biz	.mobi	.pro
.com	7.1%	10.2%	10.4%	9.0%	9.3%	13.2%	11.0%	6.7%
.net	1.8%	0.6%	1.6%	1.2%	1.5%	2.0%	2.0%	5.7%
.org		0.4%	0.8%	2.9%	0.6%	0.6%	0.6%	0.2%
.info		0.1%	0.1%	0.1%	0.7%	0.2%	0.1%	0.1%
.biz		0.1%	0.1%	0.1%	0.1%	0.7%	0.1%	0.1%
.pro								2.2%
.mobi							1.0%	
.de	4.7%	0.4%	0.7%	0.7%	1.3%	1.5%	1.3%	0.3%
.uk	0.5%	0.3%	0.3%	0.3%	0.3%	0.5%	0.4%	
.nl	2.6%	0.2%	0.1%	0.2%	0.3%	0.3%	0.4%	0.3%
.eu	1.1%		0.1%	0.1%	0.1%	0.1%		0.2%
.fr	0.5%	0.1%	0.2%	0.2%	0.1%	0.3%	0.1%	1.3%
.it	0.6%	0.1%	0.1%	0.1%	0.2%	0.2%	0.1%	0.1%
.se			0.1%	0.1%	0.1%	0.2%	0.1%	
.pl	1.4%		0.1%	0.1%	0.1%	0.2%		0.2%
Other	3.9%	0.8%	1.2%	1.1%	1.2%	1.5%	1.6%	0.9%
Total	24.2%	13.3%	15.9%	16.0%	15.9%	21.5%	18.7%	18.3%

Table 4 – The percentage of websites in the portfolios of the TLDs in the top row that are redirected to another top-level domain

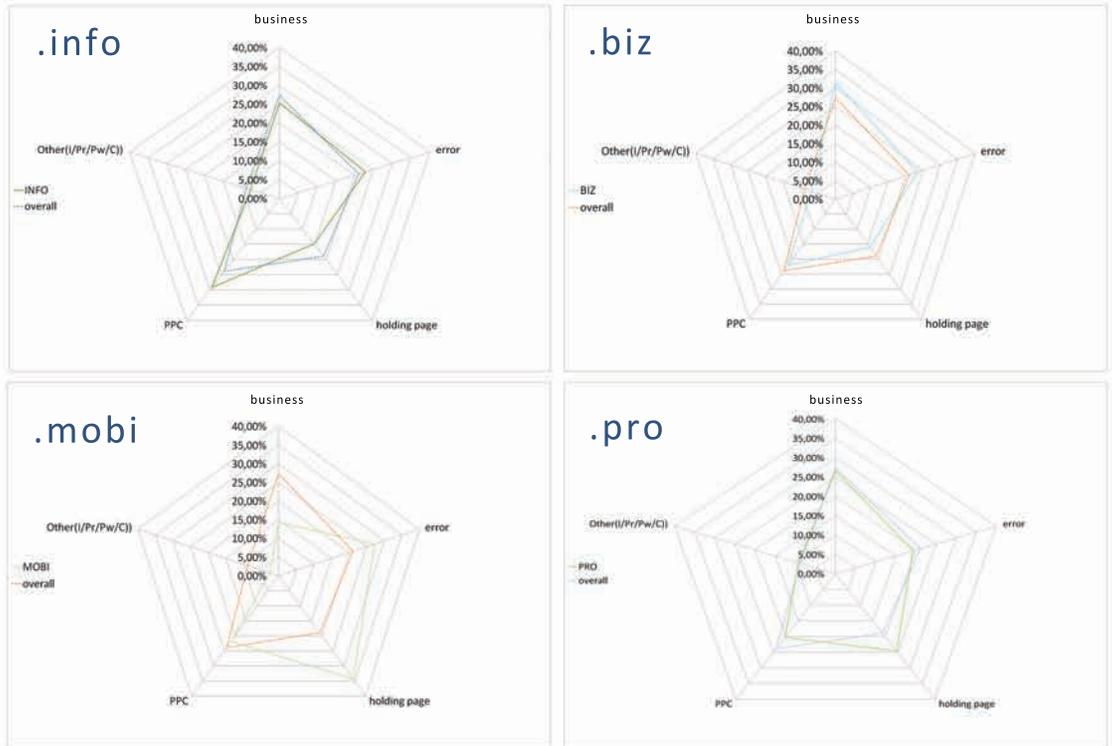


Figure 3 – Profile graphs depicting the category distribution of .info, .biz, .mobi and .pro compared with the average distribution documented in Table 1

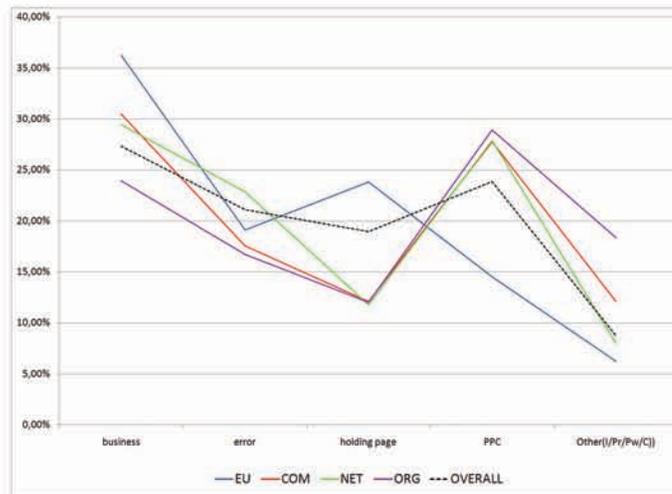


Figure 4 – Comparison of the profiles of .eu, .com, .net and .org with the overall profile

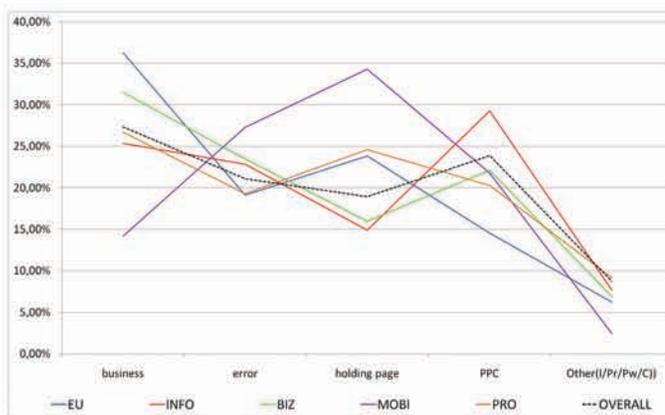


Figure 5 – Comparison of the profiles of .eu, .info, .biz, .mobi, .pro with the overall profile

## Errors

The number of non-resolving sites is similar for all TLDs and centres around 20%. This number is confirmed by the automated scan of large samples of all TLDs. In order to understand the nature of these errors, a more in depth analysis was done on a limited selection of domain names in this category. This assessment reveals that there are different categories of errors:

- The host names did not resolve; no IP address could be found for the name servers
- The host names resolved to an IP address, but no name server was listening
- The listed name servers existed but did not respond
- The name servers were not authoritative for the queried domain name
- The name servers responded, but no web server was listening at the listed IP address
- The web server was wrongly configured
- The IP address of the name server or web server was not reachable.

The level of errors in the total portfolio of domain names is so high that further research is needed to find out more about all possible error types and the share each one represents, so that adequate measures can be taken to improve the quality of the Domain Name System (DNS).

## Business remains the main activity

Although it has a similar distribution pattern to pay-per-click sites, business remains the main activity on the Internet. Community websites on the other hand still represent a marginal fraction of the categories. The use of a website for personal use (typically blogging under a proper domain name) is virtually non-existent.

## Further research

It would be interesting to do a similar test for the most important ccTLDs. Unfortunately, this cannot be done as the ccTLD zone files are not available. Some anecdotal data suggests that we might expect the ccTLDs to have a somewhat different profile, but further research on ccTLDs is needed before any conclusion can be reached.

## Learn more

The latest statistics on .eu performance and other .eu Insights reports are available at: <http://link.eurid.eu/insights>.

## About EURid

EURid is the not-for-profit organisation appointed by the European Commission to operate the .eu domain name. Set up in 2003, EURid started general registration of .eu domain names in April 2006. More than 3 million domain names have been registered to date. To find out more about .eu and EURid, please go to [www.eurid.eu](http://www.eurid.eu). You can contact us directly in any official EU language by email at [info@eurid.eu](mailto:info@eurid.eu).

## Credits

This report was prepared by EURid in cooperation with Leuven Statistics Research Centre.

# .eu