

ICANN WHOIS REGISTRANT  
IDENTIFICATION PROJECT

Appendix A:  
Exploratory Analysis Report

PRESENTED TO:  
ICANN

PRESENTED BY:  
NORC at the  
University of Chicago

DECEMBER 20, 2012



*at the* UNIVERSITY *of* CHICAGO

## Introduction

NORC has been contracted by the Internet Corporation for Assigned Names and Numbers (ICANN) to conduct the **WHOIS Registrant Identification Study**; an exploratory study to classify domains into a variety of categories such as registrant type, domain user type, and commercial activity.

In creating the data we have collected, we have kept in mind the three focus questions of this project:

- 1) What differences exist between how domains users that are natural persons identify themselves, versus how domain users that are legal persons identify themselves?
- 2) What differences exist in how domains are actually used for domains registered by natural persons versus domains registered by legal persons versus domains registered via proxy?
- 3) What differences exist in how domains with any type of potentially commercial activity are identified in WHOIS versus domains with no observed potentially commercial activity?

We start the report with some background on these three questions, including how we recoded variables in the dataset for analysis. Our analysis is organized by how these three questions are answered for different subject variables. The first three analysis sections are the variables from which we have formed the three questions. Here are the subject variables for which we have analysis sections:

- A. Apparent domain user type
- B. Apparent registrant type
- C. Potentially commercial activity variables
- D. Business Structure of Domain User
- E. Domain name extension (gTLD)
- F. Registrant country/region of the world
- G. Relationship of domain user to registrant
- H. Other coded behavior variables
- I. Blacklist variables
- J. Whitelist variables

Our key tool for our analyses has been the chi-square test of independence.<sup>1</sup> Since this is an exploratory data analysis, we mainly interpret the frequencies rather than create more complex analysis such as

---

<sup>1</sup> A chi-square test of independence is a statistical test for assessing whether two categorical variables are independent (not associated). The null hypothesis of the test is that the two categorical variables are independent. If the observed chi-square test statistic, which is based on the difference between observed and expected cross-classified frequencies, is unusually large assuming the null hypothesis of independence is true, then we conclude

building regression models. Follow-up analyses can be done with the clearer focus that will come out of this project.

One important note is that all of our analyses except the one-way frequencies of variables are weighted. In a representative sample of 1,600 domains, we would have studied only 98 \*.info and 26 \*.biz domains, but we set sample sizes of 100 for each. We did this so that we could have a sufficient number of \*.info and \*.biz domains for analysis. This results in a slight undersampling of \*.com, \*.net, and \*.org domains, and oversampled \*.info domains and especially oversampled \*.biz domains. So we applied weights to each gTLD as shown in Table 1.

**Table 1: Weighting by gTLD for the Registrant ID Study Domain Sample**

<b>gTLD</b>	<b>Global Proportion</b>	<b>Sample Size</b>	<b>Sample Proportion</b>	<b>Weight = Global/Sample Proportion</b>	<b>Sum of Weights = Sample Size *Weight</b>
*.com	74.3%	1,128	70.5%	1.0534	1188.2
*.net	10.7%	165	10.3%	1.0412	171.8
*.org	7.2%	107	6.7%	1.0813	115.7
*.info	6.1%	100	6.3%	0.9830	98.3
*.biz	1.6%	100	6.3%	0.2600	26.0
<b>TOTAL</b>	<b>100.0%</b>	<b>1,600</b>	<b>100.0%</b>		<b>1,600.0</b>

that the two categorical variables are associated (dependent upon one another). If the p-value—the probability, under the null hypothesis, of observing a test statistic value greater than or equal to the one obtained from the sample, is small, then the observed test statistic is considered unusually large. If you want at least 95 percent confidence for statistical test results, p-values less than 0.05 (5 percent) are considered too small. In this sense, we state that the chi-square test results are statistically significant.

## The Three Questions

### *Apparent Domain User Type: Legal and Natural Persons*

For each of the 1,600 domain names, we tried to determine if the domain user could be considered a legal person or a natural person. Table 2 shows that for most domain names, we could not make such a determination because almost half the domains were parked domains or had no online content at all. Only 11.5 percent of the domains had content, but had an unknown apparent domain user type. To code apparent domain user type, NORC staff reviewed all of the downloaded domain content for each domain during phase I of the Domain User variable coding. The overall procedure can be summarized as follows.

First, the downloaded web content was accessed to determine if the downloaded web content contained any usable data to conduct manual coding. If the data did not contain enough usable information, it was considered having No Usable Content and the Domain User variables relying on web content for coding were coded to their corresponding unknown codes. An example of this scenario is if the downloaded content consisted of a single webpage which only contained the following HTML data: `<html><body><p>Under Construction</p></body></html>`.

For the domains with usable data, we evaluated the downloaded content to determine if it consisted solely of common domain parking content. For example, if the full set of downloaded content consisted of a single landing page and this landing page only contained HTML content consistent with GoDaddy parking services, the apparent domain user type was coded as Unknown – Domain Parked. In some cases, it was not clear whether we should classify a domain as Domain Parked or No Online Content. Some of the No Online Content domains actually have a little content, and sometimes even some potentially commercial activity. For example, a site could have a simple index.html with an Under Construction page with a simple banner ad. There were not enough such sites to create a separate "Little Online Content" category.

All the domains which were not coded by the two procedures listed above were evaluated on a case-by-case basis to determine the phase I Domain User variables. The Apparent Domain User type was coded as a Natural Person when the Domain User was clearly a real living individual or small group of individuals and not a virtual entity such as a corporation or non-profit entity of any other named entity that is not a real living person. All other entities were coded as Legal Persons or Unknown.

To ensure that the data was accurately coded, each case underwent multiple rounds of manual coding by independent coders. The results of these multiple rounds of coding were adjudicated and all differences

detected during adjudication were collaboratively reviewed by a supervisory team to make a final determination of the Domain User variables.

---

**Table 2: Apparent Domain User Type**

---

<b>Type</b>	<b>Frequency</b>	<b>Percent</b>
Natural Person	87	5.4
Legal Person	586	36.6
Domain Parked	328	20.5
No Online Content	416	26.0
Unknown	183	11.5

---

A finer categorization of Natural Person was done to separate the variable into individuals versus small groups of related individuals, for example, a family. We found that of the 87 Natural Persons shown in Table 2, 78 are individuals and nine are small groups. Further analysis of the group of nine domains would not provide statistically meaningful results, so we will not split the Natural Person category in subsequent analyses. Analyses will only compare the three generic entity types: legal persons, natural persons, and unknown.

### *Registrants: Natural and Legal Persons and use of Privacy/Proxy Services*

Apparent registrant type was coded as to whether we could place the registrant into categories defined in ICANN's *Revised Terms of Reference for WHOIS Registrant Identification Studies* (<http://gns0.icann.org/issues/whois/tor-whois-registrant-id-studies-20may11-en.pdf>). Initially, only WHOIS information and independent searches of public databases were considered in the classification. For example, we searched known lists of privacy and proxy providers to place sampled domains into these categories, and reverse WHOIS email counts were used to help determine multiple domain name holders. Manual coding was used to code the remainder of the domains where Apparent Registrant Type could not be classified using automated means. The Apparent Registrant Type was coded during phase I of the Domain User Coding process. This manual coding process consisted of a concise set of rules to arrive at Apparent Registrant Type. The manually coded cases underwent the same quality control process consisting of multiple rounds of independent coding and an adjudication process. While investigating the domain user, the coder may have gained insights on the registrant of the domain, such as situations where the domain user is the same as the registrant. Thus, additional information was used to

correct initial categorizations or add granularity to the process. Table 3 is a summary of the final coding outcomes for Apparent Registrant Type:

**Table 3: Apparent Registrant Type Summary**

<b>Apparent Type</b>	<b>Frequency</b>	<b>Percent</b>
Registrant Name appears to be a natural person; no organization is named	447	27.9
Registrant Organization is specified; registrant name is also specified – registrant name or organization contains legal person	320	20.0
Registrant Organization appears to be a Proxy registration service	310	19.4
Registrant Organization is specified and appears to be a legal person; no registrant name is specified	183	11.4
Registrant Name and Organization are completely missing	93	5.8
Registrant Organization is specified; registrant name is also specified – both appear to be a natural person	73	4.6
Registrant Organization appears to be a multiple domain name holder	62	3.9
Registrant Name appears to be a legal person; no organization is named	52	3.3
Registrant Name and Organization look to be patently false	25	1.6
No Registrant Name or Organization available because Pending Reactivation or Deletion	11	0.7
Registrant Organization appears to be a Privacy registration service	10	0.6
Unable to classify / requires additional review	7	0.4
Registrant Organization is specified and appears to be a natural person; no registrant name is specified	5	0.3
Registrant Name and Organization are incomplete	2	0.1

With respect to the questions that are the key focus of this study, domains that are registered using Privacy or Proxy services are of particular interest. As shown in Table 3, there are 310 proxy-registered domains, but only 10 privacy-registered domains. With such a small category size, further analysis that attempts to cross-classify the privacy group with subject variables, such as commercial activities, would not be meaningful. Therefore, our analyses combine privacy and proxy registered domains together, though it is almost a comparison between proxy and non-proxy registered domains.

In order to simplify analyses of Apparent Registrant Type, we collapse the categories in Table 3 to the following four revised categories:

- Registrant appears to be a Legal Person – domains with WHOIS data which appears to identify a legal person as the Registrant (includes multiple domain holders, but not Privacy/Proxy registered domains)
- Registrant appears to be a Natural Person – domains with WHOIS data which appears to identify a natural person as the Registrant
- Registrant appears to reference a Privacy/Proxy Service – domains with WHOIS data which appears to identify a Privacy/Proxy service
- Unknown – domains with WHOIS data which could not be classified (includes data completely missing, patently false or incomplete WHOIS, and domains pending reactivation or deletion)

In what follows, the term Apparent Registrant Type refers to these revised categories. Table 4 is a summary of Apparent Registrant Type revised.

**Table 4: Apparent Registrant Type Summary (Revised)**

<b>Apparent Type</b>	<b>Frequency</b>	<b>Percent</b>
Registrant appears to be a Legal Person	617	38.6
Registrant appears to be a Natural Person	525	32.8
Registrant appears to use a Privacy/Proxy Service	320	20.0
Unknown	138	8.6

### *Potentially Commercial Activity*

There are several variables related to potentially commercial activity in the domain content section of the dataset. These variables measured whether there was any apparent activity that might be considered commercial in some countries: whether there were membership dues for online content or offline content, whether there was promotional content offline or online, whether there were banner ads and whether these banner ads were for the hosting provider or registrar, and whether there were only pay-per-click ads and whether these pay-per-click ads were for the hosting provider or registrar. We created a variable measuring Potentially Commercial activity in any of these variables. All of these variables are binary, so these tables only present the percentage of domains in each subgroup with each of these characteristics. Table 5 shows the overall percentage for each binary variable.

**Table 5: Summary of Potentially Commercial Activity Variables**

<b>Commercial Activity Variable</b>	<b>No</b>	<b>Yes</b>	<b>Percent Yes</b>
<b>E-Commerce</b>	1489	111	6.9
<b>Membership (Online Content)</b>	1572	28	1.8
<b>Membership (Offline Content)</b>	1544	56	3.5
<b>Promotional Content (Offline)</b>	1305	295	18.4
<b>Promotional Content (Online)</b>	1507	93	5.8
<b>Host Promotional Content (Online)</b>	1461	139	8.7
<b>Third Party Banner Ads</b>	1496	104	6.5
<b>Host Banner Ads</b>	1398	202	12.6
<b>Pay-Per-Click Ads</b>	1131	469	29.3
<b>Host Pay-Per-Click Ads</b>	1539	61	3.8
<b>Any Potentially Commercial Activity</b>	695	905	56.6
<b>Excluding Pay-Per-Click Ads</b>	883	717	44.8

A further explanation of coding these variables is described below:

#### *E-Commerce*

This classification variable allows for e-commerce activities to be noted for any site, even if the site is not primarily an “e-commerce” website. For instance, ESPN.com, while classified as an “informational” website, would here receive a value of “1” (true) since ESPN.com provides pages where website readers can purchase goods from ESPN.com.

#### *Membership (Online Content)*

Membership fees will typically require a user name and password for logging in to view privileged online content.

However, many websites will ask for users to create user names without charging a membership fee; the user name creation allows these websites to gather information on its users and communicate better with these users, thereby increasing traffic to the website. These types of membership are NOT marked as having commercial online membership.

To determine if member logins first require the payment of membership fees, we went to the login page of the website to see if membership is offered for a price. Sometimes, fees are not immediately apparent; for instance, the New York Times allows specific computers to access New York Times online content



ten times per month before requiring a membership fee-based login to access its content. Because of mechanisms like this, we had to carefully assess the membership requirements of the site.

#### *Membership (Offline Content)*

As opposed to online membership, offline membership refers to fees paid through the website for goods or services provided offline. For example, a gym may offer a portal through which gym members pay their monthly membership fees so that they may continue to use the physical gym.

#### *Promotional Content (Offline)*

Promotional content encourages website visitors to purchase goods or services of the website owner, either in a physical location or through some other vendor, instead of through the website itself. Promotional content is distinct from e-commerce activity because the commercial activity is merely being promoted, but cannot be transacted, on the website in question.

An example of a website with promotional content would be a small bookstore website that advertises its latest book arrivals on its website, but which does not have a web portal through which these books can be purchased online; a customer must go to the physical location of the bookstore in order to purchase the books.

#### *Promotional Content (Online)*

If a website is promoting their goods but these goods are sold on a online retailer site like Amazon or Ebay, then this is an example of PROMO-ON.

#### *Host Promotional Content (Online)*

Same as promotional content described above, but there is evidence that the promotional content was placed on the website by the hosting provider.

#### *Third Party Banner Ads*

Banner ads are graphics on websites which advertise goods or services and which act as links to pages where these goods or services can be purchased online. The placement of these ads on third party sites allows the domain users of these sites to earn revenue from the companies placing the ads. Note that these banner ads are shown regardless of the type of site visitor or the type of content they are viewing. This is opposed to pay-per-click ads, which generally appear in response to specific queries by site visitors.

Discerning whether the domain user or the hosting provider placed the banner ads on the website can be difficult. Generally, websites that appear to be administered or designed by the domain user will be more

likely to have ads that were placed by the domain user (since the domain user is exercising a large amount of control over the domain).

Conversely, if the site template is provided by the host, or if the hosting appears to be free, it is likely that the host is placing ads on the site (this would be part of the agreement for free hosting).

This variable asks simply whether banner ads are present on the site.

#### *Third Party Banner Ads Host*

A determination of whether banner ads placed by hosting providers are present on a website, following the distinctions from the preceding variable description.

#### *Pay-Per-Click Ads*

Pay-per-click ads, unlike banner ads, appear in response to site visitor queries or the type of content the visitors view. This occurs because pay-per-click ads generate revenue for domain users or hosting providers based on “performance” (number of clicks) rather than “impressions” (number of views).

Some websites appear to exist only to generate these types of ads; this variable tracks this type of website.

#### *Host Pay-Per-Click Ads*

Same as Pay-Per-Click Ads but there is evidence that the hosting provider placed the ads on the website.

Analyzing the table as a whole, since the sum of the individual Yes variables is 1,558, there are many domains with more than one type of potentially commercial activity (average of 1.72 activities for those with at least one). The most common activity in Table 5 is pay-per-click ads, which might not be considered to be potentially commercial activity by some. Therefore, we also calculated a version of the potentially commercial activity excluding domains with only pay-per-click ads. This excluded 188 domains, lowering the estimate to 44.8 percent.

## A. Apparent Domain User Type

### *Apparent Registrant Type*

**Table A.1: Apparent Domain User Type by Apparent Registrant Type  
Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Apparent Registrant Type								Total	Percent
	Natural Person	Legal Person	Privacy/Proxy	Unknown						
<b>Natural Person</b>	53.7	60.4	10.8	12.1	20.3	22.8	4.2	4.7	88.9	5.6
<b>Legal Person</b>	147.6	25.1	324.5	55.1	89.0	15.1	27.3	4.6	588.4	36.8
<b>Domain Parked</b>	116.5	35.4	95.7	29.1	98.8	30.0	18.2	5.5	329.2	20.6
<b>No Online Content</b>	128.9	31.2	130.5	31.6	75.0	18.2	78.1	18.9	412.5	25.8
<b>Unknown Type</b>	68.6	37.9	60.4	33.4	39.2	21.6	12.9	7.1	181.1	11.3
<b>Total Percent</b>	515.3	32.2	621.8	38.9	322.3	20.1	140.7	8.8	1600	100

There is a strong relationship between apparent domain user type and apparent registrant type, with a p-value for the relationship of less than .0001. Overall, 32.2 percent of registrants are apparently natural persons, but for apparent natural person domain users, this percentage is 60.4. Overall, 38.9 percent of registrants are apparently legal persons, but for apparent legal person domain users, this percentage is 55.1 percent. Only 12.1 percent of the apparently natural persons use domains registered by apparently legal persons. Overall, 20.1 percent of the domains are apparently registered using a privacy/proxy service. This percentage is highest for the domain parked domains (30.0 percent) and lowest for legal person domain users (15.1 percent). Overall, 8.8 percent of the domains have unknown registrant types, but this percentage is 18.9 percent for domains with no online content. (Note: Online content was not used to determine registrant type.)

*Potentially Commercial Activity*

**Table A.2: Apparent Domain User Type by Potentially Commercial Activity Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Potentially Commercial Activity				Total	Percent
	Not Detected	Detected				
Natural Person	56.2	63.2	32.7	36.8	88.9	5.6
Legal Person	119.1	20.2	469.3	79.8	588.4	36.8
Domain Parked	31.8	9.7	297.4	90.3	329.2	20.6
No Online Content	384.9	93.3	27.6	6.7	412.5	25.8
Unknown Type	97.4	53.8	83.7	46.2	181.1	11.3
<b>Total Percent</b>	689.4	43.1	910.6	56.9	1600	100

There is a strong relationship between apparent domain user type and Potentially Commercial activity, with a p-value for the relationship of less than .0001. Overall, 56.9 percent of domains show Potentially Commercial activity, but this is highest for domain parked domains (90.3 percent) and apparently legal person domain users (79.8 percent). Potentially Commercial activity was detected for only 6.7 percent of the domains with no online content (page 4 gives a fuller explanation of an Under Construction page with no online content other than a simple banner ad).

## B. Apparent Registrant Type

*Apparent Domain User Type*

**Table B.1: Apparent Registrant Type by Apparent Domain User Type Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Apparent Domain User Type										Total	Percent
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type							
Natural Person	53.7	10.4	147.6	28.6	116.5	22.6	128.9	25.0	68.6	13.3	515.3	32.2
Legal Person	10.8	1.7	324.5	52.2	95.7	15.4	130.5	21.0	60.4	9.7	621.8	38.9
Privacy/Proxy	20.3	6.3	89.0	27.6	98.8	30.6	75.0	23.3	39.2	12.2	322.3	20.1
Unknown	4.2	3.0	27.3	19.4	18.2	12.9	78.1	55.5	12.9	9.2	140.7	8.8
<b>Total Percent</b>	88.9	5.6	588.4	36.8	329.2	20.6	412.5	25.8	181.1	11.3	1600	100

As we discussed for Table A.1, there is a strong relationship between apparent registrant type and apparent domain user type, with a p-value for the relationship of less than .0001. This table is just Table A.1 with the rows and columns reversed. Only 5.6 percent of the domain users are apparently natural persons, but this percentage is almost doubled (10.4 percent) for registrants that are apparently natural persons. The lowest percentage of domain users that are apparently natural persons are for registrants than are apparently legal persons (1.7 percent). Overall, 36.8 percent of the domain users are apparently legal persons, but this percentage is 52.2 percent for registrants that are apparently legal persons. Overall, 20.6 percent of the domains were parked (preventing further user classification), and this percentage is highest for privacy/proxy registered domains (30.6 percent) and lower for registrants who are apparently legal persons (15.4 percent). Overall, 25.8 percent of the domains had no online content, but this percentage is 55.5 percent for unknown registrant types. Roughly ten percent of the domain users have an unknown type, regardless of the apparent registrant type.

*Potentially Commercial Activity*

**Table B.2: Apparent Registrant Type by Potentially Commercial Activity Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Potentially Commercial Activity				Total	Percent
	Not Detected	Detected				
Natural Person	229.6	44.6	285.6	55.4	515.3	32.2
Legal Person	249.5	40.1	372.3	59.9	621.8	38.9
Privacy/Proxy	114.0	35.4	208.3	64.6	322.3	20.1
Unknown	96.2	68.4	44.4	31.6	140.7	8.8
<b>Total Percent</b>	689.4	43.1	910.6	56.9	1600	100

There is a strong relationship between apparent registrant type and Potentially Commercial activity, with a p-value for the relationship of less than .0001. Overall, 56.9 percent of domains show Potentially Commercial activity, but this percentage is higher for any apparent registrant type other than unknown, which only shows Potentially Commercial activity for 31.6 percent. The differences between the other three apparent registrant types are not large.

## C. Potentially Commercial Activity Variables

### *Apparent Domain User Type*

**Table C.1: Summary of Potentially Commercial Activity Variables by Apparent Domain User Type**

Commercial Activity Variable	Percent Yes					p-value
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type	
<b>E-Commerce</b>	3.5	15.0	1.0	0.0	6.9	<.0001
<b>Membership (Online Content)</b>	0.0	3.0	1.9	0.0	1.7	0.0056
<b>Membership (Offline Content)</b>	1.2	7.7	0.3	0.0	4.1	<.0001
<b>Promotional Content (Offline)</b>	14.5	42.4	1.0	0.5	14.7	<.0001
<b>Promotional Content (Online)</b>	6.9	10.4	3.8	0.5	4.2	<.0001
<b>Host Promotional Content (Online)</b>	1.2	4.0	33.6	0.0	1.7	<.0001
<b>Third Party Banner Ads</b>	5.9	12.3	2.6	0.3	9.4	<.0001
<b>Host Banner Ads</b>	1.2	5.8	49.7	0.6	0.0	<.0001
<b>Pay-Per-Click Ads</b>	12.9	22.8	79.7	5.3	25.0	<.0001
<b>Host Pay-Per-Click Ads</b>	1.2	2.0	13.5	0.1	0.6	<.0001
<b>Potentially Commercial Activity</b>	36.8	79.8	90.3	6.7	46.2	<.0001
<b>Excluding Pay-Per-Click</b>	31.0	72.1	61.9	1.8	30.0	<.0001

All of the p-values are less than 0.0001, indicating that there are very significant differences among the apparent domain user types on the potentially commercial activity variables. All potentially commercial activity variables are significantly more likely among legal persons, except for host banner ads and the two pay-per-clicks variables, where the highest potentially commercial activity is among the domain parked domains.

**Table C.2: Potentially Commercial Activity by Apparent Domain User Type Weighted Cross-classified Frequency Counts**

Potentially Commercial Activity	Apparent Domain User Type										Total	Percent
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type							
<b>Not Detected</b>	56.2	8.2	119.1	17.3	31.8	4.6	384.9	55.8	97.4	14.1	689.4	43.1
<b>Detected</b>	32.7	3.6	469.3	51.5	297.4	32.7	27.6	3.0	83.7	9.2	910.6	56.9
<b>Total Percent</b>	88.9	5.6	588.4	36.8	329.2	20.6	412.5	25.8	181.1	11.3	1600	100

Table C.2 is the transpose of Table A.2, showing how the apparent domain user distribution differs whether the domain shows Potentially Commercial activity or not. There is a strong relationship between apparent domain user type and Potentially Commercial activity, with a p-value for the relationship of less than .0001. Overall, 5.6 percent of the domain users are apparently natural persons, but this percentage is 3.6 percent for domains with Potentially Commercial activity and 8.2 for those without Potentially Commercial activity. Overall, 36.8 percent of the domain users are apparently legal persons, but this percentage is 51.5 percent for domains with Potentially Commercial activity and only 17.3 for those without Potentially Commercial activity. Overall, 20.6 percent of the domain users are parked domains, but this percentage is 32.7 percent for domains with Potentially Commercial activity and only 4.6 for those without Potentially Commercial activity. Overall, 25.8 percent of the domain users had no online content, but this percentage is only 3.0 percent for domains with Potentially Commercial activity and 55.8 for those without Potentially Commercial activity. Overall, 11.3 percent of the domain users were of an unknown type, but this percentage is 9.2 percent for domains with Potentially Commercial activity and only 14.1 for those without Potentially Commercial activity.

*Apparent Registrant Type*

**Table C.3: Summary of Potentially Commercial Activity Variables by Apparent Registrant Type**

Commercial Activity Variable	Percent Yes				p-value
	Natural Person	Legal Person	Privacy/Proxy	Unknown	
E-Commerce	7.8	6.5	6.9	3.0	0.2383
Membership (Online Content)	0.9	1.9	1.6	3.7	0.1335
Membership (Offline Content)	2.8	3.5	5.2	1.5	0.1541
Promotional Content (Offline)	18.5	21.6	16.4	8.2	0.0019
Promotional Content (Online)	6.8	6.2	4.1	2.2	0.1044
Host Promotional Content (Online)	10.8	7.5	9.8	3.0	0.0173
Third Party Banner Ads	5.8	7.9	7.2	1.5	0.0365
Host Banner Ads	12.7	11.4	17.6	5.2	0.0017
Pay-Per-Click Ads	27.6	29.1	40.3	15.9	<.0001
Host Pay-Per-Click Ads	3.5	3.8	4.3	2.2	0.7255
<b>Potentially Commercial Activity</b>	54.6	59.5	64.0	31.6	<.0001
<b>Excluding Pay-Per-Click</b>	46.1	48.2	46.7	20.9	<.0001

Only five potentially commercial activity variables have p-values that indicate a significant difference among the apparent registrant types (i.e., offline promo content, host online promotional content, third-party/host banner ads, pay-per-click ads). If the unknowns are ignored, there are three variables with statistically significant differences between registrants who are apparently natural or legal persons on the one hand and privacy/proxy registered domains on the other hand. The privacy/proxy registered domains have a statistically significantly less online promotional content, but statistically significantly more host banner ads and pay-per-click ads.

**Table C.4: Potentially Commercial Activity by Apparent Registrant Type Weighted Cross-classified Frequency Counts**

Potentially Commercial Activity	Apparent Registrant Type							Total	Percent	
	Natural Person	Legal Person	Privacy/Proxy	Unknown						
Not Detected	229.6	33.3	249.5	36.2	114.0	16.5	96.2	14.0	689.4	43.1
Detected	285.6	31.3	372.3	40.9	208.3	22.9	44.4	4.9	910.6	56.9
<b>Total Percent</b>	515.3	32.2	621.8	38.9	322.3	20.1	140.7	8.8	1600	100

Table C.4 is the transpose of Table B.2, showing how the apparent registrant distribution differs whether the domain shows Potentially Commercial activity or not. There is a strong relationship between



apparent registrant type and Potentially Commercial activity, with a p-value for the relationship of less than .0001. Overall, 32.2 percent of the registrants are apparently natural persons, and this percentage differs little for domains with Potentially Commercial activity (31.3) and those without Potentially Commercial activity (33.3). Overall, 38.9 percent of the registrants are apparently legal persons, and this percentage differs little for domains with Potentially Commercial activity (40.9) and those without Potentially Commercial activity (36.2). Overall, 20.1 percent of the registrants are privacy/proxy registered domains, but this percentage is 22.9 percent for domains with Potentially Commercial activity and only 16.5 for those without Potentially Commercial activity; this difference is statistically significant. Overall, 8.8 percent of the registrants were of an unknown type, but this percentage is only 4.9 percent for domains with Potentially Commercial activity and 14.0 for those without Potentially Commercial activity.

## D. Business Structure of Domain User

Generic business structure of the domain user was coded based on observed domain content that included HTML content and images extracted from “www.domainname”. Coders made direct observations on the domain user's business structure and indirect observations on other aspects, such as the domain user's business function, that may provide additional clues to the domain user's business structure. Their recorded observations were then categorized into 11 major types as described below. We searched in the coder observations for keywords that best characterize each category. When a record is associated with keywords corresponding to multiple business structure types, certain rules were applied to finalize it to a best fit category. Less than 3 percent of cases that were not suitable for automation were manually reviewed and finalized. Our main goal for this variable was to determine if the domain user could be considered a for-profit business, a non-profit business, or not a business at all. We split the for-profit businesses into sole proprietorships, partnerships, and corporations if we could. Some domains in languages other than English were clearly businesses, but were not classifiable. Domains with no content, as well as parked domains and under construction domains were assigned to unclear business structure categories. One other category was created for when some business activity was detected, but it was not clear whether or not the domain was a business. The remaining domains with no clear domain user type were assigned to an Undetermined category. Table D.1 shows the full frequency for the generic business structure of the domain user:

**Table D.1: Generic Business Structure of Domain User**

<b>Description</b>	<b>Frequency</b>	<b>Percent</b>
Undetermined	940	58.9
For Profit: Corporation	268	16.8
Not a Business (natural person, blog)	102	6.2
Unclear Business Structure: No Content (domain parked, under construction)	62	3.9
Unclear Business Structure: Unable to determine	49	3.1
For Profit: Partnership	38	2.4
Not For Profit (Nonprofit, governments, political, education, religious, or community groups)	37	2.3
For Profit: Sole Proprietor	32	2.0
Non-U.S. Business	31	1.9
For Profit: Other	29	1.8
Unclear Formal Structure: Apparent Business Activities	12	0.8

For analysis, we combined the eleven categories shown in Table D.1 into four categories. We combined all four “For Profit” categories together, kept the “Non-Profit” and “Not a Business” categories, and combined the other five categories into “Unclear Business Structure.”

Therefore, our commercial business analysis variable has four levels as shown in Table D.2.

**Table D.2: Business Structure of Domain User Variable Used in Analyses**

<b>Description</b>	<b>Frequency</b>	<b>Percent</b>
Domain User appears to be for-Profit Business	410	25.6
Domain User appears to be non-Profit Business	37	2.3
Domain User is not a Business	102	6.4
Domain User has unclear Business Structure	1,051	65.7

*Apparent Domain User Type*

**Table D.3: Apparent Domain User Type by Business Structure Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Business Structure of Domain User							Total	Percent	
	For Profit		Non-profit		Not a Business		Unclear Business Structure			
Natural Person	0	0	0	0	88.9	100	0	0	88.9	5.6
Legal Person	410.4	69.7	39.2	6.7	14.7	2.5	124.2	21.1	588.4	36.8
Domain Parked	0	0	0	0	0	0	329.2	100	329.2	20.6
No Online Content	0	0	0	0	0	0	412.5	100	412.5	25.8
Unknown Type	0	0	0	0	1.1	0.6	180.0	99.4	181.1	11.3
<b>Total Percent</b>	410.4	25.6	39.2	2.5	104.6	6.5	1045.9	65.4	1600	100

The relationship between apparent domain user type and the domain user's business structure is statistically significant with a chi-squared p-value of less than 0.0001. All of the For Profit and Non-Profit businesses have been classified as legal person domain users while all of the apparently natural person domain users have been classified as not a business. All of the domains parked and domains with no online content have an unclear business structure, while almost all of the unknown type domain users also have an unclear business structure. Looking at the row with domain users who are apparently legal persons, almost 70 percent of the domains appear to be for-profit businesses, while under 7 percent appear to be non-profit businesses and only 2.5 percent do not appear to be businesses at all. It should be noted that the sample size of the apparently non-profit business category is too small for analysis.

*Apparent Registrant Type*

**Table D.4: Apparent Registrant Type by Business Structure Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Business Structure of Domain User							Total	Percent	
	For Profit		Non-profit		Not a Business		Unclear Business Structure			
Natural Person	100.4	19.5	12.7	2.5	60.9	11.8	341.2	66.2	515.3	32.2
Legal Person	229.7	36.9	20.1	3.2	15.0	2.4	357.0	57.4	621.8	38.9
Privacy/Proxy	62.4	19.4	6.4	2.0	23.4	7.3	230.1	71.4	322.3	20.1
Unknown	17.8	12.7	0	0	5.2	3.7	117.6	83.6	140.7	8.8
<b>Total Percent</b>	410.4	25.6	39.2	2.5	104.6	6.5	1045.9	65.4	1600	100

The relationship between apparent domain registrant type and business structure of domain user is statistically significant with a chi-squared p-value of less than 0.0001. Overall, 25.6 percent of the domain users have a for-profit business structure, but this percentage is 36.9 for domains registered by apparently legal persons. Except for unknown registrant types, two or three percent of the domain users have a non-profit business structure. Only 6.5 percent of the domains are used by an entity that could be classified as a non-business, but this percentage is almost double (11.8 percent) for domains registered to apparently natural persons and less than half (2.4 percent) for domains registered to apparently legal persons. Most of the domains in all registrant types, though, do have an unclear domain user's business structure.

### *Potentially Commercial Activity*

**Table D.5: Potentially Commercial Activity by Business Structure of Domain User**

	Percent Yes				p-value
	For Profit	Non-Profit	Not a Business	Unclear Business Structure	
<b>Potentially Commercial Activity</b>	83.8	53.8	39.3	48.2	<.0001

The relationship Potentially Commercial activity and domain user's business structure is statistically significant with a chi-squared p-value of less than 0.0001. The For-Profit business structure domains had the highest percentage of Potentially Commercial activity (83.8 percent)<sup>2</sup>, but the other business structures also showed a lot of Potentially Commercial activity (overall, 56.9 percent of the domains have shown Potentially Commercial activity).

## **E. Domain Name Extension (gTLD)**

Table 1 above shows the top five generic top-level domains and the distribution of the domains in our sample across these gTLDs. We compare all five domain name extensions as much as possible below.

<sup>2</sup> Note that business structure was coded independently of potentially commercial activity, so the presence of potentially commercial activity is not the reason a domain user was classified as a for-profit business.

### Apparent Domain User Type

**Table E.1: Apparent Domain User Type by Domain Name Extension Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Domain Name Extension										Total	Percent
	*.com		*.net		*.org		*.info		*.biz			
Natural Person	68.5	77.0	13.5	15.2	2.2	2.4	3.9	4.4	0.8	0.9	88.9	5.6
Legal Person	451.9	76.8	58.3	9.9	46.5	7.9	22.6	3.8	9.1	1.5	588.4	36.8
Domain Parked	246.5	74.9	36.4	11.1	21.6	6.6	19.7	6.0	4.9	1.5	329.2	20.6
No Online Content	281.3	68.2	50.0	12.1	32.4	7.9	41.3	10.0	7.5	1.8	412.5	25.8
Unknown Type	140.1	77.4	13.5	7.5	13.0	7.2	10.8	6.0	3.6	2.0	181.1	11.3
<b>Total Percent</b>	1188.2	74.3	171.8	10.7	115.7	7.2	98.3	6.1	26.0	1.6	1600	100

There is enough of a relationship between apparent domain user type and generic top-level domain (gTLD) name extension for a significant chi-square p-value of 0.0381. However, it does not appear to be a strong relationship. Overall, 74.3 percent of all domains are \*.com domains, and only the No Online Content domains differ (68.2 percent). Overall, 10.7 percent of all domains are \*.net, with the highest rate among the domain users who are apparently natural persons (15.2 percent) and the lowest rate among the unknown domain user types (7.5 percent). About seven percent of all domain user types are \*.org except the apparently natural person domain users (2.4 percent). The most variable rates occur for the \*.info gTLD. Overall, 6.1 percent of the domains are \*.info domains, but the no online content domains have a 10.0 percent rate while the apparently natural person domain users (4.4 percent) and apparently legal person domain users (3.8 percent) have lower rates. The \*.biz gTLD represents about one or two percent of domains in all domain user types.

### Apparent Registrant Type

**Table E.2: Apparent Registrant Type by Domain Name Extension Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Domain Name Extension										Total	Percent
	*.com		*.net		*.org		*.info		*.biz			
Natural Person	381.3	74.0	55.2	10.7	34.6	6.7	32.4	6.3	11.7	2.3	515.3	32.2
Legal Person	455.1	73.2	76.0	12.2	57.3	9.2	24.6	4.0	8.8	1.4	621.8	38.9
Privacy/Proxy	238.1	73.9	29.2	9.0	19.5	6.0	31.5	9.8	4.2	1.3	322.3	20.1
Unknown	113.8	80.9	11.5	8.1	4.3	3.1	9.8	7.0	1.3	0.9	140.7	8.8
<b>Total Percent</b>	1188.2	74.3	171.8	10.7	115.7	7.2	98.3	6.1	26.0	1.6	1600	100

The relationship between domain name extensions and apparent registrant type is significant with a chi-squared p-value of 0.0124. Overall, 74.3 percent of all domains are \*.com domains, and only the Unknown Registrant Type domains differ (80.9 percent). Overall, 10.7 percent of all domains are \*.net, with the highest rate among the registrants who are apparently legal persons (12.2 percent) and the lowest rates among the privacy/proxy registered domains (9.0 percent) and the unknown registrant types (8.1 percent). Overall, 7.2 percent of all domains are \*.org domains, but the percentage of registrants that are apparently legal persons is 9.2 while the percentage for Unknown Registrant Types is only 3.1 percent. Overall, 6.1 percent of the domains are \*.info domains, but the privacy/proxy registered domains have a 9.8 percent rate while the apparently legal person registrants only have a 4.0 percent rate. Overall, 1.6 percent of all domains are \*.biz domains, but this rate is higher for registrants who are apparently natural persons (2.3 percent) and lower for Unknown Registrant Types (0.9 percent).

### *Potentially Commercial Activity*

**Table E.3: Potentially Commercial Activity by Domain Name Extension**

	Percent Yes					p-value
	*.com	*.net	*.org	*.info	*.biz	
<b>Potentially Commercial Activity</b>	59.0	55.8	47.7	47.0	50.0	.0315

The relationship Potentially Commercial activity and domain name extension is statistically significant with a chi-squared p-value of 0.0315. Compared with other tables, the differences are not that large, but the \*.com and \*.net domains do show more Potentially Commercial Activity than the \*.org and \*.info domains.

## **F. Registrant Country/Region of the World**

Through our research, we were able to identify the registrant country for all but 82 of the domain names. For one domain name, there was conflicting information as to whether it was in Japan or Australia; for the remaining 81 missing registrant countries, no WHOIS information existed to be used to determine the registrant country. Table F.1 shows the countries represented by at least one domain name in our sample.

**Table F.1: Countries Represented in the Registrant ID Study Domain Sample**

Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
United States	864	54.0	864	54.0
China	76	4.8	940	58.8
United Kingdom	76	4.8	1,016	63.5
Germany	56	3.5	1,072	67.0
Australia	50	3.1	1,122	70.1
Canada	50	3.1	1,172	73.3
Spain	34	2.1	1,206	75.4
France	31	1.9	1,237	77.3
Japan	29	1.8	1,266	79.1
The Netherlands	26	1.6	1,292	80.8
Italy	22	1.4	1,314	82.1
Turkey	20	1.3	1,334	83.4
India	17	1.1	1,351	84.4
Switzerland	11	0.7	1,362	85.1
Russia	11	0.7	1,373	85.8
Indonesia	9	0.6	1,382	86.4
Brazil	8	0.5	1,390	86.9
Hong Kong	8	0.5	1,398	87.4
Vietnam	8	0.5	1,406	87.9
Singapore	7	0.4	1,413	88.3
Belgium	6	0.4	1,419	88.7
Cayman Islands	6	0.4	1,425	89.1
Norway	6	0.4	1,431	89.4
Sweden	6	0.4	1,437	89.8
Thailand	6	0.4	1,443	90.2
Czech Republic	4	0.3	1,447	90.4
Ireland	4	0.3	1,451	90.7
South Korea	4	0.3	1,455	90.9
Mexico	4	0.3	1,459	91.2
South Africa	4	0.3	1,463	91.4
Bermuda	3	0.2	1,466	91.6
Denmark	3	0.2	1,469	91.8
Finland	3	0.2	1,472	92.0
Greece	3	0.2	1,475	92.2
Philippines	3	0.2	1,478	92.4
Poland	3	0.2	1,481	92.6
Saudi Arabia	3	0.2	1,484	92.8
Bosnia and Herzegovina	2	0.1	1,486	92.9
Hungary	2	0.1	1,488	93.0
Israel	2	0.1	1,490	93.1
Iran	2	0.1	1,492	93.3
Malaysia	2	0.1	1,494	93.4

Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
New Zealand	2	0.1	1,496	93.5
Venezuela	2	0.1	1,498	93.6
British Virgin Islands	2	0.1	1,500	93.8
United Arab Emirates	1	0.1	1,501	93.8
Argentina	1	0.1	1,502	93.9
Austria	1	0.1	1,503	93.9
Bolivia	1	0.1	1,504	94.0
Bahamas	1	0.1	1,505	94.1
Chile	1	0.1	1,506	94.1
Cyprus	1	0.1	1,507	94.2
Egypt	1	0.1	1,508	94.3
Croatia	1	0.1	1,509	94.3
Jordan	1	0.1	1,510	94.4
Lebanon	1	0.1	1,511	94.4
Nicaragua	1	0.1	1,512	94.5
Peru	1	0.1	1,513	94.6
Puerto Rico	1	0.1	1,514	94.6
Qatar	1	0.1	1,515	94.7
Serbia	1	0.1	1,516	94.8
Ukraine	1	0.1	1,517	94.8
Uruguay	1	0.1	1,518	94.9
Ambiguous	1	0.1	1,519	94.9
Unknown (no data available)	81	5.1	1,600	100.0

For countries with at least fifty (50) domain names (United States, China, United Kingdom, Germany, Australia, and Canada), we have analyzed them separately. We have combined the other countries by region as follows: Other Europe, Other Asia/Pacific, and Other (North America excluding the U.S. and Canada, South America, Caribbean Islands, and Africa). Table F.2 shows the frequency for the analysis variable we used to represent country/region of the world. We concentrated on the nine subgroups with data available to analyze.



**Table F.2: Countries/Regions of the World Used in Analyses**

Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
United States	864	54.0	864	54.0
China	76	4.8	940	58.8
United Kingdom	76	4.8	1,016	63.5
Germany	56	3.5	1,072	67.0
Australia/New Zealand	52	3.3	1,124	70.3
Canada	50	3.1	1,174	73.4
Other Europe	170	10.6	1,344	84.0
Other Asia/Pacific	136	8.5	1,480	92.5
Other	38	2.4	1,518	94.9
Ambiguous/Missing	82	5.1	1,600	100.0

*Apparent Domain User Type*

**Table F.3: Apparent Domain User Type by Country/Region of the World Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Registrant Country											
	United States		China		United Kingdom		Germany		Australia/New Zealand		Canada	
Natural Person	44.4	50.0	2.1	2.4	7.3	8.2	8.6	9.7	2.1	2.4	0	0
Legal Person	320.7	54.9	33.7	5.8	25.5	4.4	24.8	4.2	16.7	2.9	19.2	3.3
Domain Parked	217.9	67.3	11.6	3.6	14.9	4.6	6.3	1.9	16.1	5.0	13.7	4.2
No Online Content	202.0	59.3	20.2	5.9	11.1	3.3	9.3	2.7	13.1	3.8	9.8	2.9
Unknown Type	85.2	47.9	11.6	6.5	16.0	9.0	5.5	3.1	4.1	2.3	8.4	4.7
<b>Total Percent</b>	870.3	57.4	79.2	5.2	74.9	4.9	54.6	3.6	52.2	3.4	51.1	3.4

  

Apparent Domain User Type	Other							Total	Percent
	Other Europe		Other Asia		Other				
Natural Person	11.8	13.3	12.5	14.1	0	0	88.9	5.9	
Legal Person	75.3	12.9	49.6	8.5	18.7	3.2	584.2	38.5	
Domain Parked	19.1	5.9	18.0	5.6	6.3	2.0	323.9	21.4	
No Online Content	37.1	10.9	29.8	8.7	8.5	2.4	340.8	22.5	
Unknown Type	18.4	10.3	24.3	13.7	4.3	2.4	177.9	11.7	
<b>Total Percent</b>	161.6	10.7	134.2	8.9	37.8	2.4	1516	100	

There is a strong relationship between apparent domain user type and country/region of the world, with a p-value of less than .0001. Overall, 57.4 percent of the domains have a United States registrant, but this percentage is 67.3 percent for parked domains and is only 50.0 percent for domain users that are apparently natural persons (and 47.9 percent for unknown domain user types). Overall, 5.2 percent of the

domains have Chinese registrants, but this percentage is 6.5 percent for unknown domain user type domains while this percentage is only 2.4 percent for domain users that are apparently natural persons (and 3.6 percent for parked domains). Overall, 4.9 percent of the domains have United Kingdom registrants, but this percentage is 8.2 for domain users that are apparently natural persons (and 9.0 percent for unknown domain user types) while this percentage is only 3.3 percent for domains with no online content. Overall, 3.6 of the domains have a German registrant, but this percentage is 9.7 percent for domain users that are apparently natural persons and is only 1.9 percent for parked domains. Overall, 3.4 of the domains have an Australia or New Zealand registrant, but this percentage is 5.0 percent for parked domains and is only 2.4 percent for domain users that are apparently natural persons (and 2.3 for unknown domain user types). Canadian registrants make up three to five percent of the domains in each domain user type category except that there are no Canadian registrants for domain users that are apparently natural persons. The overall percentage for other European countries (besides the United Kingdom and Germany) is 10.7, but this percentage is higher for domain users that are apparently natural persons (13.3 percent) and domain users that are apparently legal persons (12.9 percent), but lower for parked domains (5.9 percent). The overall percentage for other Asian and Pacific countries (besides China) is 8.9, but this percentage is 14.1 percent for domain users that are apparently natural persons (and 13.7 percent for unknown domain user types), but lower for parked domains (5.6 percent). Registrants from other countries and regions make up two to three percent of the domains in each domain user type category except that there are no registrants from these other countries/regions for domain users that are apparently natural persons.

It seems clear from the above that domain users who are apparently natural persons differ the most from the other categories in the distribution by country/region of the world.

### *Apparent Registrant Type*

**Table F.4: Apparent Registrant Type by Country/Region of the World Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Registrant Country											
	United States		China		United Kingdom		Germany		Australia/New Zealand		Canada	
Natural Person	239.9	46.8	48.7	9.5	33.3	6.5	27.8	5.4	8.7	1.7	11.6	2.3
Legal Person	366.7	59.4	15.8	2.6	33.0	5.3	25.7	4.2	12.5	2.0	19.2	3.1
Privacy/Proxy	238.0	74.3	3.1	1.0	3.4	1.1	1.0	0.3	29.9	9.3	20.3	6.3
Unknown	25.7	39.0	11.6	17.6	5.3	8.0	0	0	1.1	1.6	0	0
<b>Total Percent</b>	870.3	57.4	79.2	5.2	74.9	4.9	54.6	3.6	52.2	3.4	51.1	3.4

<b>Apparent Registrant Type</b>	<b>Other Europe</b>		<b>Other Asia</b>		<b>Other</b>		<b>Total</b>	<b>Percent</b>
<b>Natural Person</b>	72.6	14.2	58.0	11.3	11.5	2.1	512.1	33.8
<b>Legal Person</b>	72.9	11.8	54.1	8.8	17.6	2.9	617.6	40.7
<b>Privacy/Proxy</b>	5.5	1.7	14.8	4.6	4.2	1.3	320.2	21.1
<b>Unknown</b>	10.7	16.2	7.4	11.2	4.3	6.4	65.8	4.3
<b>Total Percent</b>	161.6	10.7	134.2	8.9	37.6	2.4	1516	100

There is a strong relationship between apparent registrant type and country/region of the world, with a p-value of less than .0001. Overall, 57.4 of the domains have a United States registrant, but this percentage is 74.3 percent for privacy/proxy registered domains and is only 46.8 percent for registrants that are apparently natural persons (and 39.0 percent for unknown registrant types). Overall, 5.2 percent of the domains have Chinese registrants, but this percentage is 17.6 percent for unknown registrant type domains while this percentage is only 1.0 percent for privacy/proxy registered domains (and only 2.6 percent for registrants that are apparently legal persons). Overall, 4.9 percent of the domains have United Kingdom registrants, but this percentage is 6.5 for registrants that are apparently natural persons (and 8.0 percent for unknown registrant types) while this percentage is only 1.1 percent for privacy/proxy registered domains. Overall, 3.6 of the domains have a German registrant, but almost all are registrants that are apparently natural persons (5.4 of apparently natural person registrants) and registrants that are apparently legal persons (4.2 percent of all apparently legal person registrants) while almost none are privacy/proxy registered domains (0.3 percent of privacy/proxy registered domains) or unknown registrant types (none of the 66 unknown registrant type registrants). Overall, 3.4 of the domains have an Australia or New Zealand registrant, but this percentage is 9.3 percent for privacy/proxy registered domains and two percent for all other apparent registrant types). Overall, 3.4 of the domains have a Canadian registrant, but this percentage is 6.3 percent for privacy/proxy registered domains while there were no unknown registrant cases (out of 66 total unknown registrant cases) with Canadian registrants. The overall percentage for other European countries (besides the United Kingdom and Germany) is 10.7, but this percentage is much lower for privacy/proxy registered domains (1.7 percent) and higher for registrants that are apparently natural persons (14.2 percent) and unknown registrant type registrants (16.2 percent). The overall percentage for other Asian and Pacific countries (besides China) is 8.9, but this percentage is 11.3 percent for registrants that are apparently natural persons (and 11.2 percent for unknown registrant types), but lower for privacy/proxy registered domains (4.6 percent). The overall percentage for all other countries and regions is 2.4, but this percentage is higher (6.4 percent) for unknown registrant types and lower (1.3 percent) for privacy/proxy registered domains.

It seems clear from the above that privacy/proxy registered domains differ the most from the other categories in the distribution by country/region of the world.

*Potentially Commercial Activity*

**Table F.5: Potentially Commercial Activity by Country/Region**

	Percent Yes					
	United States	China	United Kingdom	Germany	Australia/ New Zealand	Canada
<b>Potentially Commercial Activity</b>	63.8	50.5	62.6	39.1	58.9	60.3

	Percent Yes			p-value
	Other Europe	Other Asia	Other	
<b>Potentially Commercial Activity</b>	51.7	50.3	69.1	.0003

There is a strong relationship between apparent registrant type and country/region of the world, with a p-value of .0003. Ignoring the “Other” category, the United States has the highest rate of Potentially Commercial activity (63.8 percent) while Germany has the lowest rate (39.1 percent). The United Kingdom has the second highest rate (62.6 percent) while China and the Other Asia region have rates around 50 percent.

## G. Relationship of Domain User to Registrant

The relationship between the Domain User and the Registrant was coded during the second phase of the Domain User manual coding process. The entity listed in the WHOIS data Registrant Name and Registrant Organization fields were compared to the Domain User and the type of the relationship existing between the two entities was recorded. Here is a frequency:

**Table G.1: Relationship between Domain User and Registrant**

<b>Relationship Description</b>	<b>Frequency</b>	<b>Percent</b>
No Apparent Relationship: Unable to determine relationship	868	54.3
Domain User is Customer of Registrant: Privacy or Proxy service registered domain	327	20.4
Domain User same as Registrant both Legal Person	198	12.4
Domain User is Employer of Registrant	79	4.9
Domain User same as Registrant, both Natural Person	67	4.2
Domain User is Customer of Registrant: Web Developer/Development /Consulting company registered domain	27	1.7
Domain User is Customer of Registrant: Hosting or Domain provider	19	1.2
Other Specify	13	0.8
No Apparent Relationship: Registrant appears fictitious or falsified	2	0.1

For our analysis purposes, we collapsed these nine categories into four categories with the Other Specify categorized based on the text description. We combined the two categories where the Domain User is also the Registrant, whether Natural or Legal person (plus three Other Specify cases); we kept the Domain User is Customer of Privacy/Proxy Registered Domain separate, but we combined the two other “Domain User is Customer” categories together (plus one Other Specify case); and we combined the “Domain User is Employer” category with two Other Specify cases where the Domain User was the Employee of the Registrant. The remaining two “No Apparent Relationship” categories were combined with the remaining seven Other Specify cases to make the “Unknown” category. Table G.2 shows the frequency of the Relationship variable used in our analyses:

**Table G.2: Relationship Variable Used in Analyses**

<b>Relationship Description</b>	<b>Frequency</b>	<b>Percent</b>
Domain User Same as Registrant	268	16.8
Domain User is Customer of Privacy/Proxy Registered Domain (PRIVACY/PROXY)	327	20.4
Domain User is Customer of Other Registrant (OTHER CUSTOMER)	47	3.0
Domain User is Employer/Employee of Registrant (EMPLOYER/EMPLOYEE)	81	5.1
Unable to Determine Relationship	877	54.8

*Apparent Domain User Type*

**Table G.3: Relationship of Domain User and Registrant by Domain User Type Weighted Cross-classified Frequency Counts**

Apparent Domain User Type	Relationship of Domain User to Registrant										Total	Percent
	Domain User Same as Registrant	Privacy/Proxy	Other Customer	Employer/Employee	Unable to Determine Relationship							
Natural Person	62.0	69.8	19.2	21.6	2.1	2.4	1.1	1.2	4.5	5.0	88.9	5.6
Legal Person	208.6	35.5	76.5	13.0	37.6	6.4	75.7	12.9	190.0	32.3	588.4	36.8
Domain Parked	3.2	1.0	122.2	37.1	3.2	1.0	0	0	200.6	61.0	329.2	20.6
No Online Content	0	0	71.9	17.4	1.1	0.3	0	0	339.5	82.3	412.5	25.8
Unknown Type	0	0	38.1	21.1	4.5	2.5	2.1	1.2	136.4	75.3	181.1	11.3
<b>Total Percent</b>	273.8	17.1	327.9	20.5	48.4	3.0	78.9	4.9	871.0	54.4	1600	100

The relationship between apparent domain user type and the relationship of domain user to registrant is highly significant with a chi-squared p-value of less than 0.0001. Overall, the percentage of domain users who are the same entity as the registrant is 17.1, but this percentage is much higher for domain users who are apparently natural persons (69.8 percent) and domain users who are apparently legal persons (35.5 percent) while very few for the less defined domain user types (parked domains, no online content and unknown domain user type). Overall, the percentage of domain users who are clients of privacy/proxy registered domains is 20.5 percent, but this percentage is higher for parked domains (37.1 percent) and lower for domains with no online content (17.4 percent) and domain users who are apparently legal persons (13.0 percent). Overall, the percentage of domain users who are clients of other registrants (not privacy/proxy registered domains) is 3.0 percent, but this percentage is higher for domain users who are apparently legal persons (6.4 percent) and lower for domains with no online content (0.3 percent) and parked domains (1.0 percent). Almost all of the employer/employee relationships between the domain user and registrant were for domain users who are apparently legal persons (12.9 percent of domain users who are apparently legal persons), with all other domain user types having such a relationship only zero or one percent of the time. Overall, we were unable to determine the relationship for 54.4 of the domains, but this percentage was especially low (5.0 percent) for domain users who are apparently natural persons, lower (32.3 percent) for domain

users who are apparently legal persons and highest for domains with no online content (82.3 percent) and unknown domain user type domains (75.3 percent).

*Apparent Registrant Type*

**Table G.4: Relationship of Domain User and Registrant by Registrant Type Weighted Cross-classified Frequency Counts**

Apparent Registrant Type	Relationship of Domain User to Registrant										Total	Percent
	Domain User Same as Registrant	Privacy/Proxy		Other Customer		Employer/Employee		Unable to Determine Relationship				
<b>Natural Person</b>	88.1	17.1	9.6	1.9	7.3	1.4	35.8	6.9	374.5	72.7	515.3	32.2
<b>Legal Person</b>	173.1	27.8	16.0	2.6	31.7	5.1	41.0	6.6	360.0	57.9	621.8	38.9
<b>Privacy/Proxy</b>	3.2	1.0	299.2	92.8	8.3	2.6	2.1	0.7	9.5	2.9	322.3	20.1
<b>Unknown</b>	9.5	6.7	3.1	2.2	1.0	0.7	0	0	127.0	90.3	140.7	8.8
<b>Total Percent</b>	273.8	17.1	327.9	20.5	48.4	3.0	78.9	4.9	871.0	54.4	1600	100

The relationship between apparent registrant type and relationship of domain user to registrant is highly significant with a chi-squared p-value of less than 0.0001. We expect that the privacy/proxy registered domains will have their domain users all be customers, and this is almost true. Overall, the percentage of domain users who are the same entity as the registrant is 17.1, but this percentage is 27.8 for domain users who are apparently legal persons, while this percentage is only 1.0 percent for privacy/proxy registered domains (and is only 6.7 percent for unknown registrant type registrants). Overall, the percentage of domain users who are customers of privacy/proxy registered domains is 20.5 percent, but this percentage is much higher for privacy/proxy registered domains (92.8 percent) and much lower (less than eight percent) for all three of the other registrant types. Overall, the percentage of domain users who are customers, but are not privacy/proxy registered domains, is 3.0 percent, but this percentage is much higher for domain users who are apparently legal persons (5.1 percent) and lower for domain users who are apparently natural persons (1.4 percent). Overall, the percentage of domains with an employer/employee relationship between the domain user and registrant was 4.9 percent, but almost all of these relationships were for domain users who are apparently natural persons (6.9 percent of domain users who are apparently natural persons) and for domain users who are apparently legal persons (6.6 percent of domain users who are apparently legal persons) with privacy/proxy registered domains and unknown registrant types having such a relationship less

than one percent of the time. Overall, we were unable to determine the relationship for 54.4 of the domains, but this percentage was especially low (2.4 percent) for privacy/proxy registered domains and higher for registrants who are apparently natural persons (72.7 percent) and for unknown registrant types (90.3 percent).

*Potentially Commercial Activity*

**Table G.5: Potentially Commercial Activity by Relationship between Registrant and the Domain User**

	Percent Yes					p-value
	Domain User Same as Registrant	Privacy/ Proxy	Other Customer	Employer/ Employee	Unable to Determine Relationship	
<b>Potentially Commercial Activity</b>	67.5	65.9	80.4	83.5	46.5	<.0001

The relationship between Potentially Commercial activity and the relationship between the registrant and the domain user is statistically significant with a chi-squared p-value of less than 0.0001. The relationships that showed the most Potentially Commercial activity occurs when the user and registrant have an employer/employee relationship or a (non-privacy/proxy) customer relationship, while the lowest Potentially Commercial activity rate was among those domains where we were unable to determine the relationship between the domain user and registrant. This low rate may be related to the fact that we weren't able to determine the relationship for domains with no online content.

## H. Other Coded Behavior Variables

Two other coded behavior variables were used to indicate whether any alleged illegal or harmful activity was detected and whether any explicit sexual imagery was found (this differs from the analysis below on whether a domain could be matched to any blacklists). These allegedly illegal or harmful activities were coded during the Domain Content manual coding process by manually reviewing the web content for evidence of each of the activities listed in Table H.1. During the training process, coders were supplied with definitions of each of the activities, and a few examples of websites engaging in the activities were provided. However, it should be noted that the coders were not experts in Internet crime and detecting the



presence of these activities on web pages. Table H.1 shows the frequency of our allegedly illegal or harmful activity variable:

**Table H.1: Allegedly Illegal or Harmful Activities: Manually Coded**

<b>Allegedly illegal or harmful Activity</b>	<b>Frequency</b>	<b>Percent</b>
No allegedly illegal or harmful activities detected	1,582	98.9
Spam	4	0.3
Advance fee fraud (aka 419 scams)	4	0.3
Phishing	3	0.2
Cybersquatting/Typosquatting	3	0.2
Counterfeit merchandise (i.e., domain website appears to sell CM)	2	0.1
Trademark infringement (i.e., domain website appears to...)	1	0.1
Malware	1	0.1
Intellectual property theft	0	0.0
Child sexual images	0	0.0
Identity theft	0	0.0
Money laundering	0	0.0

Allegedly illegal or harmful activities were only observed for 18 out of the 1,600 domains (1.1 percent). In our analyses, we converted this variable to a binary variable of whether any alleged illegal activity was detected. Table H.2 shows the frequency of whether explicit sexual images were at the domain:

**Table H.2: Explicit Sexual Images: Manually Coded**

<b>Explicit Sexual Images</b>	<b>Frequency</b>	<b>Percent</b>
No	1,584	99.0
Yes	16	1.0

Even though both of these variables were rarely yes, we still carried out analyses to see if these two behaviors were more likely among certain subgroups.

*Apparent Domain User Type*

**Table H.3: Coded Behavior Variables by Apparent Domain User Type**

Coded Variable	Percent Yes					p-value
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type	
Allegedly illegal or harmful Activity	1.2	2.1	1.0	0	1.2	0.0653
Explicit Sexual Images	2.4	1.6	0.6	0	1.7	0.0611

While the p-values are close to significant, the p-values are not significant even though one of the apparent domain user types (no online content) could not show these coded behaviors. For allegedly illegal or harmful activity, there is a slightly higher rate (2.1 percent) among the domain users who are apparently legal persons. Few of the parked domains showed explicit sexual images (0.6 percent) while there was a slightly higher rate for domain users who are apparently natural persons.

*Apparent Registrant Type*

**Table H.4: Coded Behavior Variables by Apparent Registrant Type**

Coded Variable	Percent Yes				p-value
	Natural Person	Legal Person	Privacy/ Proxy	Unknown	
Allegedly illegal or harmful Activity	1.6	0.5	1.6	1.5	0.0580
Explicit Sexual Images	0.6	1.0	1.6	1.5	0.5173

The p-value for explicit sexual images shows no significant differences between the apparent registrant types. The p-value for allegedly illegal or harmful activity shows that the difference between registrants who apparently are legal persons (0.5 percent) and all other apparent registrant types (1.5-1.6 percent) is almost statistically significant. Meanwhile, the percentages of explicit sexual images are lower for registrants who apparently are natural persons (0.6 percent) and for registrants who apparently are legal persons (1.0 percent), but the differences in the percentage of domains with explicit sexual images could be due to random error.

*Potentially Commercial Activity*

**Table H.5: Coded Behavior Variables by Potentially Commercial Activity**

Coded Variable	Percent Yes		p-value
	No Potentially Commercial Activity	Potentially Commercial Activity	
Allegedly illegal or harmful Activity	0.8	1.5	0.5509
Explicit Sexual Images	0.9	1.2	0.6416

For both of these coded behavior variables, the domains with Potentially Commercial activity have a higher rate of the coded behavior, but the differences are not large enough to be statistically significant.

## I. Blacklist Variables

In an effort to determine allegedly illegal or harmful activities, DNSBL lists were scanned for each sample member. The DNSBL strategy was to obtain all the “ARECORDS” associated with the domain for each sample member. For each ARECORD, the returned IP address was checked against a series of DNSBLs. After running this process, we reviewed the frequency of responses received from each DNSBL. Many of the DNSBLs did not return a response, so they were removed from our analysis. For the remaining DNSBLs which returned a response, NORC conducted a review of the site to determine the relevancy of the list. Many of the lists contained an abundance of historic DNSBL listings or were no longer actively maintained, so these were removed from the analysis. Some of the response octates returned by the DNSBLs provided a trustworthiness score of the listing to indicate how sure the DNSBL is that the listing is accurate. Scores of low trustworthiness were removed from the analysis. Table I.1 is a summary of the allegedly illegal or harmful activity categories as determined by the top-ranked blacklists. It is possible for a domain to be categorized in more than one way, so the categories in the summary table are not distinct. The total number of domains associated with any top-ranked blacklist activity is provided at the bottom of the table.

**Table I.1: Allegedly illegal or harmful Activities: Domains Found on Top-Ranked Blacklists**

Description	Frequency	Percent*
Abusive	2	0.1
Abusive host	5	0.3
Abusive host & anonymous-state	28	1.8
Backscatter	28	1.8
Ddos attacks	1	0.1
Dynamic-ip	7	0.4
Spam	82	5.1
Spam abuse vulnerability	6	0.4
Spam bad host, no cookie	1	0.1
Suspicious	5	0.3
Suspicious & comment spammer	1	0.1
Tor network	1	0.1
Trojan/virus/bot	2	0.1
On Any Top-Ranked Blacklist	141	8.8

In the following analyses, we restrict our analyses to the most common four allegedly illegal or harmful activities: any of the top-ranked blacklists (141 cases), abusive host and anonymous-state (28 cases), backscatter (28 cases), and spam (82 cases).

*Apparent Domain User Type*

**Table I.2: Summary of Blacklist Variables by Apparent Domain User Type**

Blacklist Variable	Percent Yes					p-value
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type	
On Any Top-Ranked Blacklist	11.8	12.4	5.4	6.1	9.8	0.0009
Abusive host/anonymous	2.4	1.6	3.8	1.0	0.5	0.0290
Backscatter	3.5	2.9	0.3	1.0	1.2	0.0172
Spam	5.9	8.2	1.3	3.3	6.9	<.0001

All four blacklist variables show statistically significant differences between the apparent domain user types. Overall, 8.8 percent of domains appear on any top-ranked blacklist, but this percentage is higher for domains that are apparently legal persons (12.4 percent) and domains that are apparently natural

persons (11.8 percent). Parked domains (5.4 percent) and domains with no online content (6.1 percent) have the lowest rates of appearing on any top-ranked blacklist. For abusive host/anonymous blacklists, the parked domains have the highest rate (3.8 percent) of appearing on a blacklist of this type while domains with no online content (1.0 percent) and unknown domain user types (0.5 percent) have the lowest rates. For backscatter blacklists, the highest rates belong to domains that are apparently used by natural persons (3.5 percent) and domains that are apparently used by legal persons (2.9 percent), while the rates are around one percent or lower for the other three domain user types. For spam blacklists, the highest rate is for domains that are apparently used by legal persons (8.2 percent) while unknown domain user types (6.9 percent) and domains that are apparently used by natural persons (5.9 percent) also have higher rates than domains with no online content (3.3 percent) and parked domains (1.3 percent). Comparing just domains used by apparently natural persons with those that are used by apparently legal persons, they have similar overall rates of appearing on any top-ranked blacklist, but domains used by apparently legal persons have a higher spam blacklist rate while domains that are apparently used by natural persons have slightly higher rates in the two larger categories with enough positive matches to separate out (abusive host/anonymous and backscatter).

### *Apparent Registrant Type*

**Table I.3: Summary of Blacklist Variables by Apparent Registrant Type**

Blacklist Variable	Percent Yes				p-value
	Natural Person	Legal Person	Privacy/Proxy	Unknown	
<b>On Any Top-Ranked Blacklist</b>	11.6	8.0	7.9	6.7	0.0981
<b>Abusive host/anonymous</b>	2.2	2.0	1.3	0.7	0.5826
<b>Backscatter</b>	1.9	1.9	1.0	2.2	0.6971
<b>Spam</b>	7.9	3.9	3.9	4.4	0.0138

Overall, domains that are apparently registered by natural persons have a higher rate of appearing on any top-ranked blacklist than other registrant types, but the difference is not statistically significant. Domains that are apparently registered by natural persons do have a significantly higher rate of appearing on spam blacklists, however, with a rate (7.9 percent) that is about double the other registrant types (around four percent). The differences in abusive host/anonymous and backscatter blacklists are not significant, but the privacy/proxy registered domains have low rates for both.

*Potentially Commercial Activity*

**Table I.4: Summary of Blacklist Variables by Potentially Commercial Activity**

Blacklist Variable	Percent Yes		p-value
	No Potentially Commercial Activity	Potentially Commercial Activity	
<b>On Any Top-Ranked Blacklist</b>	8.3	9.5	0.3832
<b>Abusive host/anonymous</b>	1.4	2.2	0.2265
<b>Backscatter</b>	1.6	1.9	0.6433
<b>Spam</b>	4.6	5.7	0.3234

There are no significant differences in blacklist appearance between the domains with and without Potentially Commercial activity, but the rates are higher for domains with Potentially Commercial activity for all four variables shown.

## J. Whitelist Variables

Similar to the blacklists consulted, we also checked all IPs associated with the ARECORDS for the 1,600 domains against the whitelist hosted by [www.dnswl.org](http://www.dnswl.org) and two additional whitelists. If a response was returned, this signified presence on a whitelist. The response octate of the dnswl.org gave additional information on the category of the entry on the whitelist. Table J.1 is a summary of the octate results returned by the whitelists. It is possible for a domain to be identified by more than one whitelist, so the categories in the summary table are not distinct. The total number of domains associated with any of the four whitelists is provided at the bottom of the table.

**Table J.2: Domains Found on Whitelists**

Description	Frequency	Percent*
Retail/Wholesale Serices	1	0.1
Service/Network Providers	130	8.1
Email Service Providers	2	0.1
No Whitelist Octate	96	6.0
On Any Whitelist	204	12.8

It is natural to wonder if any of the domains were found on any of the whitelists and any of the blacklists, so Table J.2 answers this question:

**Table J.2: Domains Found on Whitelists and Blacklists  
Weighted Cross-classified Frequency Counts**

On Any Blacklist	On Any Whitelist				Total	Percent
	No	Yes				
No	1265.9	190.0			1455.9	91.0
Yes	130.5	13.6			144.1	9.0
<b>Total Percent</b>	1396.4	87.3	203.7	12.7	1600.0	100.0

According to Table J.2, almost one percent of the 1,600 domains were found on at least one top-ranked blacklist as well as at least one whitelist. Of the 204 domains matched to a whitelist, 6.7 percent also matched to a top-ranked blacklist compared to 9.3 percent of those that didn’t match to a whitelist. Of the 141 domains matched to a top-ranked blacklist, 9.4 percent also matched to a whitelist compared to 13.1 percent of those that didn’t match to a top-ranked blacklist.

*Apparent Domain User Type*

**Table J.3: Domains Found on Whitelists by Apparent Domain User Type**

Whitelist Variable	Percent Yes					p-value
	Natural Person	Legal Person	Domain Parked	No Online Content	Unknown Type	
<b>On Any Whitelist</b>	9.5	14.2	24.7	4.5	6.5	<.0001
<b>Service/Network Providers</b>	5.9	8.4	15.8	3.7	3.6	<.0001
<b>No Whitelist Octate</b>	3.6	8.0	11.4	1.1	2.9	<.0001

All three of these variables show highly significant differences. Parked domains have the highest rate of being on any whitelist, and they also have the highest rates in the two larger categories with enough positive matches to separate out (service/network providers and no whitelist octate). The next two highest rates for each of the three variables are for domains that are apparently used by legal persons and domains that are apparently used by natural persons. For all three variables, domains that are apparently used by legal persons have higher rates than domains that are apparently used by natural persons. The lowest rates for all three variables belong to domains with no online content and unknown domain user types.

*Apparent Registrant Type*

**Table J.4: Domains Found on Whitelists by Apparent Registrant Type**

Whitelist Variable	Percent Yes				p-value
	Natural Person	Legal Person	Privacy/Proxy	Unknown	
On Any Whitelist	14.6	13.4	12.4	3.7	0.0070
Service/Network Providers	9.2	9.0	7.1	1.5	0.0166
No Whitelist Octate	7.0	6.6	5.3	2.2	0.1701

Overall, 12.8 percent of the domains were matched to any whitelist, but this percentage is significantly lower (3.7 percent) for unknown registrant type domains, as shown by a p-value of 0.0070. Similarly, the rate of unknown registrant type domains on a service/network provider whitelist (1.5 percent) is significantly lower than for the other three registrant types (seven to nine percent), as shown by a p-value of 0.0166. The same pattern appears for the no whitelist octate, but the differences are not statistically significant. For all three variables, the privacy/proxy registration rate is slightly lower than registrants who are apparently natural or legal persons.

*Potentially Commercial Activity*

**Table J.5: Domains Found on Whitelists by Potentially Commercial Activity**

Whitelist Variable	Percent Yes		p-value
	No Potentially Commercial Activity	Potentially Commercial Activity	
On Any Whitelist	5.6	18.1	<.0001
Service/Network Providers	4.0	11.1	<.0001
No Whitelist Octate	1.9	9.3	<.0001

All three whitelist variables show very statistically significant differences between domains with and without Potentially Commercial activity. Domains with Potentially Commercial activity are much more likely to appear on any whitelist, as well as either of the two whitelist categories with enough positive matches to be separated out (service/network providers and no whitelist octate).