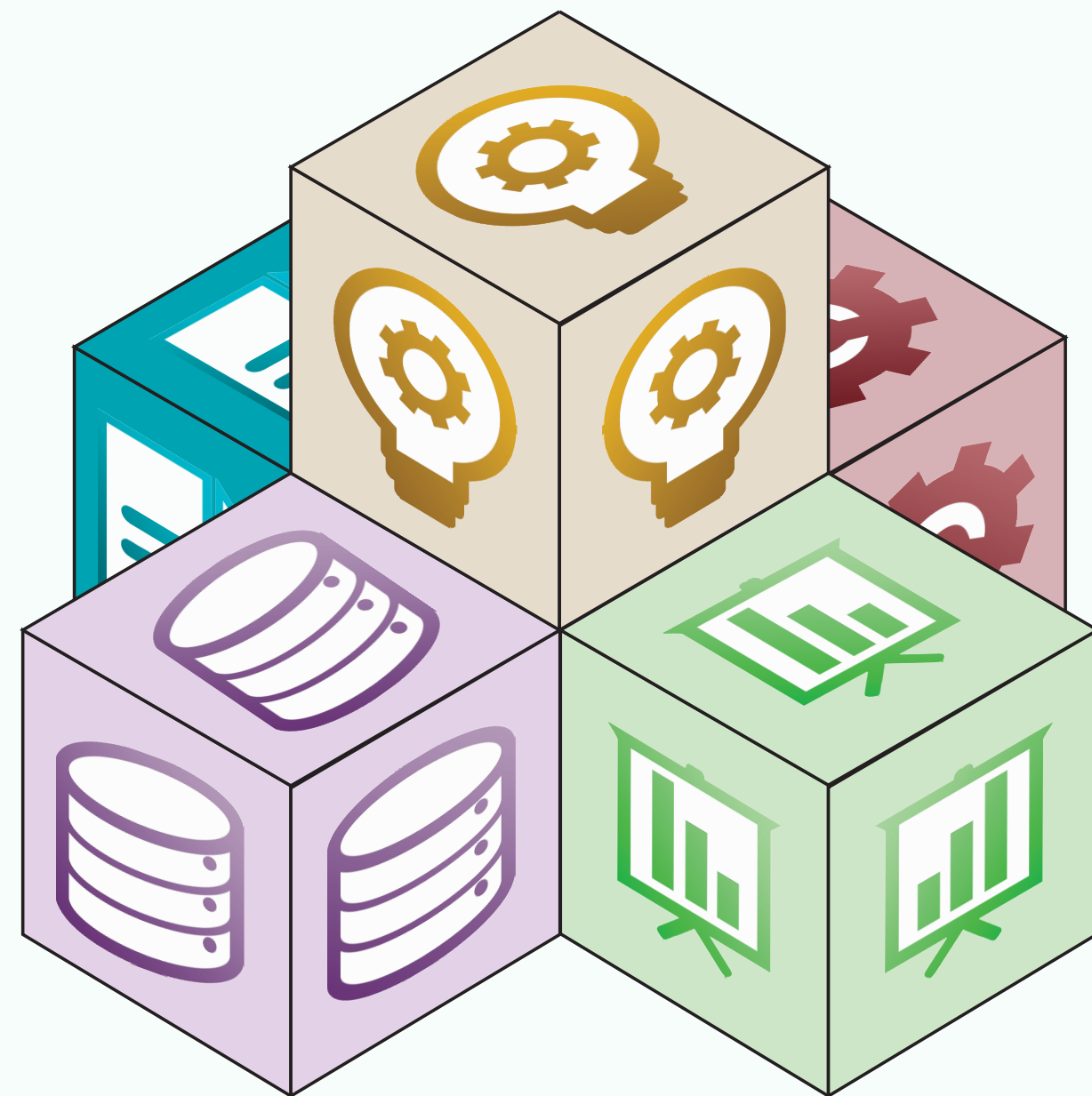
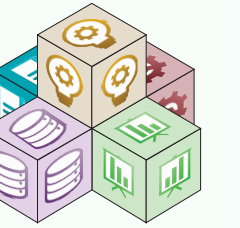
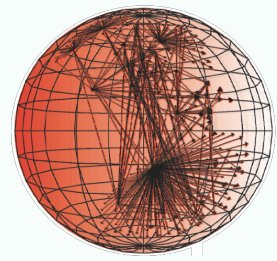


# resource catalog

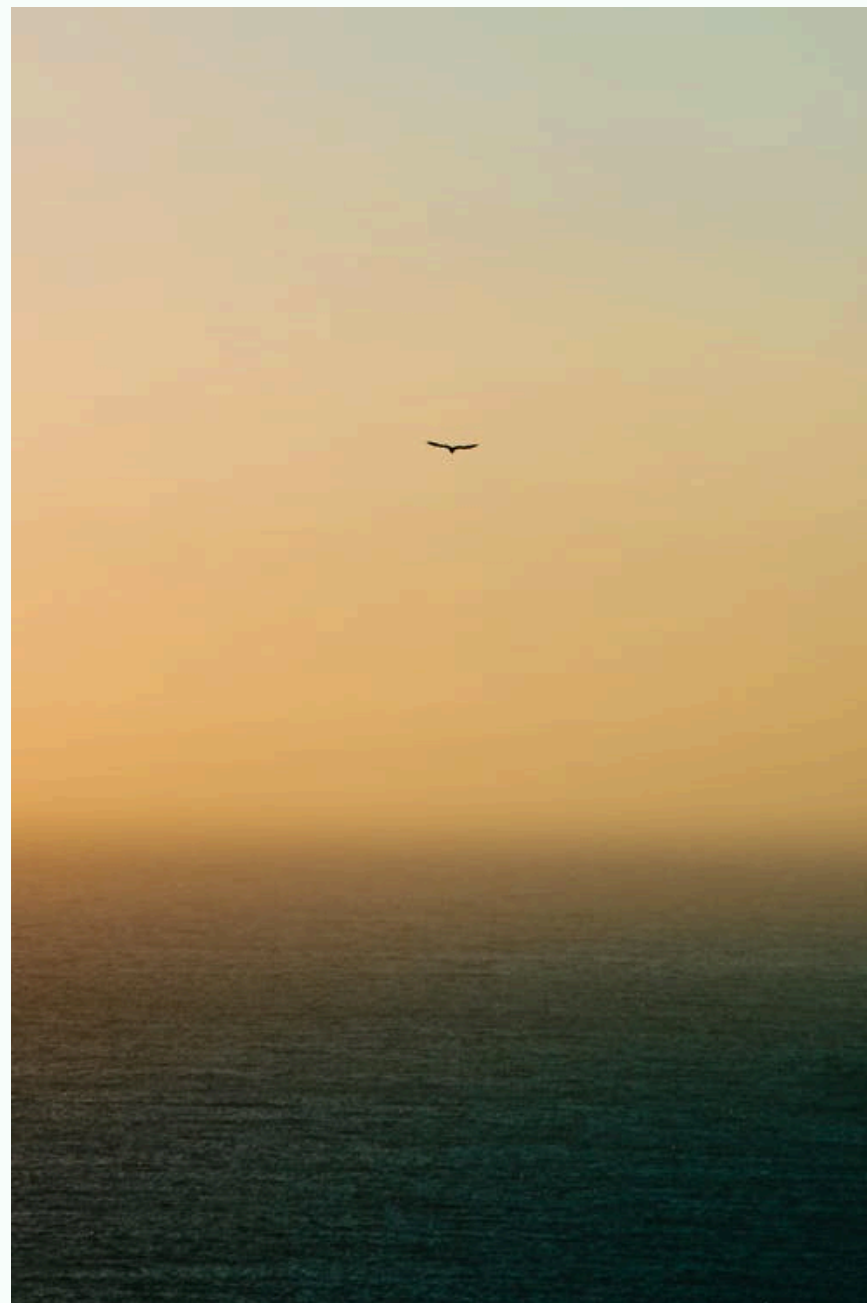


## **CAIDA's resource catalog**

[catalog.caida.org](http://catalog.caida.org)



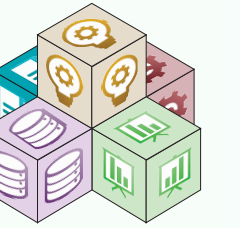
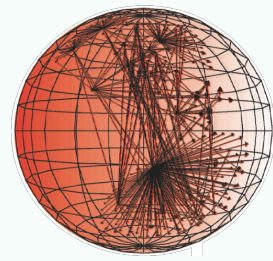
# problem: Internet infrastructure research



relevant data sets intractable for researchers to create and maintain

existing data sets have high bar to effectively use (and share)

consequence: stunted progress and low reproducibility in field

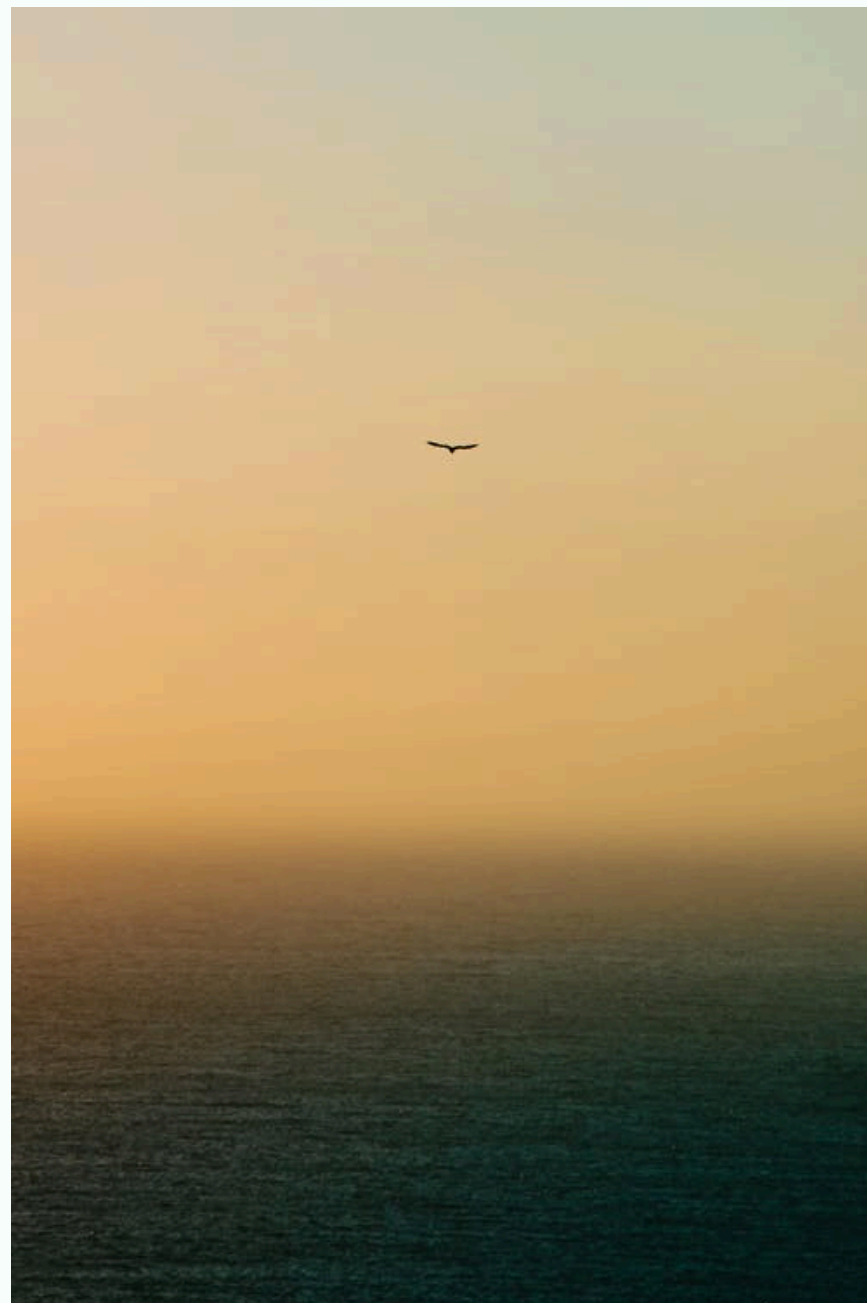


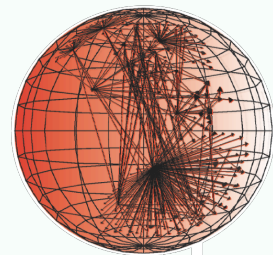
## problem: CAIDA-specific version

10-20 year longitudinal data sets  
about Internet infrastructure

researchers and students still often  
“reinvent wheel” when using it

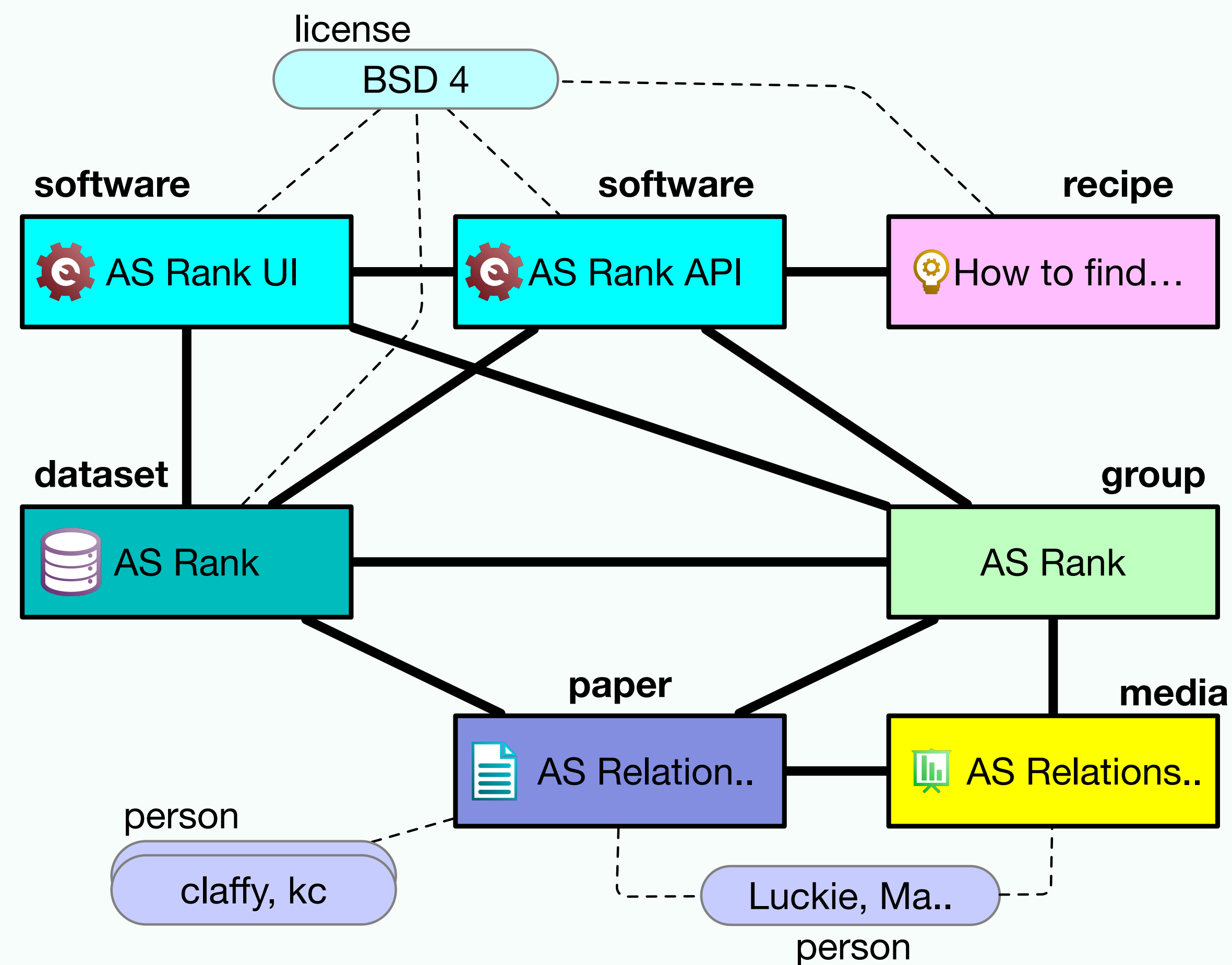
consequence: stunted progress  
and low reproducibility in field

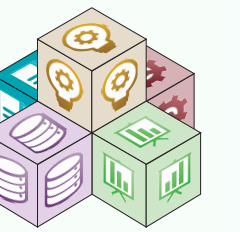
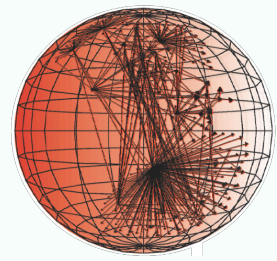




# our solution

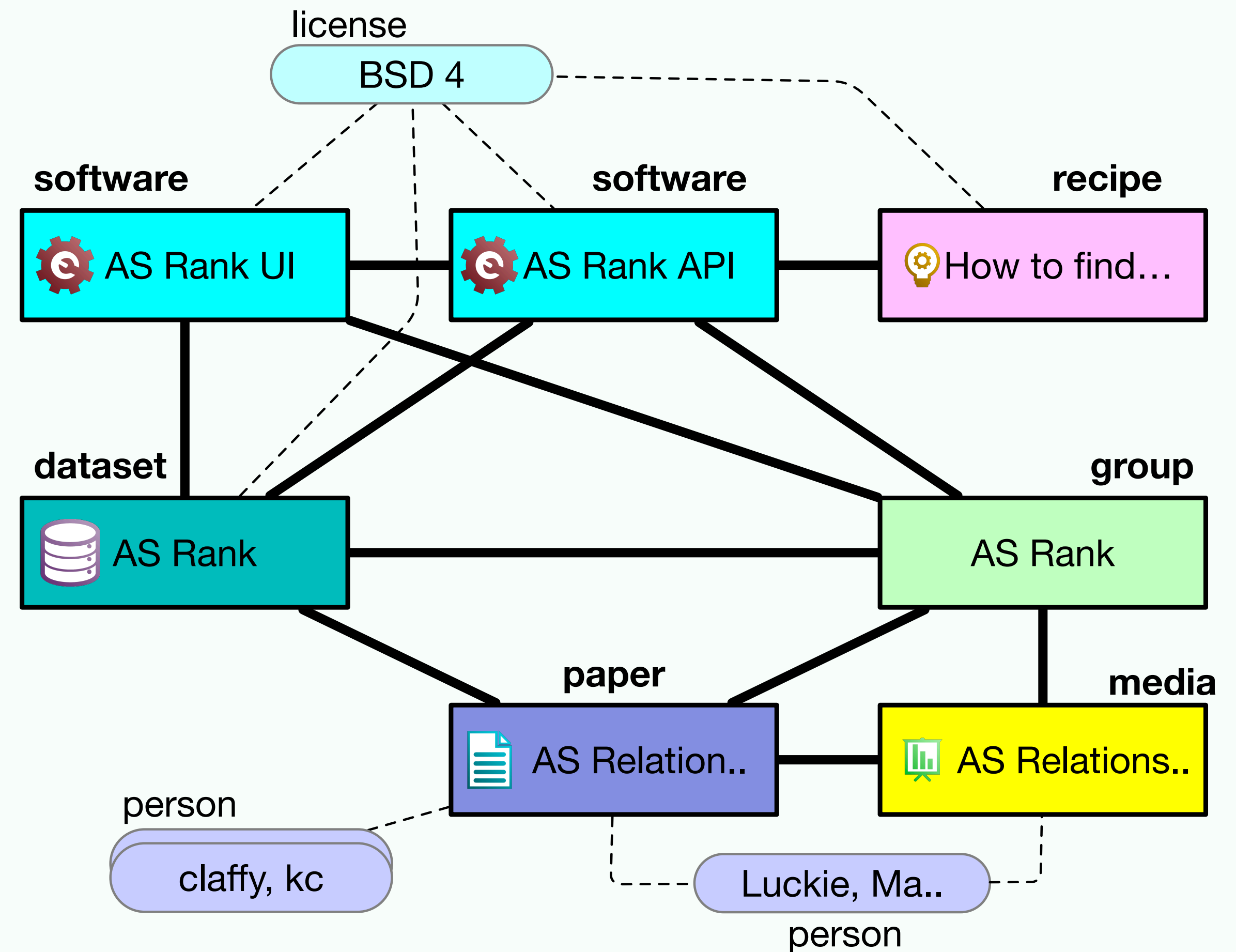
rich context  
graph



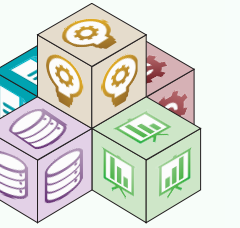
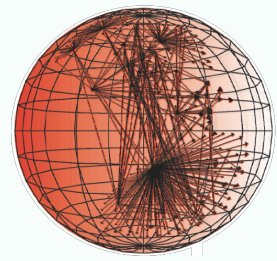


# what is a rich context graph?

a graph that  
**shows the  
relationship**  
between data  
and other  
things







# what are our goals?

discovery

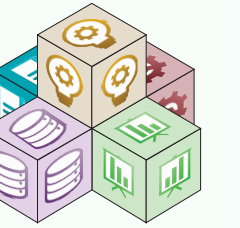
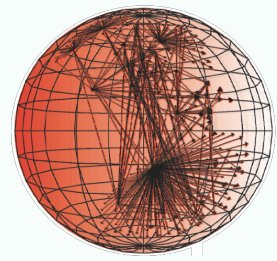
help to find data

understanding

help to understand data

inspiration

help to find new uses for data



# how does a graph do that?

discovery

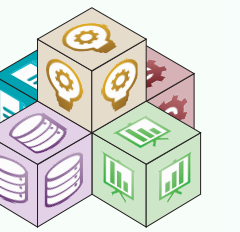
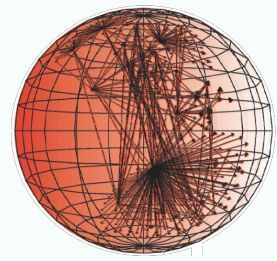
ability to search by word, tag,  
related objects

understanding

link data to sample code and  
explanations

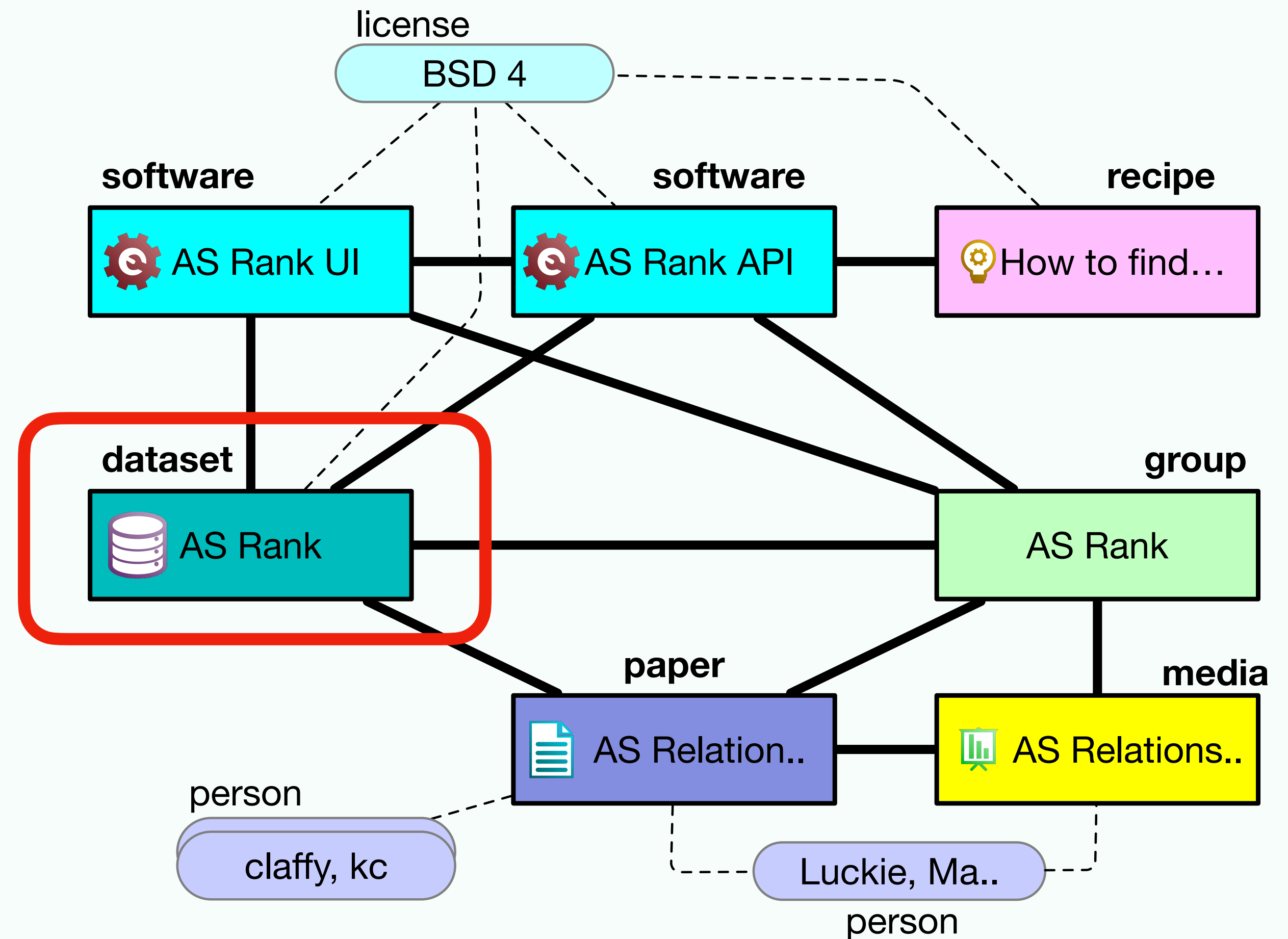
inspiration

links to existing works, helps  
inspire new works

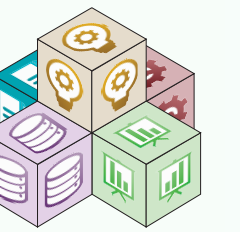
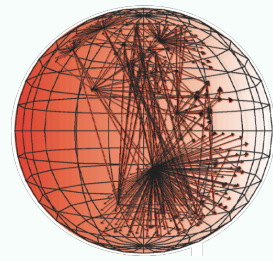


# what is in our graph?

-  dataset
-  paper
-  media
-  software
-  recipe







# what counts as a dataset?



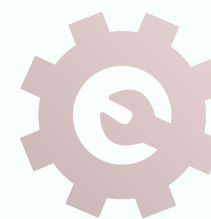
dataset



paper




media



software








recipe



### Internet eXchange Points (IXPs) Dataset

RELATED TO

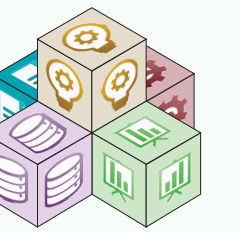
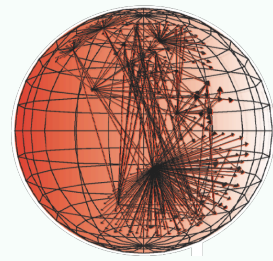


An Internet eXchange Point (IX or IXP) is a physical infrastructure used by Internet service providers (ISPs) and content delivery networks (CDNs) to exchange Internet traffic between their networks (Autonomous Systems - ASes). An IXP can be distributed and located in numerous data centers (aka faci...

[Topology](#)  
[Ixp](#)

LAST UPDATED	STATUS	RESOURCES
2018.02	Ongoing	Public

- collection of related information
- single file, database, or collection of files



# what counts as a paper?

 dataset

 paper

 media

 software

 recipe



## Residential Links Under the Weather

RELATED TO



Weather is a leading threat to the stability of our vital infrastructure. Last-mile Internet is no exception. Yet, unlike other vital infrastructure, weather's effect on last-mile Internet outages is not well understood. This work is the first attempt to quantify the effect of weather on residential...

[Measurement  
Methodology](#)  
[Data  
Routing](#)  
[Active Data Analysis](#)

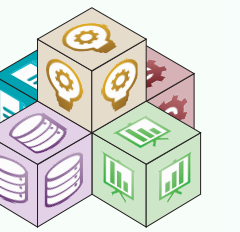
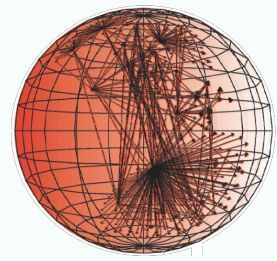
DATE PUBLISHED

2019/08/01

AUTHORS

R. Padmanabhan, A. Schulman,  
D. Levin, N. Spring

- papers, research reports, articles
- it doesn't need to be published
  - just linked to an object



# what counts as media?


 dataset

 paper

 media






 software

 recipe



### IPv4 History Visualization Interactive

RELATED TO



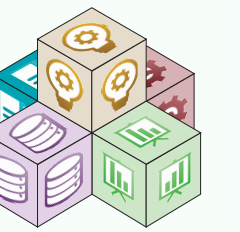
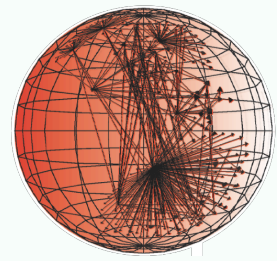
An animated slideshow presentation of the history of IPv4, using a Sankey diagram to visualize the flow of IPv4 address allocations since inception in 1977, through various address management organizations and policies.

[Ipv4 Sankey Presentation](#)

DATE PUBLISHED  
2020.08

PRESENTER

- presentations, videos, visualizations



# what counts as software?

 dataset

 paper

 media

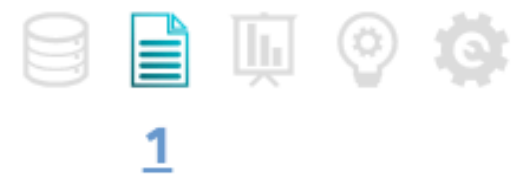
 **software**

 recipe



## Archipelago

RELATED TO



CAIDA deploys and maintains a globally distributed measurement platform we call Archipelago (Ark). We grow the infrastructure by distributing hardware measurement nodes (2nd gen. Raspberry Pi) with as much geographical and topological diversity as we can to improve our view of the global Internet. ...

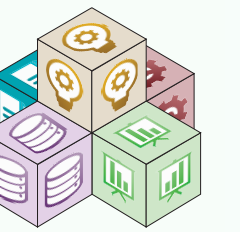
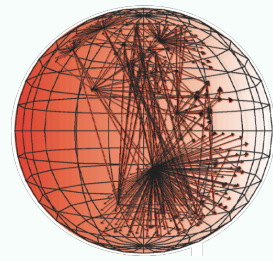
[Topology](#)  
[Active Measurement](#)

LAST UPDATED  
None Provided

RESOURCES  
Restricted

- scripts, executable, web UIs, APIs
- creates, shares, or processes data





# what counts a recipe?


 dataset

 paper






 media

 software

 recipe



Parse CAIDA's ITDK for a router's IPs, ASN, neighbors, and geographic location.

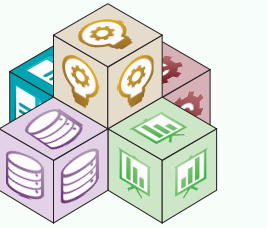
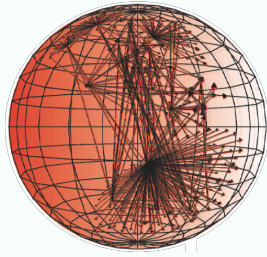
RELATED TO				
 <a href="#">1</a>				
LAST UPDATED		RESOURCES		
None Provided		N/A		

[Measurement Methodology](#)  
[Topology](#)  
[Software/Tools](#)  
[Asn](#)  
[Geolocation](#)  
[Link](#)  
[Node\\_id](#)

- overview, methodologies, code snippets
- help to quickly understand how to use the data







# overview recipe



## How to search the catalog

Overview of the more advanced catalog search features.

### search query

"types=dataset topology", "asn", "recipe=paper, recipe tag:topology"

One of the primary ways people can interact with the catalog is with a search query. A search query is an unordered collection of object ids, key value pairs, and words that returns a matching set of objects. The search query is generated from the search field. The search query can be understood to be a set of AND operations. An object matches a search query if it matches all parts of the search query. The search query is captilization insensitive.

### key/value, ids, and words

First the search field is split into tokens on white space. Tokens that contain the = character are processed as key value pairs (types=dataset). Tokens that contain a : character are processed as object ids (dataset:asrank). All other tokens are processed as words.

type	meannging	examples
key/value (key)=(value)	a key and a value pair	types=dataset,recipe
id (type):(shortName)	an object id	dataset:asrank , tag:asn
word	anything that doesn't match the above	

First the words are processed to find the set of objects that contain all the words in a combination of their fields and placed into the matching set. If no words are provided, all objects considered to match set.

If object ids are found in the search query, objects are removed from the matching set if they do not have a direct link to all the objects with a matching object id. It's important to note that an object's id is not its type and name, but its type and shortName. For example, the dataset "How to Parse CYMRU Bogan Data"'s short name is "bogons" so it's id is "dataset:bogons".

### key/value pairs

Currently we only support the key word types. If types is provided, then it's value is split on the , character into a set of types and stored in the types set.

If the types key is not provided, then all types are placed into the types set.

Objects are removed from the matching set if their type is not in the types set.

key	value
types	comma separated list of target types

introduction

breakdown



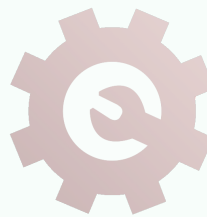
dataset



paper



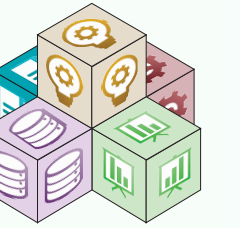
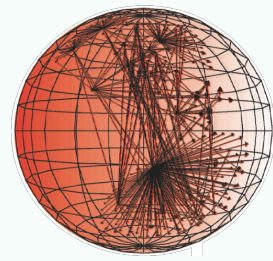
media



software



recipe



# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

The screenshot shows a data catalog entry for 'The IPv4 Routed /24 Topology Dataset'. The title is highlighted with a red box. Below the title is a 'Public/Restricted' button. The description states: 'This dataset contains information useful for studying the topology of the Internet. Data is collected by a globally distributed set of Ark monitors. The monitors use team-probing to distribute the work of probing the destinations among the available monitors. We collect data by sending scamper probes continuously to destination IP addresses. Destinations are selected randomly from each routed IPv4 /24 prefix on the Internet such that a random address in each prefix is probed approximately every 24 hours (one probing cycle). Because team-probing distributes the probing work across all monitors, a single destination /24 will be probed by only one monitor in each probing cycle.' The tags are 'active', 'topology', 'traceroute', and 'ipv4'. On the right, there is a 'Related Objects Quick Glance' section with a list of related datasets: 'Vela', 'scamper', '2013 AS Core: IPv4 vs IPv6 comparison', '2010 AS Core: IPv4 vs IPv6 comparison', 'Macroscopic Internet Topology Data Kit (ITDK)', and '2009 AS Core: IPv4 vs IPv6 comparison'.

**The IPv4 Routed /24 Topology Dataset**

Public/Restricted

This dataset contains information useful for studying the topology of the Internet. Data is collected by a globally distributed set of Ark monitors. The monitors use team-probing to distribute the work of probing the destinations among the available monitors. We collect data by sending scamper probes continuously to destination IP addresses. Destinations are selected randomly from each routed IPv4 /24 prefix on the Internet such that a random address in each prefix is probed approximately every 24 hours (one probing cycle). Because team-probing distributes the probing work across all monitors, a single destination /24 will be probed by only one monitor in each probing cycle.

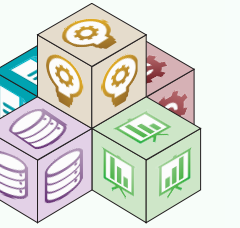
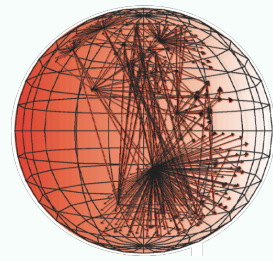
TAGS: [active](#) [topology](#) [traceroute](#) [ipv4](#)

Related Objects Quick Glance

- [Vela](#)
- [scamper](#)
- [2013 AS Core: IPv4 vs IPv6 comparison](#)
- [2010 AS Core: IPv4 vs IPv6 comparison](#)
- [Macroscopic Internet Topology Data Kit \(ITDK\)](#)
- [2009 AS Core: IPv4 vs IPv6 comparison](#)

- do not need to be unique
- The IPv4 Routed /24 Top...
- How to find an IP's ASN.





# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

**The IPv4 Routed /24 Topology Dataset**

`dataset:ipv4_routed_24_topology`

Public/Restricted

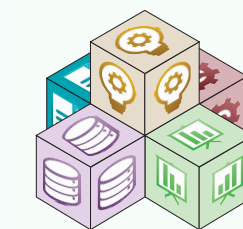
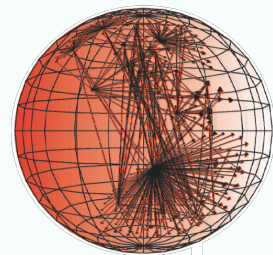
This dataset contains information useful for studying the topology of the Internet. Data is collected by a globally distributed set of Ark monitors. The monitors use team-probing to distribute the work of probing the destinations among the available monitors. We collect data by sending scamper probes continuously to destination IP addresses. Destinations are selected randomly from each routed IPv4 /24 prefix on the Internet such that a random address in each prefix is probed approximately every 24 hours (one probing cycle). Because team-probing distributes the probing work across all monitors, a single destination /24 will be probed by only one monitor in each probing cycle.

TAGS: [active](#) [topology](#) [traceroute](#) [ipv4](#)

**Related Objects Quick Glance**

- [Vela](#)
- [scamper](#)
- [2013 AS Core: IPv4 vs IPv6 comparison](#)
- [2010 AS Core: IPv4 vs IPv6 comparison](#)
- [Macroscopic Internet Topology Data Kit \(ITDK\)](#)
- [2009 AS Core: IPv4 vs IPv6](#)

- object's (type) : (short name)
- case incentive, [\_a-z]+  
software:asrank\_ui  
paper:as\_relationships



# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

**The IPv4 Routed /24 Topology Dataset**  
dataset:ipv4\_routed\_24\_topology

**Public/Restricted**

This dataset contains information useful for studying the topology of the Internet. Data is collected by a globally distributed set of Ark monitors. The monitors use team-probing to distribute the work of probing the destinations among the available monitors. We collect data by sending scamper probes continuously to destination IP addresses. Destinations are selected randomly from each routed IPv4 /24 prefix on the Internet such that a random address in each prefix is probed approximately every 24 hours (one probing cycle). Because team-probing distributes the probing work across all monitors, a single destination /24 will be probed by only one monitor in each probing cycle.

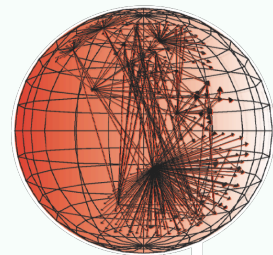
TAGS: [active](#) [topology](#) [traceroute](#) [ipv4](#)

**Related Objects Quick Glance**

- [Vela](#)
- [scamper](#)
- [2013 AS Core: IPv4 vs IPv6 comparison](#)
- [2010 AS Core: IPv4 vs IPv6 comparison](#)
- [Macroscopic Internet Topology Data Kit \(ITDK\)](#)
- [2009 AS Core: IPv4 vs IPv6](#)

**external links**  
links to a paper's PDF  
download directory





# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

**The IPv4 Routed /24 Topology Dataset**  
dataset:ipv4\_routed\_24\_topology

Public/Restricted

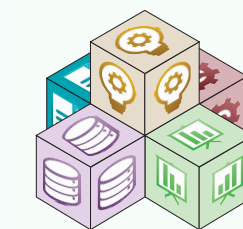
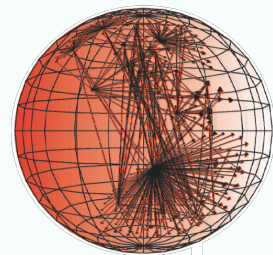
This dataset contains information useful for studying the topology of the Internet. Data is collected by a globally distributed set of Ark monitors. The monitors use team-probing to distribute the work of probing the destinations among the available monitors. We collect data by sending scamper probes continuously to destination IP addresses. Destinations are selected randomly from each routed IPv4 /24 prefix on the Internet such that a random address in each prefix is probed approximately every 24 hours (one probing cycle). Because team-probing distributes the probing work across all monitors, a single destination /24 will be probed by only one monitor in each probing cycle.

TAGS: [active](#) [topology](#) [traceroute](#) [ipv4](#)

Related Objects Quick Glance 3 11 2

- [Vela](#)
- [scamper](#)
- [2013 AS Core: IPv4 vs IPv6 comparison](#)
- [2010 AS Core: IPv4 vs IPv6 comparison](#)
- [Macroscopic Internet Topology Data Kit \(ITDK\)](#)
- [2009 AS Core: IPv4 vs IPv6](#)

• short description of the object

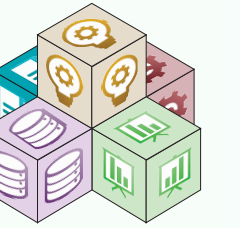
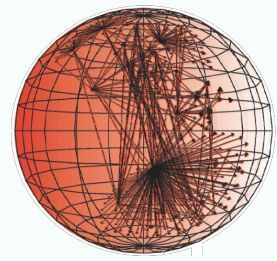


# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

The screenshot shows the dataset page for 'The IPv4 Routed /24 Topology Dataset'. The page includes a description of the dataset, a 'Public/Restricted' button, and a 'TAGS' section. The 'TAGS' section is highlighted with a red box, showing the following tags: `active`, `topology`, `traceroute`, and `ipv4`. A callout box is overlaid on the screenshot, containing the following text:

- all objects can be tagged
- including tags
  - tag:topology
  - tag:ipv4\_path



# which fields are common to all objects?

- name
- id
- resources
- description
- tags
- links

- summary of related objects
- more detailed related objects

Related Objects Quick Glance 3 11 2

- [Vela](#)
- [scamper](#)
- [2013 AS Core: IPv4 vs IPv6 comparison](#)
- [2010 AS Core: IPv4 vs IPv6 comparison](#)
- [Macroscopic Internet Topology Data Kit \(ITDK\)](#)
- [2009 AS Core: IPv4 vs IPv6 comparison](#)
- [IPv4 DNS names](#)

## Related Objects Detailed View

RESULTS FOR: [dataset:ipv4\\_routed\\_24\\_topology](#)



[Vela](#)

RELATED TO



1



1

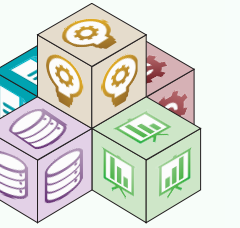
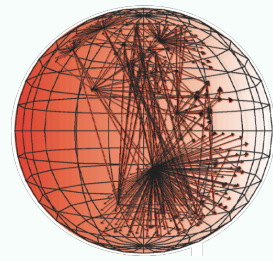
Vela provides an easy way for researchers to conduct on-demand topology measurements on Ark. Users can conduct ping and traceroute measurements in IPv4 and IPv6 using ICMP, UDP, or TCP from any Ark monitor. There are two interfaces to Vela, a command-line interface and a web-based interface. The co...

LAST UPDATED  
2020.08

RESOURCES  
UIAPIAPI MIDARAPI  
Aliasqpython Code

[Topology](#)  
[Geolocation](#)  
[UI](#)  
[API](#)





# where should you start?

search field

CAIDA  
RESOURCE CATALOG

Explore a Catalog of  
CAIDA Research  
and Knowledge

Papers, Presentations, Datasets, Tools, Questions...

SEARCH

Search through a library of Publications, Datasets, Software, Solutions, and Media  
to view descriptions, metadata, related resources, and resource links

- white space delimited
- case insensitive
- unordered collections:  
ids, key/value, words

Datasets  
30 ENTRIES

CAIDA collects several types of data, and makes this data available to the research community while protecting the privacy of individual organizations who provide data or network access.

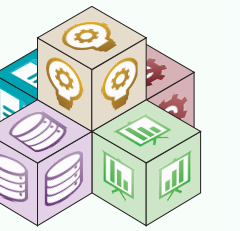
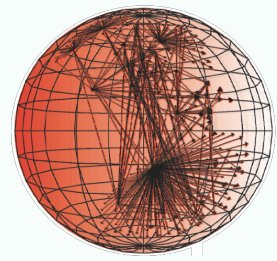
BROWSE DATASETS »

BROWSE MEDIA »

BROWSE PAPERS »

BROWSE RECIPES »

BROWSE SOFTWARE »



# what are ids, key/values, words?

## ids

**(type):(value)** extracts all objects linked to id

- filters on topology of the context graph

tag:bgp recipe:how\_to\_parse\_bgp

## key/value

**types=(types)** filters by object type

- comma-separated list of object types

types=dataset types=recipe,dataset

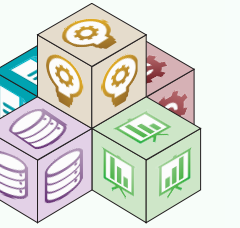
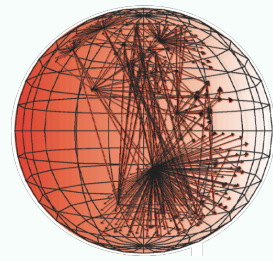
## words

all remaining tokens

- words in object fields

asn bgp huffaker ucsd ip security encrypted





# how you can help

use

we built it, please use it

[catalog.caida.org](https://catalog.caida.org)

share

with your friends and colleagues



contribute

your recipes, datasets, and papers

[github.com/CAIDA/catalog-data](https://github.com/CAIDA/catalog-data)