Internet Atlas: A Geographical Database of the Physical Internet

Active Internet Measurement Systems Workshop (AIMS)
February 6-8, 2013

Ram Durairajan
Computer Sciences
University of Wisconsin

Motivation

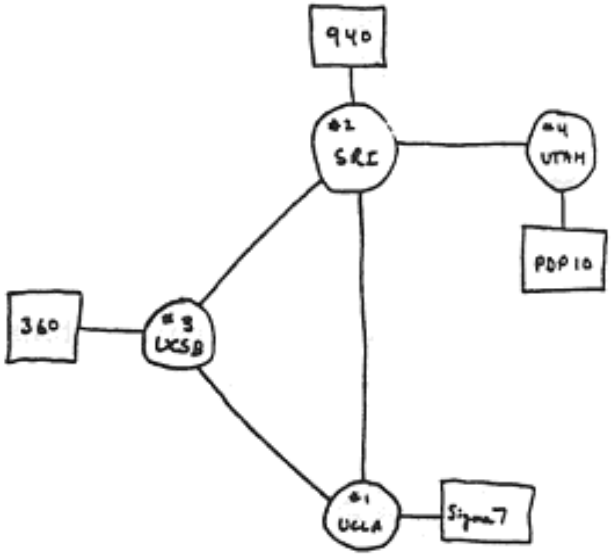


FIGURE 6.2 Drawing of 4 Node Network
(Courtesy of Alex McKenzie)

Objectives of our work

- **Create and maintain a comprehensive catalog of the *physical Internet***
  - **Geographic locations of nodes (buildings that house PoPs, IXPs etc.) and links (fiber conduits)**
- **Deploy portal for visualization and analysis**
- **Extend with relevant related data**
  - **Active probes, BGP updates, Twitter, weather, etc.**
- **Apply maps to problems of interest**
  - **Robustness, performance, security**

Related work

- Many prior Internet mapping efforts
  - S. Gorman studies from early 2000's
  - CAIDA
  - DIMES
- Commercial activities
  - TeleGeography
  - Renesys
  - Lumeta
- Internet Topology Zoo

Compiling a physical repository

- **Step #1: Identification**
  - **Utilize *search* to find maps of physical locations**
- **Step #2: Transcription**
  - **Multiple methods to automate data entry**
- **Step #3: Verification**
  - **Ensure that data reflects latest network maps**
- **Our hypothesis is that physical sites are limited in number and fixed in location**
  - **But the raw number is still large!**

Challenges

- **Accuracy**
  - **How accurate are the node locations?**
  - **How accurate are the link paths and connections?**
- **Completeness**
  - **How much of the physical Internet is in the catalog?**
- **Varying data formats**
  - **requires varying approaches for processing**
- **Verification problems**
  - **networks change, data entry errors due to manual annotations**

Internet Atlas @ UW

- **Effort began in September ’11**
  - **Capture everything from maps discovered by search**
  - **Use all relevant data sources (ISP maps, colocation, data centers, NTP, traceroute, etc.)**
- **Data extraction tools**
- **Comprehensive database**
  - **Developed using MySQL**
- **Alpha web portal – http://atlas.wail.wisc.edu**
  - **Includes ArcGIS for visualization and analysis**

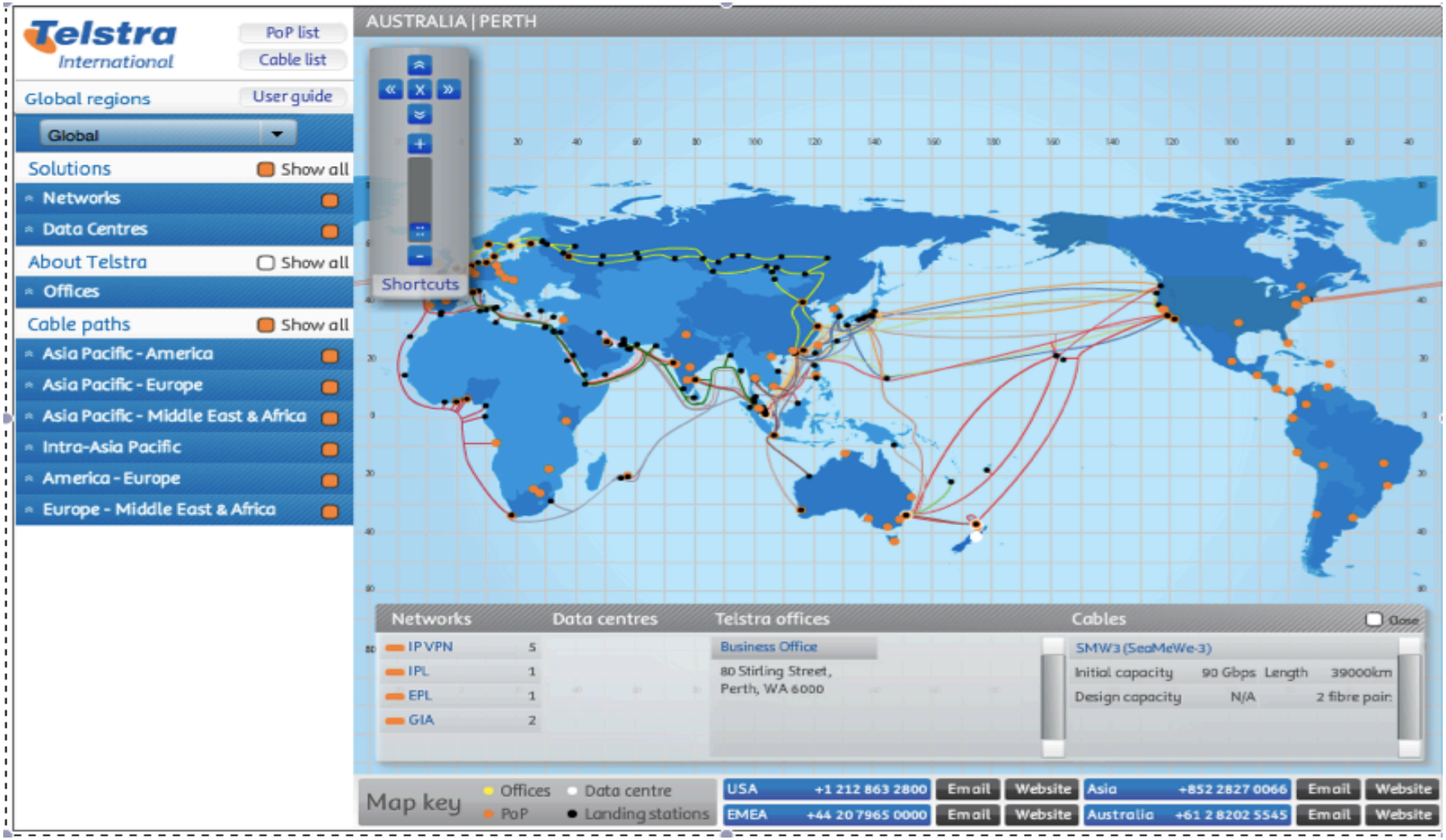
Current DB

- Number of networks: 372
- Number of tier 1 networks: 10 (all)
- Number of data centers: 2,179
- Number of NTP servers: 744
- Number of traceroute servers: 221
- Number and type of other nodes: IXP (358), DNS root (282)
- Total number of nodes: 13,734
- Number of unique locations of nodes: 7,932
- Maximum overlap at any one node: 90
- Total number of links: 13,228

Identifying relevant data

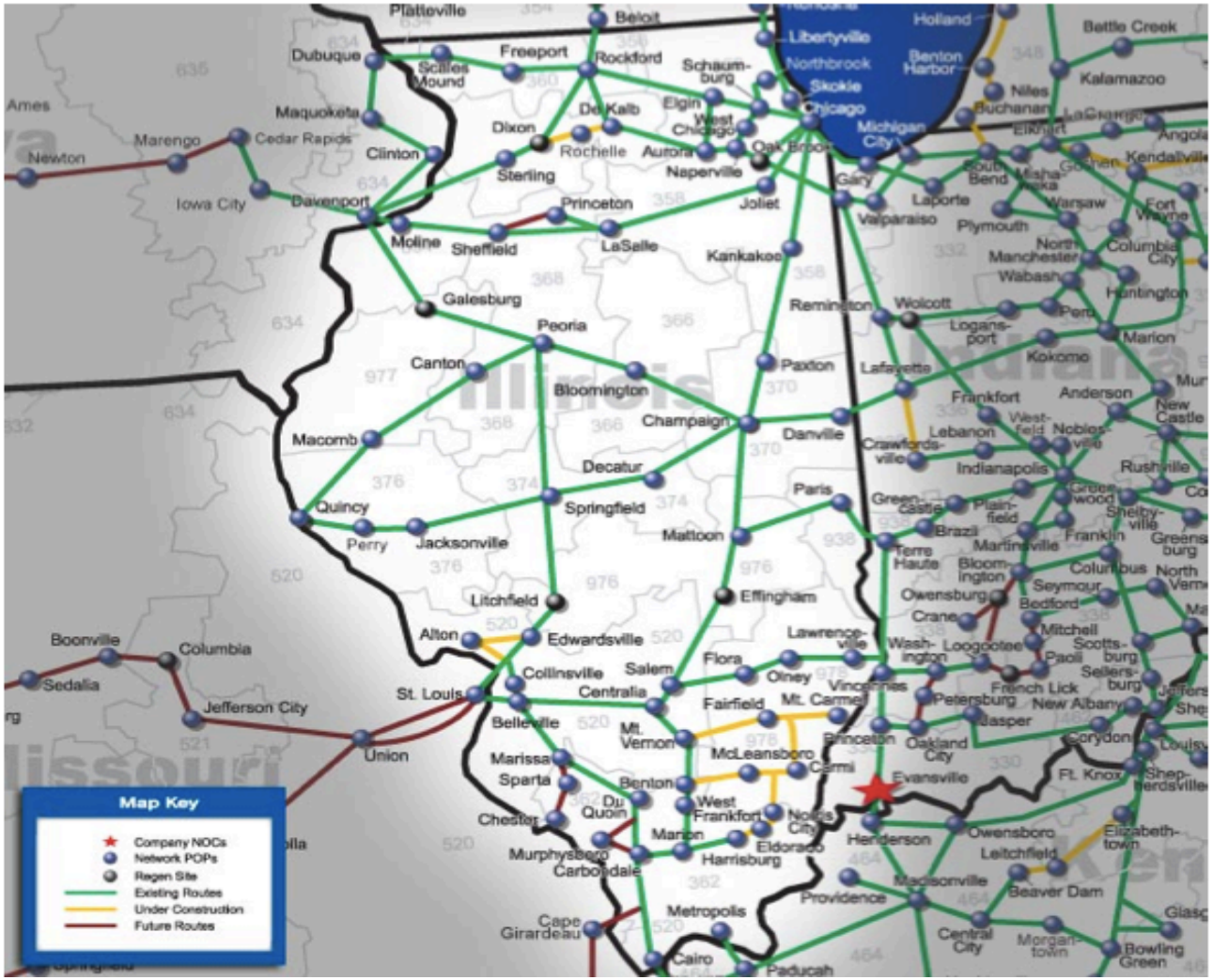
- **Internet search reveals significant information**
  - **ISP's and data center hosts routinely publish maps and locations of their infrastructure**
  - **Other elements such as NTP list precise locations**
- **Creating a corpus of search terms**
  - **Geography is important**
- **Timely representations require repetition**

Example: Telstra world wide



Example: Sprint IP network (US)

Example: Regional fiber



Illinois POP List

ALTON
Address: 1805 Washington Ave Zip: 62002
Type: CO Status: FUTURE
CLLI: ALTNILAK

BELLEVILLE
Address: 211 Kretschmer Ave Zip: 62220
Type: CO Status: ACTIVE
CLLI: BLVLILAD

BLOOMINGTON
Address: 110 E Monroe St Zip: 61701
Type: CO Status: ACTIVE
CLLI: BLTNILXD

Address: 110 E Monroe St Zip: 61701
Type: CO Status: DOUBLE
CLLI: BLTNILXD

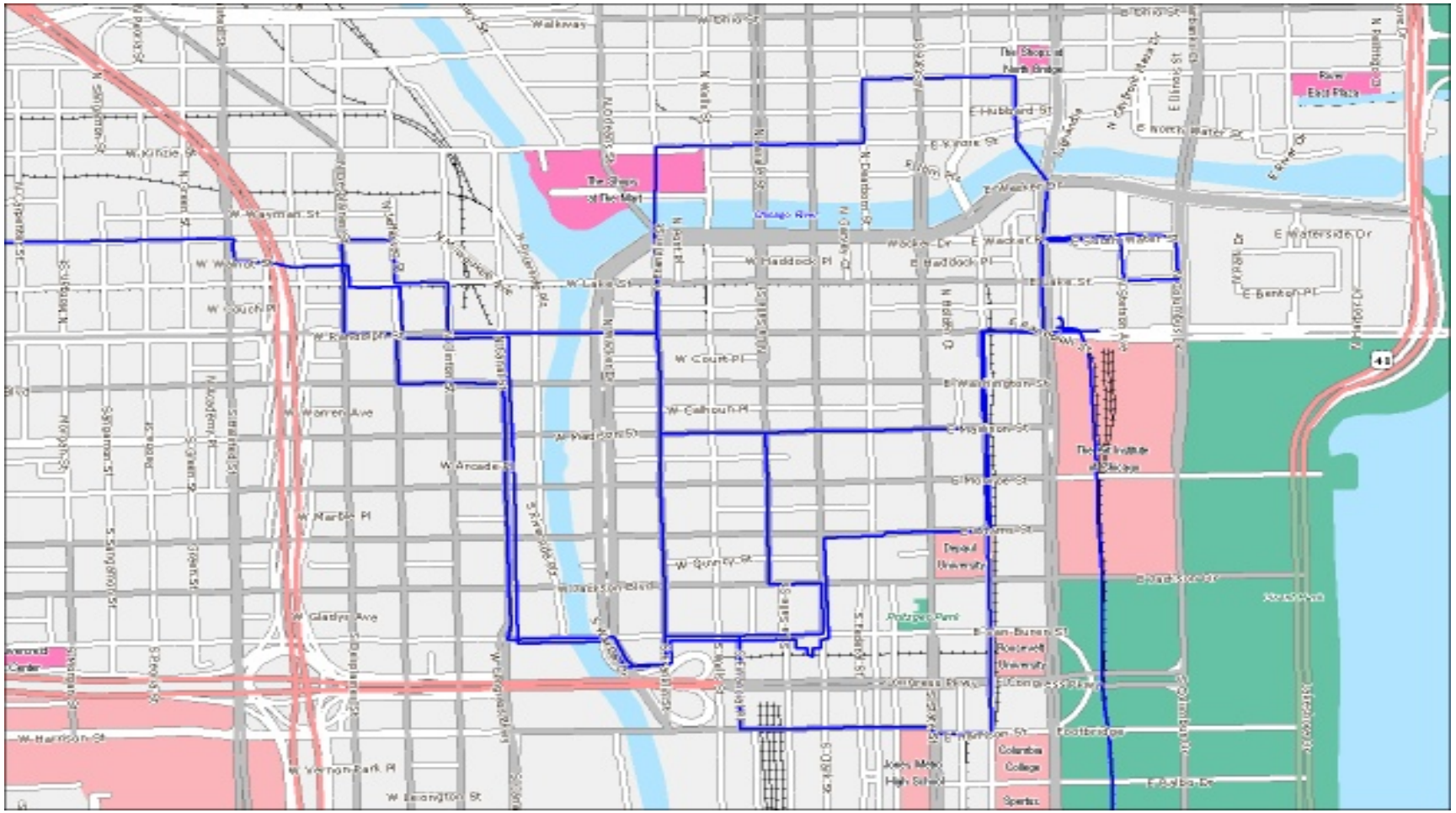
CAIRO
Address: 221 15th St Zip: 62914
Type: CO Status: ACTIVE
CLLI: CAIRILCF

CANTON
Address: 75 W Pine St Zip: 61520
Type: CO Status: ACTIVE
CLLI: CNTNILCN

CARBONDALE
Address: 208 W Monroe St Zip: 62901
Type: CO Status: ACTIVE
CLLI: CRDLILXE

CARMI
Address: 200 W Cherry St Zip: 62821

Example: Metro fiber maps



Automating transcription

- **Web pages contain Internet resource information in a variety of formats**
  - **Text, flash, images, Google maps-based, etc.**
- **Our goal is to extract information and enter it into our DB *automatically***
  - **Requires identification of relevant page**
- **Library of parsing scripts for various formats**
- **Sometimes manual entry and annotation is necessary**

Geo-coding node locations

- **Physical locations of nodes from search**
  - **Lat/Lon**
  - **Street address**
  - **City**
- **All locations decomposed in DB to Lat/Lon**
  - **Google geocoder**
  - **http://maps.googleapis.com/maps/api/geocode/xml?address="+address+"&sensor=false**

Geo-accurate link transcription

- Transcribing geographic information for links is much more challenging than for nodes
- Step #1: Copy images
  - Max zoom required for max accuracy
- Step #2: Image patching via feature matching
- Step #3: Link image extraction from base map
- Step #4: Geographic projection
  - Key step uses ArcGIS registration functionality
- Step #5: Link vectorization

Structure in link maps

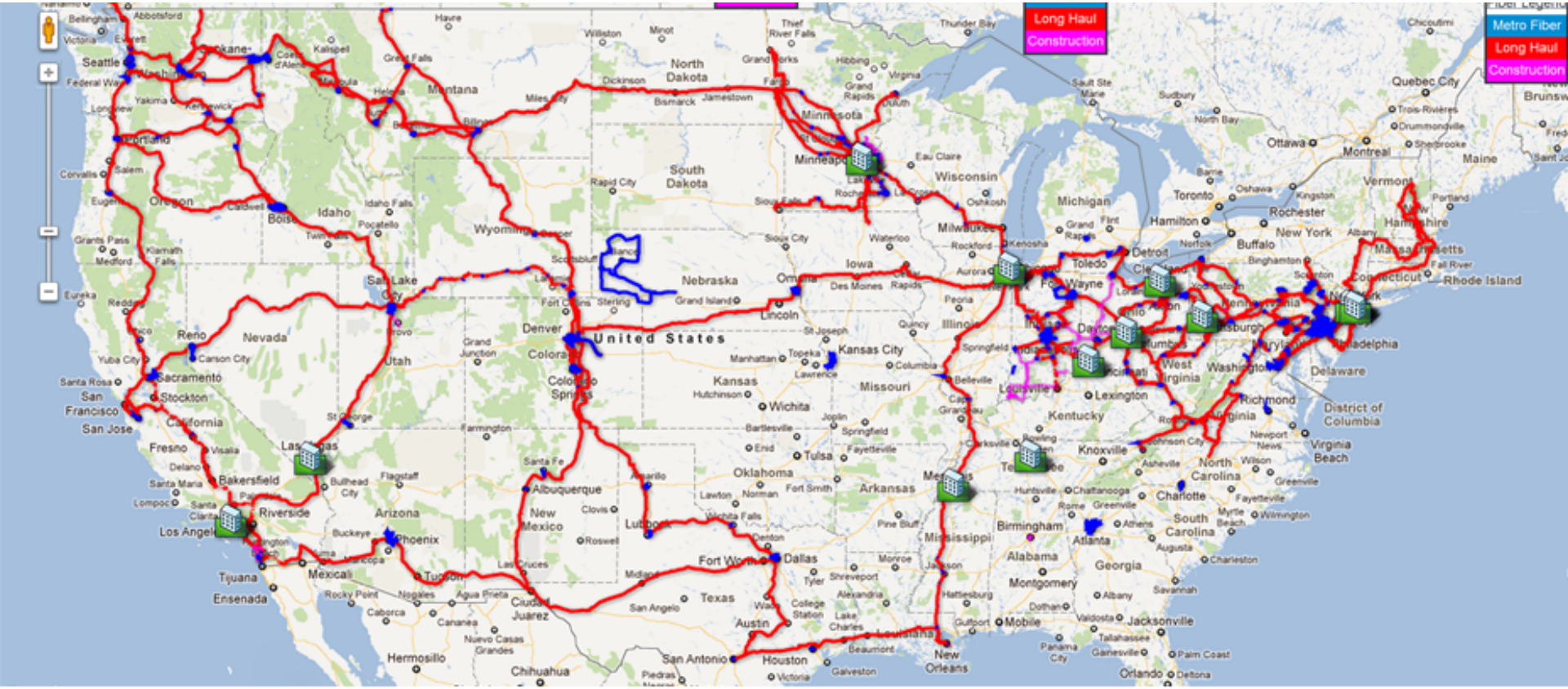
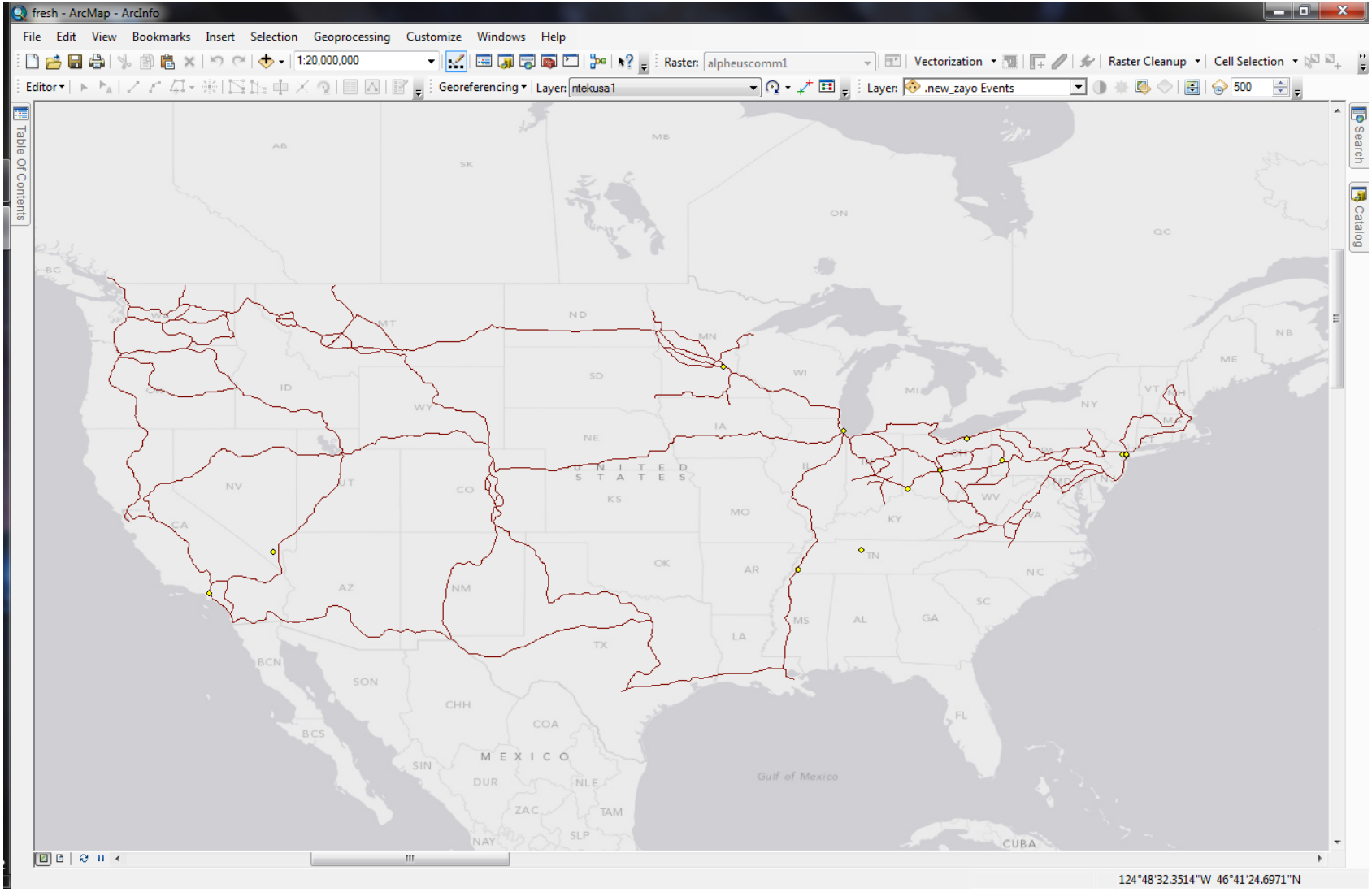
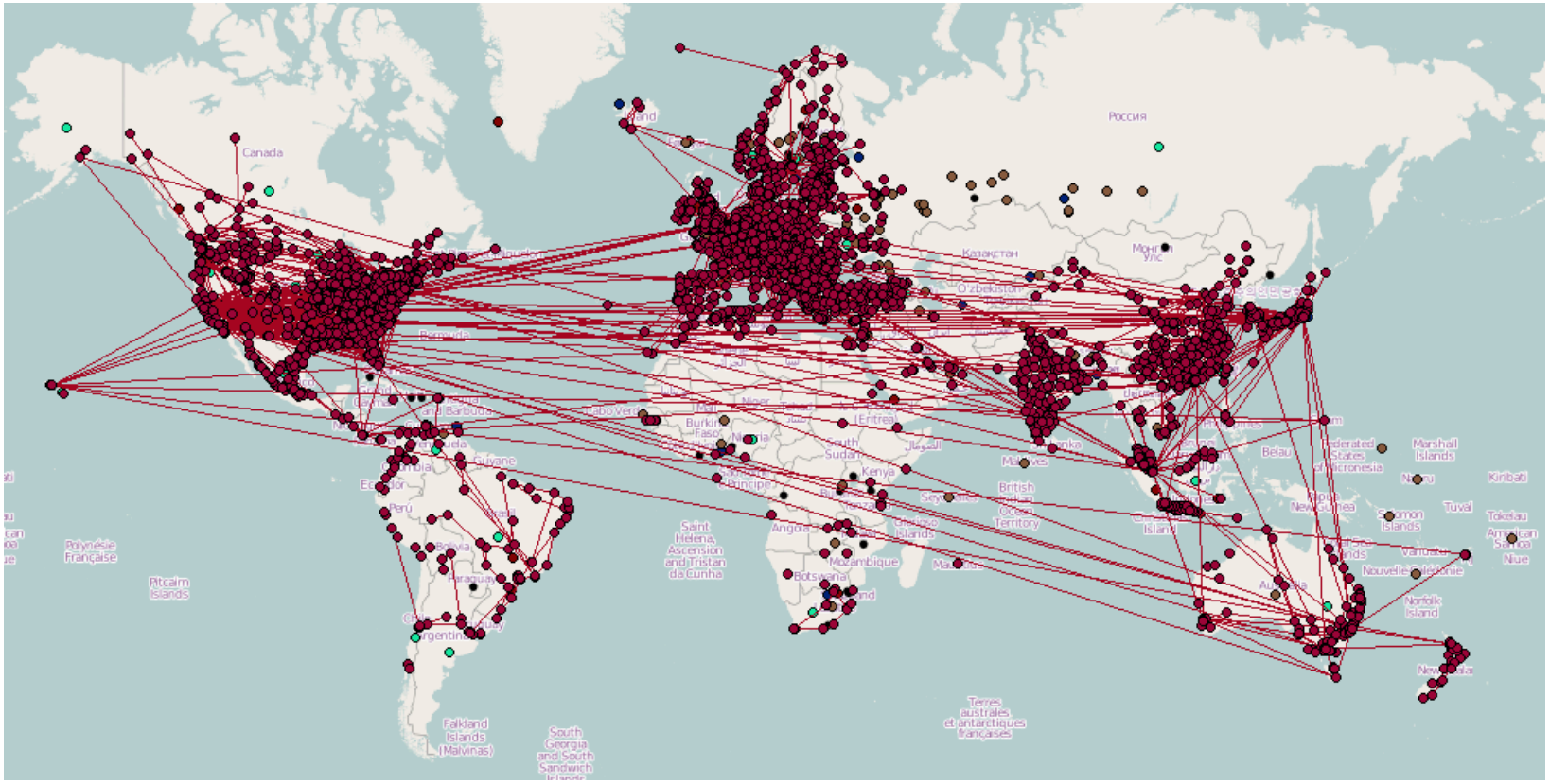


Image extraction

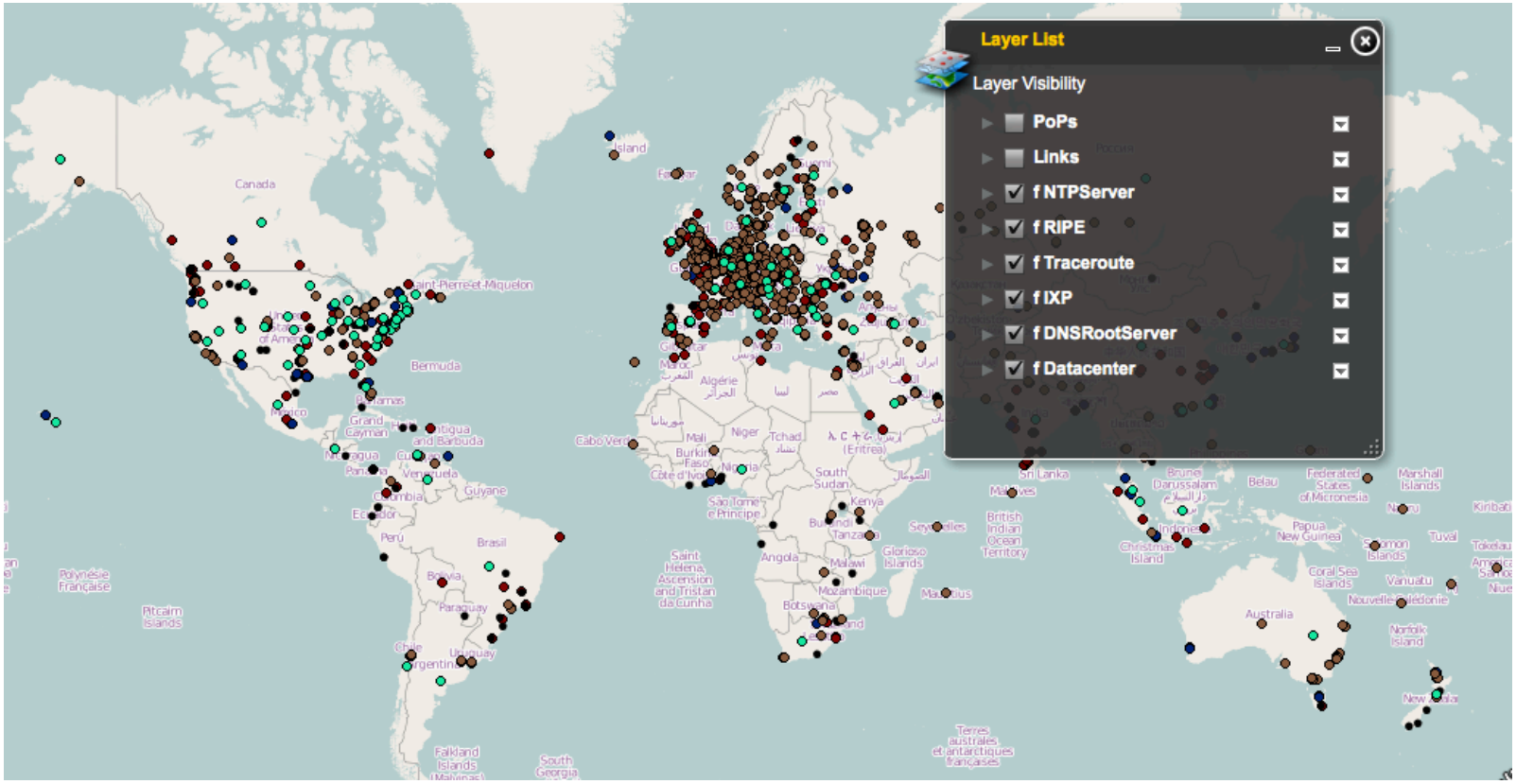
Geo-specific link encoding



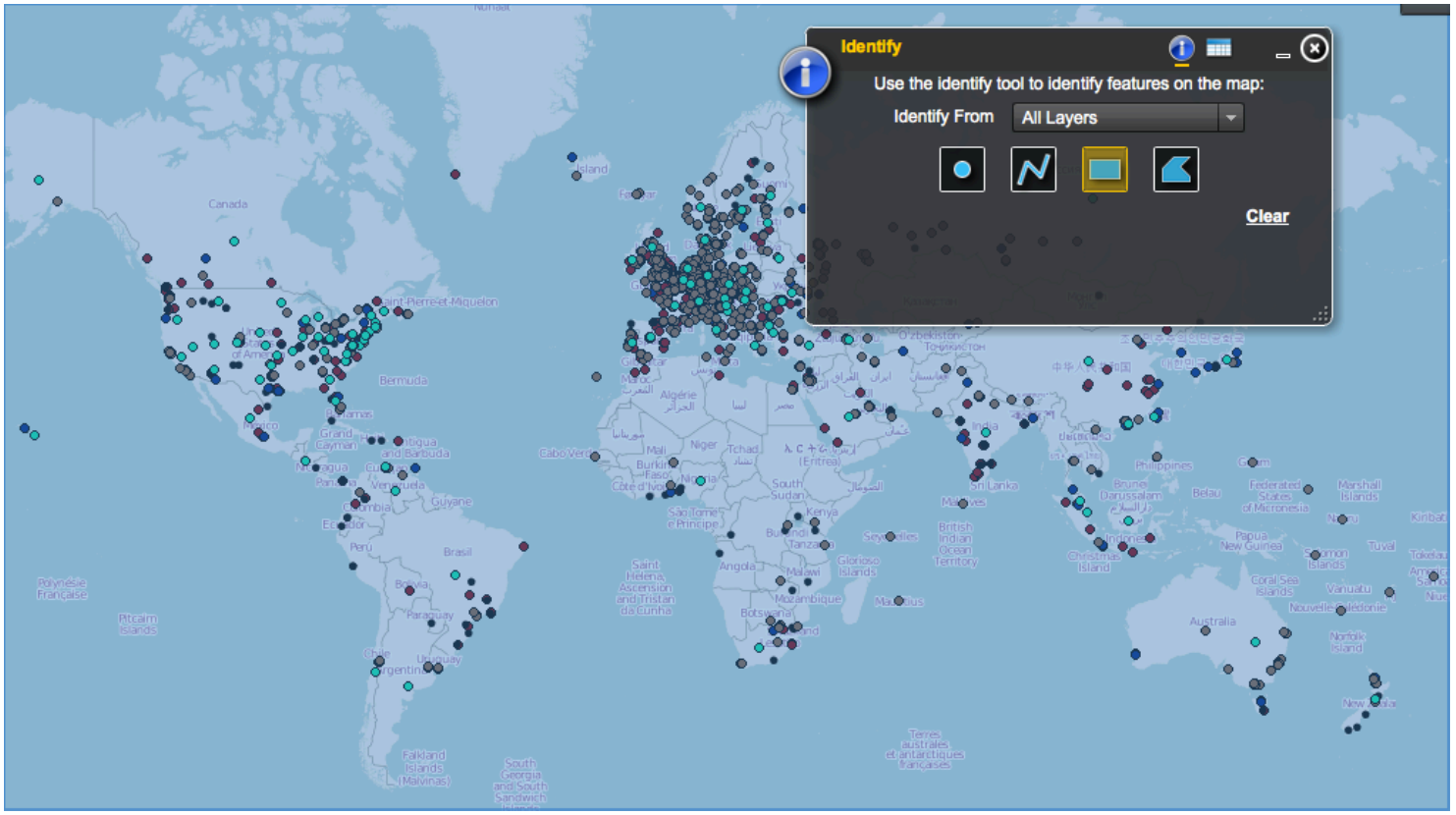
Internet Atlas – Full View



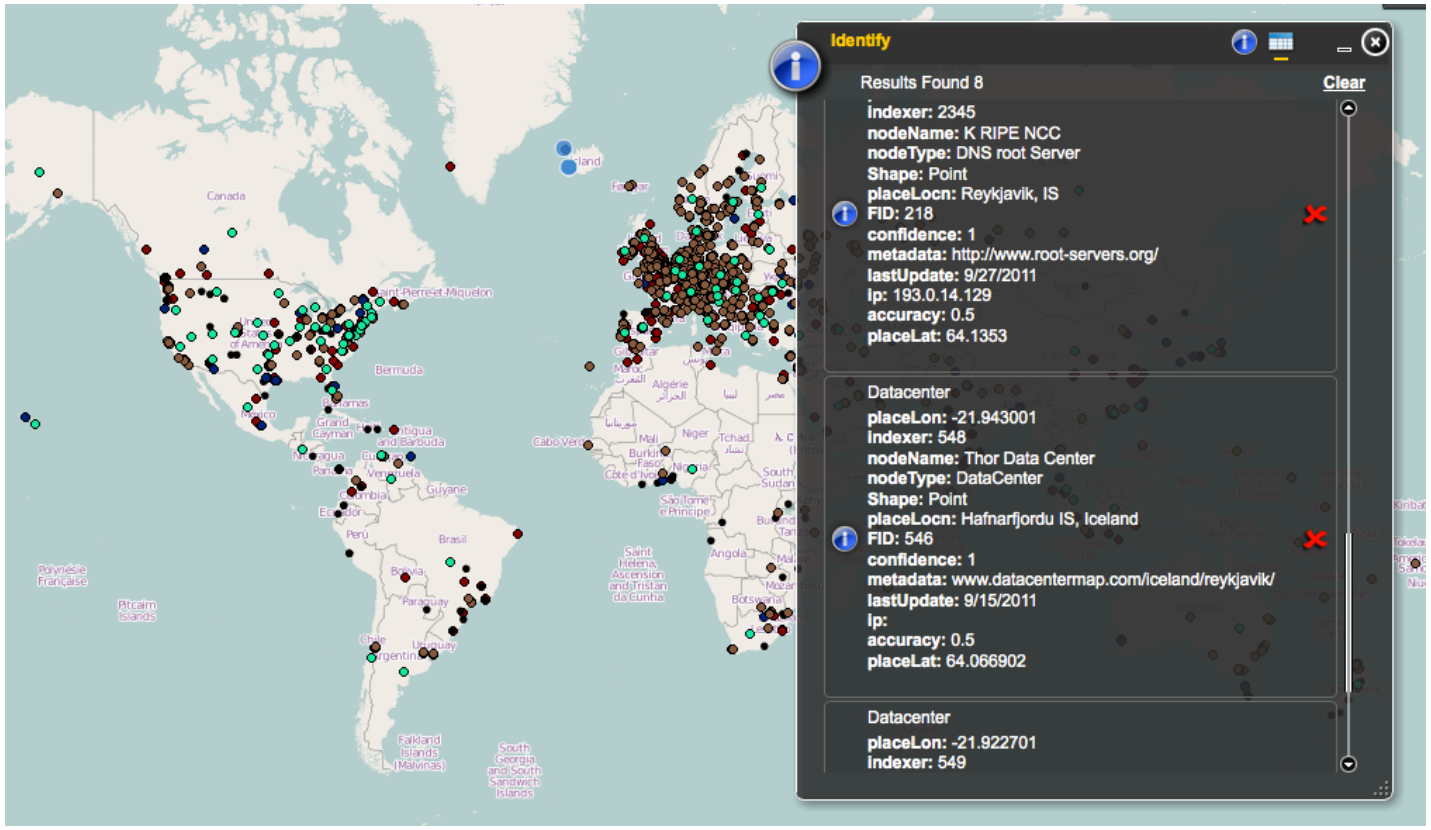
Internet Atlas – Layers



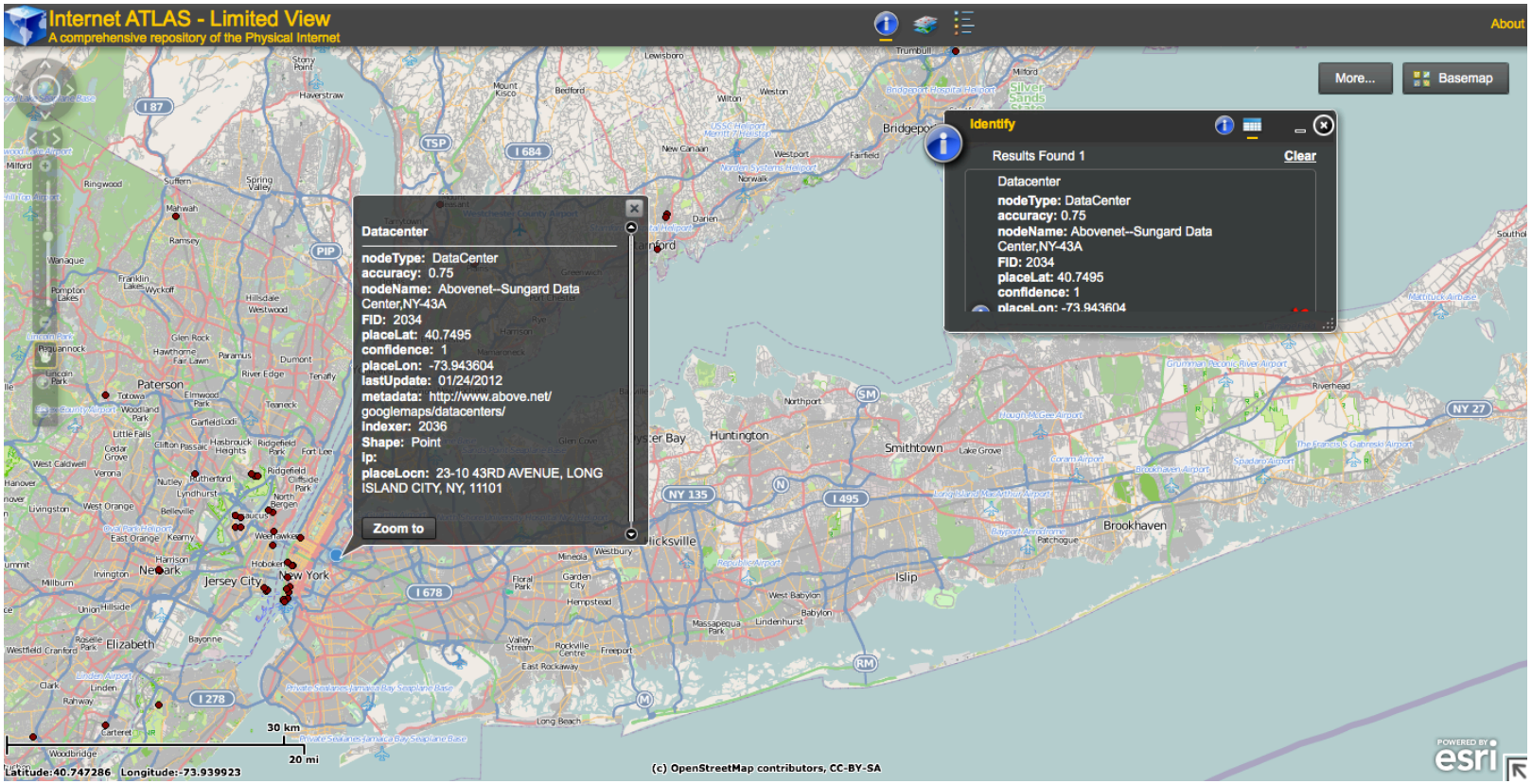
Internet Atlas – Identify



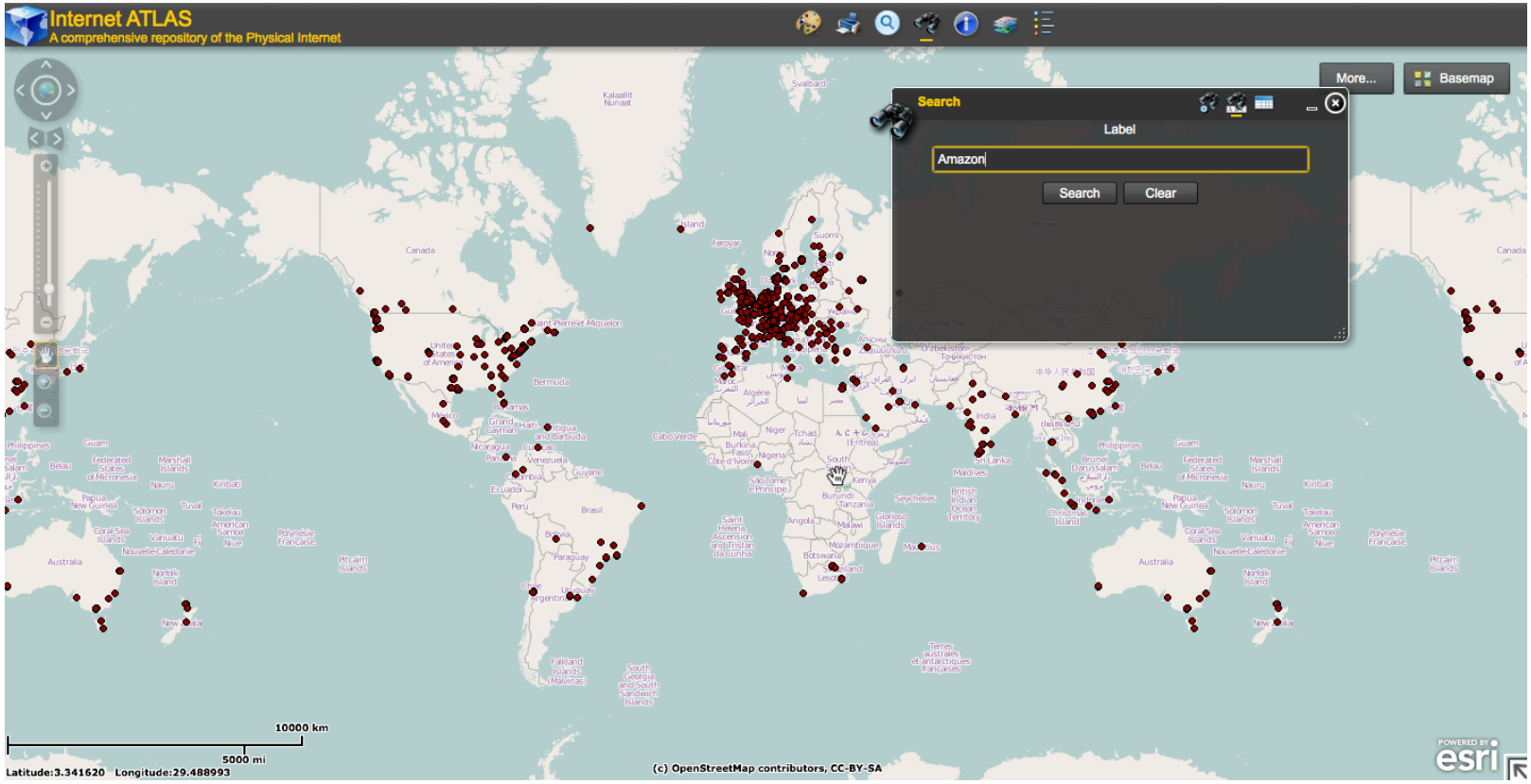
Internet Atlas – Identify



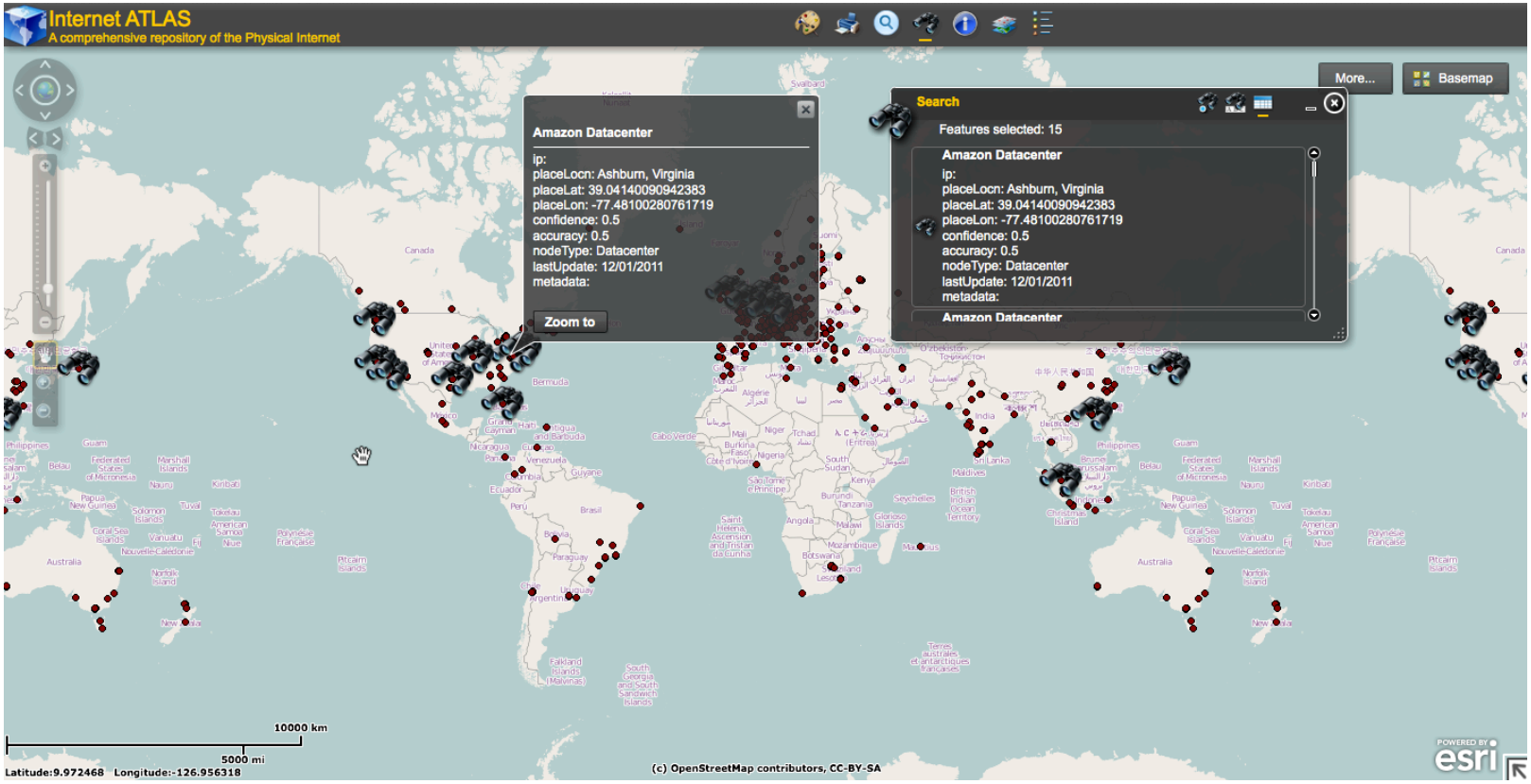
Internet Atlas – Zoom



Internet Atlas – Search



Internet Atlas – Search



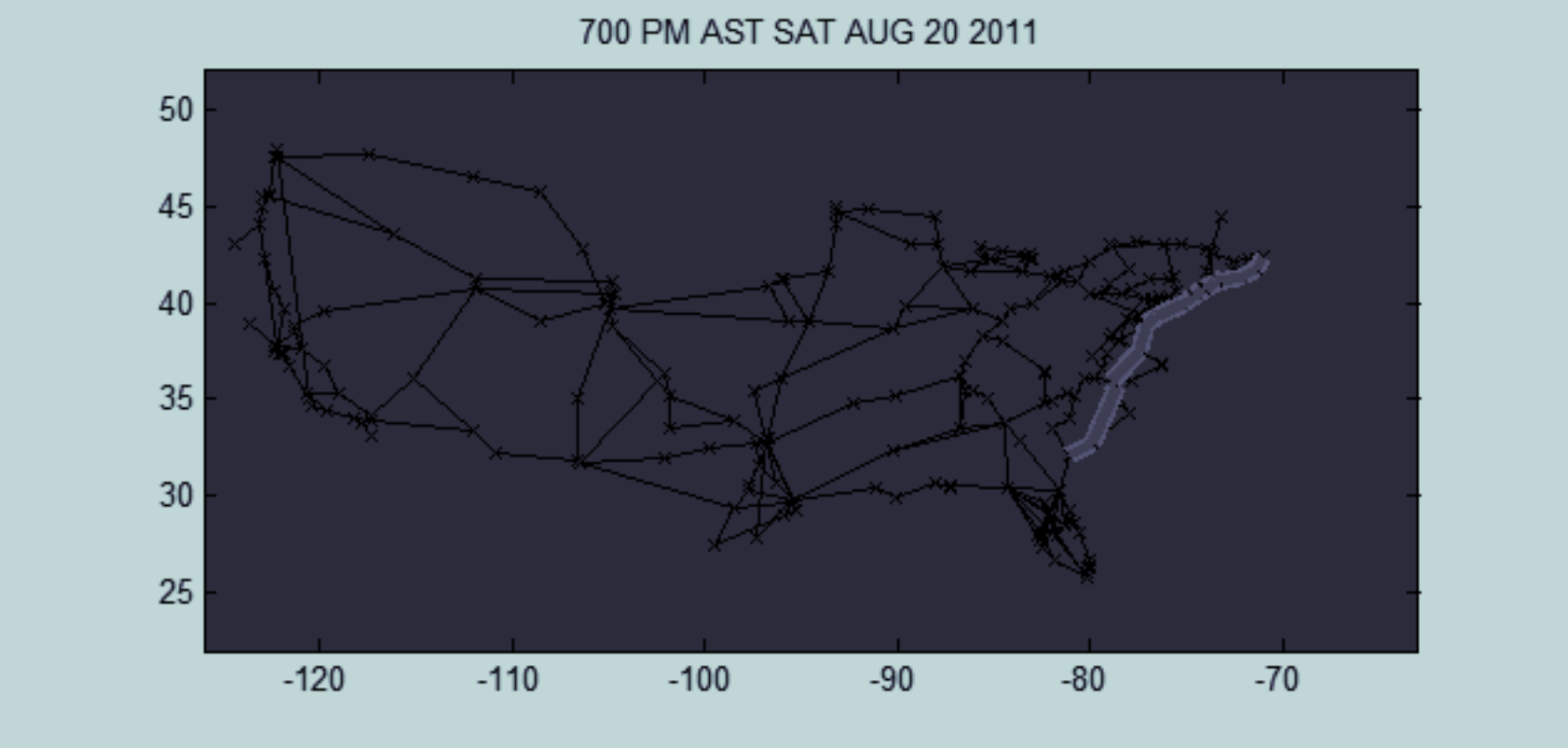
Target applications

- Many potential applications for an accurate, but incomplete graph of the physical Internet
- Application 1: link characterization
  - What are the physical distances of links?
- Application 2: robustness
  - Are there vulnerabilities in the current infrastructure?
- Application 3: intra-domain routing
  - Given peering relationships, can we identify inefficiencies?

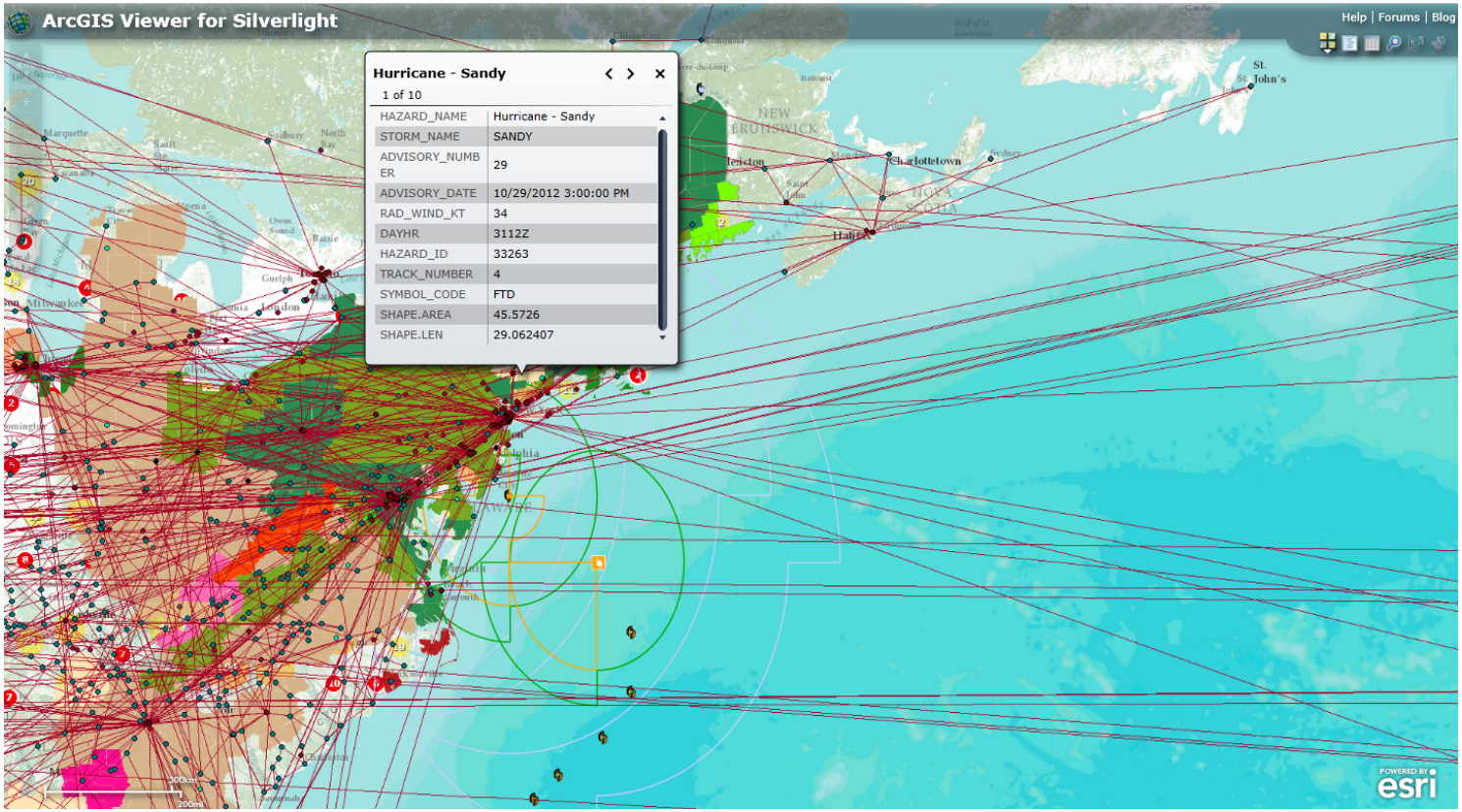
Improving network availability

- **Given outage event risk profile, how can network availability be improved?**
  - **Backup routes within an infrastructure**
  - **Additional provisioning to extend infrastructure**
- **RiskRoute optimization framework**
  - **Identifies backup routes and provisioning options**
  - **Considers historical and/or real time outage events**
- **Case study using networks and disaster event data from US**
  - **Many opportunities to reduce risk!**

Level3 and Hurricane Irene



Internet Atlas – Risk Analysis



Data Sharing

- NO!
- Questions? Enquiries?
  - Prof. Barford (pb@cs.wisc.edu)
- Accounts?
  - Prof. Barford (pb@cs.wisc.edu)
  - Ram Durairajan (rkrish@cs.wisc.edu)

Thank you!

- Paul Barford
- Brian Eriksson
- Xin Tang
- Subhadip Ghosh