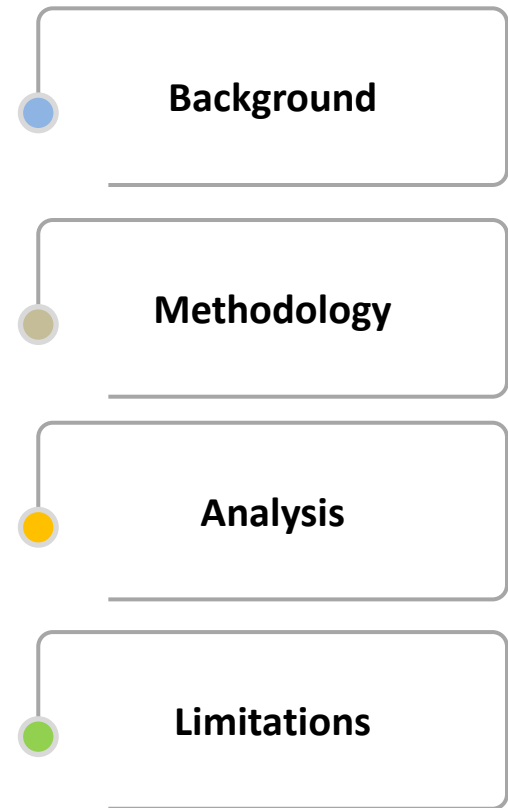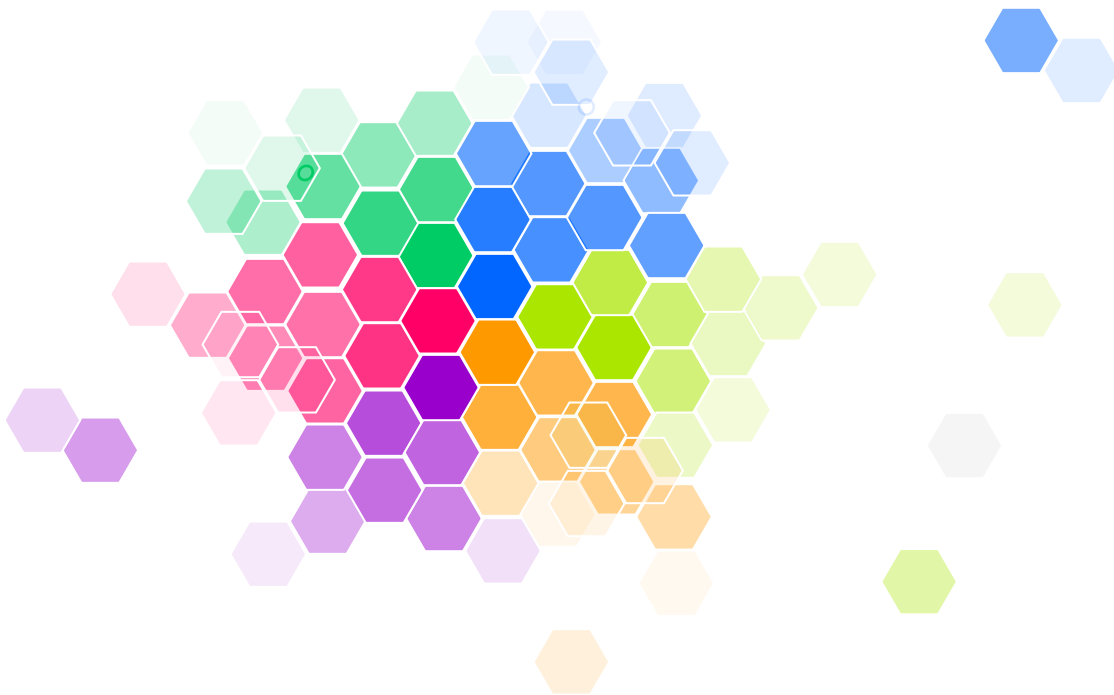# Whois Object Inventory Statistical Analysis

Ning Kong, Linlin Zhou, Guangqing Deng (CNNIC)
Steve Sheng, Francisco Arias (ICANN)
Ray Bellis (Nominet), Frederico Neves (.BR)

WEIRDS-IETF 84-Vancouver

August 1st 2012

# Agenda

Background

Methodology

Analysis

Limitations

# Background

- The data framework for the domain name space is one of the biggest elephants in the room and that making sure we converge to a common understanding is critical to success of the working group. [mailing list- thread with subject "Divide and Conquer"]

- We collected, analyzed and built an inventory of Whois output elements from 124 TLDs.

- The inventory may help to move forward the working group's work.

# Methodology

- A script to collect port 43 responses from 294 ccTLDs
  - Using nic.ccTLD as the query string which is usually registered in a registry.
  - Responses of 106 ccTLDs were received.
- Collected port 43 responses of 16 gTLDs and 2 gTLDs according to the contracts.
- Manually collected web Whois data from 124 registries that have port 43 responses.
- Manually re-verify 8% of the data using domains other than nic.TLD.

# Methodology

- In accordance with gTLD Applicant Guidebook and draft-kong-dnrd-ap-response-json-00, we separate the data into different categories which include common data part and extension data part.

- Data elements with the same meaning, but using different labels, are grouped together. The numbers of registries that support the data elements is calculated in the total count column.

| Data Element | Number of Labels | Total Count | Label | Super Set | Port 43 | Web |
|---|---|---|---|---|---|---|
| Domain Name | 6 | 118 | Domain Name | 73 | 66 | 52 |
| | | | Domain | 32 | 27 | 17 |
| | | | Query | 19 | 19 | 12 |
| | | | Nombre de Dominio | 2 | 1 | 2 |
| | | | Name | 1 | 1 | 0 |
| | | | Complete Domain Name | 2 | 1 | 1 |

# Analysis

- Whois data is collected from 124 registries, including 106 ccTLDs and 18 gTLDs.

- 124 registries support domain query. Among 124 registries, 8 ccTLDs and 15 gTLDs support contact query. 10 ccTLDs and 18 gTLDs support name server query. 4 ccTLDs and 18 gTLDs support registrar query.

- Domain Whois data contains 68 data elements that use a total of 550 labels.

- There are total 392 extensions for Domain Whois data. (not classified)

The raw data can be accessed with the following link:
https://docs.google.com/open?id=0B96TtoK8a--MTTRuVUt3UHZMdEk

# Analysis - Domain Whois data

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Domain Name | Domain Name | 118 | 6 |
| Name Server | Name Server | 114 | 63 |
| Creation Date | Created | 106 | 24 |
| Domain Status | Status | 95 | 8 |
| Sponsoring Registrar | Registrar | 84 | 6 |
| Expiration Date | Expiration Date | 81 | 21 |
| Updated Date | Modified | 70 | 20 |
| Registrant Name | Registrant Name | 65 | 7 |
| Admin Street | Address | 64 | 19 |
| Admin Name | Admin Name | 60 | 9 |
| Registrant Email | Registrant Email | 59 | 7 |
| Tech Email | Tech Email | 59 | 9 |
| Tech ID | Tech ID | 55 | 16 |
| Admin Email | Admin Email | 54 | 12 |
| Admin ID | Admin ID | 52 | 16 |
| Registrant ID | Registrant ID | 50 | 12 |
| Registrant Phone | Registrant Phone | 48 | 6 |
| Tech Name | Tech Name | 47 | 6 |
| Billing Name | Name | 47 | 5 |
| Tech Phone | Tech Phone | 45 | 10 |
| Tech Fax | Tech Fax | 45 | 9 |
| Registrant Fax | Registrant Fax | 44 | 6 |
| Admin Fax | Admin Fax | 44 | 8 |

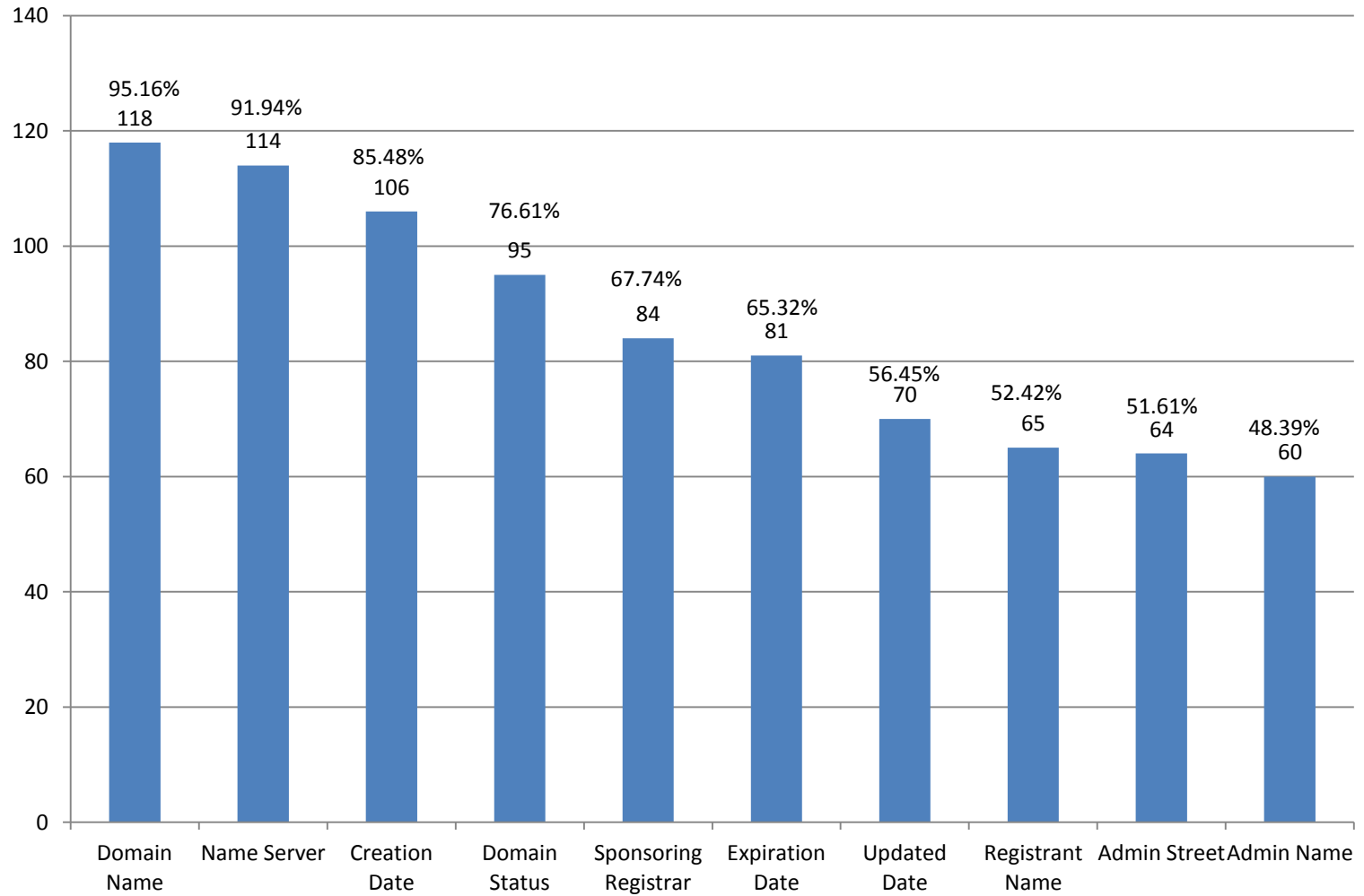| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Admin Phone | Admin Phone | 43 | 9 |
| Billing Fax | Fax | 43 | 6 |
| Registrant Organization | Registrant Organization | 42 | 4 |
| Registrant Country Code | Country | 42 | 6 |
| Admin Organization | Admin Organization | 42 | 9 |
| Admin Country Code | Country | 42 | 7 |
| Billing Email | Email Address | 42 | 7 |
| Tech Country Code | Country | 40 | 4 |
| Tech Organization | Tech Organization | 39 | 7 |
| Registrant City | Registrant City | 38 | 4 |
| Billing Country Code | Country | 38 | 4 |
| Registrant Postal Code | Registrant Postal Code | 37 | 5 |
| Tech City | Tech City | 36 | 4 |
| Tech Postal Code | Tech Postal Code | 36 | 7 |
| Admin City | Admin City | 35 | 5 |
| Admin Postal Code | Admin Postal Code | 35 | 7 |
| Domain ID | Domain ID | 34 | 5 |
| Billing Phone | Phone Number | 34 | 6 |
| Registrant State/Province | Registrant State/Province | 32 | 4 |
| Registrant Street | Registrant Street1 | 31 | 16 |
| Tech State/Province | Tech State/Province | 30 | 4 |
| Admin State/Province | Admin State/Province | 28 | 5 |

# Analysis - Domain Whois data

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Billing ID | Billing ID | 28 | 9 |
| Billing Organization | Billing Organization | 28 | 5 |
| Billing City | Billing City | 28 | 4 |
| Tech Street | Tech Street1 | 27 | 16 |
| Billing Postal Code | Billing Postal Code | 27 | 4 |
| Billing State/Province | Billing State/Province | 21 | 4 |
| Registrant Country | Registrant Country | 19 | 4 |
| Billing Street | Billing Street1 | 19 | 13 |
| Registrant Phone Ext. | Registrant Phone Ext. | 18 | 2 |
| Tech Country | Tech Country | 18 | 5 |
| Tech Fax Ext | Tech Fax Ext | 18 | 3 |
| Registrant Fax Ext | Registrant Fax Ext | 17 | 2 |
| Admin Country | Admin Country | 17 | 5 |
| Admin Phone Ext. | Admin Phone Ext. | 17 | 3 |
| Admin Fax Ext. | Admin Fax Ext. | 17 | 3 |
| Created by Registrar | Created by | 14 | 3 |
| DNSSEC | DNSSEC | 14 | 4 |
| Tech Phone Ext. | Tech Phone Ext. | 13 | 3 |
| Billing Country | Billing Contact Country/Billing Country | 13 | 5 |
| Updated by Registrar | Last Updated by Registrar | 11 | 3 |
| Billing Phone Ext. | Billing Phone Ext. | 10 | 2 |
| Billing FAX Ext | Billing FAX Ext | 10 | 2 |
| Last Transferred Date | Last Transferred Date | 4 | 3 |

# Analysis - Top 10 Data Elements



| Data Element | Percentage | Count |
|---|---|---|
| Domain Name | 95.16% | 118 |
| Name Server | 91.94% | 114 |
| Creation Date | 85.48% | 106 |
| Domain Status | 76.61% | 95 |
| Sponsoring Registrar | 67.74% | 84 |
| Expiration Date | 65.32% | 81 |
| Updated Date | 56.45% | 70 |
| Registrant Name | 52.42% | 65 |
| Admin Street | 51.61% | 64 |
| Admin Name | 48.39% | 60 |

# Analysis—Cumulative Distribution of Data Elements
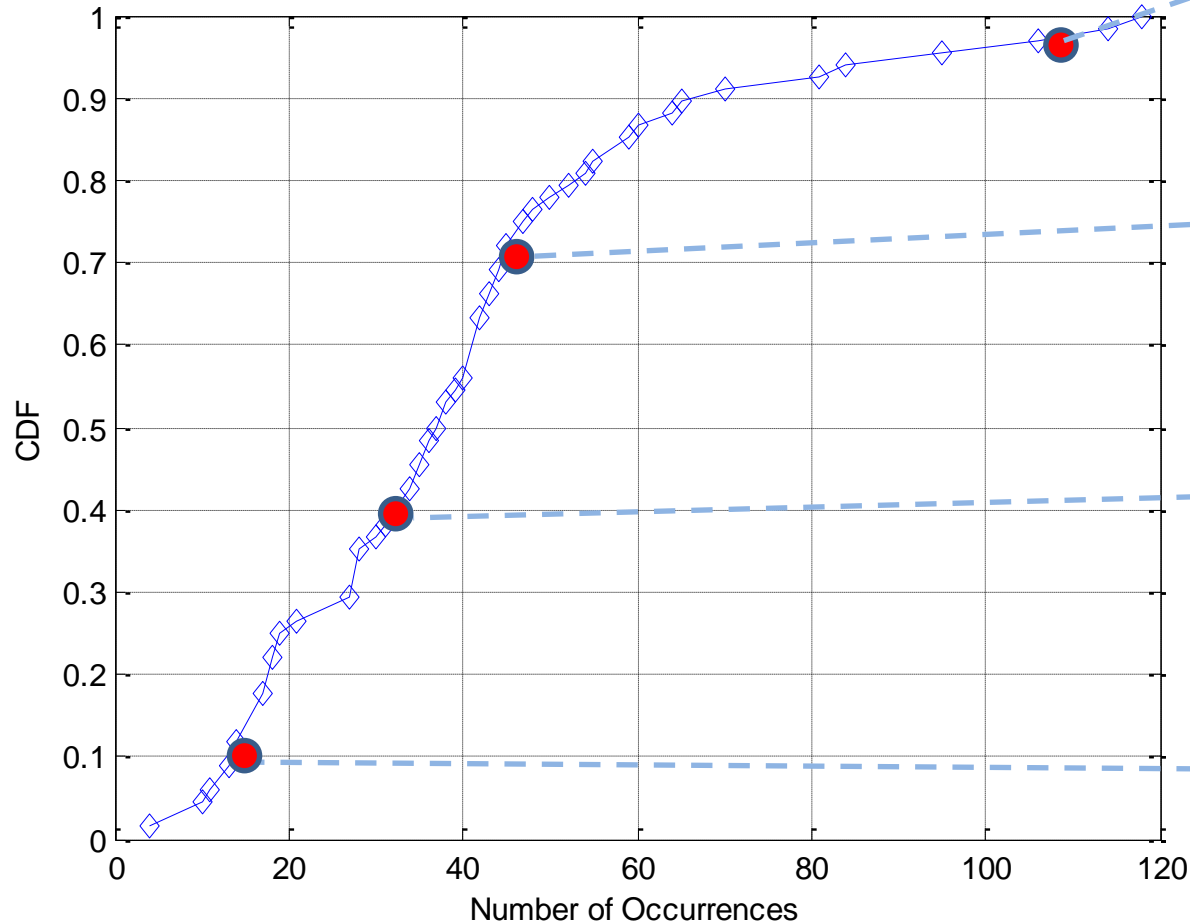


About 5% of data elements are supported by over 111 (90%) registries.

About 30% of data elements are supported by over 44(35%) registries.

About 60% of data elements are supported by over 32 (26%) registries.

About 90% of data elements are supported by over 14 (11%) registries.

# Analysis - Domain

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Domain Name | Domain Name | 118 | 6 |
| Creation Date | Created | 106 | 24 |
| Domain Status | Status | 95 | 8 |
| Expiration Date | Expiration Date | 81 | 21 |
| Updated Date | Modified | 70 | 20 |
| Domain ID | Domain ID | 34 | 5 |
| DNSSEC | DNSSEC | 14 | 4 |
| Last Transferred Date | Last Transferred Date | 4 | 3 |
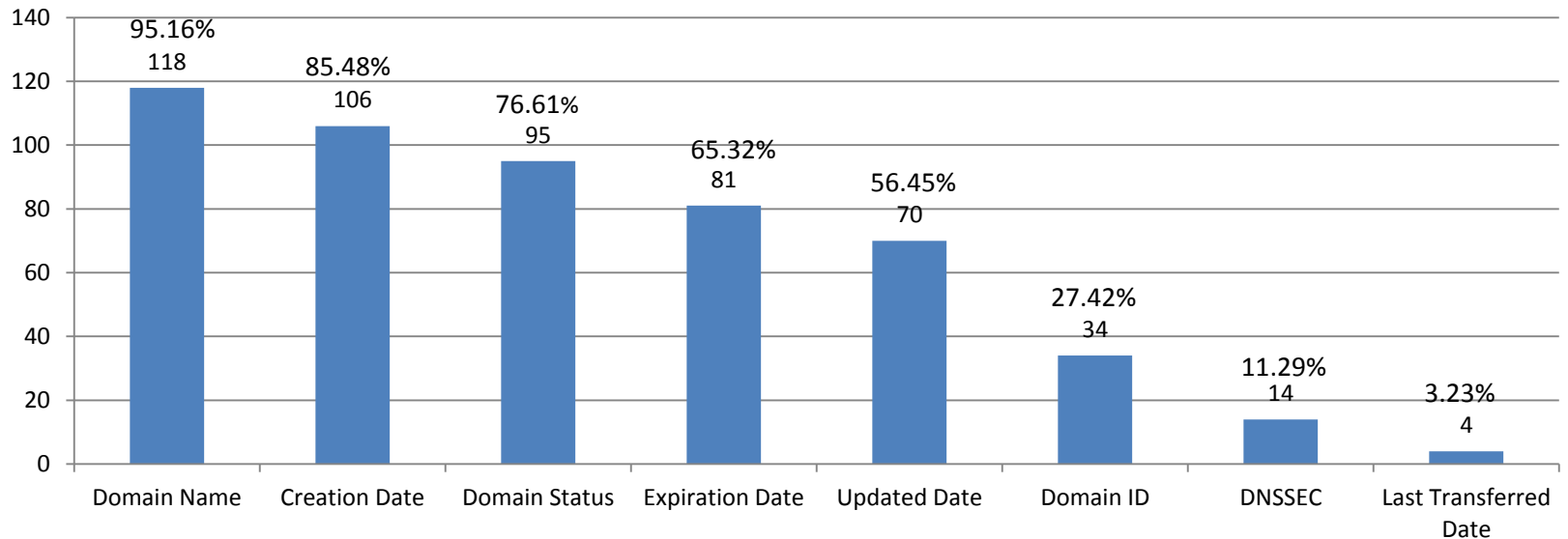
# Analysis - Registrant Contact

| Data Element | Most Popular Label | No. of TLDs | No. of  Labels |
|---|---|---|---|
| Registrant Name | Name | 65 | 7 |
| Registrant Email | Registrant Email | 59 | 7 |
| Registrant ID | Registrant ID | 50 | 12 |
| Registrant Phone | Registrant Phone | 48 | 6 |
| Registrant Fax | Registrant Fax | 44 | 6 |
| Registrant Organization | Registrant Organization | 42 | 4 |
| Registrant Country Code | Country | 42 | 6 |
| Registrant City | Registrant City | 38 | 4 |
| Registrant Postal Code | Registrant Postal Code | 37 | 5 |
| Registrant State/Province | Registrant State/Province | 32 | 4 |
| Registrant Street | Registrant Street1 | 31 | 16 |
| Registrant Country | Registrant Country | 19 | 4 |
| Registrant Phone Ext. | Registrant Phone Ext. | 18 | 2 |
| Registrant Fax Ext | Registrant Fax Ext | 17 | 2 |
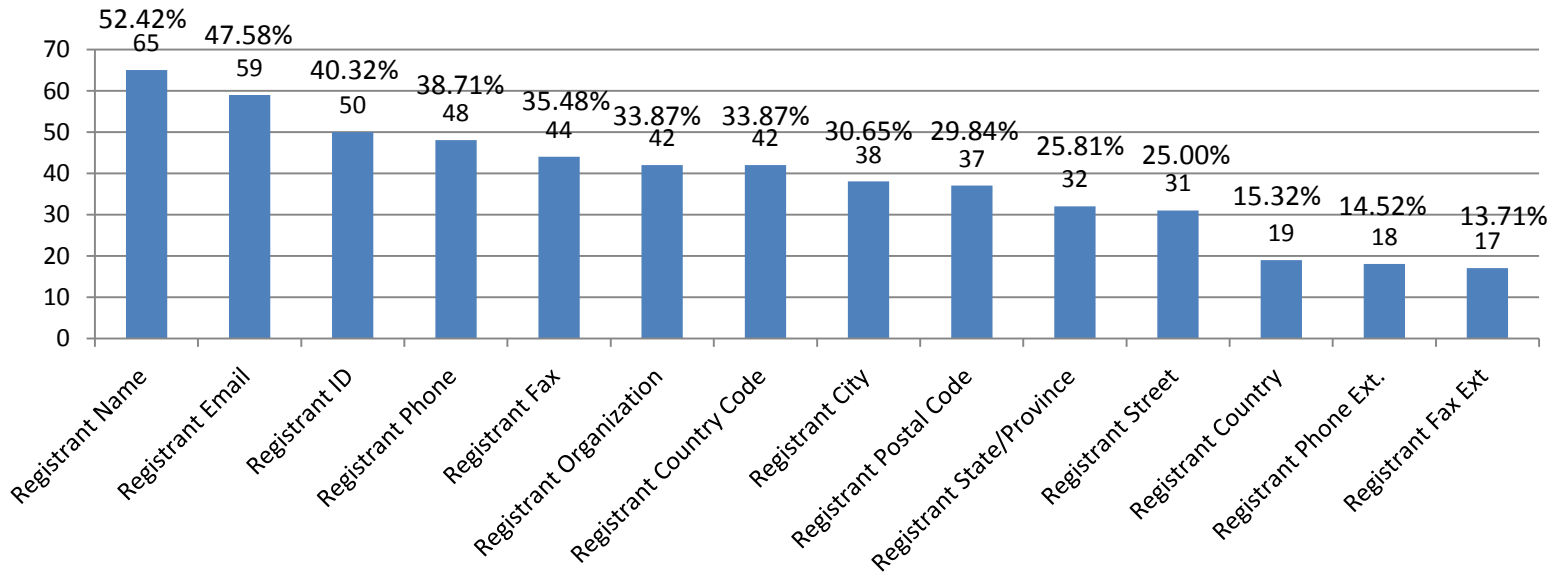
# Analysis - Admin Contact

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Admin Street | Address | 64 | 19 |
| Admin Name | Admin Name | 60 | 9 |
| Admin Email | Admin Email | 54 | 12 |
| Admin ID | Admin ID | 52 | 16 |
| Admin Fax | Admin Fax | 44 | 8 |
| Admin Phone | Admin Phone | 43 | 9 |
| Admin Organization | Admin Organization | 42 | 9 |
| Admin Country Code | Country | 42 | 7 |
| Admin City | Admin City | 35 | 5 |
| Admin Postal Code | Admin Postal Code | 35 | 7 |
| Admin State/Province | Admin State/Province | 28 | 5 |
| Admin Country | Admin Country | 17 | 5 |
| Admin Phone Ext. | Admin Phone Ext. | 17 | 3 |
| Admin Fax Ext. | Admin Fax Ext. | 17 | 3 |
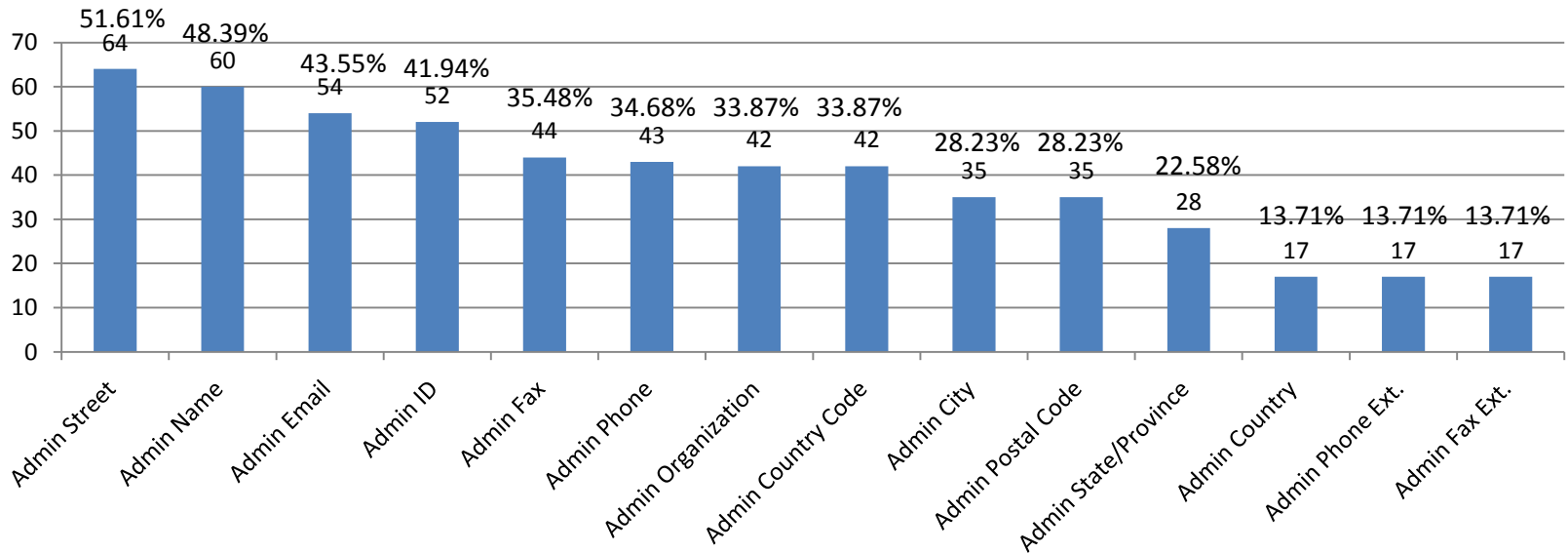
# Analysis - Tech Contact

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Tech Email | Tech Email | 59 | 9 |
| Tech ID | Tech ID | 55 | 16 |
| Tech Name | Tech Name | 47 | 6 |
| Tech Fax | Tech Fax | 45 | 9 |
| Tech Phone | Tech Phone | 45 | 10 |
| Tech Country Code | Country | 40 | 4 |
| Tech Organization | Tech Organization | 39 | 7 |
| Tech City | Tech City | 36 | 4 |
| Tech Postal Code | Tech Postal Code | 36 | 7 |
| Tech State/Province | Tech State/Province | 30 | 4 |
| Tech Street | Tech Street1 | 27 | 16 |
| Tech Country | Tech Country | 18 | 5 |
| Tech Fax Ext | Tech Fax Ext | 18 | 3 |
| Tech Phone Ext. | Tech Phone Ext. | 13 | 3 |
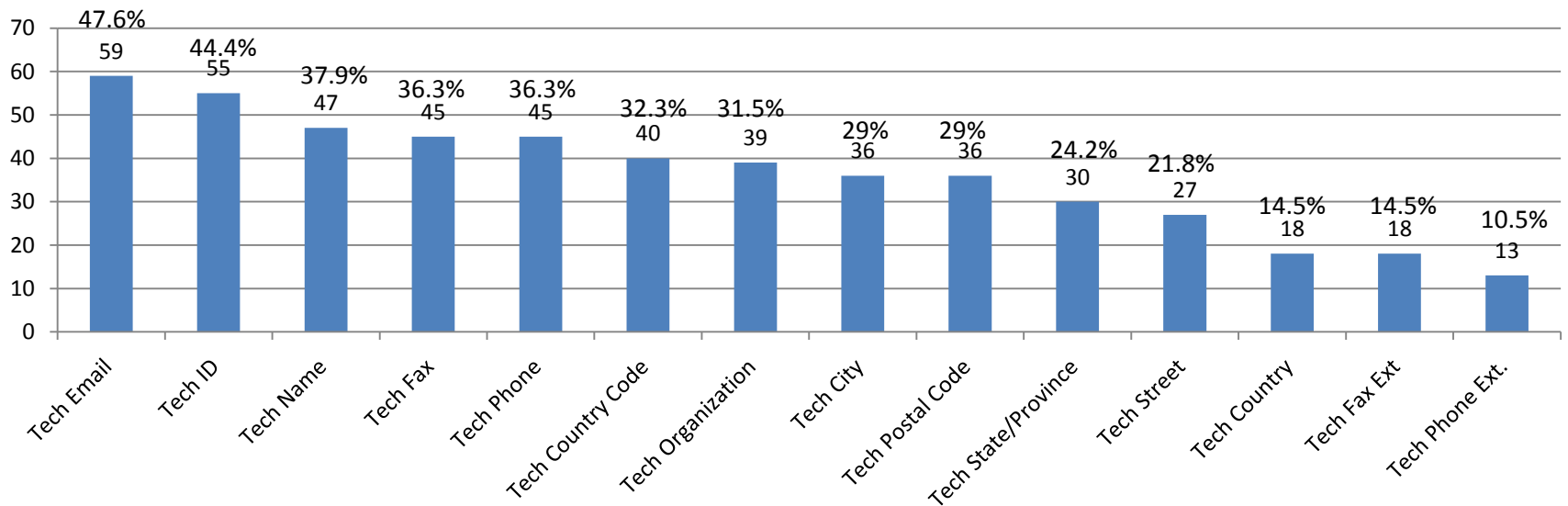
# Analysis - Billing Contact

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Billing Name | Name | 47 | 5 |
| Billing Fax | Fax | 43 | 6 |
| Billing Email | Email Address | 42 | 7 |
| Billing Country Code | Country | 38 | 4 |
| Billing Phone | Phone Number | 34 | 6 |
| Billing ID | Billing ID | 28 | 9 |
| Billing City | Billing City | 28 | 4 |
| Billing Organization | Billing Organization | 28 | 5 |
| Billing Postal Code | Billing Postal Code | 27 | 4 |
| Billing State/Province | Billing State/Province | 21 | 4 |
| Billing Street | Billing Street1 | 19 | 13 |
| Billing Country | Billing Country | 13 | 5 |
| Billing Phone Ext. | Billing Phone Ext. | 10 | 2 |
| Billing FAX Ext | Billing FAX Ext | 10 | 2 |
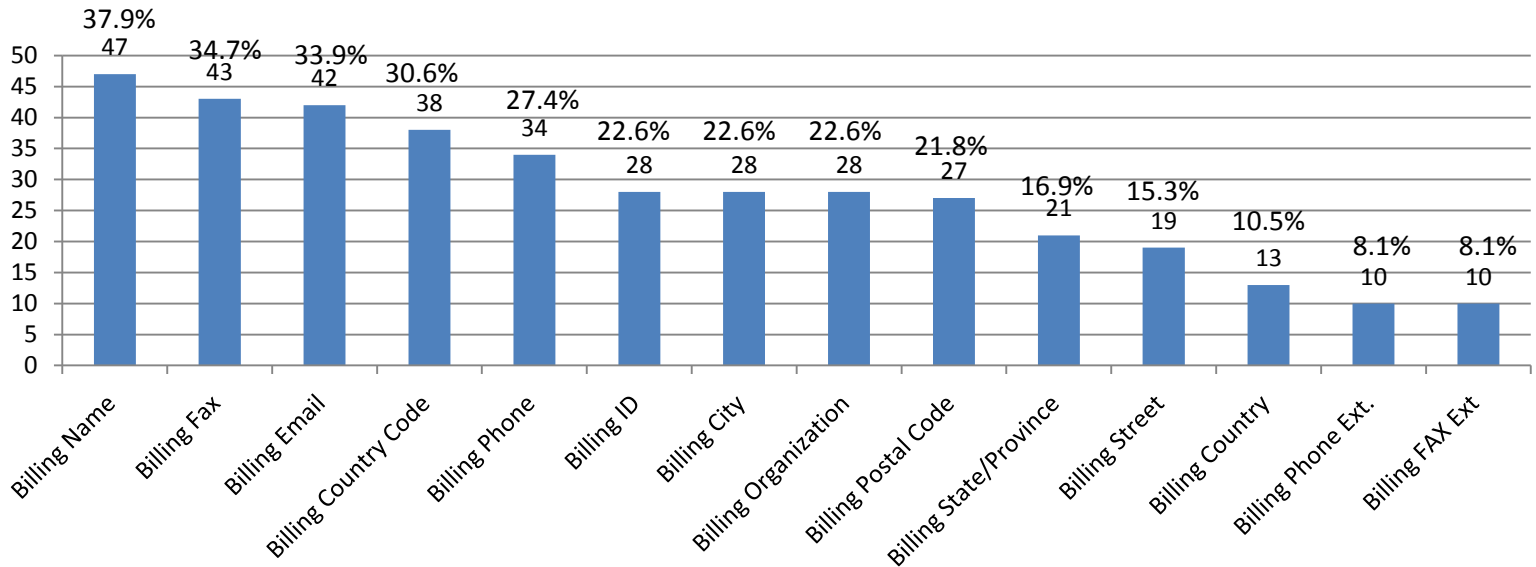
# Analysis - Name Server & Registrar

| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Name Server | Name Server | 114 | 63 |



| Data Element | Most Popular Label | No. of TLDs | No. of Labels |
|---|---|---|---|
| Sponsoring Registrar | Registrar | 84 | 6 |
| Created by Registrar | Created by | 14 | 3 |
| Updated by Registrar | Last Updated by Registrar | 11 | 3 |
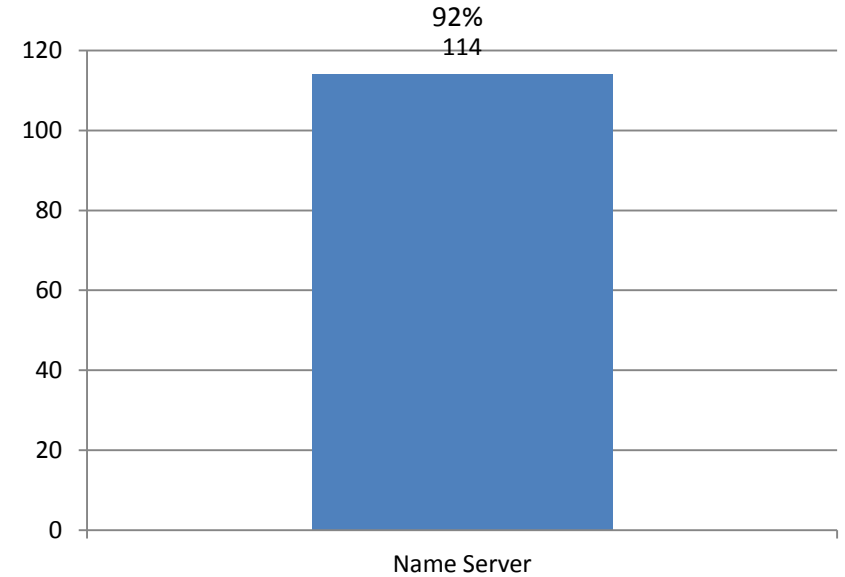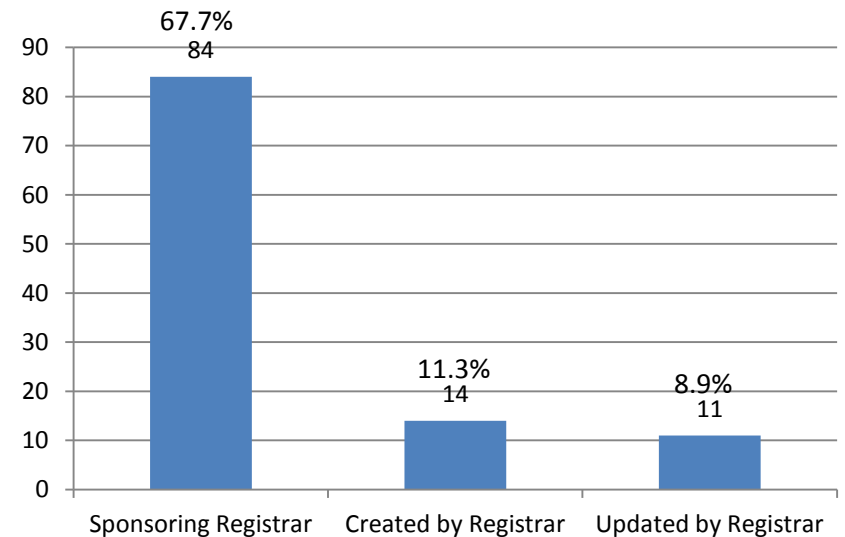
# Analysis - Top 10 No. of Labels
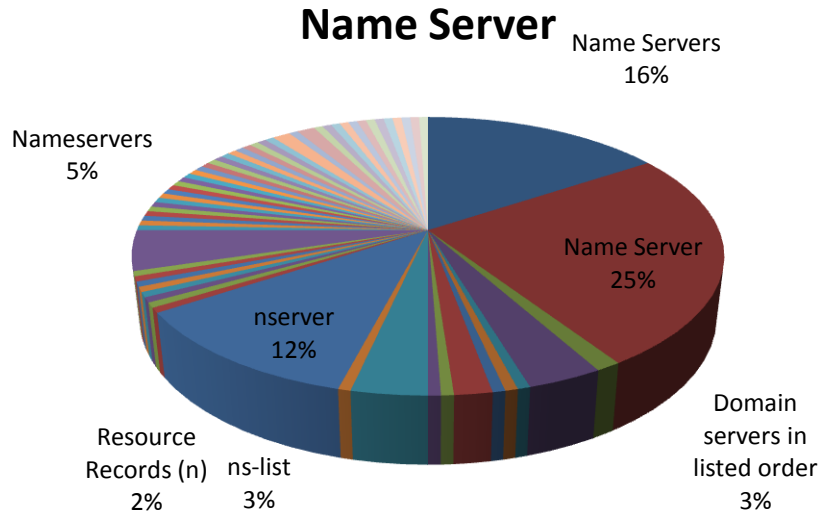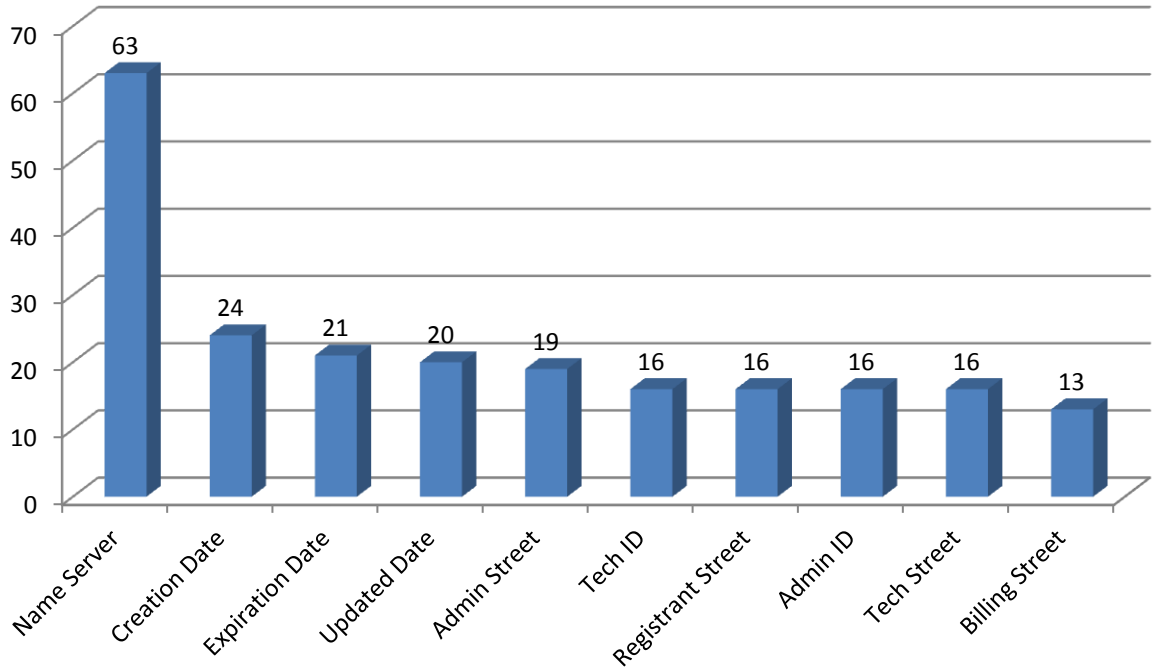
# Analysis - Cumulative Distribution of No. of Labels



About 90% of Data elements have more than 2Labels.
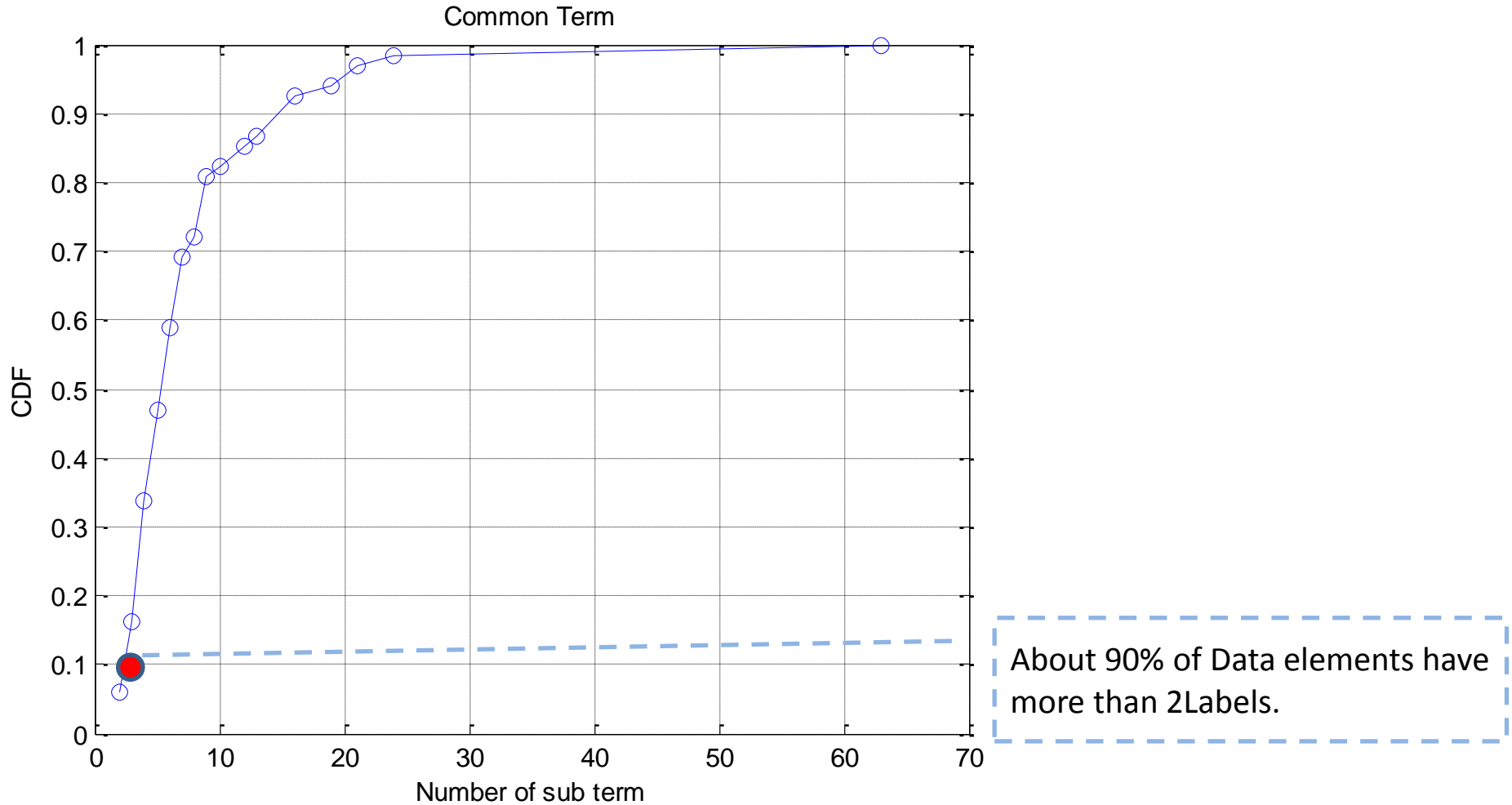
# Analysis—Top 20 Extensions
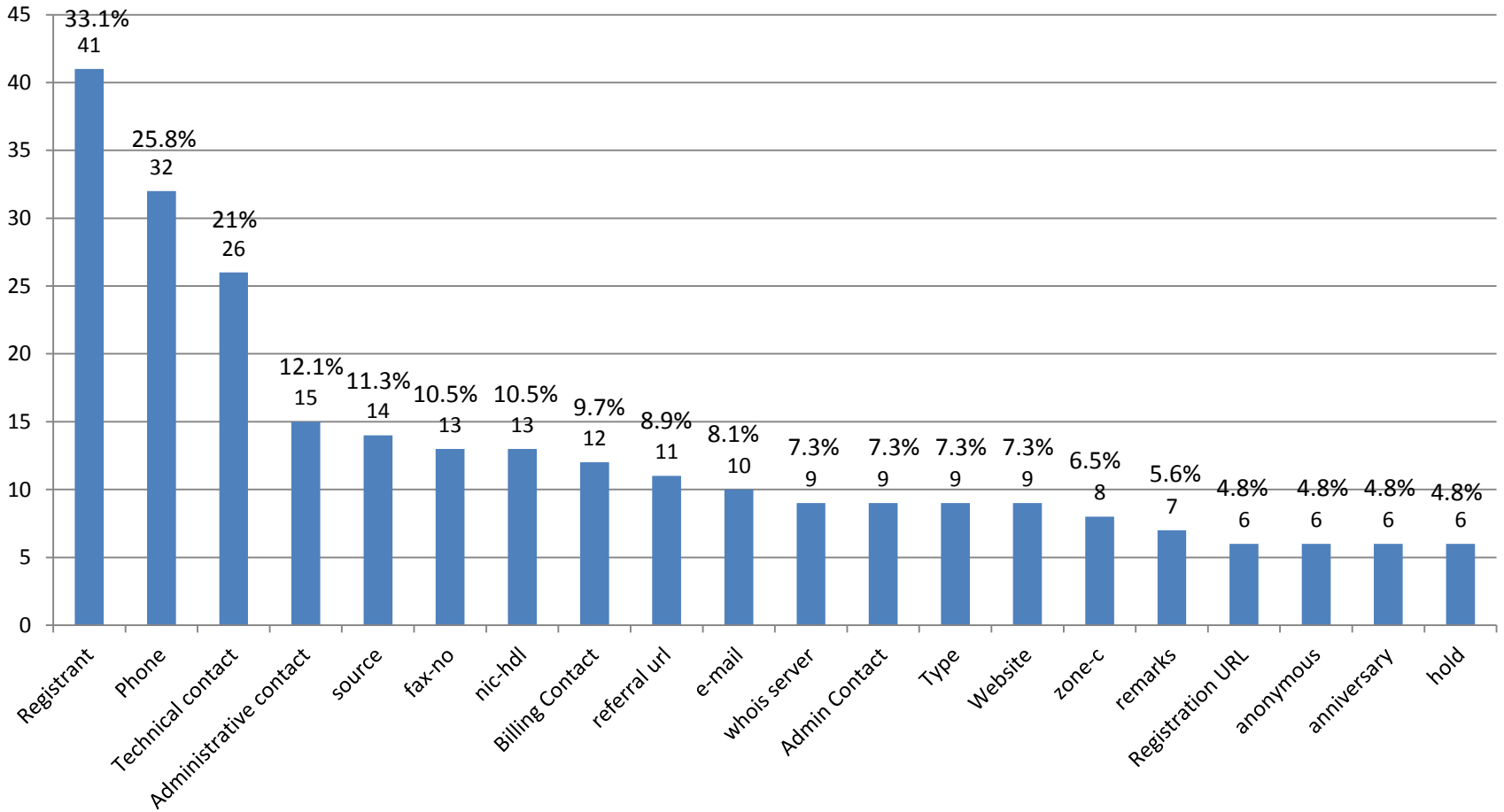
# Analysis—Part of Extensions

| Data Label | No. of TLDs |
|---|---|
| Registrant | 41 |
| Phone | 32 |
| Technical contact | 26 |
| Administrative contact | 15 |
| source | 14 |
| fax-no | 13 |
| nic-hdl | 13 |
| Billing Contact | 12 |
| referral url | 11 |
| e-mail | 10 |
| whois server | 9 |
| Admin Contact | 9 |
| Type | 9 |
| Website | 9 |
| zone-c | 8 |
| remarks | 7 |
| Registration URL | 6 |
| anonymous | 6 |
| anniversary | 6 |
| hold | 6 |
| nsl-id | 6 |
| obsoleted | 6 |
| Customer Service Contact | 5 |
| Customer Service Email | 4 |
| Registrar ID | 4 |

| Data Label | No. of TLDs |
|---|---|
| org | 4 |
| person | 4 |
| Maintainer | 4 |
| Nombre | 3 |
| Sponsoring Registrar IANA ID | 3 |
| Trademark Number | 3 |
| Trademark Country | 3 |
| descr | 3 |
| url | 3 |
| Postal address | 3 |
| Registrar URL | 3 |
| International Name | 3 |
| International Address | 3 |
| Admin Contacts | 2 |
| Contractual Language | 2 |
| Date Trademark Registered | 2 |
| Date Trademark Applied For | 2 |
| IP Address | 2 |
| Keys | 2 |
| Language | 2 |
| NIC handle | 2 |
| Record maintained by | 2 |
| Registration Service Provider | 2 |
| Registration Service Provided By | 2 |
| Registrar URL (registration services) | 2 |

# Limitations

- The input "nic.ccTLD" maybe is not a good choice.
- The classification of local script data elements may not be accurate. (11 registries give local script responses)
- The extension data elements are used randomly by different registries. It is difficult to do statistical analysis.
- Sample sizes of contact, name server and registrar queries are small.
  - We only use "Whois contactID/nameserver/registrar" as the query commands to check.
  - Some registries may not support contact, name server or registrar queries.
  - Some may not support query contact by ID.
  - Contact information of some registries may be protected.

Thank you!

Questions or Comments