OSLOMET

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

simulamet

UNIVERSITY
OF TWENTE.

Lowering the Barriers to Working with Public RIR-Level Data

Alfred Arouna[1,2] Ioana Livadariu[1] Mattijs Jonker[3]
alfred@simula.no

[1]Simula Metropolitan [2]Oslo Metropolitan University [3]University of Twente

July 24, 2023 • Applied Networking Research Workshop (ANRW) 2023 • San Francisco, USA (Hybrid)

Outline

1 Background

- RIR System
- RIR Core Functions

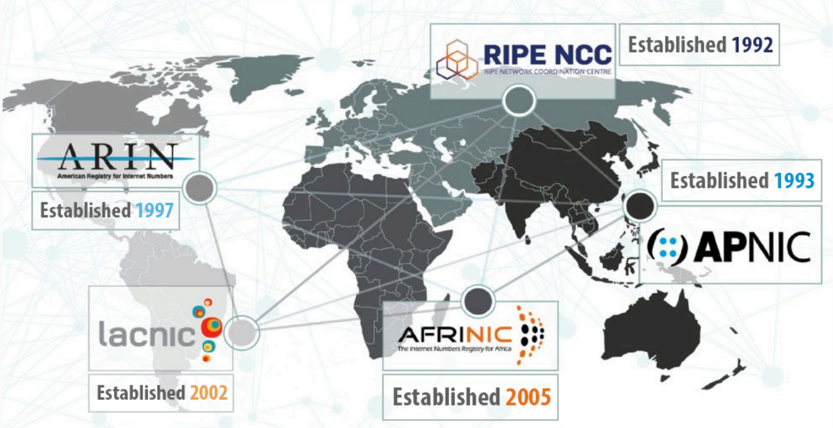
2 RIR Data

- rDNS & WHOIS Data
- Inconsistencies & Peculiarities

3 Consolidated RIR Data

- Data Format
- Example

Internet Resources Management: the RIRs System

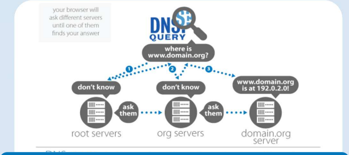

RIPE NCC
RIPE NETWORK COORDINATION CENTRE
Established 1992
ARIN
American Registry for Internet Numbers
Established 1997
Established 1993
APNIC
lacnic
Established 2002
AFRINIC
The Internet Numbers Registry for Africa
Established 2005

Five Regional Internet Registries (RIRs).

https://www.nro.net/wp-content/uploads/How-it-Works-The-RIR-System_ICANN66_Nov2019.pdf

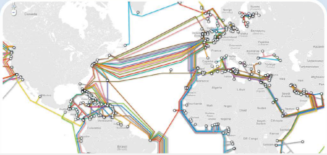
Regional Internet Registry: Core Functions

Manage, distribute and register Internet Number Resources (IPV4 & IPv6 addresses and Autonomous System Numbers (ASNs).



Maintain directory services including Whois and routing registries.

Provide reverse DNS.



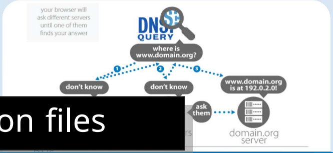
Support Internet infrastructure through technical coordination.

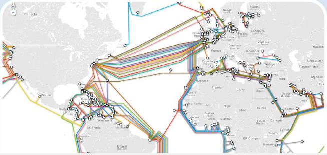
Facilitate community driven policy development process.

https://www.nro.net/wp-content/uploads/How-it-Works-The-RIR-System__ICANN66__Nov2019.pdf

Regional Internet Registry: Core Functions

WHOIS & Delegation files


your browser will ask different servers until one of them finds your answer
DNS QUERY
where is www.domain.org?
don't know
don't know
www.domain.org is at 192.0.2.0!
ask them
domain.org server



Manage, distribute and register Internet Number Resources (IPV4 & IPv6 addresses and Autonomous System Numbers (ASNs).

Maintain directory services including Whois and routing registries.

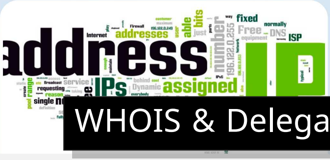
Provide reverse DNS.

Support Internet infrastructure through technical coordination.

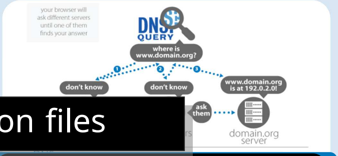
Facilitate community driven policy development process.

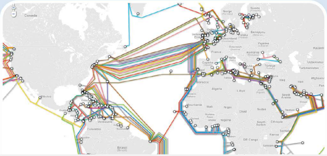
https://www.nro.net/wp-content/uploads/How-it-Works-The-RIR-System__ICANN66__Nov2019.pdf

Regional Internet Registry: Core Functions



WHOIS & Delegation files


your browser will ask different servers until one of them finds your answer
DNS QUERY
where is www.domain.org?
1
2
3
don't know
don't know
www.domain.org is at 192.0.2.0!
ask them
domain.org server



Manage, distribute and register Internet Number Resources (IPV4 & IPv6 addresses and Autonomous System Numbers (ASNs).

Maintain directory services including Whois and routing registries.

Provide reverse DNS.

rDNS Zone files

Support Internet infrastructure through technical coordination.

Facilitate community driven policy development process.

https://www.nro.net/wp-content/uploads/How-it-Works-The-RIR-System__ICANN66__Nov2019.pdf

RIR Data

rDNS zones: Support for Critical Internet Services

Example of snippets with unexpected RRs (209.in-addr.arpa from ARIN).

```
g.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2a01:4f8:c0c:72f4:0:0:0:1
g.ns.157.16.209.in-addr.arpa.    86400  IN  A     78.47.120.45
p.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2a00:a600:6:42:0:0:0:1
p.ns.157.16.209.in-addr.arpa.    86400  IN  A     209.16.157.42
y.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2001:19f0:7401:8a21:0:0:0:1
y.ns.157.16.209.in-addr.arpa.    86400  IN  A     95.179.232.160
```

Example of snippets with classless delegation (164.in-addr.arpa from RIPE).

```
191.39.215.164.in-addr.arpa.      86400 IN CNAME  191.128-191.39.215.164.in-addr.arpa.
128-191.39.215.164.in-addr.arpa.  86400 IN NS     dns1.ficolo.net.
192.39.215.164.in-addr.arpa.      86400 IN CNAME  192.192-255.39.215.164.in-addr.arpa.
192-255.39.215.164.in-addr.arpa.  86400 IN NS     ns1.shellit.org.
```

rDNS zones: Support for Critical Internet Services

Example of snippets with unexpected RRs (209.in-addr.arpa from ARIN).

```
g.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2a01:4f8:c0c:72f4:0:0:0:1
g.ns.157.16.209.in-addr.arpa.    86400  IN  A     78.47.120.45
p.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2a00:a600:6:42:0:0:0:1
p.ns.157.16.209.in-addr.arpa.    86400  IN  A     209.16.157.42
y.ns.157.16.209.in-addr.arpa.    86400  IN  AAAA  2001:19f0:7401:8a21:0:0:0:1
y.ns.157.16.209.in-addr.arpa.    86400  IN  A     95.179.232.160
```

Example of snippets with classless delegation (164.in-addr.arpa from RIPE).

```
191.39.215.164.in-addr.arpa.      86400 IN CNAME  191.128-191.39.215.164.in-addr.
    arpa.
128-191.39.215.164.in-addr.arpa. 86400 IN NS     dns1.ficolo.net.
192.39.215.164.in-addr.arpa.      86400 IN CNAME  192.192-255.39.215.164.in-addr.
    arpa.
192-255.39.215.164.in-addr.arpa. 86400 IN NS     ns1.shellit.org.
```

164.215.39.128/26

164.215.39.192/26

rDNS zones: Support for Critical Internet Services

Example of snippets with unexpected RRs (209.in-addr.arpa from ARIN).

```
g.ns.157.16.209.in-addr.arpa.     86400   IN   AAAA   2a01:4f8:c0c:72f4:0:0:0:1
g.ns.157.16.209.in-addr.arpa.     86400   IN   A      78.47.120.45
p.ns.157.16.209.in-addr.arpa.     86400   IN   AAAA   2a00:a600:6:42:0:0:0:1
p.ns.157.16.209.in-addr.arpa.     86400   IN   A      209.16.157.42
y.ns.1
y.ns.1
```

rDNS can be used to track lame delegation at the RIR level or to map authoritative nameservers to prefixes.

Example of snippets with classless delegation (164.in-addr.arpa from RIPE).

```
191.39.215.164.in-addr.arpa.     86400 IN CNAME  191.128-191.39.215.164.in-addr.
    arpa.
128-191.39.215.164.in-addr.arpa. 86400 IN NS     dns1.ficolo.net.
192.39.215.164.in-addr.arpa.     86400 IN CNAME  192.192-255.39.215.164.in-addr.
    arpa.
192-255.39.215.164.in-addr.arpa. 86400 IN NS     ns1.shellit.org.
```

WHOIS: Information on Resources Registration

ARIN object using route attribute instead of inetnum.

```
[empty line]
route:           173.245.144.0/20
origin:          AS15065
descr:           3330 State Highway 11B,
                 P.O. Box 150
                 Nicholville NY 12965
                 United States
admin-c:         ANDER639-ARIN
tech-c:          ANDER639-ARIN
tech-c:          NETWO3464-ARIN
tech-c:          NOC32314-ARIN
mnt-by:          MNT-SLICCO-1
created:         2021-08-31T21:31:33Z
last-modified:   2021-08-31T21:31:33Z
source:          ARIN
[empty line]
```

LACNIC objects with custom inetnum notation.

```
[empty line]
inetnum:     170.150.4/22
status:      allocated
city:        Andradina
country:     BR
created:     2016-06-01
changed:     2020-03-11
source:      LACNIC
[empty line]
inetnum:     190.144/14
status:      allocated
city:        Bogota
country:     CO
created:     2007-01-11
changed:     2007-01-11
source:      LACNIC
[empty line]
```

WHOIS: Information on Resources Registration

ARIN object using route attribute instead of inetnum.

```
[empty line]
route:            173.245.144.0/20
origin:           AS15065
descr:            2330 State Highway 11B
admin-c:          ANDER639-ARIN
tech-c:           ANDER639-ARIN
tech-c:           NETWO3464-ARIN
tech-c:           NOC32314-ARIN
mnt-by:           MNT-SLICCO-1
created:          2021-08-31T21:31:33Z
last-modified:    2021-08-31T21:31:33Z
source:           ARIN
[empty line]
```

LACNIC objects with custom inetnum notation.

```
[empty line]
inetnum:      170.150.4/22
status:       allocated
city:         Andradina
source:       LACNIC
[empty line]
inetnum:      190.144/14
status:       allocated
city:         Bogota
country:      CO
created:      2007-01-11
changed:      2007-01-11
source:       LACNIC
[empty line]
```

WHOIS is widely used by network researchers and operators, but it comes with some limitations.

RIRs Data Limitations: Inconsistencies and Peculiarities

WHOIS

- One-off data.
- URLs variety.
- Objects & key inconsistency.

	Prefixes	Mnt.	Name	Created	Status
RIPE	inetnum	mnt-by	netname	created	status
ARIN	route	mnt-by	desc	created	N.A.
LACNIC	inetnum	N.A.	N.A.	created	status
APNIC	inetnum	mnt-by	netname	last-modified	status
AFRINIC	inetnum	mnt-by	netname	changed[0]	status

rDNS Zones

- One-off data.
- Unexpected RRs.
- Not compliant with RFC 1035.

Consolidated RIR Data

Addressing Limitations: Consolidated Data

Consolidated and common format, interoperable and optimised (tiered – year, month, day – hierarchy) for large-scale analysis tool.

WHOIS + Statistics

- Longest prefix matching.
- Identifier: start and end address.
- Complementary data from delegation files.

rDNS Zones

- Domain to prefix.
- Identifier: start and end address.
- Classfull (`<octet>`) vs. classless (`CNAME`).

The data is available and further documented at https://rir-data.org

Examples: WHOIS and rDNS Consolidated Records

🟧 Identifier (WHOIS/rDNS) 🟩 Complentary (Delegation) 🟦 Flag (Classless vs. classfull) ⬛ Original (WHOIS/rDNS)

```
{”serial”: 748705, ”use_route”: true, ”prefixes”: [”23.219.183.0/24”], ”af”: 4,
    ”start_address”: ”23.219.183.0”, ”end_address”: ”23.219.183.255”, ”descr”: ”Akamai
     Technologies”, ”origin”: 20940, ”mnt-by”: ”MNT-AKAMAI”, ”source”: ”ARIN”, ”
    created”: 1555027200, ”last-modified”: 1555027200, ”status”: ”ALLOCATED”, ”
    netname”: null, ”country”: ”US”}
```

Example of WHOIS data.

```
{”prefixes”: [”23.219.0.0/16”], ”start_address”: ”23.219.0.0”, ”end_address”: ”23.219.25
    5.255”, ”rfc_2317”: false, ”timestamp”: 1684357200, ”source”: ”ARIN”, ”af”: 4, ”
    rdns”: {”name”: [”219.23.in-addr.arpa.”], ”origin”: [”23.in-addr.arpa.”], ”ttl
    ”: 86400, ”rdclass”: ”IN”, ”rdatasets”: {”NS”: [”ns{1-8}.reverse.deploy.
    akamaitechnologies.com.”]}}}
```

Example of rDNS data.

WHOIS

- Identifier (start/end).
- Longitudinal.

rDNS

- Public (since Nov 1, 2022).
- Interoperable and efficient.

The data is available and further documented at https://rir-data.org

Thanks

Backup Slides

Data Access

Direct Files Download

a) Scraping files from base urls:
https://data.rir-data.org/rir-data/rirs-rdns-formatted/type=enriched/
https://data.rir-data.org/rir-data/whois-formatted/type=enriched/

b) Files are organized in tiered (YYYY,MM, DD) hierarchy:
`/year=YYYY/month=MM/day=DD/hour={00|20}/`

c) File names are as follows:

- `/all_rdns__pytricia_YYYYMMDD00_YYYYMMDD23_without_RRSIG_NSEC_DNSKEY.jsonl.bz2`
- `/all_objects_pytricia_inetnum_YYYYMMDD00_YYYYMMDD23.jsonl.bz2`

```
# Using Wget
wget https://data.rir-data.org/rir-data/rirs-rdns-formatted/type=enriched/year=2023/
     month=01/day=01/hour=00/all_rdns__pytricia_2023010100_2023010123
     _without_RRSIG_NSEC_DNSKEY.jsonl.bz2
# Using Curl
curl -O https://data.rir-data.org/rir-data/whois-formatted/type=enriched/year=2023/
     month=01/day=01/hour=20/all_objects_pytricia_inetnum_2023010100_2023010123.
     jsonl.bz2
```

PySpark for large-scale data analysis

```
# Requirement: A running Hadoop cluster.
# Import required modules.  Note that pySpark need s3a access modules.
# Create Spark configuration and initialize  Spark Session.
spark = SparkSession.builder.config(conf=sparkConf).getOrCreate()
sc = spark.sparkContext

# Read RIR rDNS data into DataFrame
my_rir_data_df = spark.read.format(”json”).option(”basePath”, ”s3a://rir-data/rirs-rdns-formatted/type=enriched”).load(
    [”s3a://rir-data/rirs-rdns-formatted/type=enriched/year=2023/month=05/day=31/”]
)
my_rir_data_df = my_rir_data_df.persist(pyspark.StorageLevel.MEMORY_AND_DISK)
my_rir_data_df.printSchema()

# Read WHOIS data into DataFrame
my_whois_data_df = spark.read.format(”json”).option(”basePath”, ”s3a://rir-data/whois-formatted/type=enriched”).load(
    [”s3a://rir-data/whois-formatted/type=enriched/year=2023/month=05/day=31/”]
)
my_whois_data_df = my_whois_data_df.persist(pyspark.StorageLevel.MEMORY_AND_DISK)
my_whois_data_df.printSchema()
```