

Supporting Climate Applications over Named Data Networking (NDN)

Christos Papadopoulos

Cathie Olschanowsky

David Randall

Colorado State University

Internet2 Summit, April 7, 2014

#13410999: NSF CC-NIE Integration award

IP



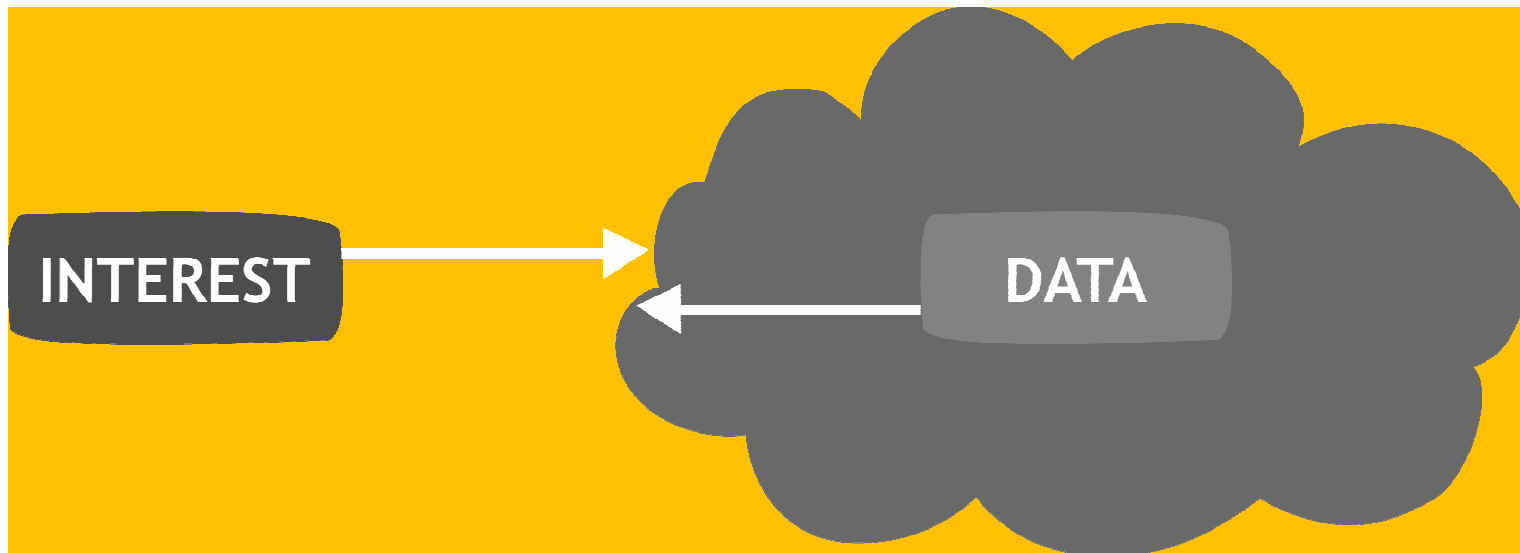
Host-centric
addressing

NDN



Data-centric
addressing

NDN: Focus on Data



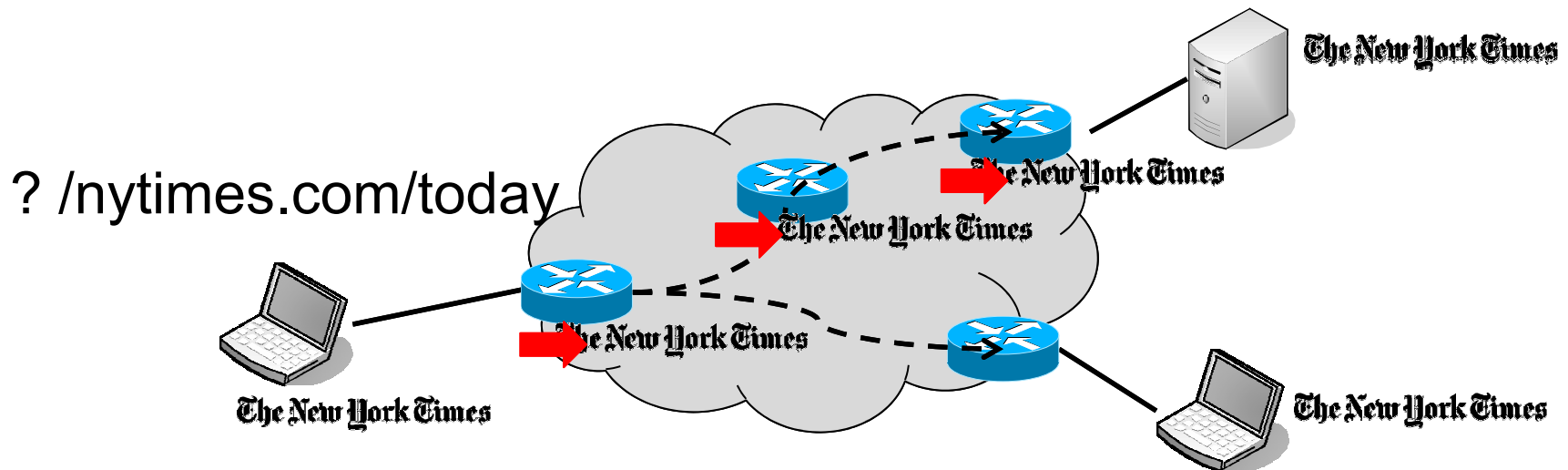
Abstracting away the notion of “host”

Superset of host-to-host communication model

Problems with Current Internet

- Focus is on end-point communication
 - Artifact of original thinking: share resources, not content
 - Login to fast machine, access to the tape drive, the printer, etc.
 - **NDN solution: Focus on the what not the where**
- Security
 - To get data, you build a secure path
 - Once you authenticated with the server, you trust the content
 - **NDN solution: Secure the data not the container**

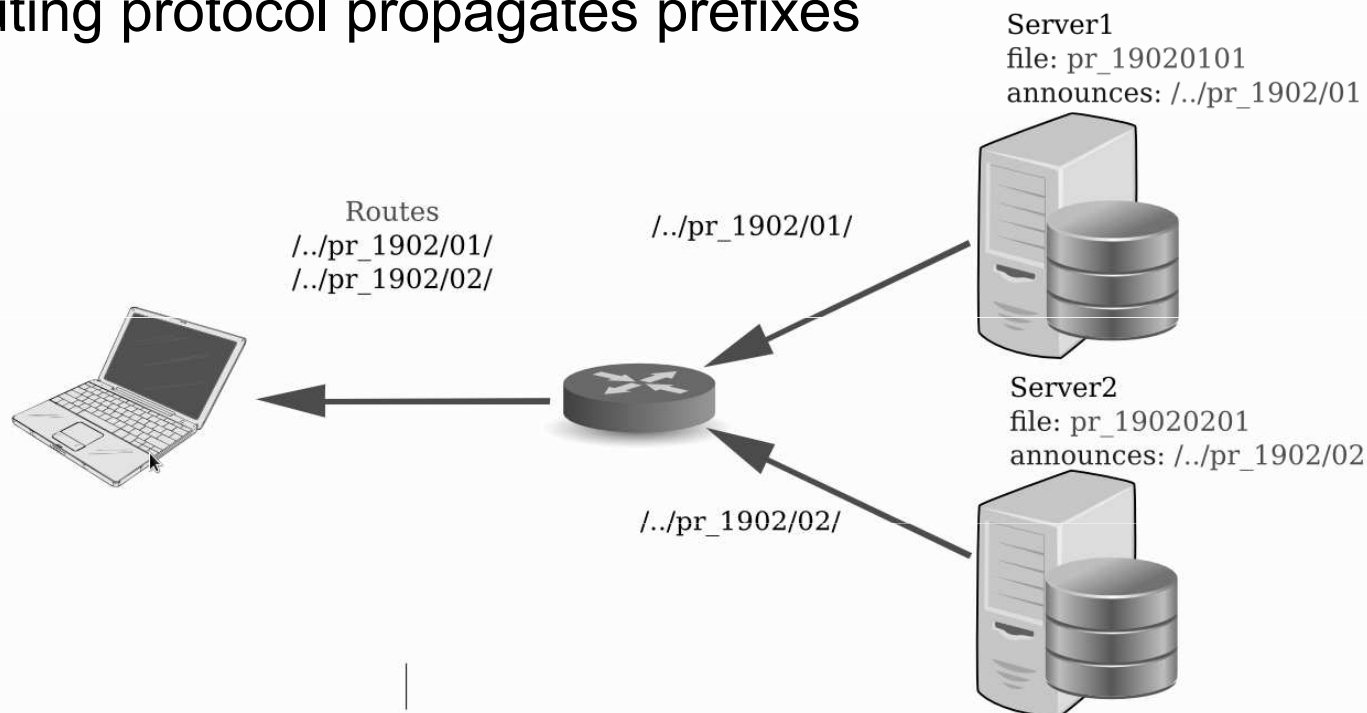
NDN Operation



- Route on content names
- Content from anywhere: not just the producer
- “Breadcrumbs” & de-duplication of requests
- Cache retrieved data in Content Store (CS)

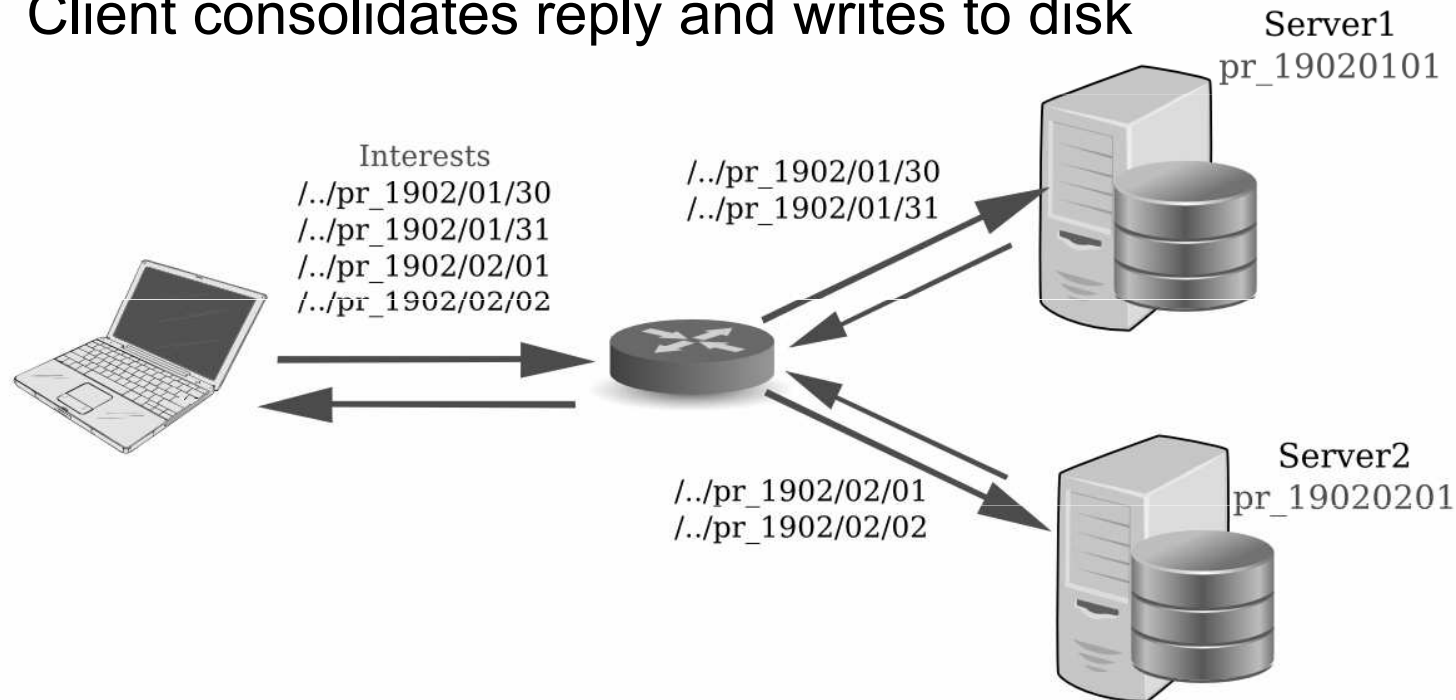
NDN and Climate Data

- Servers have `.nc` files, each `.nc` file have one month's data
- Route announcements in network are based on filename
- Each server advertises one prefix for a file
 - Server having file `pr_19020101.nc` announces `../../pr_1902/01/`
- Routing protocol propagates prefixes



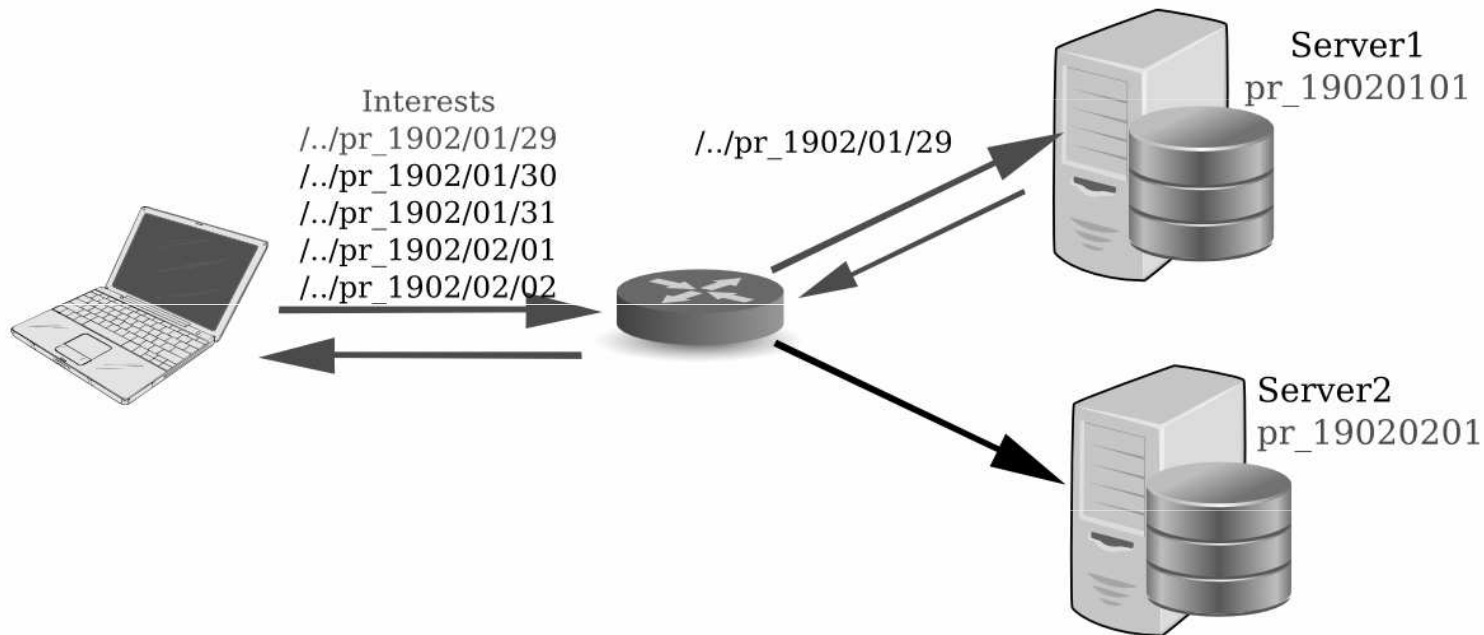
A Data Request

- Want data for Jan 30 – Feb 02
- Client expresses interests, one for each day
- Interests for Jan 30-31 go to server1
- Interests for Feb 01-02 go to server2
- Data is dynamically generated and sent back
- Client consolidates reply and writes to disk



Partially Cached Data

- What happens if we ask for Jan 29 – Feb 2 ?
- Request for data not cached goes to server
- Rest is answered from cache



Naming In Climate Applications

- Climate community has recognized the importance of structured naming
- Reference documents are beginning to appear
- Mostly targeted to file and dataset naming
 - Including dataset publication
- Provide consistency across distributed archives

Example: CMIP5: Coupled Model Inter-comparison Project, phase 5

- “.. standard experimental protocol for studying the output of coupled atmosphere-ocean general circulation models.”
- International effort, started in 1995
- “CMIP5 is meant to provide a framework for coordinated climate change experiments.”

CMIP5 Data Reference Syntax (DRS) and Controlled Vocabularies

- “.. a common naming system to be used in files, directories, metadata, and URLs to identify datasets wherever they might be located within the distributed CMIP5 archive. It defines controlled vocabularies for many of the components comprising the data reference syntax (DRS).”
- Current version is 1.3.1, June 2012

DRS: Directory and File Structure

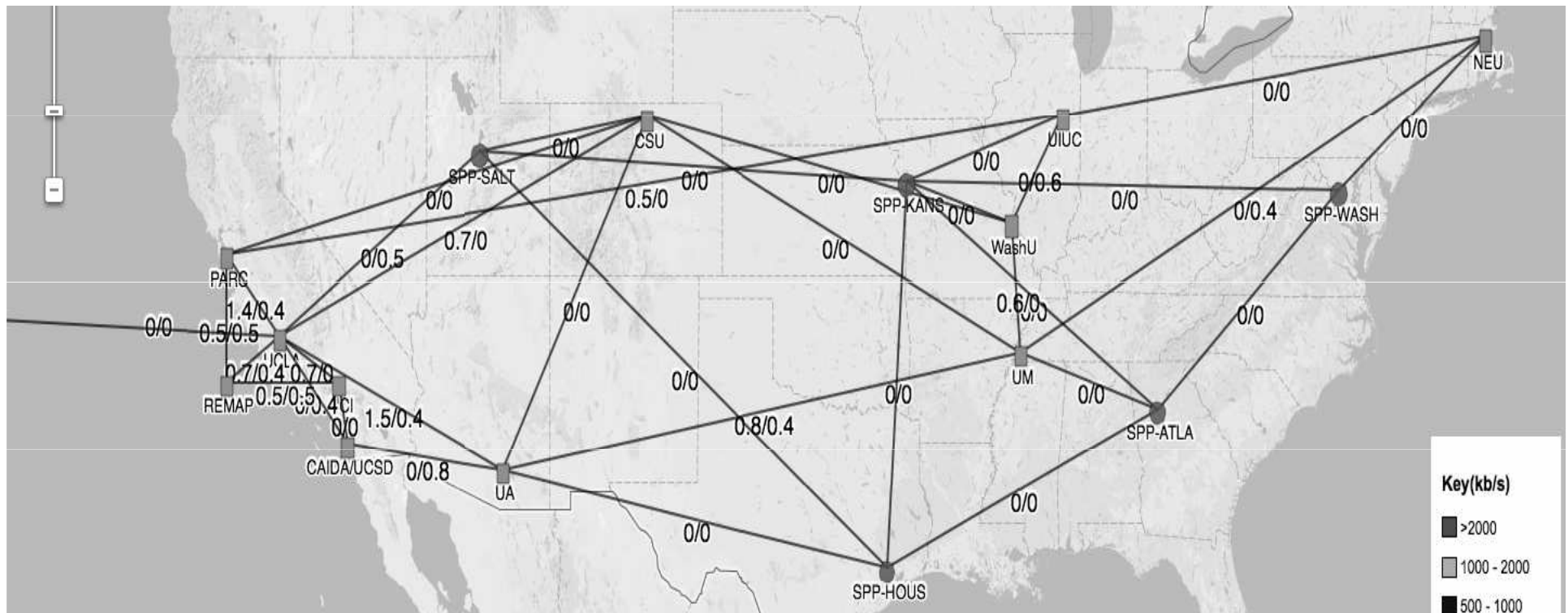
- Directory encoding:
 - /<activity>/<product>/<institute>/<model>/<experiment>/<frequency>/<modeling realm>/ <variable name>/<ensemble member>
 - NDN Example:
`/CMIP5/output/MOHC/HadCM3/decadal1990/day/atmos/tas/r3i2p1/`
- Filename encoding:
 - filename = <variable name>_<MIP table>_<model>_<experiment>_<ensemble member>[_<temporal subset>][_<geographical info>].nc
 - NDN Example:
`tas_Amon_HADCM3_historical_r1i1p1_185001-200512.nc`

Observations

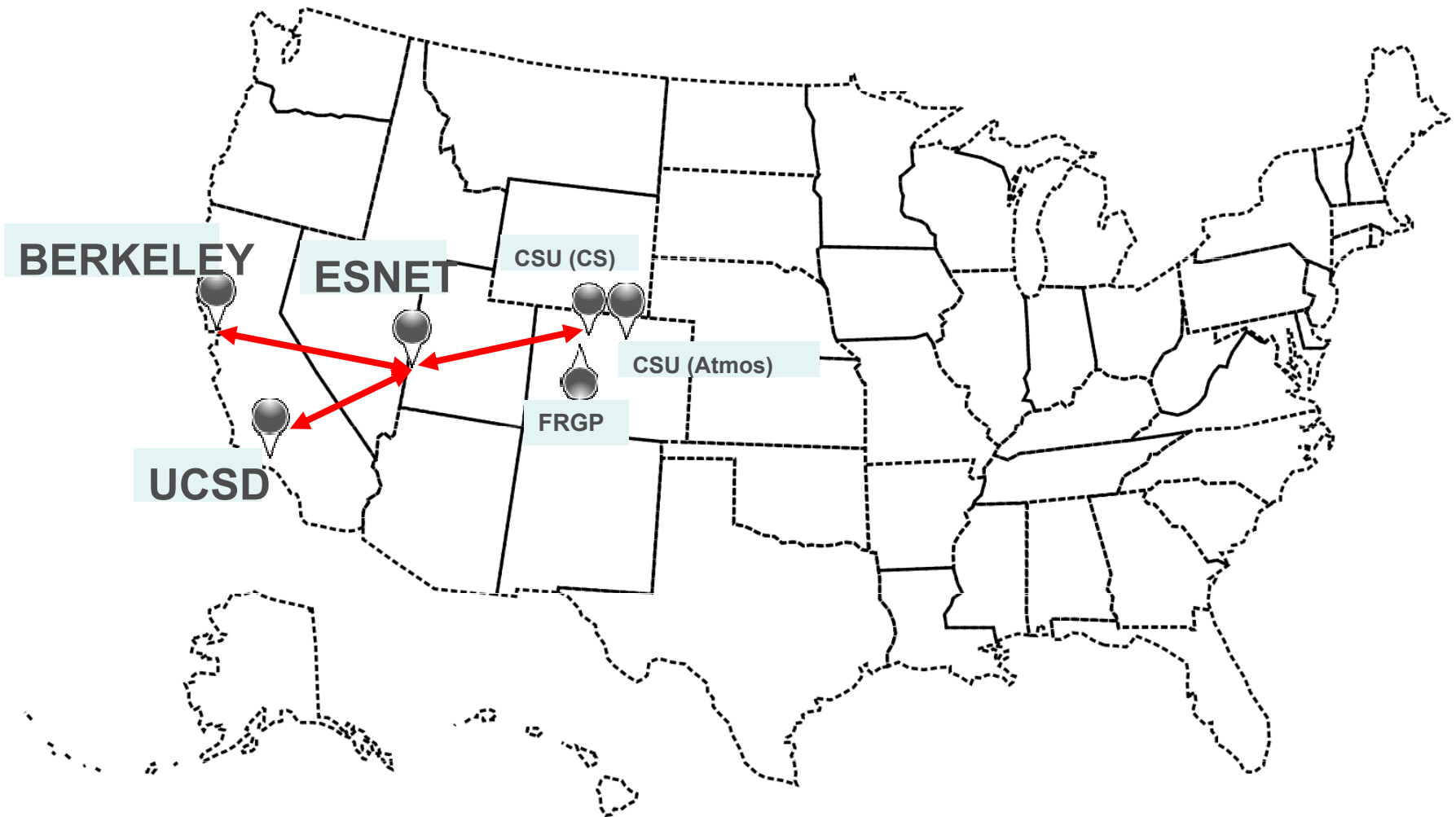
- CMIP5 DRS is NDN-compliant
 - Hierarchical
 - Clearly defined name components
 - Well-defined vocabularies
 - Organizes components from less specific to more specific
- Utilities to translate into DRS (CMOR)
- Ports readily into NDN names

Is NDN Deployed?

- Deployable now as an overlay (TCP, UDP) or on Layer 2 transport
- C, Java, Python, Javascript libraries
 - <http://github.com/named-data>
- Testbed of 15 routers, including 5 on Internet2 / GENI
- <http://ndnmap.arl.wustl.edu/>



Planned Climate Testbed



Moving Forward

- NDN offers two potential improvements to the scientific community:
 - Make data discovery far more convenient
 - Speed up data retrieval
- Data naming and discovery can be done now
- Speeding up data retrieval – forwarding, strategy layer, etc.
 - Real application to test congestion control

Current Status

Now:

- Deploying climate testbed
- Designing appropriate namespaces
- Implementing translators – two already done

Near Future:

- Integrate with existing workflow
 - discovery and retrieval
- Implement appropriate access controls
- Implement high-throughput transport protocol