

# The NDGF Distributed Tier 1

## Design and Experiences (focusing on data handling)

Gerd Behrmann

(with some slides stolen from  
Oxana Smirnova and David Cameron)

# Nordic Data Grid Facility



- Provides unique Nordic distributed WLCG Tier 1
- Involves 7 largest Nordic academic HPC centers
- ...plus a handful of University centers (Tier 2 service)
- Connected to CERN directly with GEANT 10Gbit fiber
- Inter-Nordic shared 10Gbit network from NORDUnet

# Challenges

- Computational jobs on distributed resources
- Storage on distributed resources
- Connectivity
- Stability
- Operation

# Technical differences

- Heterogeneous environments
- Heterogeneous interconnect
- Several independent tape systems
- Limited bandwidth
- High latency
- Frequent network failures
- Spanning many administrative domains

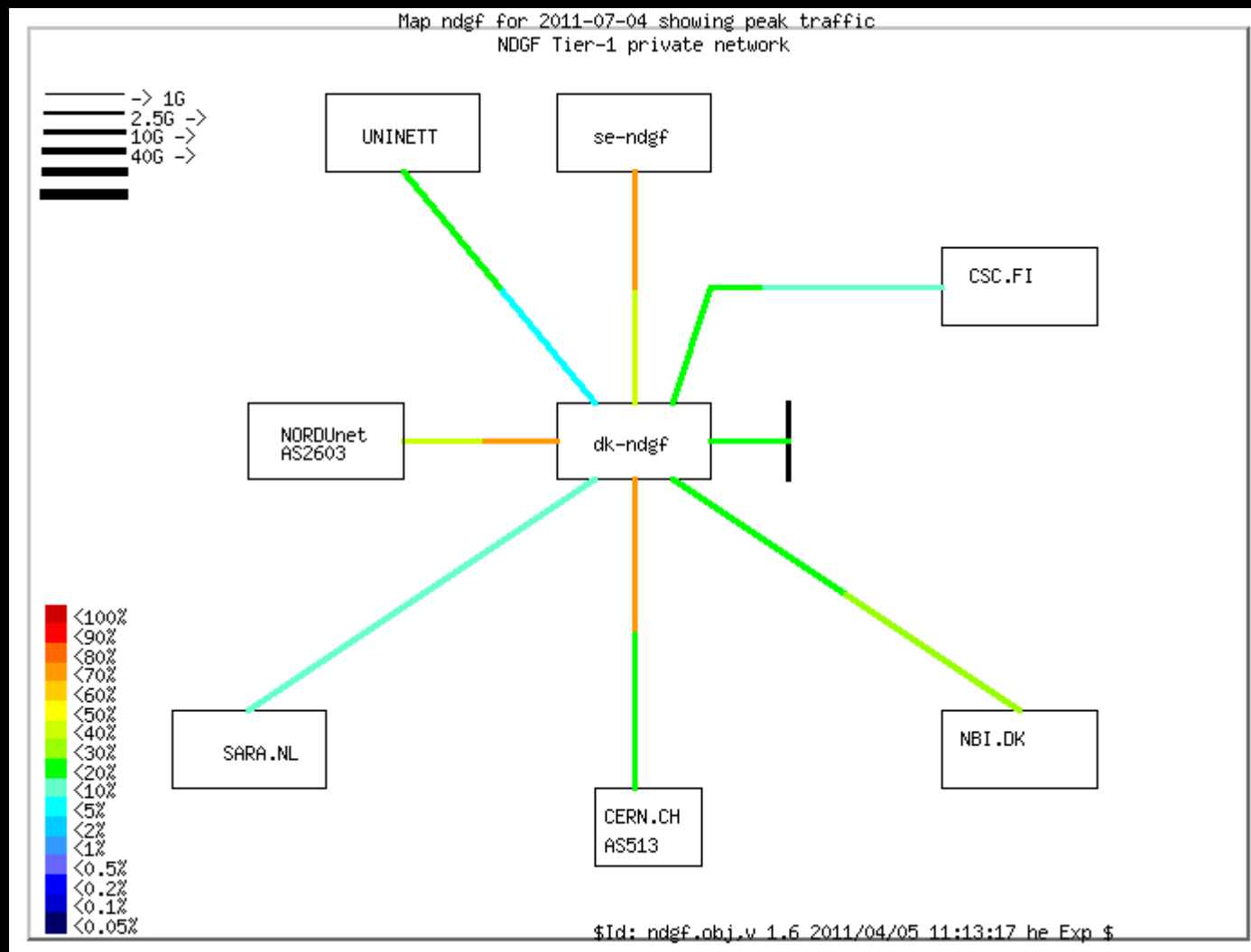
# Goals towards participating sites

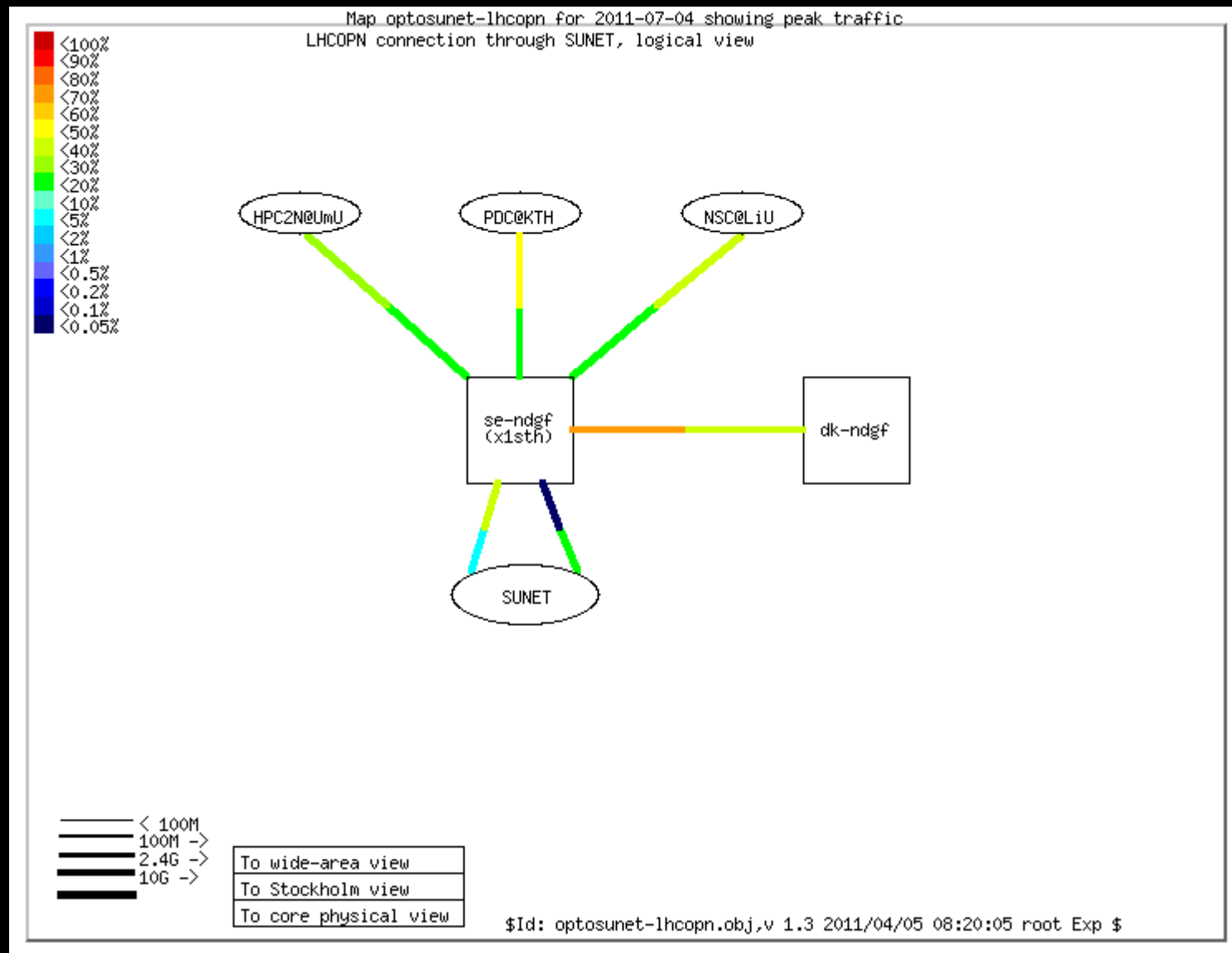
- Minimal operational requirements
- Minimal software requirements
- Free to choose OS
- Free to choose CPU architecture
- Free to choose storage architecture
- Free to choose storage density
- Free to choose HSM
- Combat fear to loose control

5 ▪ ...

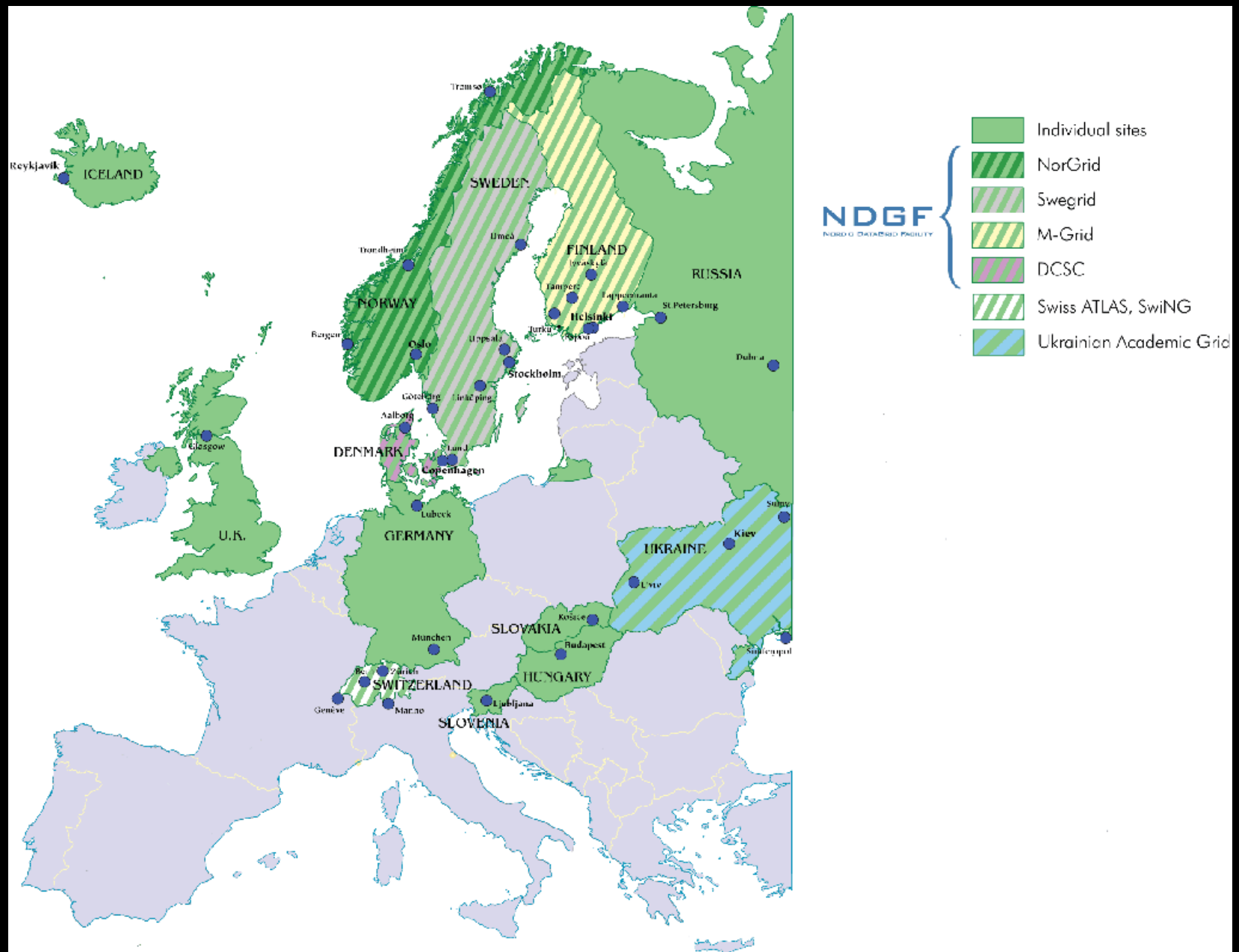
# Goals towards customers

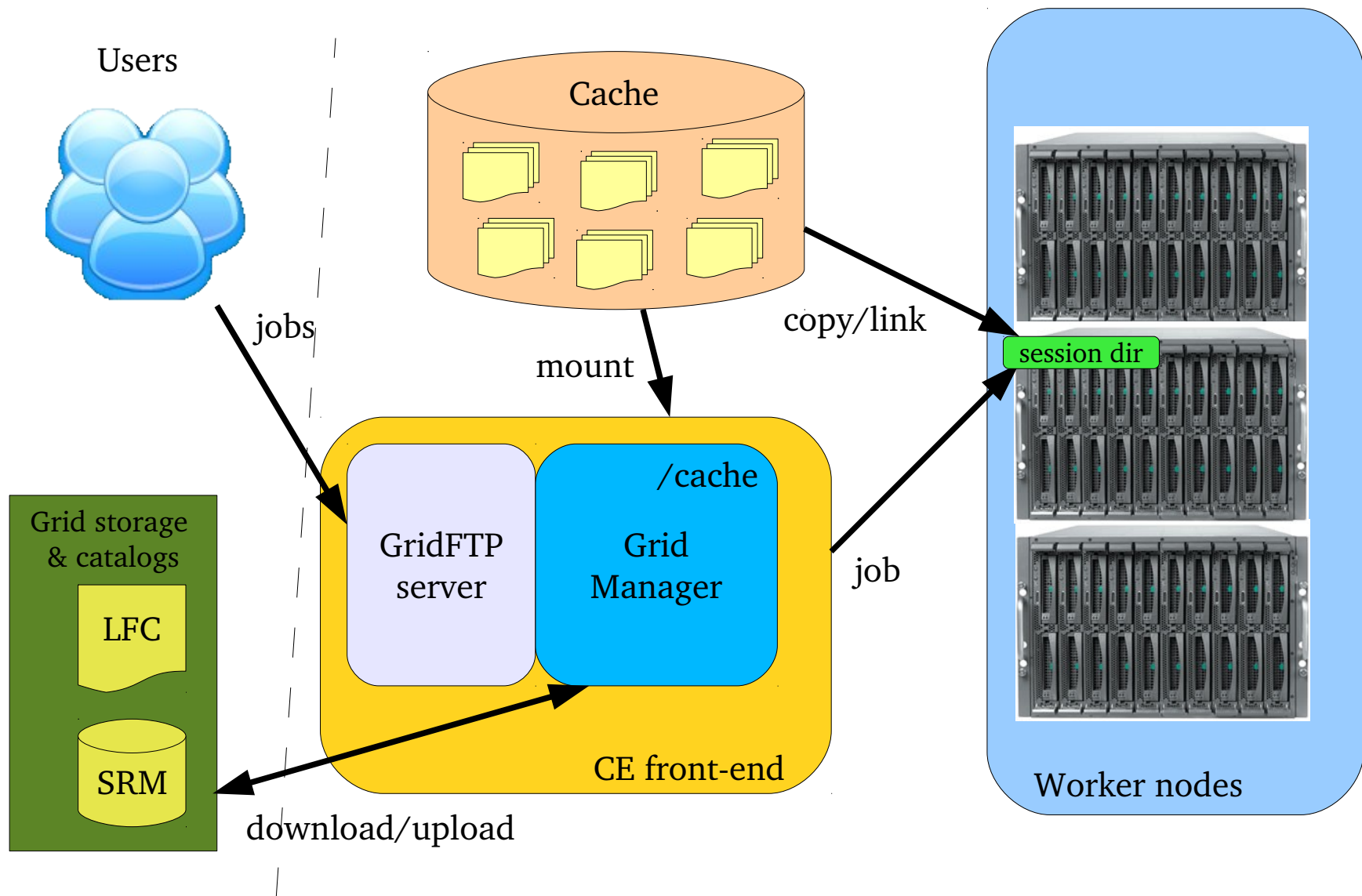
- Atlas and Alice Tier 1
- Stability
- First priority is data taking
- One operational contact point
- Hide the distributed nature
- Single SRM endpoint
- Non WLCG communities











WebDAV



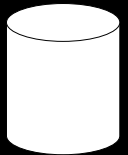
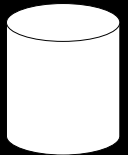
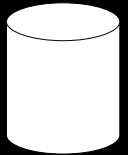
HTTP

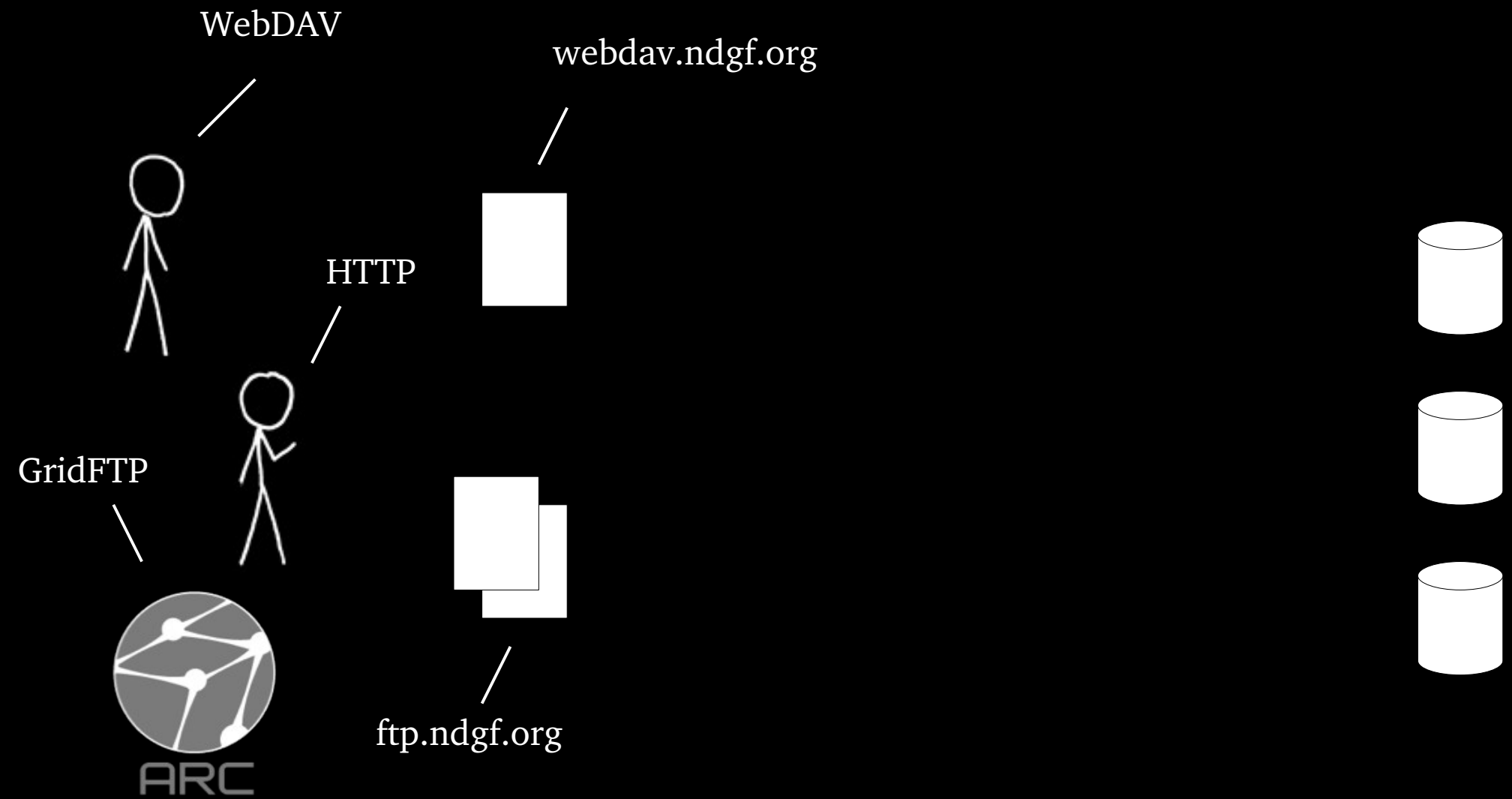


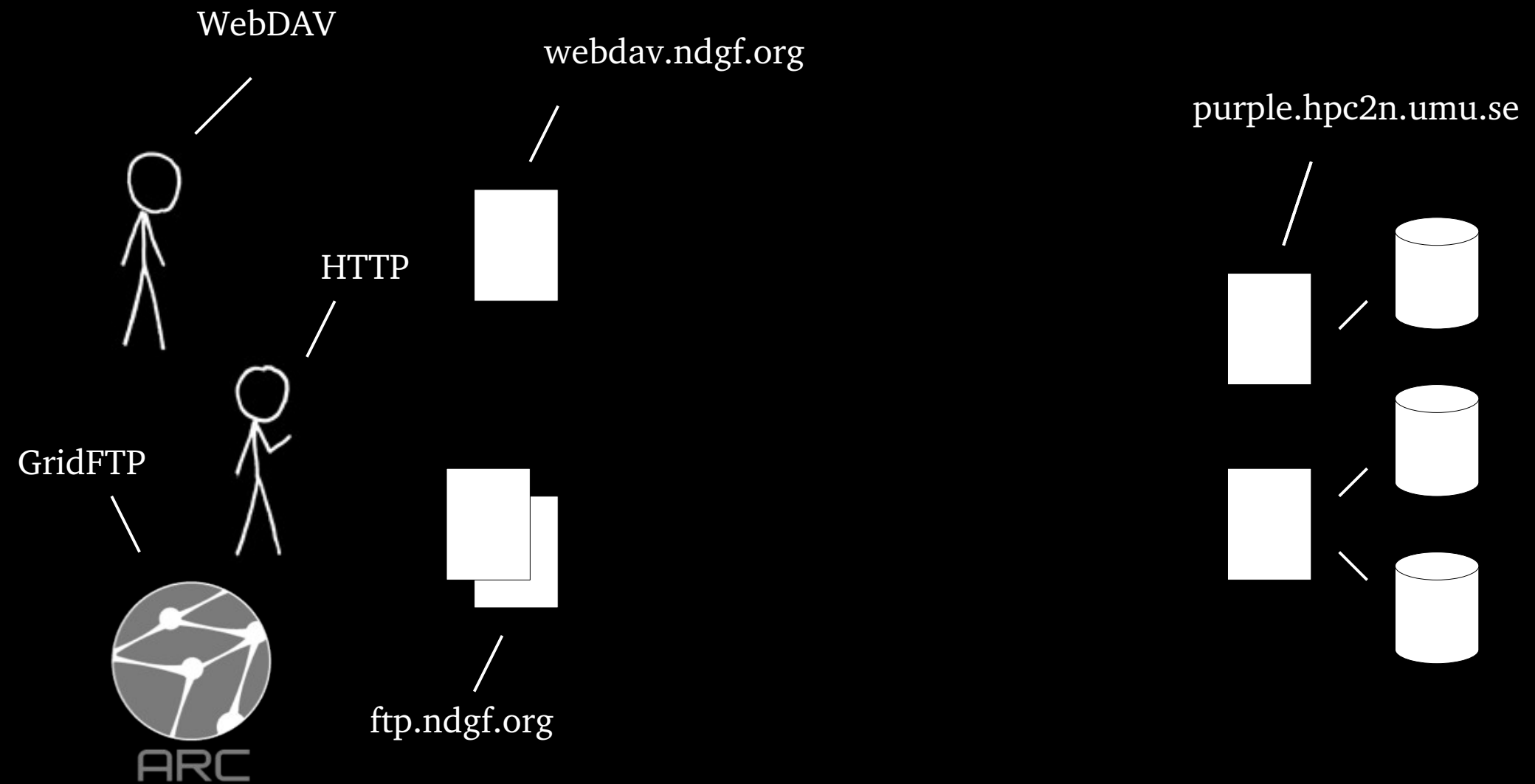
GridFTP

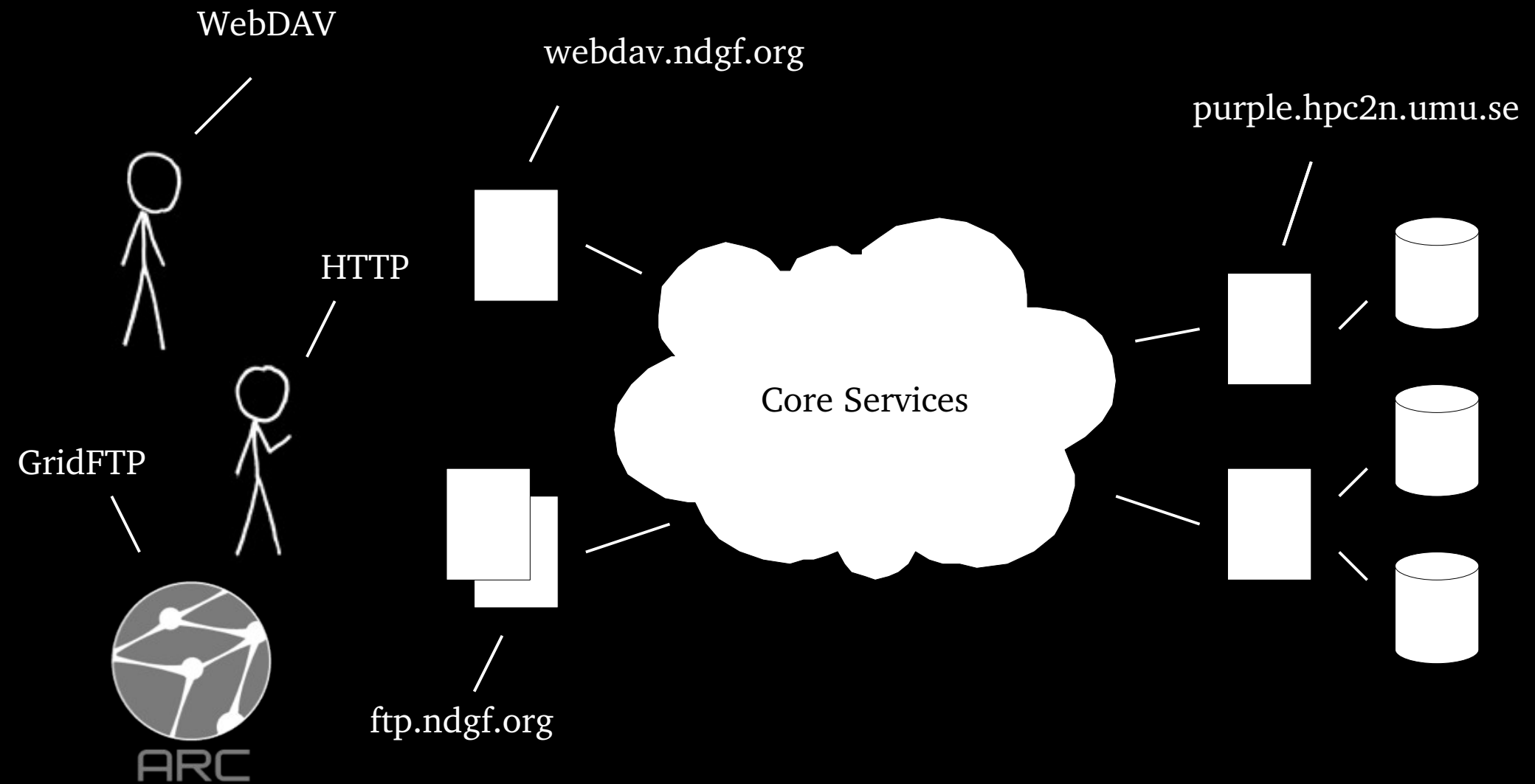


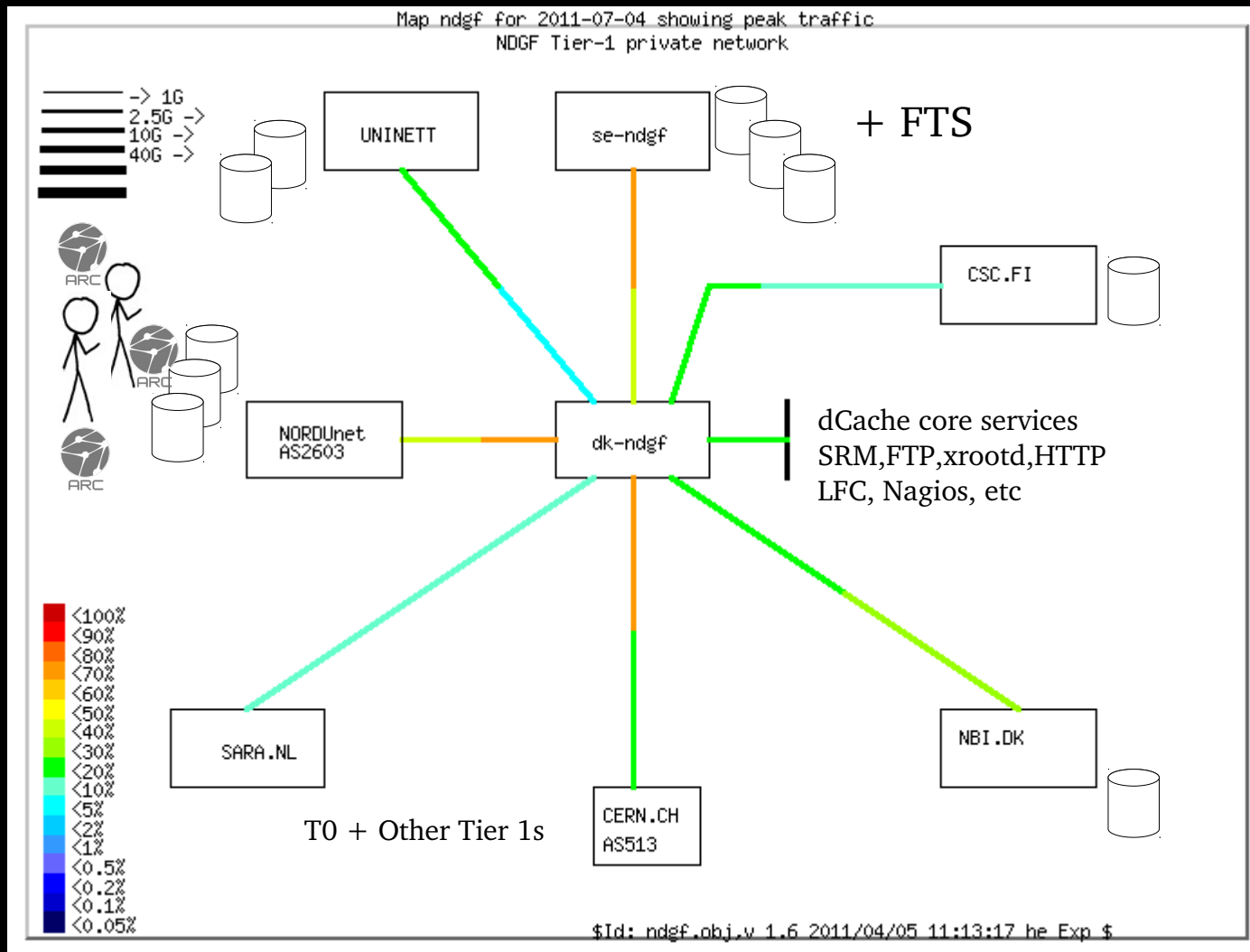
ARC

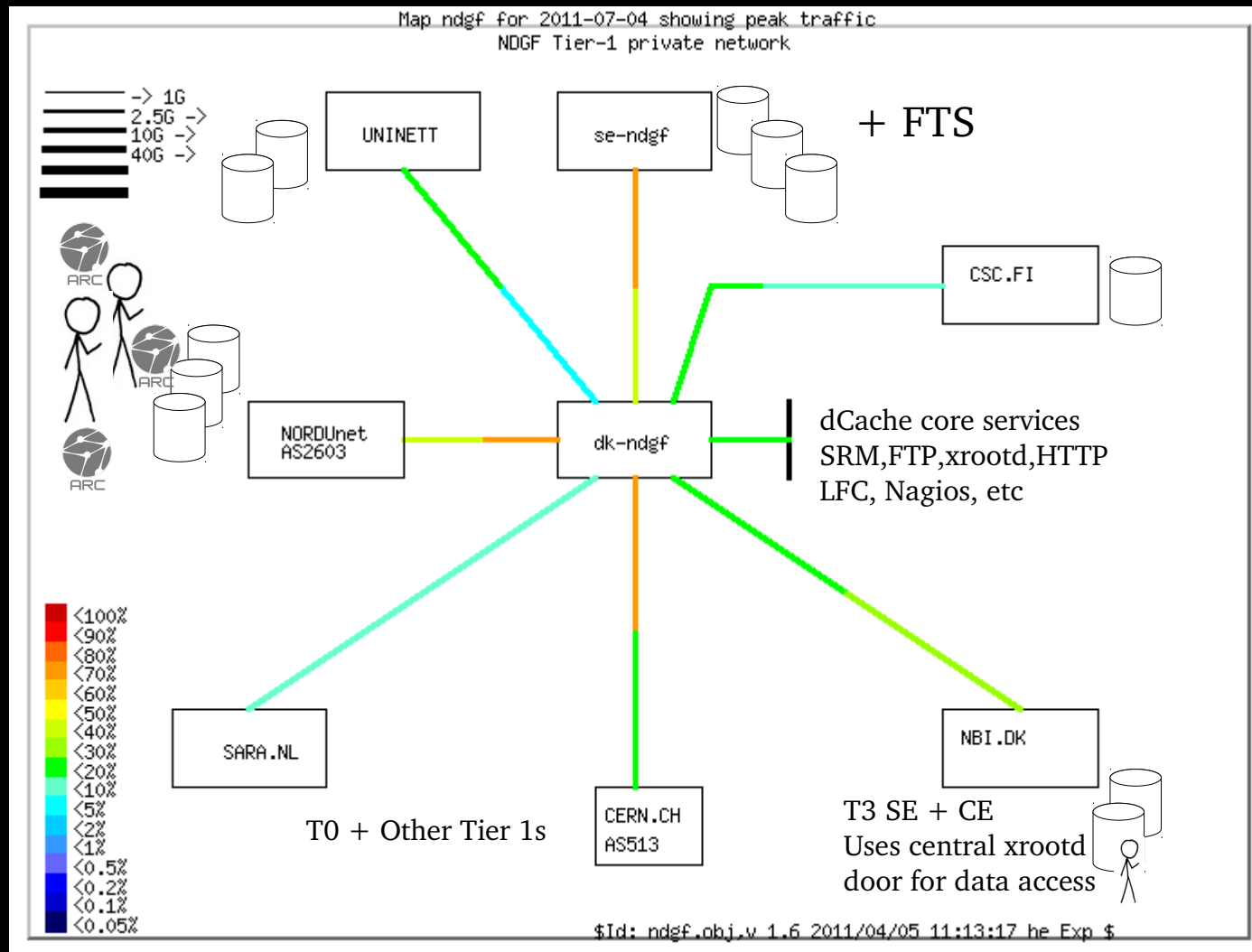














# Failure modes

- n sites means n times as many failures
- Write redundancy
  - As long as one site is up we can take data
- Limited read redundancy
  - When a site is down then data is unavailable
  - Some data is replicated
  - ARC job queue + data staging + caching means we can keep clusters filled even with multi hour pool downtimes
- Central components are critical

# Some strong points about dCache

- Orthogonality of name space, access protocol and physical placements of files
- Central management
- xrootd, SRM, GridFTP, HTTP(S)/WebDAV, etc.
- Multiple HSM systems
- Interfaces with external (local) HSM systems
- Java = minimal requirements on host environment, decent performance
- Relatively easy to adapt to new requirements

# Some weak points

- No network model
- File locality is hidden from external systems
- Limited local control over resources
- Central components that are not easily replicated
- Data may be unavailable

# Operational challenges

- Many small sites
  - Rolling upgrades are a must have
  - Response times for operational issues
  - Choosing the right hardware
  - SLA or no SLA?
  - Monitoring
- 24x7 support outsourced to NORDUnet NOC
  - Application experts only available during office hours

# Development

- Don't reinvent – join existing projects
- Choose software you can influence; hire developers to adapt and contribute – keep the software alive
- Think long term; the developers will become your application experts that stay when the operations team changes
- Interaction between development and operations important

# Future Developments

@NDGF, storage focussed

- Self-service portal
  - Fine grained authorization
  - Web interface for file management
  - Short-term password generator for non-X509 authentication
  - Integration with federated identity providers
- Move to FHS compliant packaging (& OSGi?)
- Decoupled, redundant and secure messaging
- More flexible data placement and movement