Lowering the Barriers to Working with Public RIR-Level Data

Alfred Arouna
SimulaMet / OsloMet
Oslo, Norway

Ioana Livadariu
SimulaMet
Oslo, Norway

Mattijs Jonker
University of Twente
Enschede, the Netherlands

ABSTRACT

Regional Internet Registries (RIRs) publish WHOIS, route object delegation, and reverse DNS zone files. These data are valuable resources for network researchers and engineers, yet contain inconsistencies and are not all available long-term. In this work, we consolidate and make available longitudinal RIR-level data, aiming to lower the barriers to start working with these data.

CCS CONCEPTS

• **Networks → Naming and addressing**.

KEYWORDS

RIR, WHOIS, rDNS

ACM Reference Format:
Alfred Arouna, Ioana Livadariu, and Mattijs Jonker. 2023. Lowering the Barriers to Working with Public RIR-Level Data. In *Applied Networking Research Workshop (ANRW '23), July 24, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3606464.3606473

1 INTRODUCTION

RIRs manage critical Internet number resources (e.g., IP prefixes) on behalf of the Internet Assigned Numbers Authority (IANA) and the Internet Corporation for Assigned Names and Numbers (ICANN). Each RIR stores member allocations in a RIR-level WHOIS database as well as in delegation files. Registries also publish their reverse DNS zone files. All five RIRs present comparable policies related to the sharing of such operational data [6], which are a valuable resource for networking research. The data is however not consistent across RIRs and some RIRs show inconsistencies between their own data sets. Moreover, not all data are historically available. Recent work by Streibelt *et al.* proposes Back-to-the-Future whois service, which relies on data that is inconsistent [7]. To the best of our knowledge, no other work has focused specifically on navigating inconsistencies in RIR data. We take steps towards this end. First, we share our observations in working with data from individual RIRs. Second, we make available consolidated and longitudinal data in a format optimised for analysis. Our overall goal is to lower barriers for others to start working with RIR data.

2 AVAILABLE RIR-LEVEL DATA

2.1 WHOIS

The WHOIS protocol was specified in the earlier days of the Internet in RFC 3912 [3]. Recently, the ICANN community voted to sunset the use of WHOIS at the registrar and registry level [5]. Still, each RIR publicly provides a snapshot of (only) their *current* WHOIS database. Although each RIR provides a dedicated process for *bulk* WHOIS data collection, the publicly available WHOIS databases: a) use different data models (i.e., objects and attributes); b) are not historically available; and c) are provided via a variety of URL schemes. Table 1 lists a few examples of different attributes used by RIRs for similar content. For example, `inetnum` is used across all RIRs except for ARIN, which uses `route`. The attribute `status` is missing in ARIN WHOIS data while LACNIC WHOIS does not provide the `maintainer` attribute. These omitted attributes are, however, present in RIR delegation files, thus one can rely on these to consolidate some missing information.

	Prefixes	Mnt.	Name	Created	Status
RIPE	inetnum	mnt-by	netname[1]	created	status
ARIN	route	mnt-by	desc	created	
LACNIC	inetnum			created	status
APNIC	inetnum	mnt-by	netname[1]	last-modified	status
AFRINIC	inetnum[2]	mnt-by	netname[1]	changed[0]	status

Table 1: Examples of WHOIS IPv4 attribute inconsistencies between RIRs. [1]Can also be stored in textttdesc attribute. [2]Can also be stored in textttroute attribute.

ANRW '23, July 24, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0274-7/23/07…$15.00
https://doi.org/10.1145/3606464.3606473

2.2 Delegation Files

RIRs provide daily delegation files of their resources allocation and assignment. These files follow the *RIR Statistics Exchange Format* [2] and are available in a directory on RIR-operated FTP servers that also mirror data from other RIRs. However, alternative URLs may be used. The delegation files are named `delegated-<registry>-extended-yyyymmdd` and include a RIR-specific `opaque-id` column that internally identifies a single organisation. In addition, RIPE, APNIC and AFRINIC provide a second daily delegation file in "non-extended" form, which does not include the `opaque-id`. We consider the extended-form files only and extract country and prefix allocations from them to enhance `WHOIS` records.

2.3 Reverse DNS Zones

All RIRs provide snapshots of their *current* reverse DNS zone files via FTP. The snapshots may involve separate files per subdomain (e.g., a file per `<octet>.in-addr.arpa`). Using these data, it becomes easier to track lame delegation at the RIR level [6]. We expect `NS` and `CNAME` resource records for classful and classless delegation. Indeed, RFC 1035 [1] specifies that one `SOA` RR should be present at the top of the zone. However, with the exception of ARIN, most RIRs do not publish these `SOA` RR on their rDNS zone files. Additionally, we note a lack of historical RIRs rDNS zone files.

3 DATASET CONSOLIDATION

3.1 WHOIS Data

We collect and store the currently available `WHOIS` snapshots (one per RIR) once every day. We collect and store the daily delegation files as they become available. We then consolidate the collected snapshots. By tracking serial number changes in `WHOIS` data, we deduplicate `WHOIS` records. Where necessary, we add information omitted in `WHOIS` records by taking it from the daily delegation files. We leverage the `inetnum` object (`route` attribute in the case of ARIN) from `WHOIS` data to map allocated IPv4 prefixes to organisations and country (if present). We also perform some sanitisation, as the data is not always consistent and can contain, e.g., objects with typographical errors.


```
{"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"}
```


Listing 1: Example record of consolidated WHOIS data.

We store the consolidated `WHOIS` data in a tiered *(year, month, day, hour)* hierarchy, which popular tools for data engineering can use for partition discovery as well as optimisation. The partitioned data contains per record information such as the source RIR, `WHOIS` serial number, object `created` and `last-modified` dates.

Note that, in `WHOIS` snapshots, route objects do not have an identifier per sé. A single `WHOIS` entry can mention one or more prefixes (adjacent and non-adjacent) allocated to a given maintainer, or omit a prefix altogether and mention only the start and end of the address block. Our consolidated records are keyed by start and end address, for contiguous space. We fill in either the prefix or the start and end addresses if absent, and separate entries into the subsets of adjacent prefixes where necessary. Listing 1 shows an example for 17 May 2023 for the start address `23.219.183.0` and end address `23.219.183.255`, assigned to *Akamai* by ARIN.

3.2 Reverse DNS Zones (RIR-Level)

We collect and store RIR-level reverse DNS zone files once per hour and rely on file fingerprints to detect changes.


```
{"prefixes": ["23.219.0.0/16"], "start_address": "23.219.0.0", "end_address": "23.219.255.255", "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."]}}}
```


Listing 2: Example record of consolidated rDNS data.

To consolidate the data, we convert each subdomain to a prefix. We enrich the data, e.g., by adding a classless delegation flag. The prefixes in RIR-level zones largely follow octet boundaries, but `CNAMEs` are sometimes present for classless delegation, i.e., RFC 2317 [4]. We store the consolidated rDNS data in a tiered hierarchy similar to the `WHOIS` data and key records in the same manner (see Section 3.1). Listing 2 contains an example and shows the `NS` records for the prefix `23.219.0.0/16`. The rDNS delegation is classful in this case, meaning the corresponding flag is set to false. Furthermore, we include information such as the TTL value.

3.3 Repository of Consolidated Data

We started creating daily consolidated data on November 1, 2022, and update these continually. The data is available for download and further documented at https://rir-data.org.

4 CONCLUSION

Public RIR-level data such as `WHOIS` snapshots and reverse DNS zone files are a valuable resource for networking research. However, some of these data know inconsistencies and not all are available longitudinally. We shed some light on inconsistencies and peculiarities in RIR-level data and propose a consolidated and common format, inter-operable and optimised for analysis with popular tools for data engineering. We make the data publicly available to lower barriers to entry. And we will continue to do so to facilitate longitudinal research.

REFERENCES

[1] 1987. Domain names - implementation and specification. RFC 1035. https://doi.org/10.17487/RFC1035

[2] APNIC. 2023. RIR statistics exchange format. https://www.apnic.net/about-apnic/corporate-documents/documents/resource-guidelines/rir-statistics-exchange-format/

[3] Leslie Daigle. 2004. WHOIS Protocol Specification. RFC 3912. https://doi.org/10.17487/RFC3912

[4] Havard Eidnes, Paul A. Vixie, and Geert Jan de Groot. 1998. Classless IN-ADDR.ARPA delegation. RFC 2317. https://doi.org/10.17487/RFC2317

[5] ICANN. 2023. 2023 Global Amendments to the Base gTLD Registry Agreement (RA), Specification 13, and 2013 Registrar Accreditation Agreement (RAA). https://www.icann.org/resources/pages/global-amendment-2023-en

[6] Number Resource Organization. 2023. RIR Comparative Policy Overview 2023-Q1. https://www.nro.net/wp-content/uploads/RIR-Comparative-Policy-Overview-2023-Q1-Final.pdf

[7] Florian Streibelt, Martina Lindorfer, Seda Gürses, Carlos H Gañán, and Tobias Fiebig. 2023. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets. In *Passive and Active Measurement: 24th International Conference, PAM 2023, Virtual Event, March 21–23, 2023, Proceedings*. Springer, 209–226.