Global Research Map

The need for a crowd source Global Science Infrastructure Database and our efforts to start one

Edward Balas

GlobalNOC
Global Research Network Operations Center

Research Support Motivation

- Science Research Coordinators need to understand
  - Big Science users and their needs
  - Facilities involved and how they are connected
  - How to resolve inter-domain performance problems
- Network Planners need anticipate future demand

Advocacy and Education Motivation

- Highlight the key role R&E networks play in enabling science and the latest major discoveries.
- Help people see how modern science happens
- Show how the R&E Network connect the people, instruments, compute and storage

Barriers

- Effort required
  - Manual entry is a daunting and thankless task
  - Requires inter-disciplinary skill
  - Requires persistent focus to fight change
- Resistance to share
  - Flow data has privacy concerns
  - Topology data can be sensitive for some
  - Not sure who is the authority

What is needed

- An open database we can incrementally improve over periods that transcend individual grants
- A balanced approach to data sharing
- Use of crawling / automation to scale the effort
- An crowd source approach to leverage community knowledge

Global Research Map

- Create open source platform to **facilitate** network research and **advocate** for the global infrastructure used in modern scientific endeavor
- Combine automated web-mining with **crowd sourced** editing to create an open source database and advocacy platform
- Create **useful** and **attractive experiences** for technical and non-tech users

What is it?

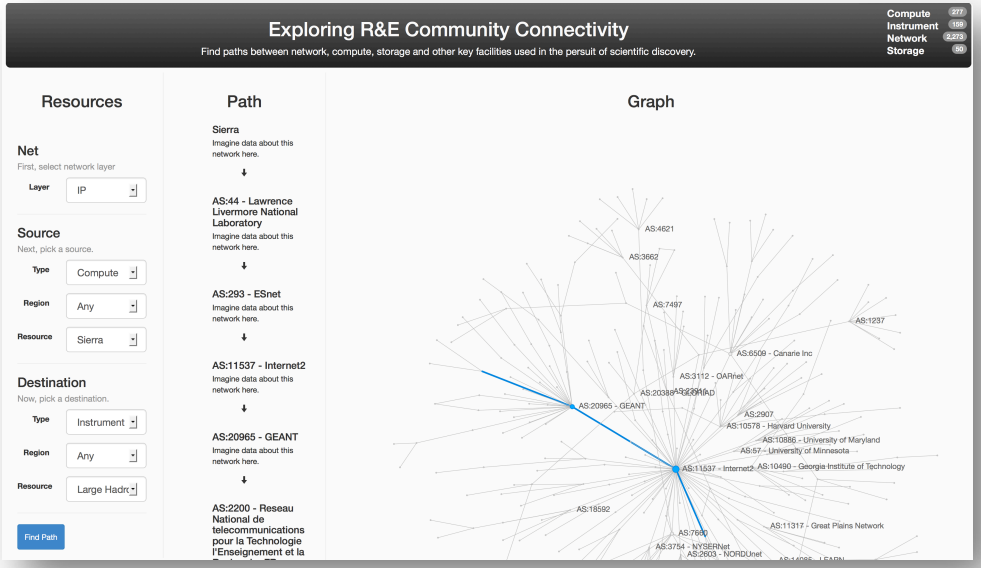
- A system that allows us to document and study the relationships between
  - Layer2 and Layer3 networks
  - Supercomputer and Storage facilities
  - Scientific Instruments
  - Research Projects
  - People involved in these projects

Modest Beginning

- Targeted crawl + manual tune
  - 277 Super Computers
  - 159 Scientific Research Instruments
  - 2,273 Layer2 and Layer3 networks
- Sources
  - BGP route dumps and OSCARs topo data
  - Wikipedia, dmoz.org, top500
  - Manual input

What can we do today?

- Layer2 or Layer3 connectivity between
  - Networks
  - Computers
  - Instruments
  - Storage



- If interested, visit the Worldview or hunt me down for a demo

GlobalNOC
Global Research Network Operations Center

What we learned

- Existing sources like Wikipedia are a great starting point
- Lack structure needed to automate this process
- Encoding challenges
  - Facility to network
  - Facility to type
  - Facility to discipline
- Semi-automated techniques based on context and string similarity show promise

How you can help

- Suggest Sources
  - Of data to use as seeds for crawling
- Suggest Use Cases
  - That we have not thought of yet
- Contribute time
  - Editing or adding manual data
  - (We are here yet but someday)

Final thoughts

- Thanks
  - Grover Browning for contributing Indiana L2 data
  - Paul Schopis for contributing Ohio L2 data
- For more Information
  - Send me an email: ebalas@iu.edu
  - Grab me sometime this week