GNSO-APPROVED WHOIS STUDIES

COMPILATION OF EXECUTIVE SUMMARIES (IN ORDER OF COMPLETION)

I. WHOIS PRIVACY & PROXY RELAY & REVEAL FEASIBILITY SURVEY

- **Research Team:** Interisle Consulting Group, LLC
- **Completion/Publication of Final Report:** March 2012

Executive Summary:

The Domain Name Registration Data Directory Service (WHOIS) is an Internet standard mechanism for providing public access to identity and contact information about domain name registrants. ICANN-accredited domain name registrars are contractually obligated to provide accurate information about all registrants via WHOIS, either directly or through a generic top-level domain (gTLD) registry. Some registrars and third-party service providers offer registrants the opportunity to limit the public disclosure of their personal contact information by offering privacy services that publish alternative contact information. Other providers act as "proxies" by registering domain names for another user, who may access and use the domain name through a separate arrangement with the proxy service provider. A recent study by the National Opinion Research Center suggested that some or all of the public contact information for at least 18% of the domain names registered under the five largest generic top-level domains might be shielded from WHOIS by a proxy or privacy service.

Over time, the public-information requirement and the use of proxy and privacy services have become a battleground on which privacy and data protection advocates have squared off against law enforcement and intellectual property interests over access to domain name registrant data. This battle has often been highly charged and emotional, and in the absence of accurate and authoritative information about the way in which registrant contact information access is affected by the use of privacy and proxy services the debate has been driven more by anecdote than by data. Recognizing this as an impediment to resolving the issue, ICANN's Generic Names Supporting Organization (GNSO) Council has commissioned several studies to collect reliable data on WHOIS deployment and use, including a study of the effect of proxy and privacy services on access to domain name registrant data.

ICANN asked Interisle Consulting Group to conduct a survey to determine whether or not the study of proxy and privacy services contemplated by the GNSO Council would in fact be feasible, and if so how such a study should be designed in order to secure the greatest possible participation from potential information sources and thereby deliver the most useful data to the WHOIS debate.

Interisle gathered information from three broadly defined constituencies: initiators of

relay/reveal requests; proxy/privacy service providers; and registrars involved in processing relay/reveal requests and responses. An initial multi-lingual online survey collected 168 responses from 73 request initiators, 25 proxy/privacy service providers, and 36 registrars. Sixteen follow-up interviews were conducted with a representative sample of stakeholders, including 5 request initiators, 3 proxy/privacy service providers, and 4 registrars. The remaining interviews were conducted with individuals who had insights into the use of WHOIS proxy and privacy services but were not directly involved in making or processing relay or reveal requests.

This report presents the results of our analysis of the survey responses and interview data, which demonstrates that:

a) a full study of WHOIS privacy and proxy could, if defined in such a way as to resolve identified barriers, provide some—but not all—of the data anticipated by the GNSO Council;

b) such a study (specifically by ICANN) would be well received by people on all sides of the WHOIS information access debate;

c) attention to issues including confidentiality and convenience in the design of the study would improve the quantity and quality of the data that it would deliver, but would not entirely overcome the asymmetric reluctance of potential participants from different constituencies to contribute; and

d) the results of a full study thus encumbered might not satisfy the expectations of the GNSO Council or the ICANN community with respect to statistical validity or independent verifiability.

The data show that roughly 40% of the principal constituencies (relay and reveal request initiators, proxy/privacy service providers, and registrars) would be interested in participating in a full study. These participants would be able to provide only summary or incomplete data concerning the incidence, processing, and disposition of relay/reveal requests. For the most part they would not, however, be able to provide data concerning specific individually identifiable instances of relay or reveal requests. Several potential full study participants said that additional non-aggregate data could be obtained from the public records of legal and arbitration proceedings, and that they would actively assist in their discovery.

Specifically, in their responses to the online survey 47% of request initiators, 40% of proxy/privacy service providers, and 39% of registrars said that they would be interested in participating in a full study; 77% of request initiators, 72% of proxy/privacy service providers, and 75% of registrars said that strong privacy guarantees for data contributed to a full study would be important; and 77% of request initiators, 40% of proxy/privacy service providers, and 47% of registrars said that the results of a full study would be valuable either to their organization or to the Internet community as a whole (or both). A full study should be designed to overcome the two most important barriers to participation that were cited by survey

respondents as either critical or very significant: the time and effort required to participate (46%) and the confidentiality of client information (44%). 75% of survey respondents said that strong confidentiality guarantees would be significant, very significant, or critical to their ability to provide data to a full study. Follow-up interviews revealed, however, that most potential participants would be willing to provide only anonymized and aggregated data to a full study regardless of how strong the confidentiality guarantees might be.

These findings suggest that a full study would have to be designed and carried out in a way that did not require participants to disclose specific details of domain names or identify registrants using privacy/proxy services. A full study that depended on the ability to track and correlate individually identifiable requests and responses would therefore be impractical. A study designed to work with anonymized or aggregated request data would be acceptable to at least some potential participants if strong assurances were provided that their data would be protected and their participation would not require substantial time and effort. Anonymized or aggregated data, however, might not support the type of detailed analysis expected by the GNSO Council. Careful consideration of this tradeoff should precede any decision to invest in a full study.

II. WHOIS REGISTRANT IDENTIFICATION STUDY

- **Research Team:** NORC at the University of Chicago
- **Completion/Publication of Final Report:** May 2013

Executive Summary of Report:

The Internet Corporation for Assigned Names and Numbers (ICANN) has contracted NORC at the University of Chicago (NORC) to conduct the **WHOIS Registrant Identification Study.** This project is an exploratory examination of WHOIS data for a representative sample of gTLD domain names, using WHOIS Registrant Name and Registrant Organization values to classify the types of entities that register domains, including natural persons, legal persons, and privacy and proxy service providers.

NORC analyzed available web/FTP content associated with each sampled domain name to classify the types of entities that appear to be using those domains and the various types of activities associated with them. Additionally, we analyzed inter-relationships between these categories, seeking to provide a foundation for answering the following questions posed by the Government Advisory Committee (GAC):

- What is the percentage of registrants that are natural versus legal persons?
- What is percentage of domain name uses that are commercial versus non-commercial?

- What is the relative percentage of Privacy/Proxy use among legal persons?
- What is the relative percentage of Privacy/Proxy use among domains with commercial use?

This report is a summary of the activities undertaken to conduct and complete this project. Of primary interest are the interpretations of the statistical analysis. In particular, we focus on analyses related to the following three questions.

- 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?

- What differences exist between how domains users that are natural persons identify themselves, versus how domain users that are legal persons identify themselves?

- 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?

In many cases, classification of the characteristics and activities were difficult to discern and often had to be coded as "unknown." Unknowns that remained even after extensive investigation is an important study finding because they illustrate the degree of the difficulty experienced by those trying to use WHOIS data and Internet content to identify domain registrants and users.

Nevertheless, NORC has produced a coded set of data that is useful for its intended purpose—an exploratory study of registrant and domain user characteristics and the types of domain use activities. With respect to answering the issues posed by the GAC:

Percentage of registrants that are natural versus legal persons: Based on our analysis of the WHOIS records retrieved from a random sample of 1,600 domains from the top five gTLDs,

- 39 percent (± 2.4 percent) appear to be registered by legal persons
- 33 percent (± 2.3 percent) appear to be registered by natural persons
- 20 percent (± 2.0 percent) were registered using a privacy or proxy service.

We were unable to classify the remaining 8 percent (± 1.4 percent) using data available from WHOIS.

Percentage of domain name uses that are commercial versus non-commercial: Per the GNSO Council's request, we attempted to categorize all observed monetary activities that in some countries might be legally considered "commercial activities," documenting a broad range of potentially commercial activities to enable multiple post-study interpretations that apply varied legal definitions. For example, because pay-per-click ads were found so frequently that

they dominated this variable. We completed our analysis with and without pay-per-click ads to enable both interpretations of potentially commercial activity. Based on our analysis of web/FTP content retrieved from a random sample of 1,600 domains from the top five gTLDs:

- When pay-per-click ads are included in the monetary activities that make up potentially commercial activity, 57 percent (± 2.4 percent) of all sampled domains were perceived to have potentially commercial activity.
- When pay-per-click ads are not included in the monetary activities that make up potentially commercial activity, approximately 45 percent (± 2.4 percent) of all sampled domains were perceived to have potentially commercial activity.

WHOIS records and the web/FTP content retrieved from a random sample of 1,600 domains from the top five gTLDs:

- 15.1 percent (± 2.9 percent) of domains used by legal persons were registered using a privacy or proxy service.

Relative percentage of Privacy/Proxy use among domains with commercial use: Based on our analysis of the WHOIS records and the web/FTP content retrieved from a random sample of 1,600 domains from the top five gTLDs:

- 22.9 percent (± 2.7 percent) of domains with potentially commercial activity were registered using a privacy or proxy service.

Additional interesting findings related to the three focus questions for this study are:

(1) Differences in how domains are used based on registrant type

Domain names registered by legal persons were

- More likely to be used by legal persons—52.2 ± 3.9 percent, as compared to the entire sample's 36.6 percent.
- Equally as likely to be used for some kind of potentially commercial activity —59.9 ± 3.9 percent, as compared to the entire sample's 56.6 percent.
- Equally as likely to have WHOIS addresses in the U.S.—59.4 ± 3.9 percent, as compared to the entire sample's 56.9 percent.
- More likely to be both registered and used by the same legal person—27.8 ± 3.5 percent, as compared to the entire sample's 16.8 percent.
- More likely to be used by a for-profit entity—39.9 ± 3.8 percent, as compared to the entire sample's 25.6 percent.

Domain names registered by natural persons were

- More likely to be used by natural persons—10.4 ± 2.6 percent, as compared to the

entire sample's 5.4 percent.

- Equally as likely to be used for some kind of potentially commercial activity as the overall sample—55.4 ± 4.3 percent, as compared to the entire sample's 56.6 percent.
- Less likely to have WHOIS addresses in the U.S.—46.0 ± 4.3 percent, as compared to the entire sample's 56.9 percent.
- More likely to have undetermined domain user/registrant relationships—72.5 ± 3.9 percent, as compared to the entire sample's 54.8 percent.
- More likely to be used by a non-business entity—11.8 ± 2.8 percent, as compared to the entire sample's 6.4 percent.

Domain names registered using a Privacy/Proxy service were

- More likely to be parked—30.7 ± 5.0 percent, as compared to the entire sample's 20.5 percent.
- More likely to be used for some kind of potentially commercial activity—64.6 ± 5.2 percent, as compared to the entire sample's 56.6 percent.
- More likely to be registered with a WHOIS address in the U.S.—74.3 ± 4.8 percent, as compared to the entire sample's 56.9 percent.
- More likely to have a user/registrant relationship of a customer of a privacy/proxy service—92.8 ± 2.8 percent, as compared to the entire sample's 20.4 percent.
- More likely to be used by an entity with an unclear business structure—71.4 ± 4.9 percent, as compared to the entire sample's 65.7 percent.

(2) Differences in how kinds of domains users identify themselves based on domain registrant type

Domain names used by legal persons were

- More likely to be registered by legal persons—55.1 ± 4.0 percent, as compared to the entire sample's 38.6 percent.
- More likely to be used for some kind of potentially commercial activity—79.8 ± 3.2 percent, as compared to the entire sample's 56.6 percent.
- Equally likely to have WHOIS addresses in the U.S.—54.9 ± 4.0 percent, as compared to the entire sample's 56.9 percent.
- More likely to also be registered by that legal person—35.5 ± 3.9 percent, as compared to the entire sample's 16.8 percent.
- More likely to be used by for-profit businesses—60.7 ± 3.7 percent, as compared to the entire sample's 25.6 percent.

Domain names used by natural persons were

- More likely to be registered by natural persons—60.4 ± 10.2 percent, as compared to the entire sample's 32.8 percent.
- Less likely to have potentially commercial activity—36.8 ± 10.1 percent, as compared to the entire sample's 56.6 percent.
- Equally likely to have WHOIS addresses in the U.S.—49.9 ± 10.4 percent, as compared to the entire sample's 56.9 percent.

- More likely to be registered by that natural person—69.7 ± 9.6 percent, as compared to the entire sample's 16.8 percent.
- Never used by a business; this is by design—when coding apparent business structure, if the user was a natural person, then the business structure was coded as not a business.

(3) Differences in domains with potentially commercial activity (pay-per-clicks ads included)

Domains with detected potentially commercial activity were

- More likely to have legal person users—51.5 ± 3.3 percent, as compared to the entire sample's 36.6 percent.
- Less likely to have user/registrant relationships that cannot be determined—44.8 ± 3.2 percent, as compared to the entire sample's 54.8 percent.
- Less likely to have an unclear business structure—55.2 ± 3.2 percent, as compared to the entire sample's 65.7 percent.

For both Apparent Registrant Type and Registrant WHOIS Address County/Region of the World differences between the relative percentage among domains with potentially commercial activity and the entire sample's percentage are small. Thus, knowing that a domain has potentially commercial activity does not provide any additional insight as to the registrant type or the WHOIS address of the registrant.

III. WHOIS PRIVACY & PROXY SERVICES ABUSE STUDY

- **Research Team:** Dr. Richard Clayton (University of Cambridge) and National Physical Laboratory (United Kingdom)

- **Completion/Publication of Final Report:** March 2014

Executive Summary:

This study is one of series that seek to establish reliable evidence for various beliefs that are held about the operation of the Internet domain name 'Whois' system, which provides the public with information about the registrants of domain names.

1.1 Aims of the study

This particular study was originally proposed by ICANN in 2010, one of several that were to examine the impact of domain registrants using privacy services (where the name of a domain registrant is published, but contact details are kept private) and proxy services (where even the domain licensee's name is not made available on the public database). The exact definitions of privacy and proxy services that we used are set out in Section 5.

The initial intention was to test the hypothesis:

"A significant percentage of the domain names used to conduct illegal or harmful

Internet activities are registered via privacy or proxy services to obscure the perpetrator's identity".

In April 2012 a contract to perform the study was awarded to the National Physical Laboratory (NPL), one of the United Kingdom's leading science and research facilities. The technical lead on the project was Dr. Richard Clayton of the University of Cambridge.

We broadened the study because it was implicit that a "significant percentage" would be one that is measured – with high statistical confidence – to be substantially greater than the equivalent percentage for entirely lawful and harmless Internet activities. Hence we also sought to examine the related hypothesis:

"The percentage of domain names used to conduct illegal or harmful Internet activities that are registered via privacy or proxy services is significantly greater than the percentage of domain names used for lawful Internet activities that employ privacy or proxy services."

We wanted to consider what other methods might be chosen by those involved in criminal activity to obscure their identities, because in the event of changes to privacy and proxy services, it is likely that they will turn to these alternatives. Accordingly, we determined experimentally whether a significant percentage of the domain names we examined have been registered with incorrect Whois contact information – specifically whether or not we could reach the domain registrant using a phone number from the Whois information.

1.2 The types of activity we considered

The approach we took was to consider different categories of harmful activity and generate robust statistics for each category. We split the work into a number of work packages:

WP1	phishing
WP2	money laundering
WP3	unlicensed pharmacies
WP4	typosquatting
WP5	child sexual abuse image websites
WP6	lawful and harmless websites
WP7	domains appearing in email spam (SURBL domains)
WP8	domains associated with malware (StopBadware domains)
WP9	domains subject to the UDRP process

For each work package we have obtained a list of relevant URLs or hostnames for the particular type of activity and then categorised the domain names involved. The scale of these activities differs considerably, but in every case we have collected data over a sufficiently long period to ensure that results are representative of each category and our results will have appropriate statistical significance.

Our study mainly addresses the use of domain names that have been implicated in illegal or harmful activities. The study also examines (particularly in WP6) some samples of lawful and harmless domain names to establish a point of reference, but it is important to understand that the selection we have made is not necessarily representative of the overall usage of domain names for lawful and harmless purposes.

For each set of domain names within the various work packages we collected and examined the Whois data for the domain names that were registered within the top five generic top level domains (gTLDs), i.e. .biz, .com, .info, .net and .org. The domain names in other top level domains (TLDs) were counted, but no further analysis was performed.

For the domain names where we had collected the Whois data we determined the proportion of these registrations that were using privacy or proxy services. If the domain was not using a privacy or proxy service we looked to see whether the Whois record contained a phone number for the domain registrant and if it did have a phone number we checked whether it passes some simple rules, so that we believe that it can be used to telephone the registrant.

We took a random sample of the domains which have these 'apparently valid' contact phone numbers and we attempted to ring up the domain registrants within this sample to have a short conversation with them, in their native language, to ascertain whether or not they acknowledged registering the domain.

1.3 WP1 (phishing) – the study in a nutshell

The overall results that we obtained can be seen with real clarity in the results of work package WP1 – where we examined domains that had occurred in URLs for phishing pages.

We split this work package into three, since we could analyse the URLs and determine whether the domain:

- was registered by a third party (e.g. companies set up to provide hosting services or URL shortening) and their services were used for criminal purposes;
- was registered by a legitimate business (or individual) whose website had been compromised and phishing web pages added without their knowledge or permission;
- appeared to have been maliciously registered for the purpose of phishing.

We found very striking differences between these categories when we considered the usage of privacy and proxy services and also whether we were successful in making contact with the registrant by phone or, conversely, had no hope of doing so:

	Using privacy or proxy services		Missing / invalid phone number		Cannot contact by phone	Phone contact succeeded
Third party domains	13.7%	+	35.9%	=	49.6%	32.3%
Compromised website domains	24.7%	+	37.0%	=	61.7%	23.7%
Maliciously registered domains	31.2%	+	61.3%	=	92.5%	1.8%

The people who maliciously registered domains for phishing chose privacy and proxy services somewhat more than people who registered domains for legitimate purposes.

However, when a privacy or proxy service was not chosen for a malicious registration a workable contact phone number was seldom given – and even if the number was apparently valid, we almost never managed to make contact with the registrant for our survey.

Conversely, even entirely legitimate 'third party' businesses that provide services to the law- abiding public – and occasionally for malicious purposes – use privacy and proxy services to a certain extent, and for almost half of the domains these businesses use there is no possibility of using the phone to reach the domain registrant. Of course there are many other ways of making contact with such businesses, and they would doubtless want people to use the information about contact pathways on their websites, rather than consulting Whois.

The compromised website category falls between these two extremes – these domain registrants use privacy and proxy services about a quarter of the time. Nearly two thirds of these registrants are impossible to contact by phone, and so we reached only a quarter of them for our survey.

1.4 Privacy or proxy service usage

The following table summarises the evidence we have of linkage between malicious registration of domains and the usage of privacy or proxy services. The main body of the report contains the detailed results and explains their statistical significance.

	Work package	Maliciously registered?	Usage of privacy or proxy services
WP6.4	Legal pharmacies	no	low
WP6.3	Law firms	no	low
WP1t	Phishing: third parties	no	low
WP6.6	Typosquatted domains	no	average
WP8	StopBadware domains	some	average
WP6.2	Executive search consultants	no	average
WP1c	Phishing: compromised sites	no	average
WP6.1	Banks	no	high
WP5	Child sexual abuse image websites	yes	high
WP1m	Phishing: malicious registration	yes	very high
WP9	Domains subject to UDRP	some	very high
WP7	SURBL domains	mostly	very high
WP6.5	Adult websites	no	very high
WP2	Advanced Fee Fraud	yes	extremely high
WP4	Typosquatting	yes	extremely high
WP3	Unlicensed pharmacies	yes	extremely high

The table clearly shows a correlation, in that maliciously registered domains have a higher usage of privacy and proxy services – but this correlation is not universal in that banks are above average users of these services, as are adult websites.

1.5 Reaching registrants by phone

The most useful way looking at the data we collected about the results of our phone survey is *not* to consider whether our survey calls were successful – there are several reasons for this not being a compelling analysis which we set out in the body of report, but one important issue was that we often reached voicemail systems, or cellphones were not answering, and so we could not determine whether or not the phone number was valid.

Instead, we considered an opposing analysis – whether from the Whois information it would be impossible to reach the party using the domain name directly by phone. The impossibility would result from the use of a privacy or proxy service, from a failure to provide a phone number that can be called, or from the provision of a phone number that reaches someone other than the registrant or licensee actually using the domain.

The results of this analysis are shown in the following table. In two thirds of cases where domains were maliciously registered it is inherently impossible to use the phone to reach the registrant of the domain. There is also a wide range of likelihoods for lawful and harmless activities – but the pattern is far clearer than just considering the usage of privacy and proxy services in isolation: one way or another, those registering domain names to be used for criminal activity seldom provide valid contact information.

Work package	**Privacy or proxy usage**	**Not possible to phone the registrant**	**Maliciously registered?**
Legal pharmacies	8.8%	24.2%	no
Law firms	13.4%	33.6%	no
Executive search consultants	22.4%	36.7%	no
Banks	28.2%	44.6%	no
Typosquatted domains	19.2%	47.1%	no
Phishing: third parties	13.7%	49.6%	no
StopBadware domains	20.4%	51.4%	some
Adult websites	44.2%	55.1%	no
SURBL domains	44.1%	58.5%	mostly
Phishing: compromised sites	24.7%	61.7%	no
Typosquatting	48.2%	67.7%	yes
Advanced Fee Fraud	46.5%	88.9%	yes
Unlicensed pharmacies	54.8%	91.8%	yes
Phishing: malicious registration	31.2%	92.5%	yes

1.6 What we believe to be true

Our study shows that it IS TRUE that:

"A significant percentage of the domain names used to conduct illegal or harmful Internet activities are registered via privacy or proxy services to obscure the perpetrator's identity".

Our study shows that it is PARTLY TRUE that:

"The percentage of domain names used to conduct illegal or harmful Internet activities that are registered via privacy or proxy services is significantly greater than the percentage of domain names used for lawful Internet activities that employ privacy or proxy services."

More helpfully, we can say:

"When domain names are registered with the intent of conducting illegal or harmful Internet activities then a range of different methods are used to avoid providing viable contact information – with a consistent outcome no matter which method is used.

However, although many more domains registered for entirely lawful Internet activities have viable telephone contact information recorded within the Whois system, a great percentage of them do not."

IV. WHOIS MISUSE STUDY

- **Research Team:** Cylab at Carnegie Mellon University
- **Completion/Publication of Final Report:** March 2014

Executive Summary:

Does public access to WHOIS-published data lead to a measurable degree of misuse[1]? This study, sponsored by the Internet Corporation for Assigned Names and Numbers (ICANN) and initiated by ICANN's Generic Names Supporting Organization (ICANN GNSO, 2010), attempts to answer this question, with a focus on the five most populous generic Top Level Domains (gTLDs). To do so, we first surveyed experts, law enforcement agents, Registrants, Registrars, and Registries, and collected their input on the prevalence of WHOIS misuse, thereby obtaining a descriptive data set. We then complemented this descriptive portion of the study with a set of experimental measurements of WHOIS misuse, which we obtained by registering 400 domains in the top five gTLDs across 16 Registrars, associating unique, synthetic WHOIS contact information with these domains, and monitoring incidents of misuse for a period of 6 months.

The main finding of the descriptive study is that there is a statistically significant occurrence of WHOIS misuse affecting Registrants' email addresses, postal addresses, and phone numbers, published in WHOIS when registering domains in these gTLDs. Overall, we find that 43.9% of Registrants experience one or more of these types of WHOIS misuse. Other types of WHOIS misuse are reported, but at a smaller, non-significant rate. Among those, a handful of reported cases appear to be highly elaborate attempts to achieve high attack impact.

[1] In this study, WHOIS misuse refers to harmful acts that exploit contact information obtained from WHOIS. Harmful acts may include generation of spam, abuse of personal data, intellectual property theft, loss of reputation or identity theft, loss of data, phishing and other cybercrime related exploits, harassment, stalking, or other activity with negative personal or economic consequences.

As a caveat, most findings of the descriptive study are affected by low response rates from the parties we surveyed. Most importantly, we are unable to draw meaningful conclusions about the geographical aspects of WHOIS misuse. Indeed, the great majority of survey responses originated from the US, even though we used a much more geographically diverse Registrant population sample, and tried to survey Registrants in their native language.

The experimental study corroborates the findings of the descriptive study. In particular, it offers quantitative insights regarding both the extent of WHOIS misuse, and the parameters affecting WHOIS misuse. A limitation of the experimental study is that the impact of geographical location on postal address misuse could not be measured, due to the prohibitively expensive cost of setting up postal boxes in countries without having an actual residence there.

Among the measurable factors analyzed by this experiment, we identify the gTLD as the sole statistically-significant characteristic that affects the occurrence of the associated misuse of phone numbers published in WHOIS. For example, the rates of WHOIS phone number misuse are negatively correlated with .ORG domains (less misuse), but positively with .BIZ and .INFO (more misuse).

Similarly, we find that the domain price is negatively correlated with the possibility of misuse of email addresses published in WHOIS (i.e., experimental domains purchased at greater cost had less email address misuse). We also discover that .COM, .NET, and .ORG domains are associated with less email address misuse, while .BIZ domains are associated with more misuse.

We also studied whether the composition of domain names themselves impacts the probability of WHOIS misuse. We find that experimental domain names representing natural person names appear to foster less email misuse, while for other experimental domain name categories (e.g., professional, randomly-generated, etc.), WHOIS misuse probability seems independent of the domain name composition.

We find that WHOIS anti-harvesting techniques, applied both at the Registry and Registrar level, is statistically significant in reducing the possibility of WHOIS email address misuse. Overall, we find that experimental WHOIS data registered with Registries/Registrars with no observable anti-harvesting countermeasures was twice more likely to result in unwanted emails compared to cases where a countermeasure was deployed. We do not offer, however, a comparative analysis of the effectiveness of specific anti-harvesting techniques against WHOIS misuse, as any differences we could observe were not statistically significant.

Finally, we do not find other statistically-significant correlations between specific Registrars used to register experimental domains and measured rates of WHOIS misuse.