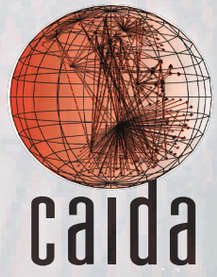
TOWARD REALTIME VISUALIZATION OF GARBAGE

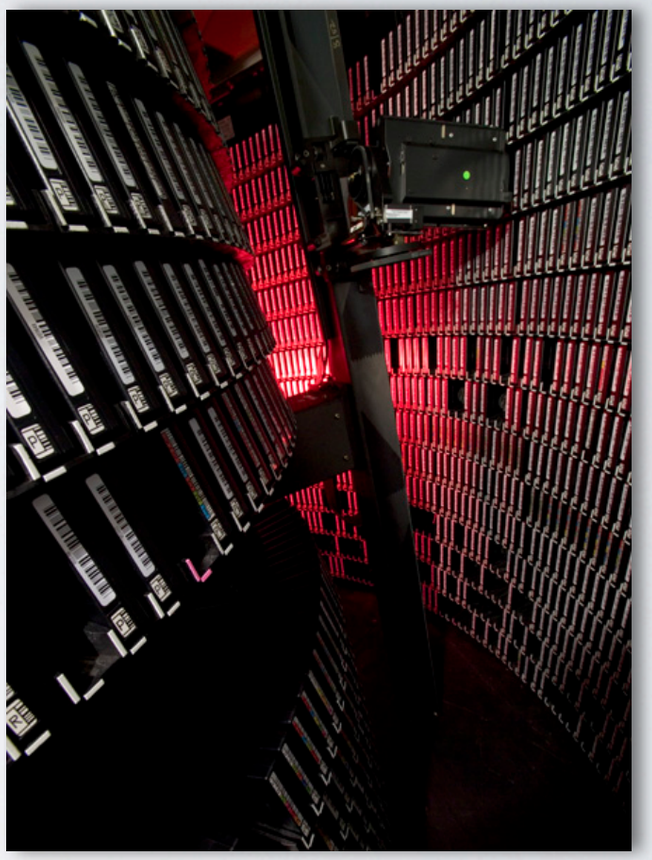
Alistair King, Alberto Dainotti
alistair@caida.org
CAIDA, UC San Diego



OUR PROBLEM

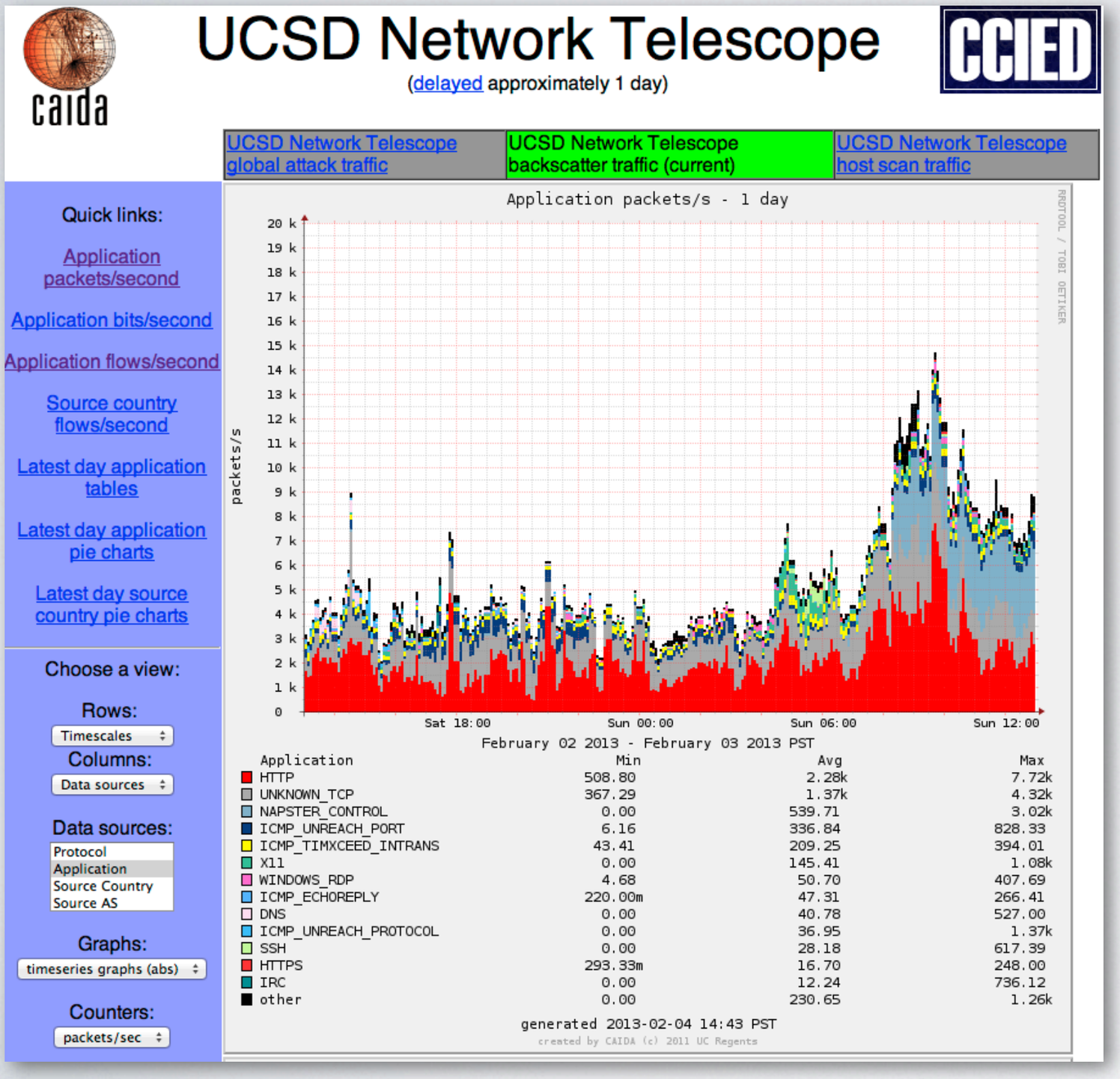
(well, one of them...)

- /8 darknet - full packet traces 24/7
- Huge amount of archived and streaming data
  - 150TB on tape, 4.5TB new data each month
- Asking 'basic' analysis questions is **time consuming**
  - e.g. plotting a graph of unique source IPs for TCP-80 over a year
- Getting **realtime** insight into the data is **nearly impossible**



caida

WHAT WE HAVE

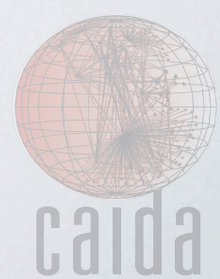


- Daily report generation
- Traffic is classified into:
  - Attack
  - Backscatter
  - Host scan
- Several metrics:
  - Protocol
  - Application
  - Country
  - AS

GOALS

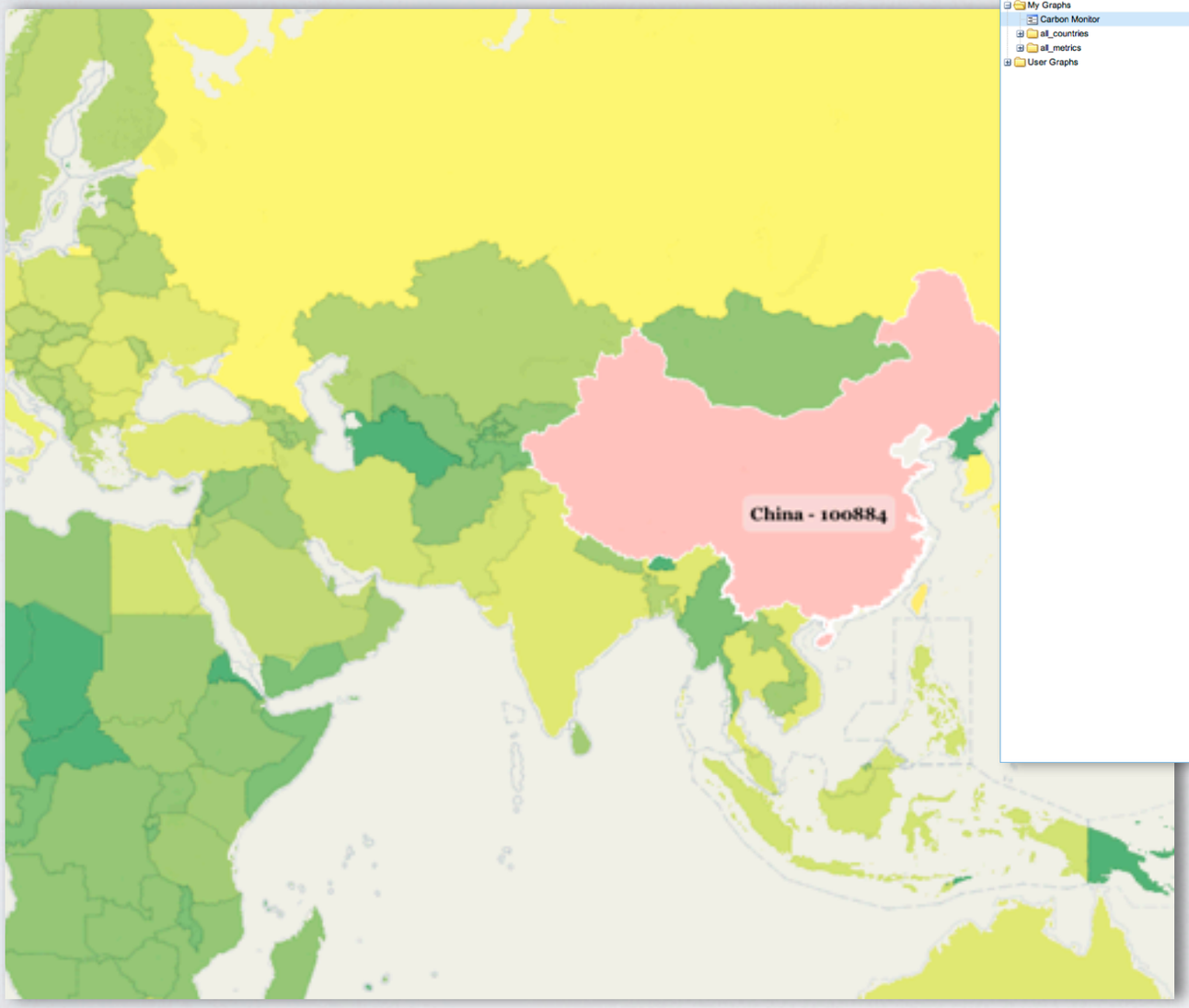
(visualizing 150TB of garbage is easy, right?)

- **Dynamic** query interface
- **Interactive** exploration of data
- Easily **extensible** for new metrics and dimensions
- **Minimize latency** between capture and viz
- Scalable to **millions of metrics**, and **years of data**



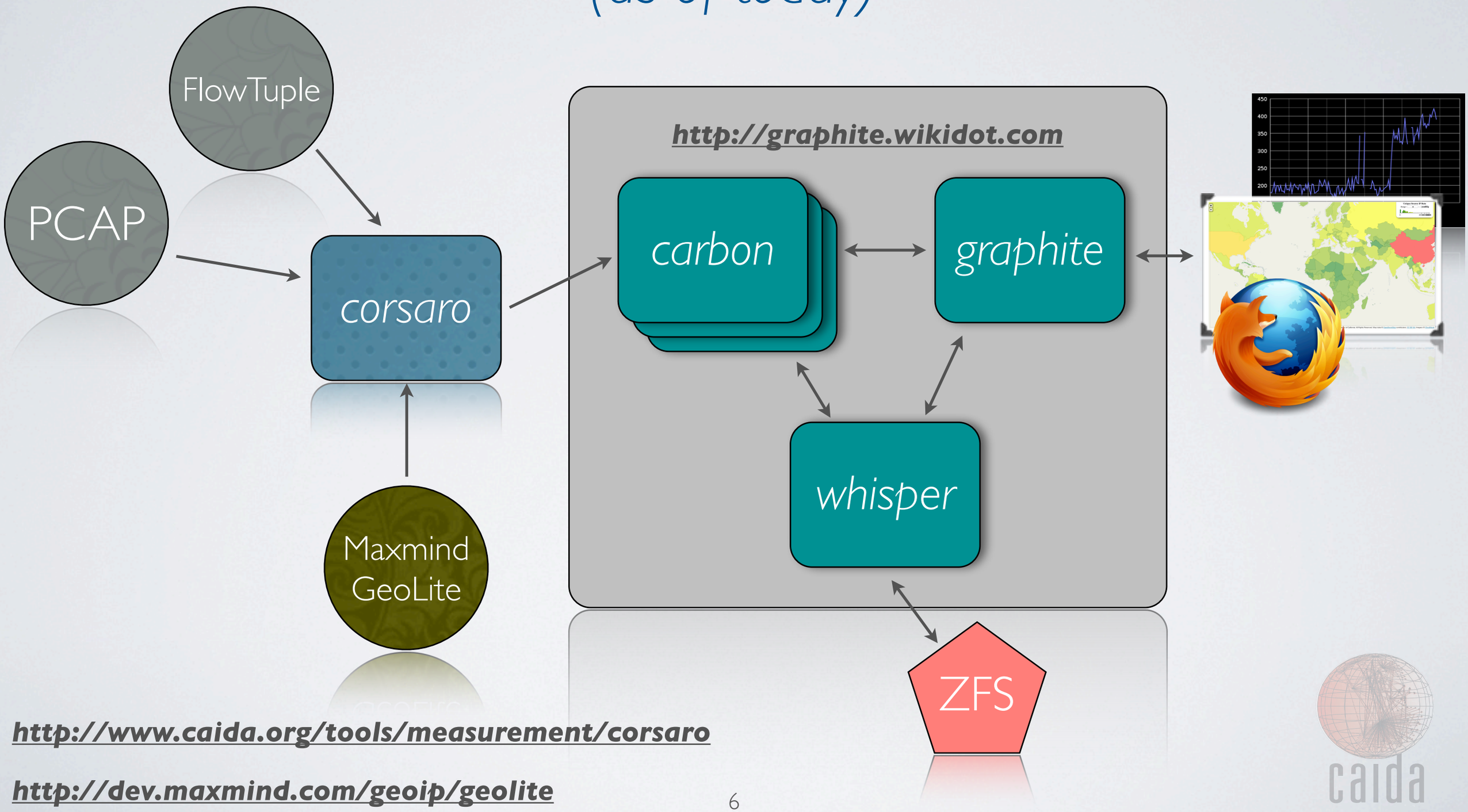
<DEMO>

(fingers crossed...)



PIPELINE

(as of today)



http://www.caida.org/tools/measurement/corsaro

http://dev.maxmind.com/geoip/geolite

NOW WHAT?

(lessons learned and next steps)

- Substantially expand the metrics/dimensions captured
  - RRDtool style databases (incl. *whisper*) are not scalable
  - Experimenting with integrating *tsdb*

    http://luca.ntop.org/tsdb.pdf
- Connect *corsaro* to the capture interface for ~1 min viz latency
- Extend world map viz (*crosslet)* to support time-series data

  http://sztanko.github.com/crosslet/

COME TALK TO ME

(please)

- Do you have a similar problem?

- Do you have any suggestions?

- For more info about our telescope data, see *http://www.caida.org/projects/network_telescope/*

- Coming soon: an educational dataset using telescope data