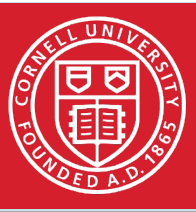


Cornell University
Center for Advanced Computing

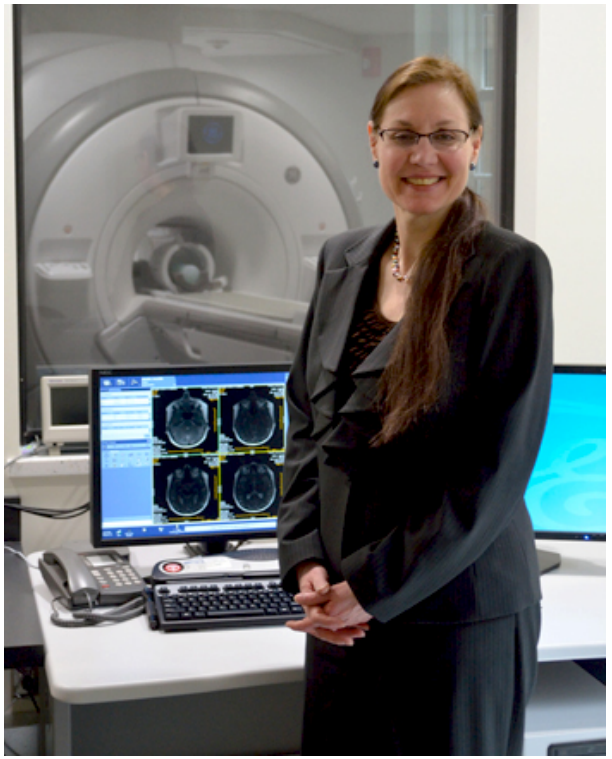
Automating Backups with Globus

Michael Padula – *mdp15@cac.cornell.edu*
Systems Consultant
Cornell University Center for Advanced Computing (CAC)

The Basics

- **Human Neuroscience Institute (HNI)**
  - Needed low-cost backup solution for research and user data

- **Key components of our solution**
  - Globus Connect Personal
  - Globus Command Line Interface (CLI)
  - CAC Archive Storage Service
  - Globus Connect Server
  - Scripts and cron



Some Details

- **GC personal installation on file server**
  - Need access to all files and directories
  - OS account dedicated to this task
- **Globus CLI**
  - ssh key authentication, dedicated Globus account access to CLI
  - CLI commands used
    - endpoint-activate
    - transfer
- **CAC Archive Storage Service**
  - Runs GC Server
  - 1/5 the cost of a file server
  - Easy to share data with collaborators via Globus

Some Statistics

- **Mid-March Backups**
  - Daily backup of user data
    - Complete in 1 hr. consistently
    - Nearly 1 million files, 2400 directories
    - Bytes transferred between 6k-21k
  - Weekly backup of shared data
    - Complete in 30 minutes consistently
    - Nearly 23 million files, 55,000 directories
    - Bytes transferred between 26GB and 1+ TB
      - In a unique case: 2 million files totaling 1.3TB

Limitations

- Stopping Globus Connect Personal
  - There's no time/cost-effective way to do this
- Activating credentials
  - Must be re-activated weekly by a person

Some Links

- Storage Services (Archive is one of them)
  - http://www.cac.cornell.edu/services/storage.aspx
- Disk Farm and Network details
  - http://www.cac.cornell.edu/services/CAC%20Disk%20Farm.pdf