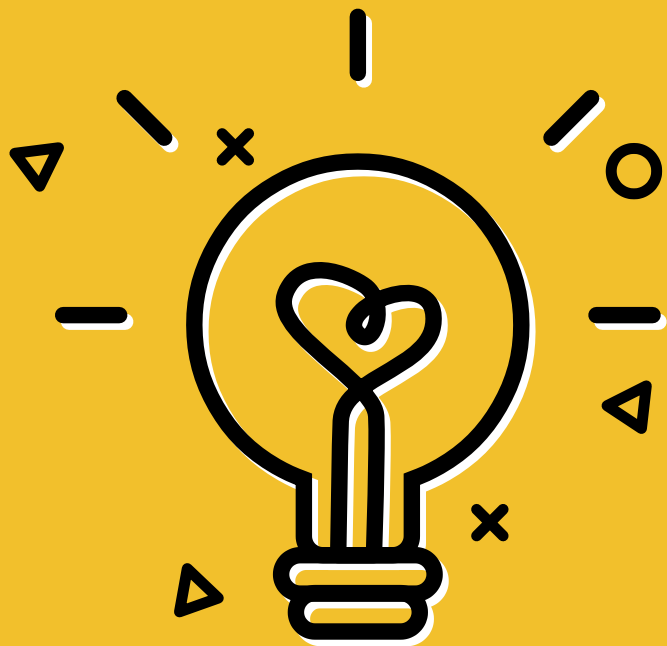




Internet Yellow Pages

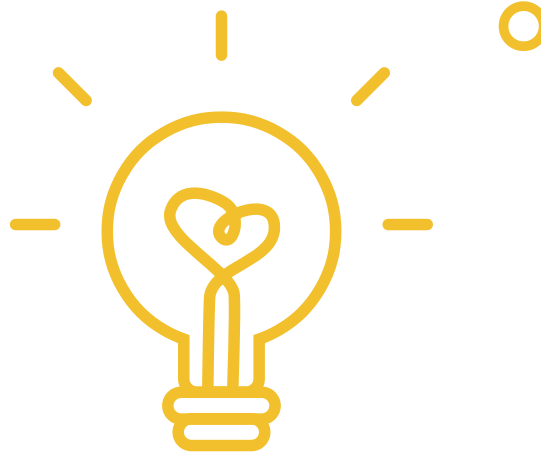
Romain Fontugne et al.

IIJ

Internet Initiative Japan



How many Internet measurement datasets have you used this year?





Internet measurements data

The good, the bad, and the ugly

Tons of open data.

Internet measurement community produces numerous open access datasets.

This is great!

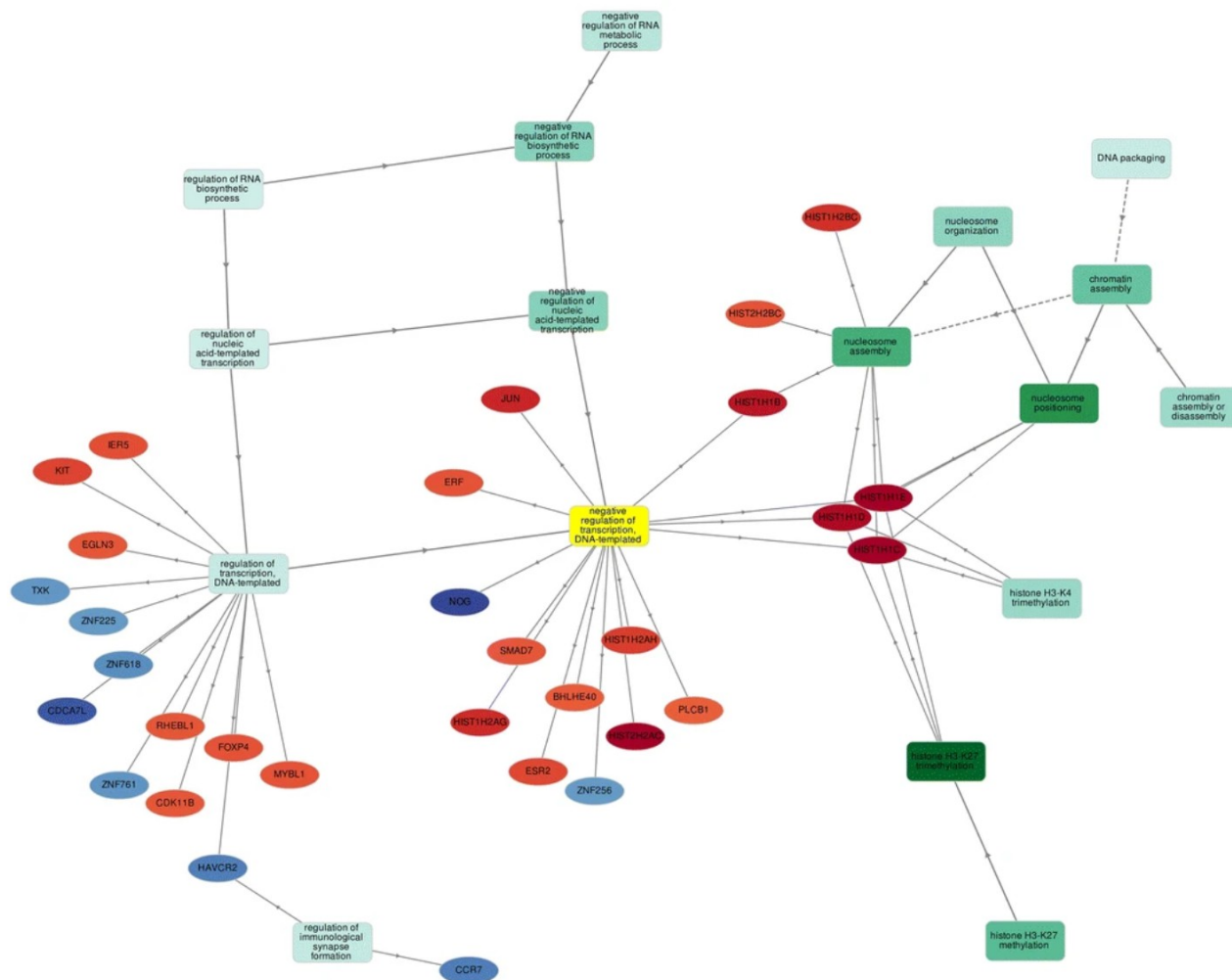
Not easy to deal with.

Each dataset using a different format, terminologies, identifiers, etc... and each researcher repeating the same data forming tasks.

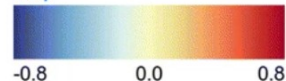
Collective knowledge?

- Currently hard to answer simple questions.
- Datasets are merged only if there is expected benefits.
- Penalize innovation and data exploration.

Gene Ontology



Help



Choose layout

Cose

Hierar

Euler

Fit to window

Remove GO parents

Remove genes

Export data

?

Pvalue threshold

$\leq 4.72e-5$

Color genes by expression ?

User supplied

GO term ID [GO:0045892](#)

Definition
negative regulation
of transcription,
DNA-templated

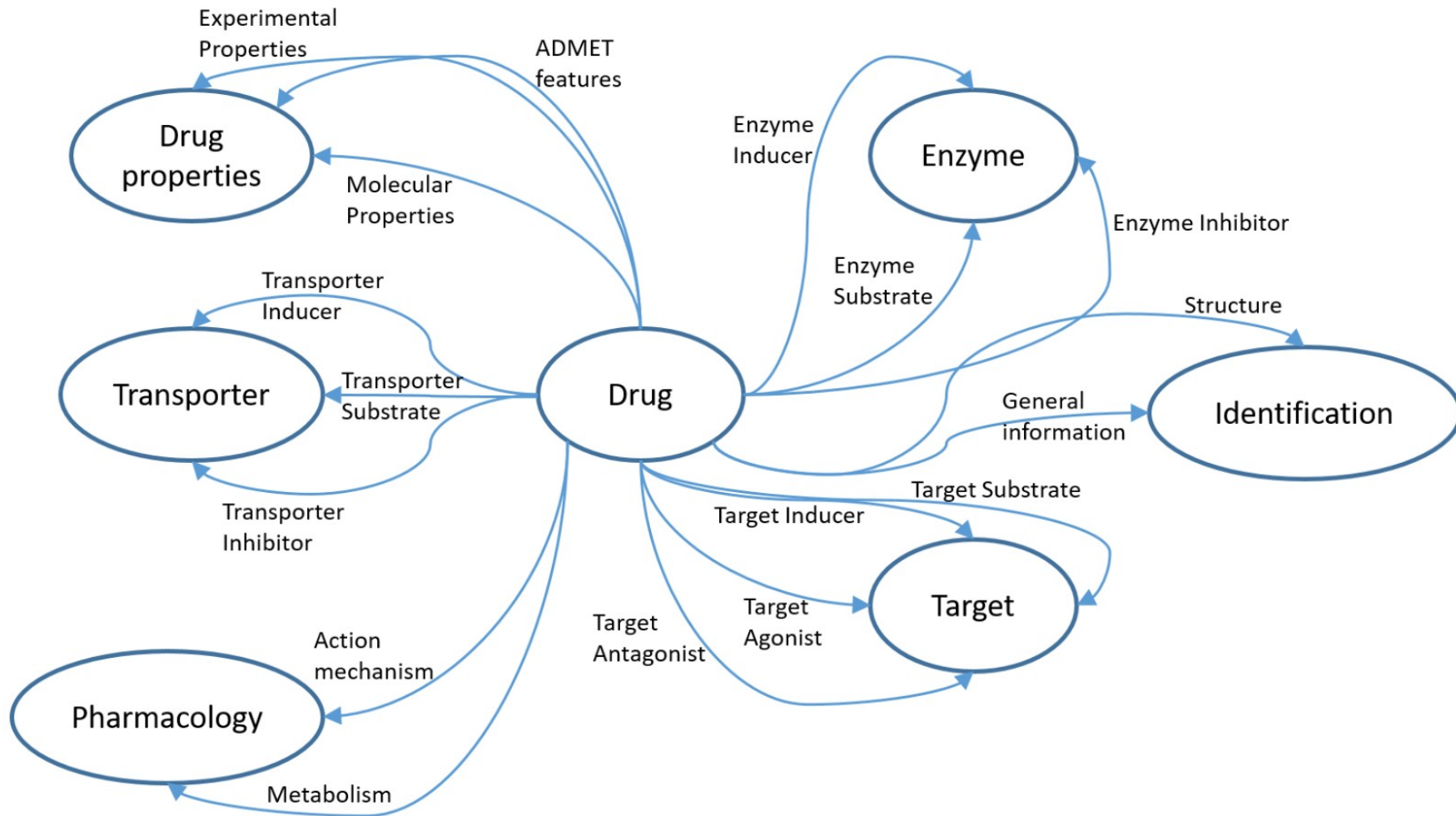
P-value
9.0e-6

P-value (FDR
adjusted)
1.6e-2

Genes: [BHLHE40](#) [SMAD7](#) [JUN](#)
[HIST1H2AG](#) [HIST1H1E](#) [HIST1H1B](#)
[HIST1H1D](#) [HIST1H1C](#) [ERF](#) [NOG](#)
[HIST2H2AC](#) [ESR2](#) [HIST1H2AH](#) [PLCB1](#)
[ZNF256](#)



Drug Bank





- Goal

Internet Yellow Pages

A knowledge graph of Internet resources:

- integrate various Internet datasets
- provide quick and homogeneous access to data
- collective knowledge of the community



Building IYP

Datasets

PeeringDB AS to Org.

asn	org_id
AS65536	123

BGPKIT pfx2asn

asn	prefix
65536	2001:0db8::/64
65536	2001:0db8::/32

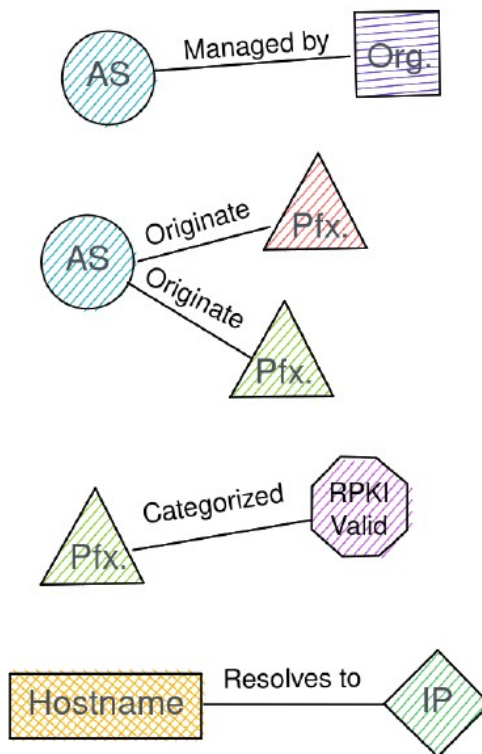
IHR ROV

prefix	RPKI
2001:DB8::/32	Valid

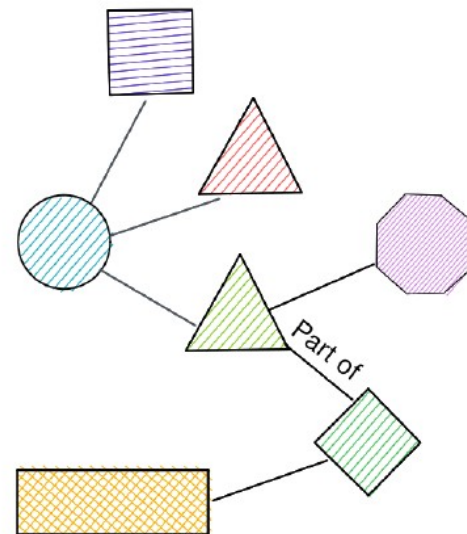
OpenINTEL Tranco

hostname	ip
example.com	2001:DB8::1

Knowledge Extraction



Fusion & Refinement





IYP in numbers

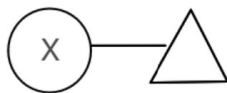
- 1 database (neo4j)
- 46 datasets from xx organizations
- 10 million nodes
- 68 million relationships
- Ontology:
 - 24 node types
 - 24 relationship types



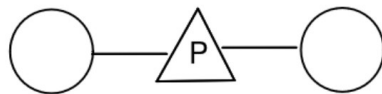
- In practice

Semantic search

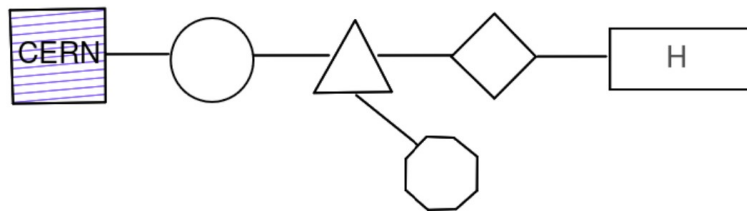
1. All originating ASes?



2. MOAS prefixes?



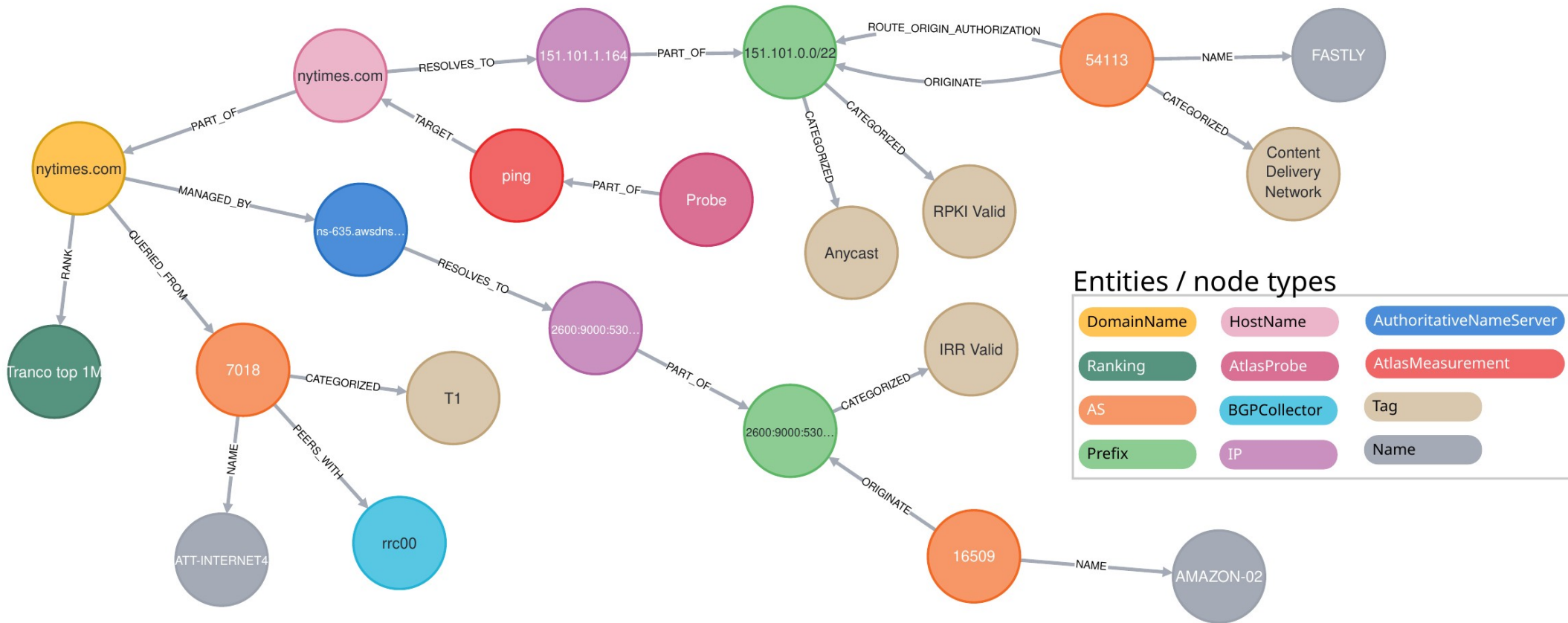
3. Popular hostnames hosted by CERN and secured by RPKI?





- In practice

Data exploration





01

Sharing queries

- Sharing queries = Sharing datasets
- Jupyter netbook = On demand reproducibility

02

Augmented datasets

- Integrate your data to IYP = connect your data to all IYP datasets



03

Local instance

- Play locally with IYP
- Integrate confidential data



04

Community engagement

- Maintain ontology
- Integrate datasets
- Entirely open source
- <https://iyp.iiijlab.net>





THANK YOU

romain@ij.ad.jp

<https://iyp.ijlab.net>