

Website usage trends among top-level domains

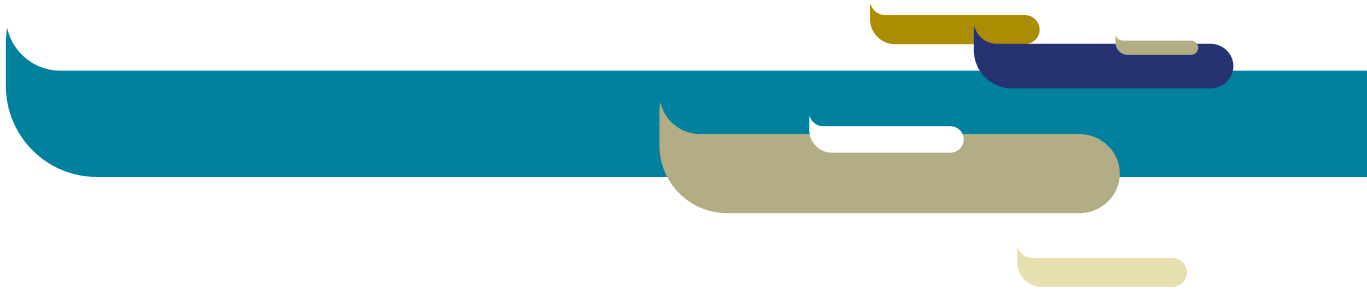
EURid, in collaboration with Leuven Statistics Research Centre, analyses 11 of the world's top-level domains. EURid illustrates how different domains have different purposes by classifying websites into eight usage categories.

November 2011

www.eurid.eu

.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.



Contents

- 1 EXECUTIVE SUMMARY04
- 2 INTRODUCTION.....06
- 3 USAGE CATEGORIES07
- 4 RESULTS08
 - 4.1 Overall categorisation results08
 - 4.2 TLD clustering.....12
 - 4.3 Comparison with previous years15
 - 4.4 Automatic categorisation.....17
- 5 CONCLUSIONS18
- 6 METHODOLOGY.....19
 - 6.1 Quality of correspondence analysis.....20
 - 6.2 Significance analysis21
- 7 MORE INFORMATION22



1 Executive Summary


For the third year running, EURid, together with the Leuven Statistics Research Centre, has investigated the websites behind some of the most popular top-level domains. This report focuses on two questions:

- 1. What is the most common usage of domain names?**
- 2. Is there a difference in how different top-level domains are used?**

To answer these questions EURid, the registry for the .eu top-level domain, carried out an extensive manual categorisation exercise, classifying the websites of 11 different top-level domains (TLDs) into eight categories.

This report shows that the most common use of a domain name is to link to a business website, since 26.5% of the categorised websites were used for business purposes.

Other categories with a large percentage of websites were: error (23.5%), holding page (20.6%) and pay-per-click (17.3%). Relatively few sites (10.0%) were used by communities and the percentages for the password-protected, institutional and pornography categories were almost negligible.



Comparing different TLDs, the report concludes that the older generic top-level domains (gTLDs) still have distinct profiles. The main example is .org, which has a very high percentage of community websites. In this, the gTLDs differ from the national country code top-level domains (ccTLDs), which are all used for very similar purposes. .eu has a lot in common with both the ccTLDs and with certain gTLDs (mainly .net but also .biz and .com).

A clear trend can be seen in the percentage of business websites: older, more established TLDs have a higher percentage of business sites than the new TLDs. Furthermore, ccTLDs have more business websites than the older gTLDs. Being one of the more recent TLDs, .eu nevertheless has a significantly higher percentage of business websites than the average, 31.4% compared with 26.5%. This is the third year in a row that .eu shows a strong business profile, confirming its unique position as a TLD for businesses and small- to medium-sized enterprises (SMEs) with a cross-border dimension.

This is the third year running that EURid conducts such an exercise. It is the most comprehensive manual research into TLD usage of which we are aware.

2 Introduction

Simply put, domain names, or more specifically, their corresponding web addresses are important because they give us an easy way to remember IP addresses, the complex sets of numbers the Domain Name System (DNS) uses to identify websites on the Internet. Domain names have proved so useful that today there is hardly a business, government body or institution that doesn't have a web address which it advertises at every opportunity.

When the DNS was introduced, a number of so-called generic top-level domains (gTLDs) were created so people or businesses that had IP addresses could associate them with an easy to remember domain name. The thought was that each gTLD would cater to a certain audience. .com would be for commercial companies, .net would be for network providers and .org for organisations and institutions. In addition, a number of country code top-level domains (ccTLDs) were delegated, each corresponding to a specific country. For example, .de for Germany and .uk for the United Kingdom.

Since then additional top-level domains have been introduced throughout the years both in the gTLD and ccTLD space, such as .info for informative sites, .pro for professionals, .mobi, for mobile websites, .biz as an alternative to .com and .eu for residents of the European Union (EU). More recently the creation of even more top-level domains has been approved with the aim of offering the Internet user more choice and further expanding the domain name market.

In this ever-evolving market context it might be interesting to find out:

- What are domain names most commonly used for?
- Do the old, original gTLDs still have distinct profiles, as initially intended, or are they now used for the same purpose?
- Are the newer TLDs used differently to the older, original TLDs?
- Are gTLDs used differently to ccTLDs?

To answer these questions EURid, the registry for the .eu TLD, conducted an extensive manual categorisation exercise in which it analysed 11 TLDs. The TLDs analysed were:

- Three of the original gTLDs: .com, .net and .org
- Three ccTLDs, all among the ten largest ccTLDs in the EU.
To protect the privacy of those involved, these will be named ccTLD1, ccTLD2 and ccTLD3
- Four of the new gTLDs: .biz, .info, .pro and .mobi
- One new ccTLD: .eu.



3 Usage categories

The website usage categories are unchanged from those defined in last year's report. However, the researchers in this year's exercise were more accurately briefed on the definition and attributes of each category. Categories were selected according to the type of information that EURid wanted to gather from this exercise.

Categories

- **Business:** A website that clearly shows commercial activity and that is designed for customer interaction. Business websites provide information about the company, including contact details, company structure, and descriptions of products or services. Some also allow customers to shop online.
- **Community or personal:** A website that clearly contains information about a community, club, association or religious institution. Personal websites contain information about an individual, family or political candidate. Although community and personal websites might contain promotional material, such as merchandising, they do not have a commercial purpose.
- **Pay-per-click (PPC):** A website that mainly contains advertising links. Advertisers only pay when a user clicks on an advert or link.
- **Holding page:** A website typically consisting of a single page that acts as a placeholder for a future website. Holding pages can have different formats:
 - Corporate holding page: A page with no information or just the company contact details. The page might indicate that more information will be published in due time. Note, if the page contained enough information about the company, its products and contact details, it was categorised as a business website.
 - Under construction: A page that tells the visitor that the website will be available soon.
 - Hosting company page: A page set up by a hosting company, often by default, when a customer has registered a domain name but not yet linked it to a website. If the holding page only contained pay-per-click links, it was categorised as a PPC.

- **Institutional:** A website that belongs to a government, government-related or government-sponsored institution, such as a city, museum, public school or university.
- **Password protected:** A website whose content can only be accessed once a password has been entered.
- **Error:** Websites displaying an error message. Although the domain name was found in the zone file, it was either wrongly configured or not configured at all and therefore it did not resolve to a 'www'.
- **Pornography:** A website displaying pornographic content. This category was chosen to evaluate the impact of pornography on the Internet.

Similar categorisation exercises are frequently carried out automatically. But an automatic categorisation often differs significantly from a manual categorisation, which will also be illustrated here. More information about the methodology used in this categorisation exercise can be found in the methodology section at the end of this report.

4 Results

4.1 Overall categorisation results

The overall results for all eight categories, based on a sample of 50 000 websites randomly selected across 11 TLDs, are illustrated in Figure 1.

On average, more than one quarter (26.5%) of the analysed websites were used for business purposes, closely followed by websites that could not resolve and resulted in error pages (23.5%). At the other end of the scale, websites with pornographic content constituted only 0.4% of all the websites sampled, indicating that the common perception, that there are a large number of pornographic websites on the Internet, is incorrect.

Holding pages, PPC and community websites, comprised 20.6%, 17.3% and 10% of the analysed websites, respectively.

Figure 1 – average usage per category, by percentage, across all 11 TLDs

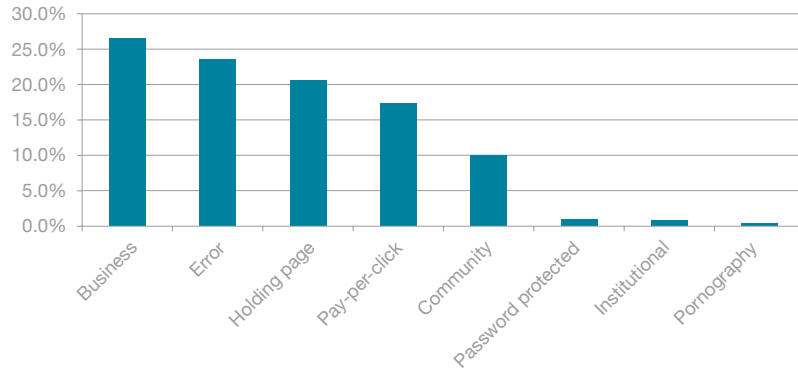


Figure 2 shows the categorisation results broken down per TLD. Business is the largest category in six of the 11 TLDs. The ccTLDs have the highest percentage of business websites (around 40%), followed by .eu (31.4%), .com (29.5%), .net (27.1%) and .biz (27.0%). Least popular with businesses are .org, .mobi and .pro. The results are also illustrated in Figure 3 where, for visual clarity, the password-protected category and six of the smaller TLDs have been excluded.

Figure 2 – distribution of sites across categories, by percentage, for all 11 TLDs

	.eu	.com	ccTLD1	ccTLD2	ccTLD3	.org	.info	.net	.biz	.mobi	.pro	Total
Business	31.4%	29.5%	39.1%	38.0%	42.9%	13.5%	17.2%	27.1%	27.0%	12.3%	12.5%	26.5%
Error	22.2%	18.2%	12.1%	15.9%	13.3%	20.0%	30.2%	21.1%	24.0%	26.9%	55.3%	23.5%
Holding page	25.4%	13.1%	19.5%	24.3%	19.5%	14.5%	14.1%	18.8%	16.1%	43.4%	17.8%	20.6%
Pay-per-click	12.8%	27.3%	11.4%	4.5%	8.2%	24.8%	27.1%	21.4%	27.7%	13.9%	10.7%	17.3%
Community	5.7%	10.1%	14.9%	15.2%	13.6%	23.6%	9.5%	9.2%	3.9%	2.3%	2.5%	10.0%
Password protected	1.0%	0.8%	1.4%	0.9%	0.8%	1.2%	0.8%	1.4%	0.8%	0.7%	0.7%	1.0%
Institutional	1.3%	0.2%	1.1%	1.2%	1.4%	2.1%	0.4%	0.5%	0.3%	0.2%	0.1%	0.8%
Pornography	0.3%	0.9%	0.5%	0.1%	0.2%	0.5%	0.6%	0.5%	0.2%	0.4%	0.4%	0.4%

■ indicates that the TLD has a significantly higher than average percentage of that particular type of website.
 ■ indicates a significantly lower than average percentage of that particular type of website.

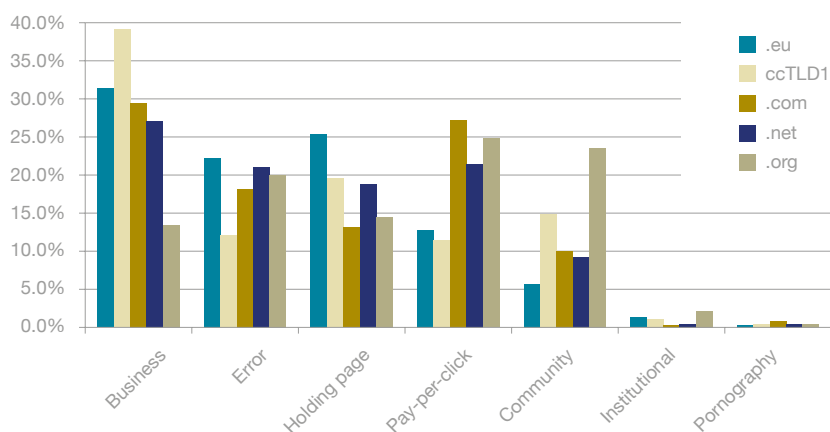
.org, although now open to all types of registrations, maintains its distinct community profile, with roughly double the amount of community and institutional websites when compared with the average TLD. 23.6% of .org websites are used for community purposes, versus 10% on average, while 2.1% of websites are institutional, compared with the average of 0.8%

.eu has a less than average number of community websites (5.7% versus 10%). This is typical for most new TLDs (.eu, .biz, .mobi and .pro). .info, though new, is the exception, with a fairly high percentage of community sites (9.5%). PPC websites are present in each TLD, but are more common for .com, .org, .info, .net and .biz. The number of PPC website sites for .eu is considerably lower than the average (12.8% versus 17.3%).

.eu has a higher than average number of holding pages (25.4% versus 20.6%). This is not unusual, as holding pages are typical for newly registered domain names and are therefore more frequent in the newer TLDs, which have the most newly registered names. The ccTLDs also have relatively high levels of holding pages. This could point to the practice of certain registrars to automatically create a holding page whenever a domain name is registered.

Password-protected websites are few and numbers are comparable across all TLDs. Pornographic websites are even less common, making up less than 1% of the websites across all TLDs. Institutional websites are also small in number across all TLDs, although they are slightly more common for .org, where they make up 2.1% of all websites, and for .eu and the other ccTLDs where they make up between 1.1% and 1.4% of all websites.

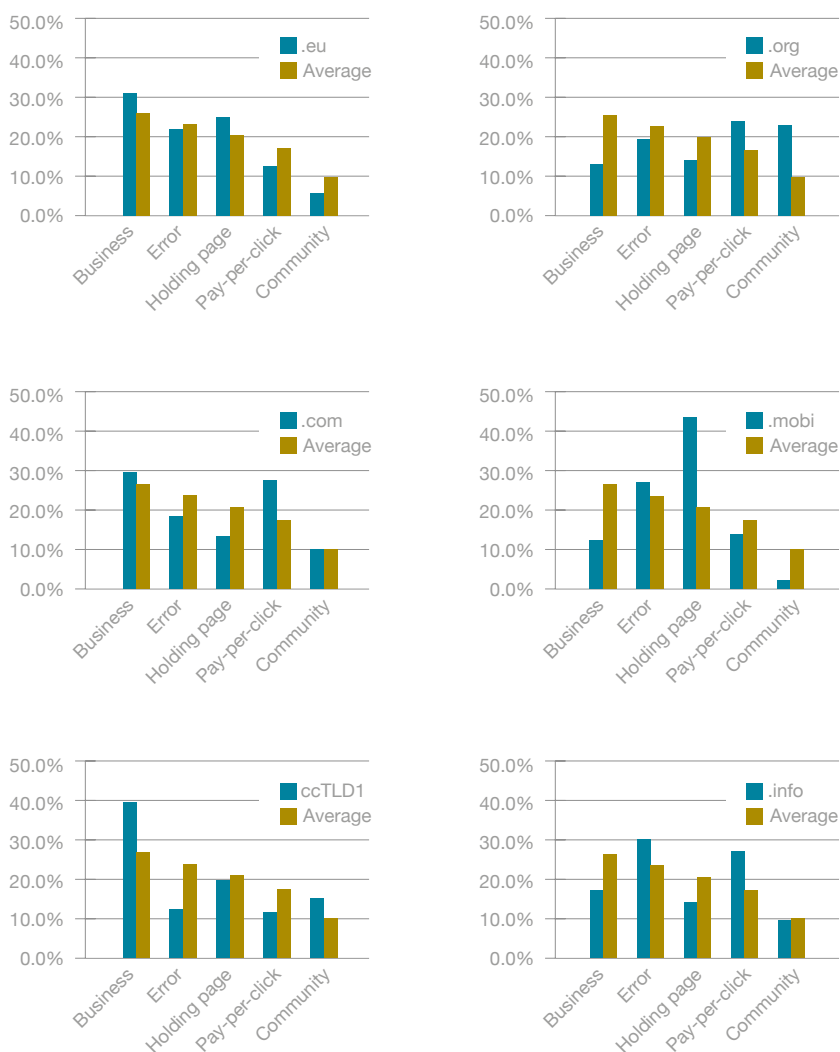
Figure 3 – Distribution of websites across categories by percentage for five TLDs



Based on the research results, we created an average TLD usage profile, as shown in Figure 4, against which we plotted the usage profiles of .eu, .org, .com, .mobi, ccTLD1 and .info.

The .eu usage profile closely matches the average profile, although it has more business websites and holding pages. The .com profile clearly has more PPC websites than average, as does .info. .info also has a higher number of error pages. ccTLD1 has a larger number of business websites, while .org maintains its institutional profile and .mobi has a remarkable number of holding pages.

Figure 4 – individual TLD usage against the average profile



4.2 TLD clustering

The categorisation exercise also examined patterns in the research data to establish whether TLDs bear any resemblance to each other. For example: whether ccTLDs and gTLDs are used for different purposes or whether more established TLDs, such as ccTLD1, ccTLD2, ccTLD3, .org, .com and .net, are used differently than the newer TLDs, such as .mobi, .info, .biz, .pro and .eu.

The percentage of business websites for each TLD was plotted against the percentage of PPC websites and community websites in Figures 5 and 6, respectively. These figures also display averages for the three ccTLDs and the gTLDs, indicated by blue dots. The results clearly show that ccTLDs have a very common profile whereas gTLDs do not. Also, there is no obvious grouping of long established TLDs versus new TLDs.

Figure 5 – TLD clustering based on fraction of business versus community for all 11 TLDs

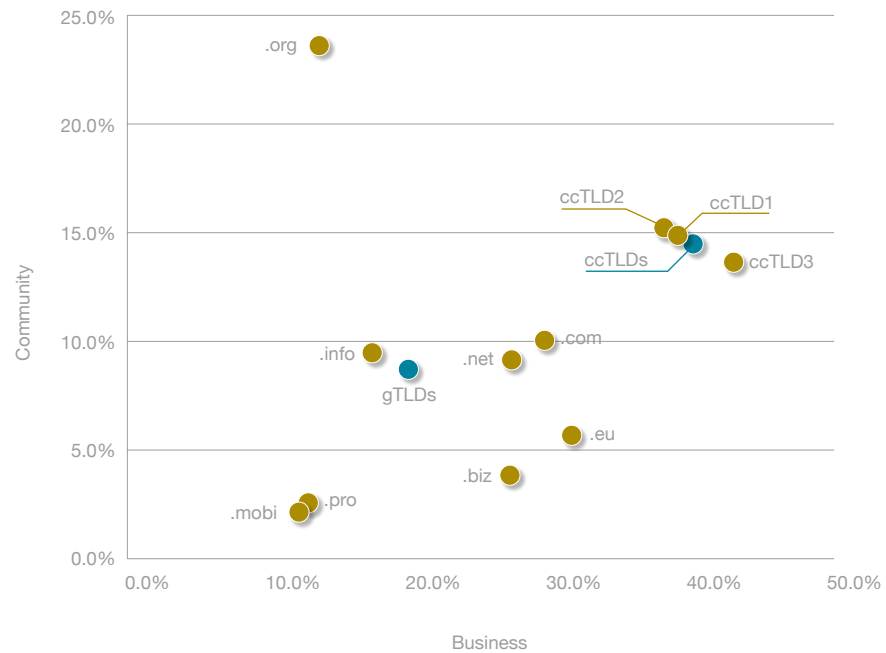
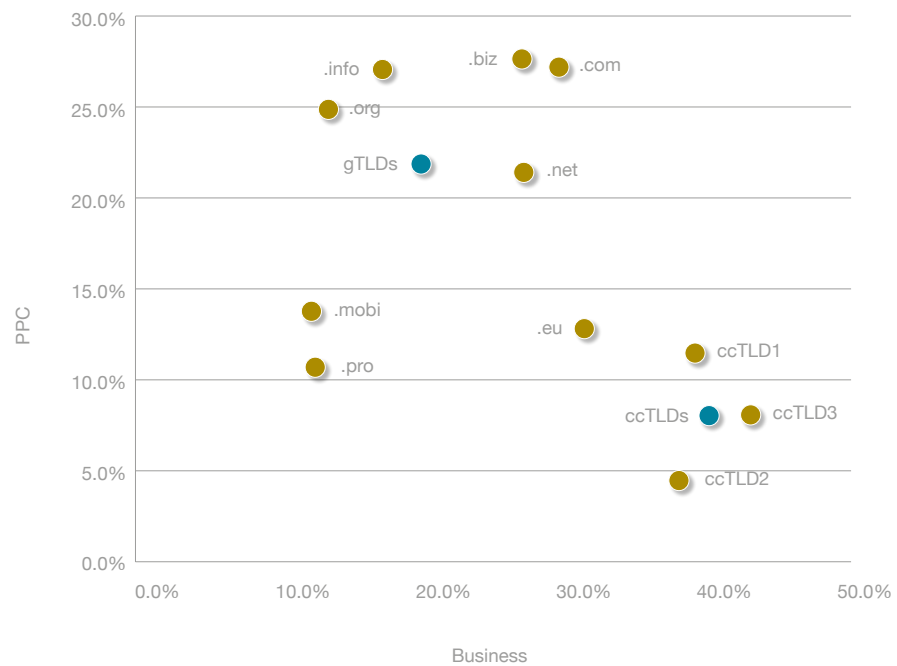


Figure 6 – TLD clustering based on fraction of business versus PPC for all 11 TLDs



Another way of establishing similarities between different TLDs is via the correlation coefficient. Figure 7 displays a table of values between 1 and -1. Each value indicates to what extent the two variables, in this case two TLDs, correlate. A value of 1 means that there is a resemblance between the two TLDs, a value of zero that they are completely independent of each other. The values highlighted in green show a strong correlation, those highlighted in blue show a weak correlation.

The national ccTLDs form a group of TLDs which are clearly used for similar purposes, whereas .com, .net and .biz form another group. .eu is most similar to .net but does have a more varied profile, with similarities to the ccTLDs, but also to .biz and .com.

Figure 7 – correlation coefficients

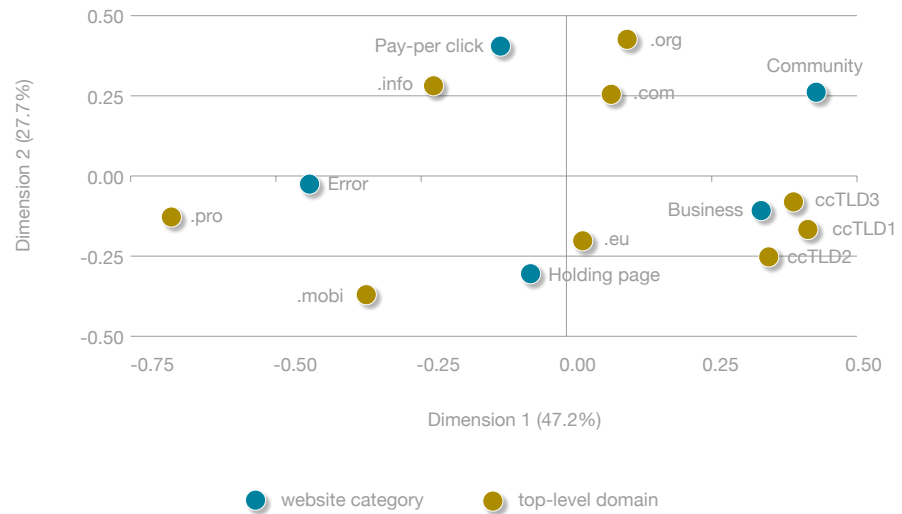
	.com	ccTLD1	ccTLD2	ccTLD3	.org	.info	.net	.biz	.mobi	.pro
.eu	0.80	0.87	0.90	0.87	0.50	0.70	0.93	0.84	0.76	0.61
.com		0.79	0.67	0.75	0.74	0.85	0.96	0.96	0.45	0.44
ccTLD1			0.97	0.99	0.48	0.48	0.83	0.70	0.43	0.25
ccTLD2				0.97	0.42	0.44	0.78	0.61	0.54	0.35
ccTLD3					0.40	0.44	0.80	0.66	0.41	0.27
.org						0.82	0.73	0.68	0.43	0.46
.info							0.87	0.92	0.61	0.78
.net								0.97	0.66	0.60
.biz									0.60	0.62
.mobi										0.65

Using the correspondence analysis technique, we also plotted the similarity of TLDs against the categories to which they most strongly relate. The technique allows a data set to be displayed in two-dimensional graph, displayed in Figure 8. The graph should be viewed as two sets of different, overlaid plots, one showing the categories, in blue, and one displaying the TLDs, in green. The closer two points are to each other, the more the categories and TLDs resemble each other.

For the current data set, 75% of the differences can be explained by these two variables (categories and TLDs). We excluded the institutional, password-protected and pornography categories, as well as the .biz and .net TLDs, since the quality of their representations in two dimensions is relatively poor.

Figure 8 shows three TLD clusters: a ccTLD cluster, a cluster of the youngest TLDs (.pro and .mobi) and a cluster of the established TLDs (.org, .info, .com). .eu is situated at the centre of the three clusters. This method shows that the ccTLDs are associated with business websites, the younger TLDs (.pro and .mobi) with error sites and .mobi with holding pages.

Figure 8 – correspondence analysis of category per TLD



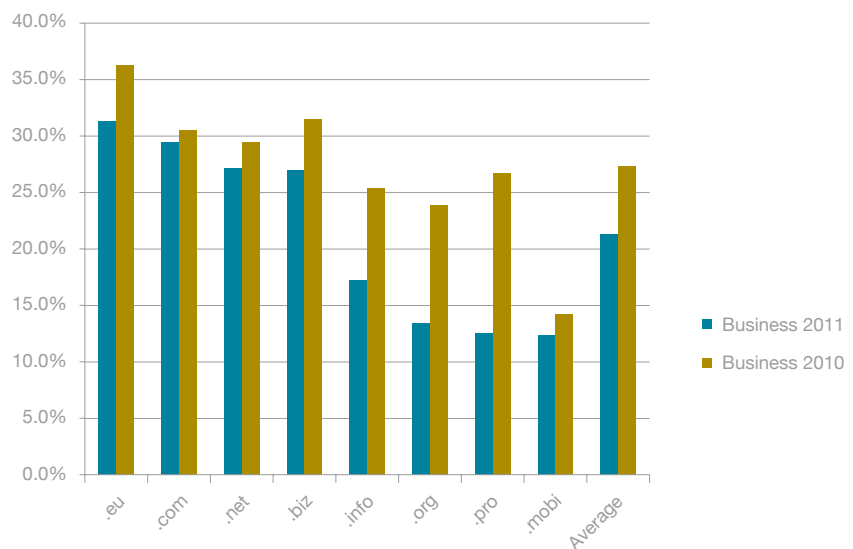
4.3 Comparison with previous years

Compared with last year, and excluding the ccTLDs as they were not categorised in 2010, all TLDs have less business and institutional websites and more community websites, as displayed in Figure 9. We attribute most of the differences to the more specific instructions given to the researchers rather than to changes in the use of domain names. The only exception is .pro, where business use dropped from 27% to 13%. This drop is too significant to be attributed to the research instructions. Comparing last year with this year, we do however see that in both years, .eu had the highest percentage of business websites (Figure 10), clearly showcasing .eu's business profile. This strong business profile was also evident in the 2009 exercise, where business was the dominant category for both .eu and .com. A more careful comparison with the 2009 results is not possible, as the 2009 exercise analysed only .eu and .com domain names and also categorised the analysed websites differently.

Figure 9 – average category comparison between 2010 and 2011, excluding ccTLDs

Categories	2010	2011
Error	21.1%	27.2%
Business	27.3%	21.3%
Pay-per-click	23.9%	20.7%
Holding page	19.0%	20.4%
Community	3.5%	8.3%
Password protected	1.3%	0.9%
Institutional	2.9%	0.6%
Pornography	1.0%	0.5%

Figure 10 – business websites in 2010 and 2011



4.4 Automatic categorisation

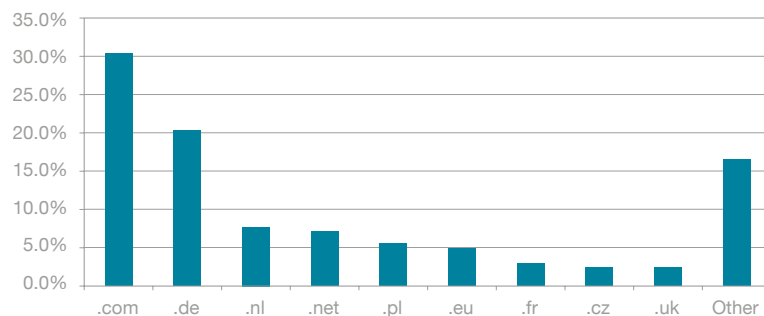
To compare the results of the manual categorisation exercise against those of an automatic categorisation exercise, the same sample of 5 410 .eu websites was categorised by computer software. The results differed for 36% of the websites, which indicates that there are still significant differences in how humans and computers categorise websites. A more careful investigation of some of the websites, where opinions differed on which category was most appropriate, found in favour of the manual category choice in the majority of cases.

The most common difference between the manual and automatic categorisation exercises was websites which were categorised as business in the manual exercise, but as error or holding page in the automatic exercise. The reason for this difference is that the computer software listed websites that were very slow to load as error pages, and websites that only had one page, but clearly gave a company description, as holding pages.

Automatic categorisation of the entire .eu zone file was also used to measure the percentage of .eu sites that automatically redirected to another TLD. In total, 26% of .eu domain names linked to a website redirected visitors to another TLD. Redirects to .com (30% of all redirects) and .de (20%) were by far the most common, comprising half of all redirects, as can be seen in Figure 11. Redirects to another .eu website made up 5% of all redirects.

Which TLD a user was redirected to depended strongly on the country of the .eu registrant. Registrants based in Cyprus (73%) or Spain (56%) most often redirected visitors to a .com website. For Poland (86%), Slovakia (76%) and the Czech Republic (75%) redirects to the national TLD were most common. Registrants in Estonia most often redirected visitors to another .eu website (73%).

Figure 11 – redirects of .eu domain names



5 Conclusions

This report shows that the most common use of a domain name is to link to a business website, since 26.5% of the categorised websites were used for business purposes. Since worldwide Internet penetration is at an all-time high, this is perhaps surprising. A possible explanation is that private individuals are using social networks like Facebook, Twitter, LinkedIn or Wordpress to maintain their online presence. Companies however, although also active on social media, still consider it important to develop their own online presence and so control the management of their brand.

Other categories with a large percentage of websites were: error (23.5%), holding page (20.6%) and PPC (17.3%). The high percentages for these categories might reflect the low cost of registering a domain name. Relatively few sites (10.0%) were used by communities and the percentages for the password-protected, institutional and pornography categories were almost negligible.

Comparing different TLDs, the report concludes that the older gTLDs still have distinct profiles. The main example is .org, which has a very high percentage of community websites. In this, the gTLDs differ from the national ccTLDs, which are all used for very similar purposes. .eu has a lot in common both with the ccTLDs and with certain gTLDs (mainly .net but also .biz and .com). An explanation of the similarity between .eu and .net could be that they are both the main alternatives to the ccTLDs and .com, respectively.

A clear trend can be seen in the percentage of business websites: older, more established TLDs have a higher percentage of business sites than the new TLDs. Furthermore, ccTLDs have more business websites than the older gTLDs. Being one of the more recent TLDs, .eu nevertheless has a significantly higher percentage of business websites than the average, 31.4% compared with 26.5%. This is the third year in a row that .eu shows a strong business profile, confirming its unique position as a TLD for businesses with a cross-border dimension.

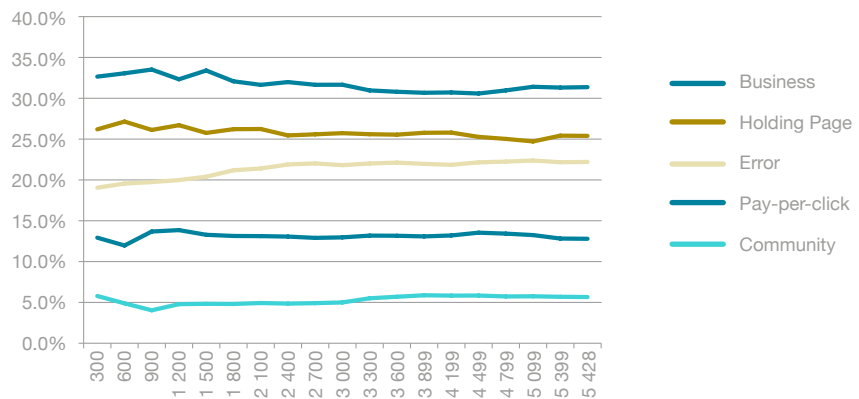
6 Methodology

The domain names used in this exercise were taken from the zone files of the 11 studied TLDs. From these zone files, a random sample was generated which was then assessed by a multilingual team of researchers. In total, around 5 000 domain names were assessed for each TLD. gTLD zone files were publically available, while the ccTLD domain name samples were contributed by their national registries.

This method does not take into account domain names that were registered without configuring name servers (the so called 'parked' domain names), which is typical of defensive registrations. Please note that in some TLDs, parked domain names are not allowed. In that case the domain name is often linked to a holding page to circumvent this limitation. 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 redirected.

As the number of analysed websites increased, the evolution of the distribution across categories was carefully monitored, as displayed in Figure 12. The results were relatively stable after 3 000 websites had been categorised, indicating that the data set was large enough to provide a realistic reflection of website usage at large. Statistical analysis shows that a sample size of 3 941 is large enough to state that a result of 36% is larger than a result of 33%, with a 95% significance level.

Figure 12 – evolution of percentages for the different categories for .eu as sample size grows

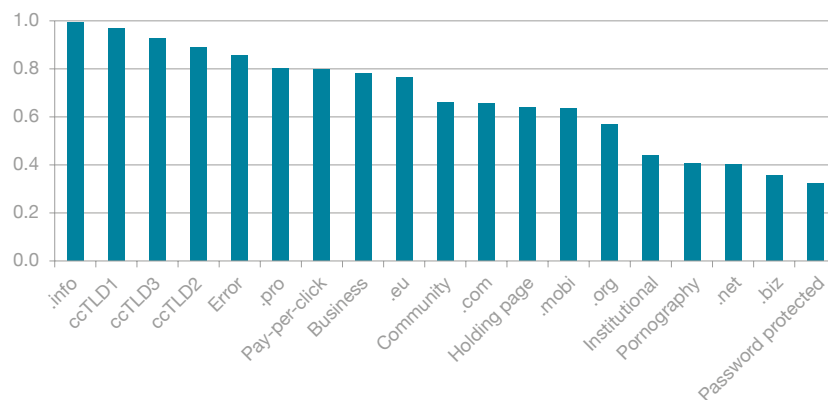


6.1 Quality of correspondence analysis

The correspondence analysis plot (Figure 8) relies on reducing the full table of categories and TLDs to just two dimensions. Naturally, a lot of information is lost by doing this. How well a TLD or category is represented is given by the quality measure, a number between zero and one, where one indicates that the category or TLD could be perfectly represented in the plot.

Figure 13 shows the quality of the representation for each TLD or category in the correspondence analysis plot (Figure 8). Most categories and TLDs have a fairly high quality value, indicating that they are well represented. The exceptions are the.biz and .net TLDs and the institutional, pornography and password-protected categories, which all have a quality value below 0.5. These TLDs and categories were therefore removed from Figure 8.

Figure 13 – quality of the representation of the TLDs and categories in the correspondence plot (Figure 8)



6.2 Significance analysis

In the statistical analysis to determine which TLDs and categories resemble each other, an overall confidence level of 95% was used.

Figure 14 indicates whether the category results for the other TLDs differ significantly from the results for .eu. For example, the difference between the percentage of .eu PPC websites (12.8%) and .com PPC websites (27.3%) is statistically significant. Conversely, the difference between the percentage of .eu holding pages (25.4%) and ccTLD2 holding pages (24.3%) is not big enough to be considered statistically significant.

Figure 14 – categories from other TLDs that differ significantly from .eu

Category	Business	Error	Holding page	Pay-per-click	Community
Overall	Significant	Not Significant	Significant	Significant	Significant
.com	Not Significant	Significant	Significant	Significant	Significant
ccTLD1	Significant	Significant	Significant	Not Significant	Significant
ccTLD2	Significant	Significant	Not Significant	Significant	Significant
ccTLD3	Significant	Significant	Significant	Significant	Significant
.org	Significant	Significant	Significant	Significant	Significant
.info	Significant	Significant	Significant	Significant	Significant
.net	Significant	Not Significant	Significant	Significant	Significant
.biz	Significant	Not Significant	Significant	Significant	Significant
.mobi	Significant	Significant	Significant	Not Significant	Significant
.pro	Significant	Significant	Significant	Significant	Significant



7 More information

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 operator of the .eu top-level domain. 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. You can contact us directly in any official EU language by email to info@eurid.eu.

Credits

This report was prepared in collaboration with the Leuven Statistics Research Centre.

Acknowledgements

EURid gratefully acknowledges the assistance of the Leuven Statistics Research Centre and the many TLD registries that generously shared information for the purpose of this report.

EURid vzw/asbl (headquarters)
Woluwelaan 150, 1831 Diegem, Belgium
T: +32(0)2 401 27 50
F: +32(0)2 401 27 51

