

Internet Yellow Pages



Romain Fontugne
IIJ Research Lab

Emile Aben
RIPE NCC

RIPE86. May 25th, 2023

Knowledge seekers

- What is this AS?
- Who announces this prefix?
- How it propagates in BGP?
- Is it in RPKI? Seen at an IXP route server?

Going to RIPEstat, bgp.he.net, PeeringDB, CAIDA, whois, etc...

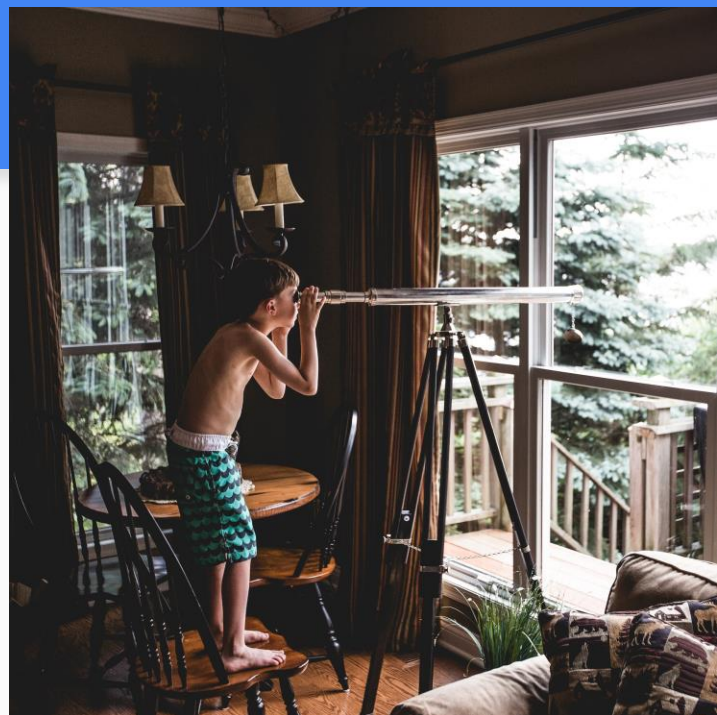
These provide complementary views, sometimes with different semantics.



Striving for better data

Our Goal:

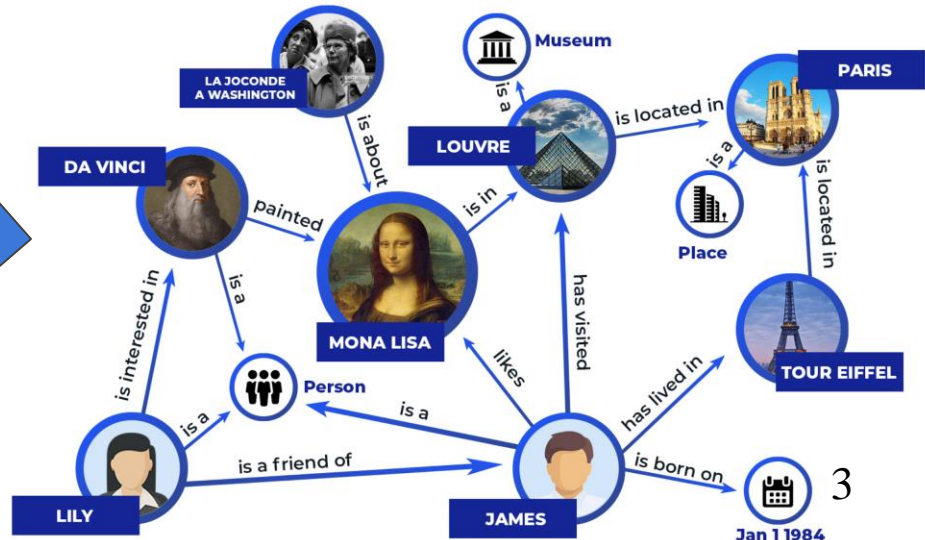
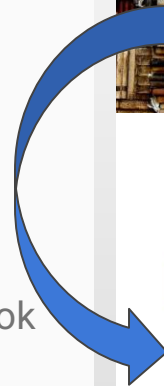
- One place, all datasets
- **Open** to anyone: Users and contributors
- **Structured**: not a repo with tons of data dumps
- **Extensible**: no fixed database schema



Knowledge graph

Structure data in a graph

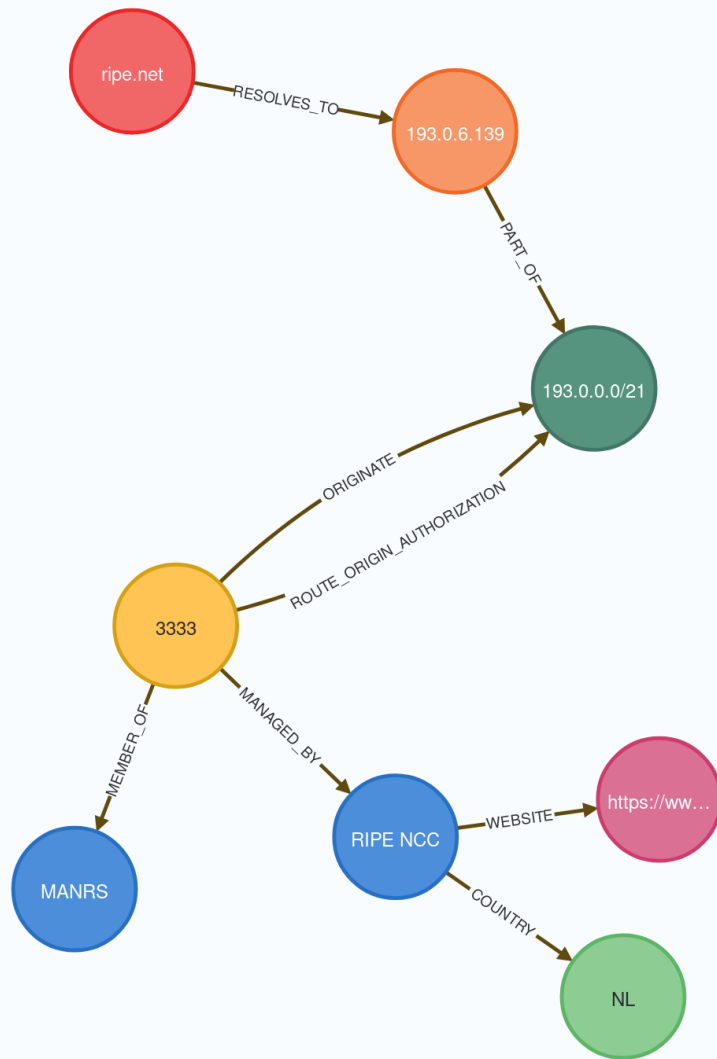
Inspired from wikidata.org, Google/Facebook knowledge graph, world factbook



Ontology

“Things, not strings”

- 19 node types
(AS, Prefix, IP, IXP, Facility, Domain Name,...)
- 22 link types
(Originate, Peer with, Resolve to, ...)



Node labels

- (8) Organization (2)
- DomainName (1) AS (1)
- Prefix (1) IP (1) Country (1)
- URL (1)

Relationship types

- (8) MEMBER_OF (1)
- ROUTE_ORIGIN_AUTHORIZATION (1)
- PART_OF (1) RESOLVES_TO (1)
- MANAGED_BY (1) ORIGINATE (1)
- COUNTRY (1) WEBSITE (1)

Displaying 8 nodes, 8 relationships.

Internet Yellow Pages: Current status

<http://iyp.ijlab.net>

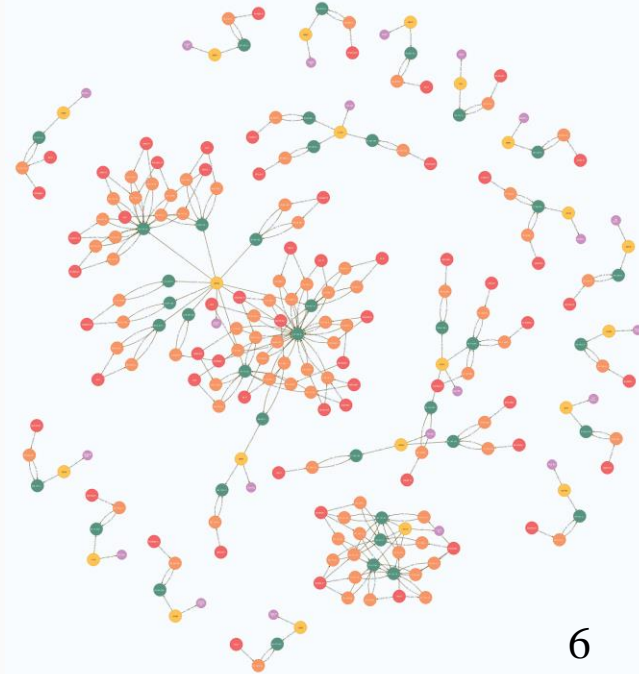
- Base on Neo4j (cypher / bolt)
- About 4M nodes, 20M links
- 16 data sources:
APNIC, BGPKIT, Bgp.tools, CAIDA, Citizen Lab, Cloudflare, IHR, InetIntel, MANRS, NRO OpenINTEL, PCH, PeeringDB, RIPE NCC, Stanford, Tranco

The screenshot displays the Internet Yellow Pages application interface, which is built on Neo4j. It is divided into several sections:

- Database Information:** Shows the database name 'neo4j' and lists various node labels (e.g., AS, PREFIX, IP, ORIGINATE, PART_OF) and relationship types (e.g., ORIGINATE, PART_OF, RESOLVES_TO).
- Neo4j Shell:** Contains a Cypher query: `neo4j$ match (n:AS {asn:12389})-[:ORIGINATE]-(:PREFIX)-(:IP)-(:DOMAIN_NAME)-[:RANK]-(:RANKING) where ...`
- Graph View:** A network graph showing nodes and relationships. The nodes are color-coded and labeled with IP addresses and ASNs. Relationships are represented by lines connecting the nodes.
- Overview Panel:** Provides a summary of the current view, including the number of nodes and relationships displayed.

Popular domain names (not in RPKI)

```
MATCH (:Ranking {name:"Tranco top 1M"})-[r:RANK]-(dn:DomainName)-  
[dnip:RESOLVES_TO]-(ip:IP)-[ippfx:PART_OF]-(pfx:Prefix)-[o:ORIGINATE]-(net:AS),  
(net)-[nn:NAME {reference_org:'PeeringDB'}]-(name:Name),  
(pfx)-[:CATEGORIZED]-(:Tag {label:'RPKI NotFound'})  
WHERE r.rank<100000 and dn.name ends with '.nl'  
RETURN dn, ip, pfx, net, dnip, ippfx, o, nn, name
```

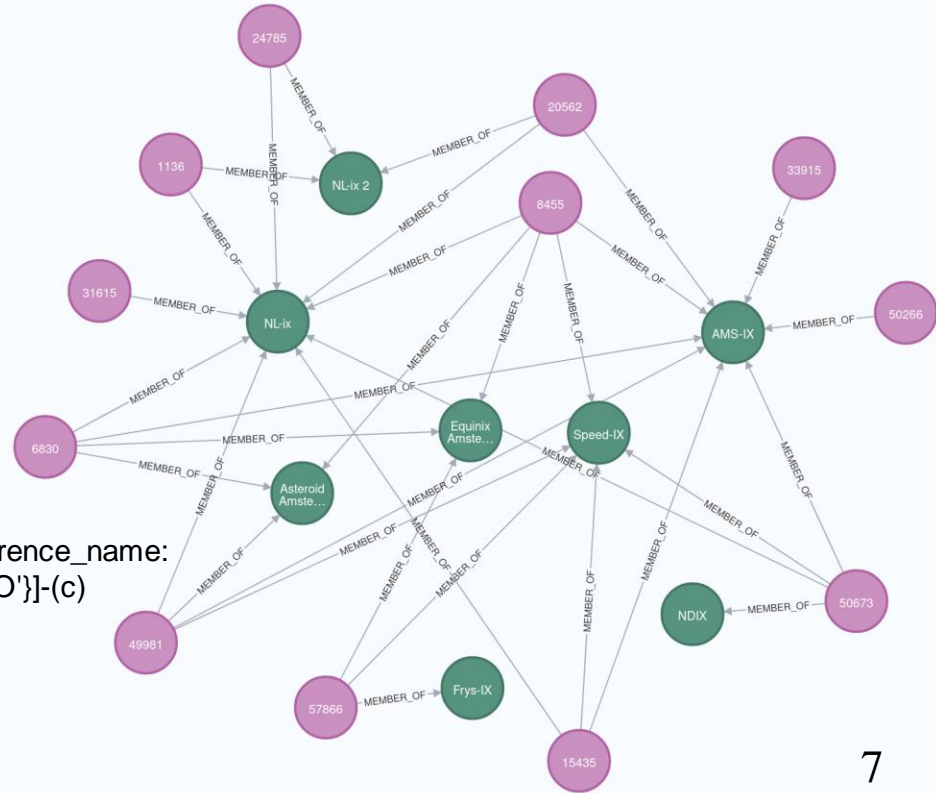


IXPs in NL?

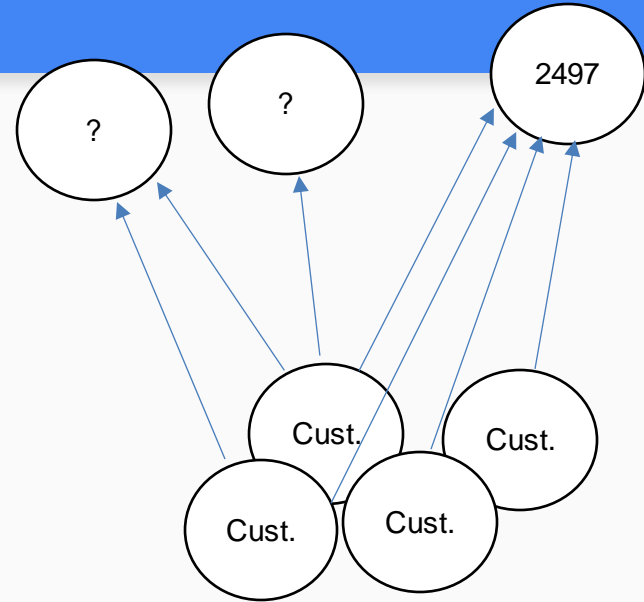
```

MATCH (c:Country {country_code:'NL'})--(r:Ranking)-[rnet:RANK {reference_name:
"ihr.country_dependency"} ]-(net:AS)-[:COUNTRY {reference_org:'NRO'}]-(c)
WHERE rnet.hege>0.02
OPTIONAL MATCH (net)-[netix:MEMBER_OF]-(ix:IXP)--(c)
RETURN net, netix, ix

```



Upstream competitors



```
MATCH (comp:AS)-[:PEERS_WITH {rel:1}]->(customer:AS)-[:PEERS_WITH {rel:1}]-(:ij:AS {asn:2497})
WITH comp, customer OPTIONAL MATCH (comp)-[:NAME {reference_org:'RIPE NCC'}]-(:comp_name:Name)
RETURN comp, comp_name, count(DISTINCT customer) AS nb_customer ORDER BY nb_customer DESC
```

Summary

- IYP is a knowledge graph for networking resources: <http://iyp.iijlab.net>
 - Weekly database dump
- Deploy your own instance and contribute to IYP
 - <https://github.com/InternetHealthReport/internet-yellow-pages>
- Let us know:
 - What dataset you use?
 - What you'd like to do with IYP?

romain@iij.ad.jp

emile.aben@ripe.net