The NDGF Distributed Tier 1

Design and Experiences
(focusing on data handling)

Gerd Behrmann

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

NDGF
Nordic DataGrid Facility

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

NDGF
Nordic DataGrid Facility

Challenges

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

Technical differences

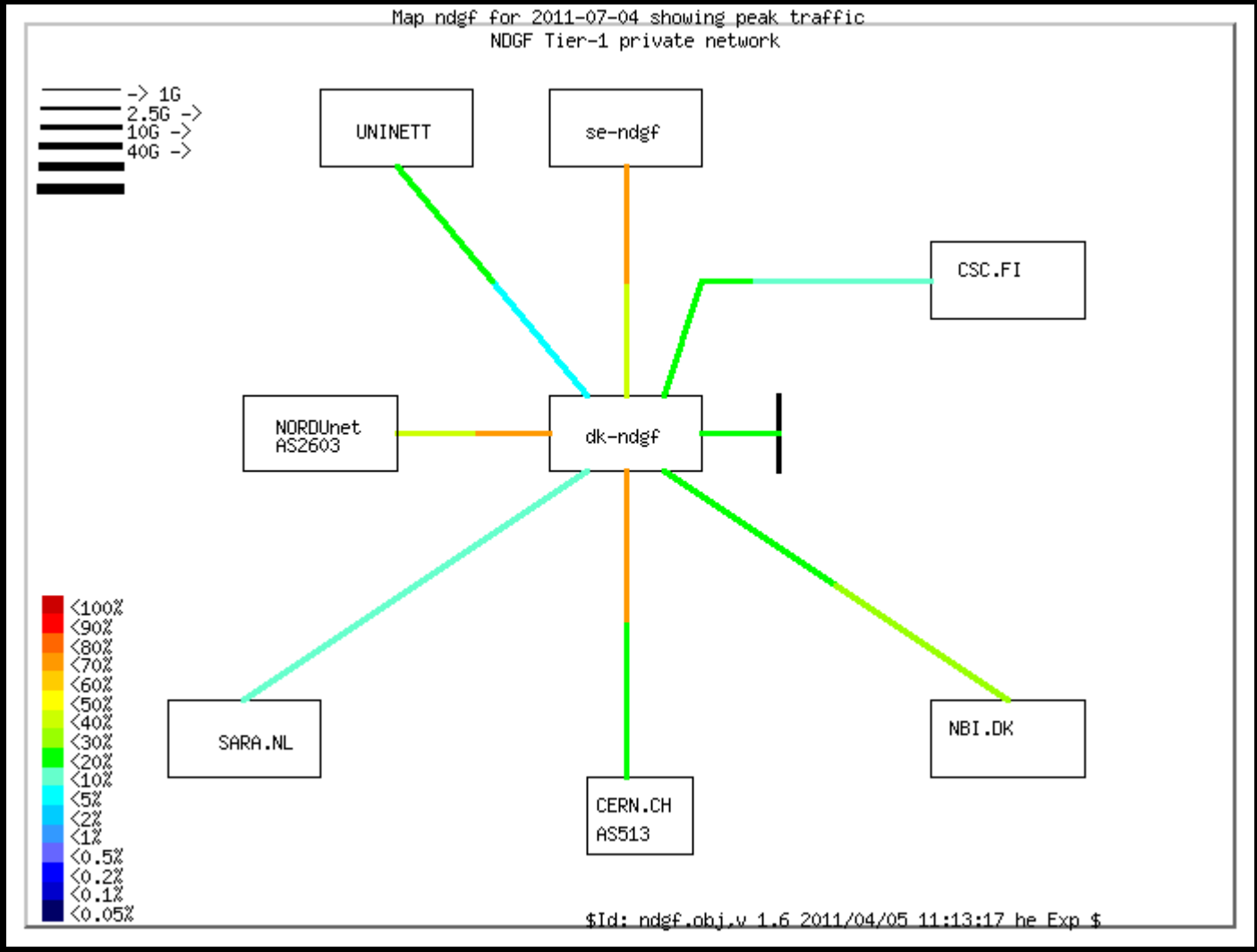
- Heterogeneous environments
- Heterogeneous interconnect
- Several independent tape systems
- Limited bandwidth
- High latency
- Frequent network failures
- Spanning many administrative domains

Goals towards participating sites

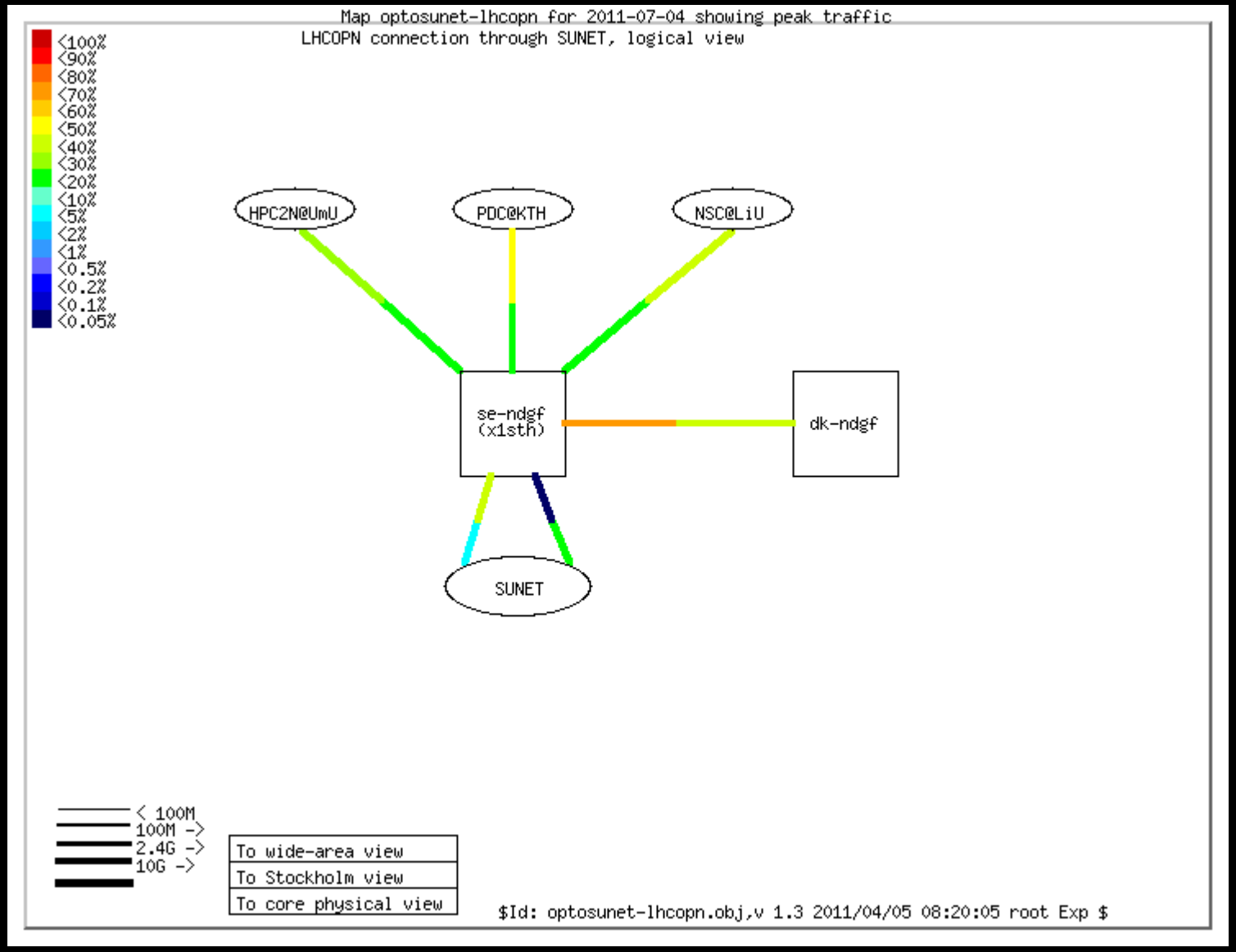
- Minimal operational requirements
- Minimal software requirements
- Free to choose OS
- Free to choose CPU architecture
- Free to choose storage architecture
- Free to choose storage density
- Free to choose HSM
- Combat fear to loose control
- …

Goals towards customers

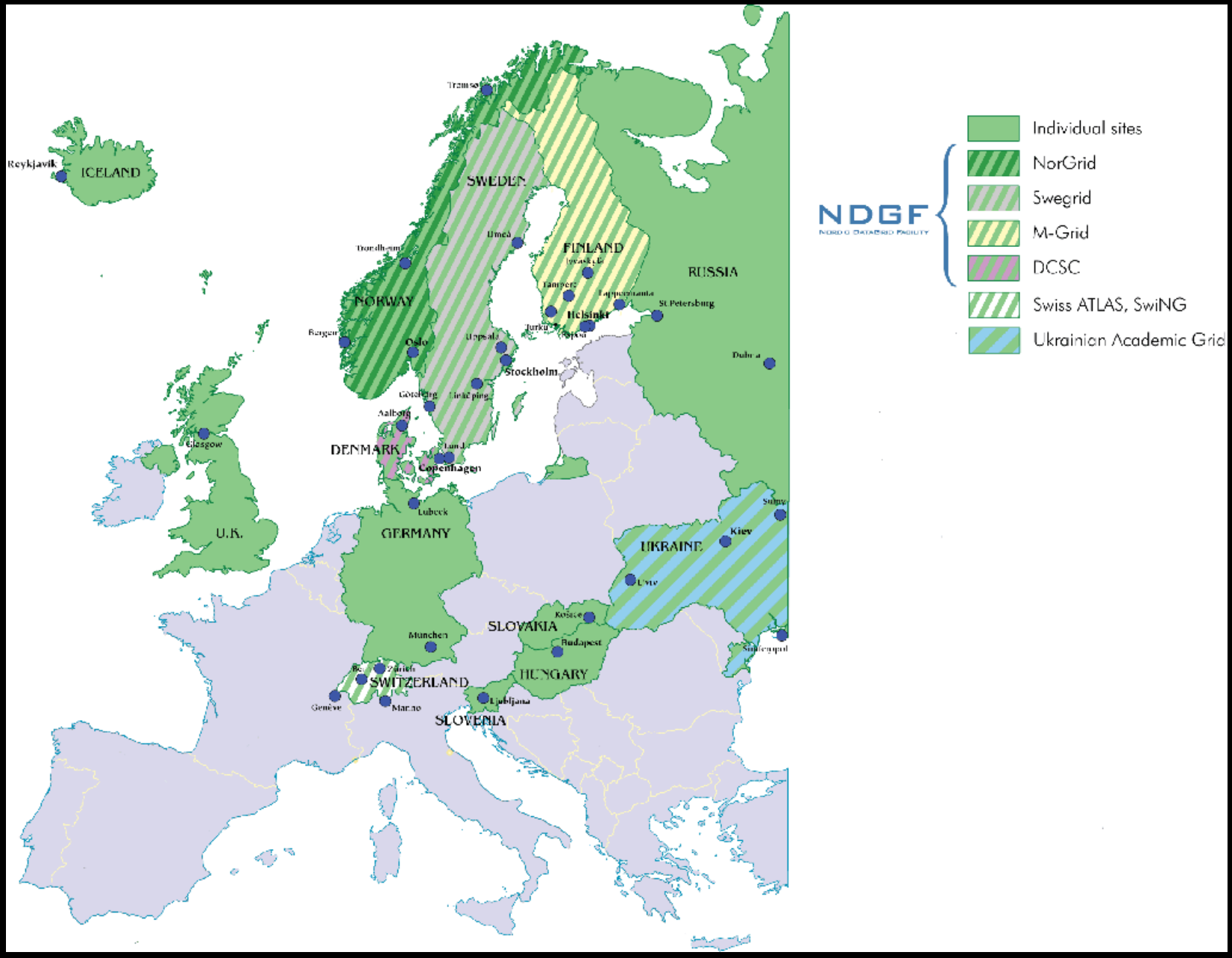
- Atlas and Alice Tier 1
- Stability
- First priority is data taking
- One operational contact point
- Hide the distributed nature
- Single SRM endpoint
- Non WLCG communities

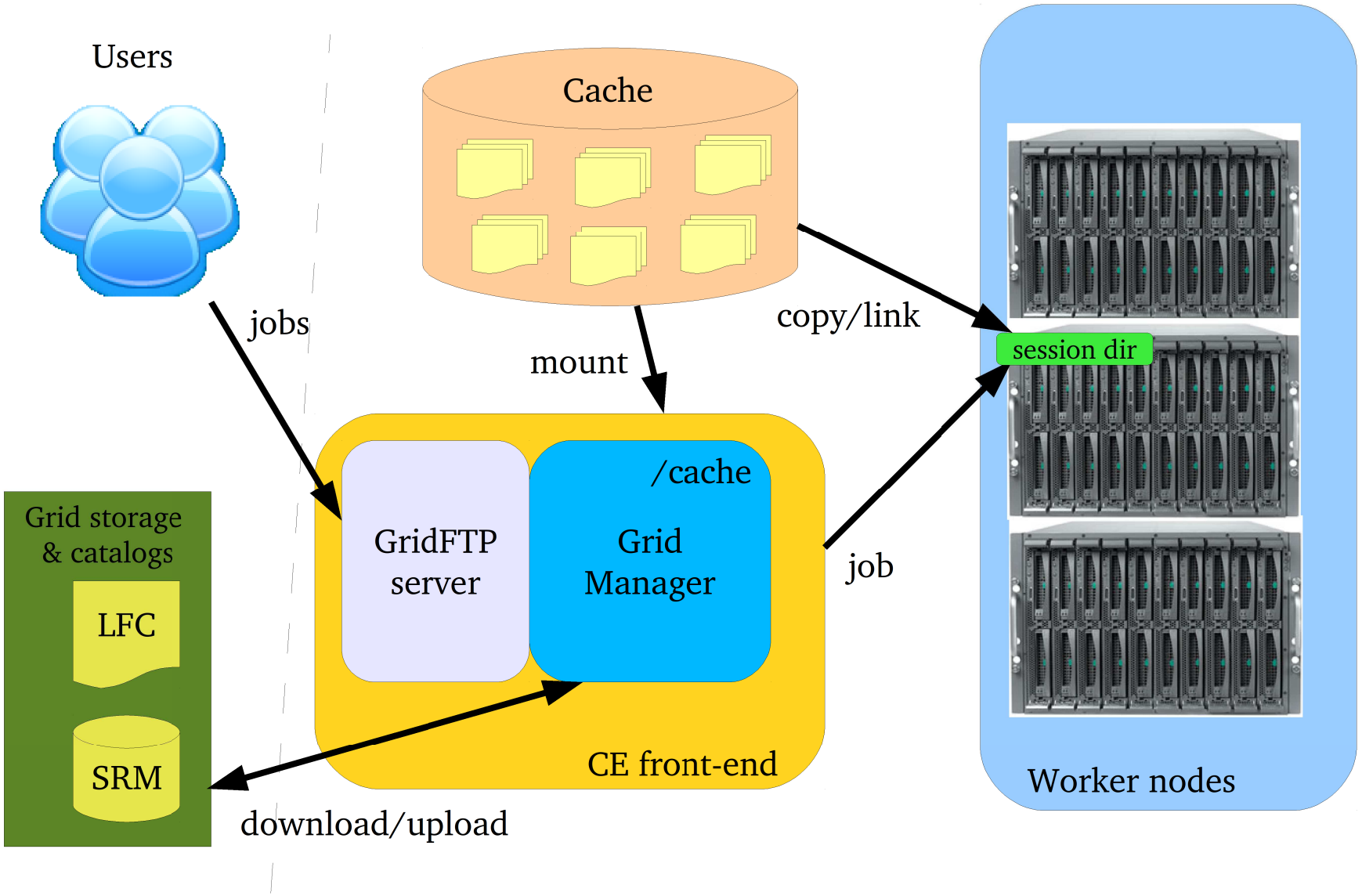


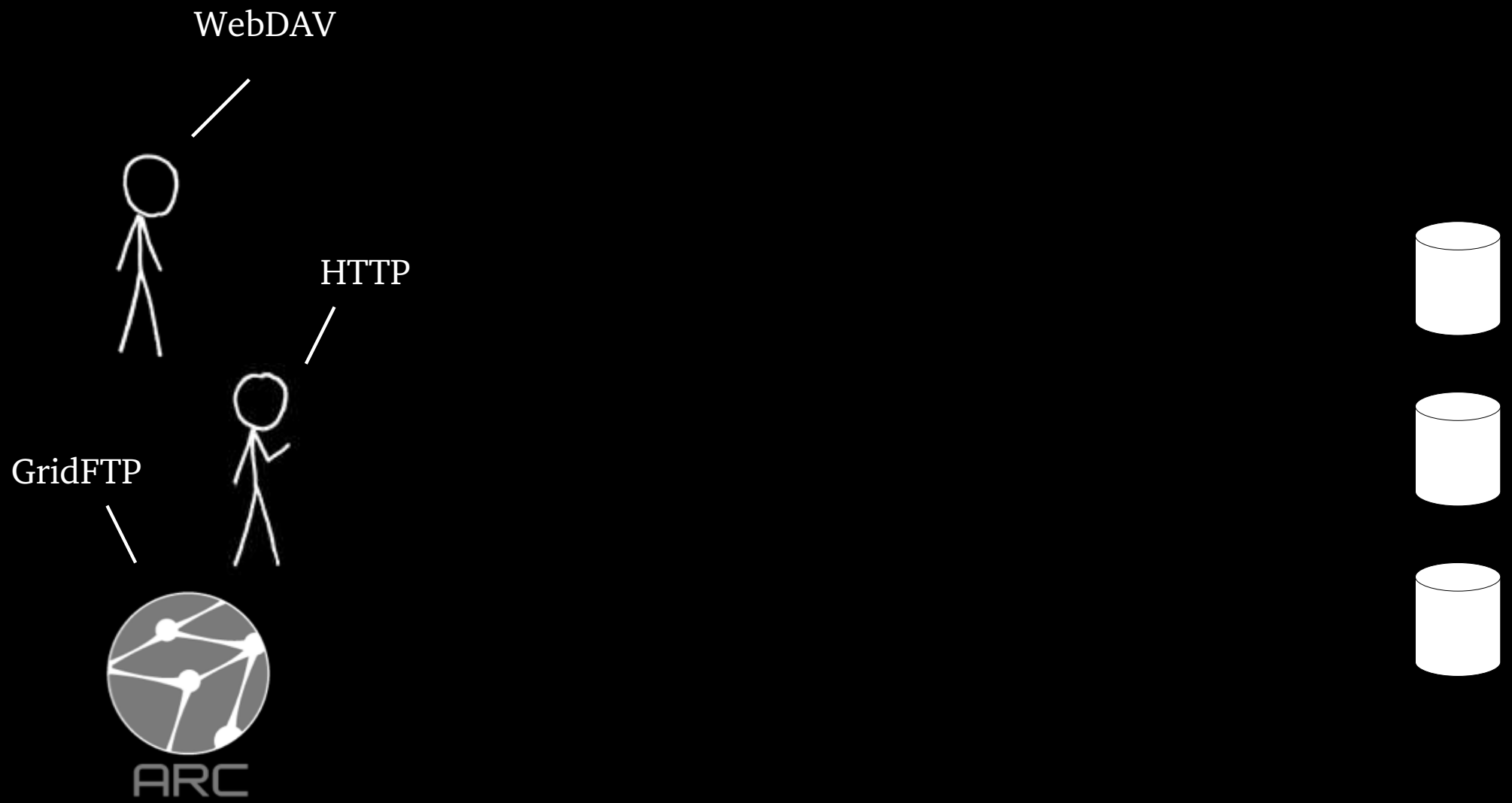
NDGF
Nordic DataGrid Facility

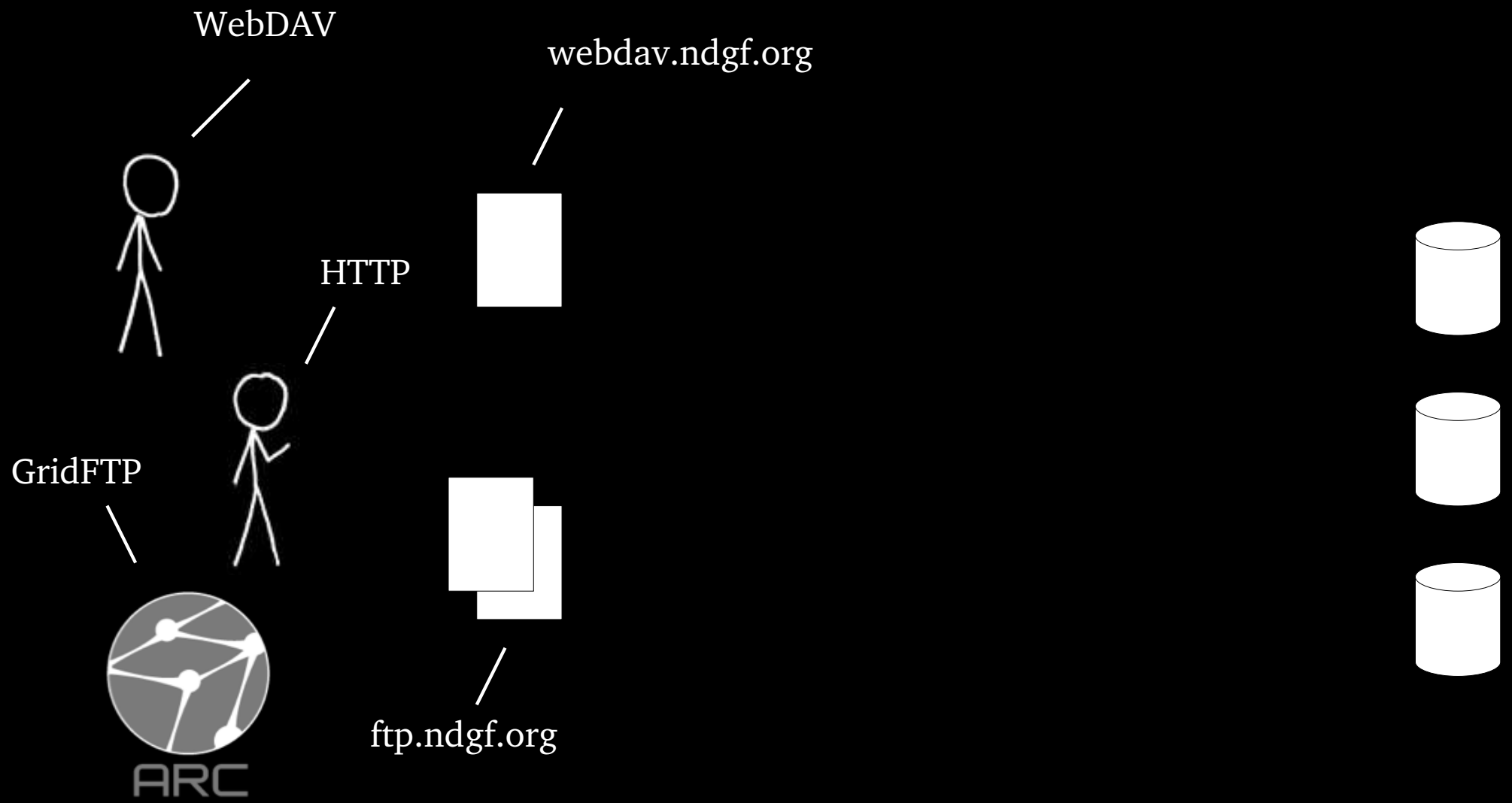


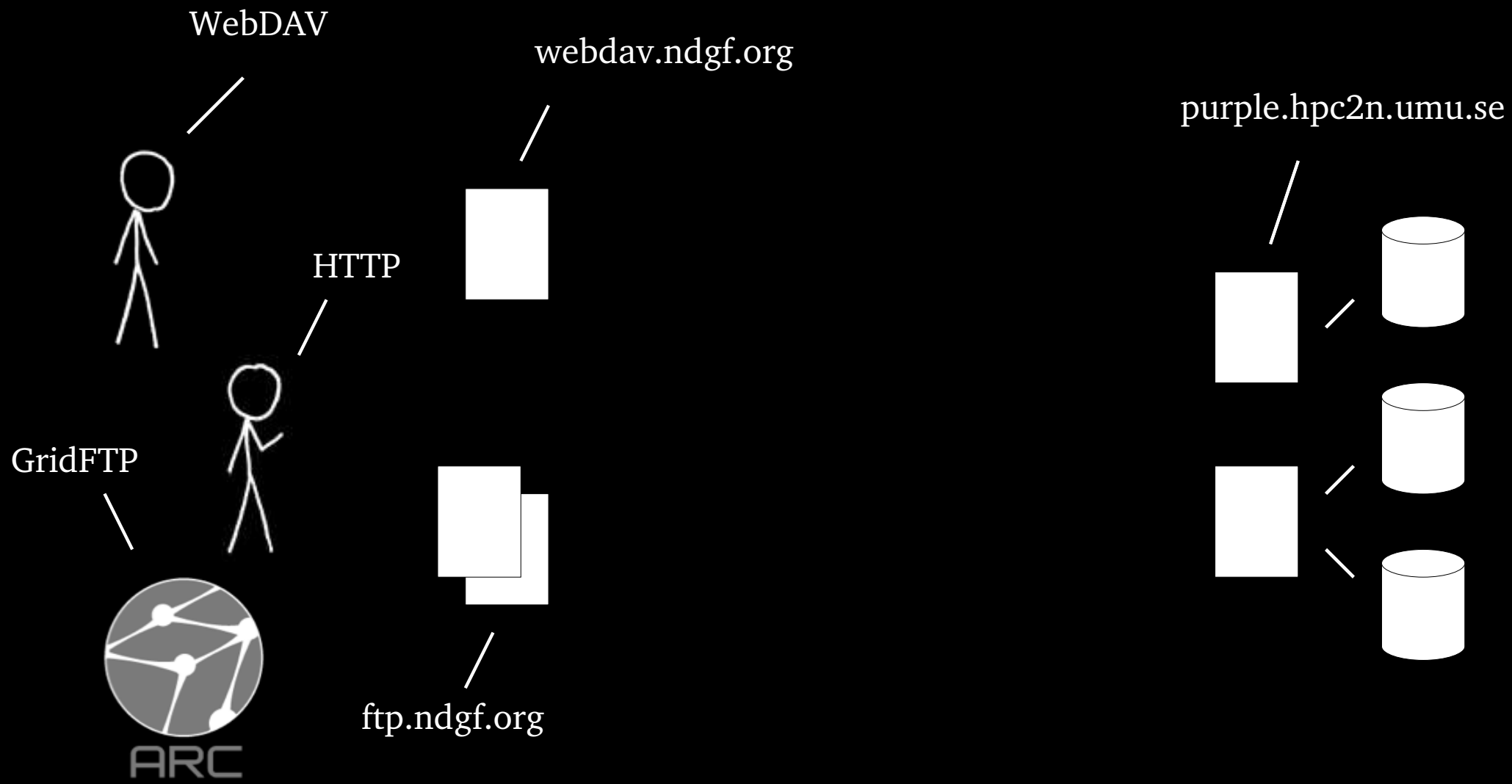
NDGF
Nordic DataGrid Facility


Reykjavik
ICELAND
Tromsø
SWEDEN
Umeå
Trondheim
FINLAND
RUSSIA
Tampere
Tapiovaara
St Petersburg
Helsinki
NORWAY
Bergen
Oslo
Uppsala
Turku
Stockholm
Linköping
Göteborg
Aalborg
Glasgow
DENMARK
Lund
Copenhagen
Lubeck
U.K.
GERMANY
Sumy
Kiev
UKRAINE
Lviv
Kosice
SLOVAKIA
München
Budapest
Simferopol
HUNGARY
Zurich
SWITZERLAND
Genève
Milano
Ljubljana
SLOVENIA
NDGF
NORDIC DATAGRID FACILITY
Individual sites
NorGrid
Swegrid
M-Grid
DCSC
Swiss ATLAS, SwiNG
Ukrainian Academic Grid

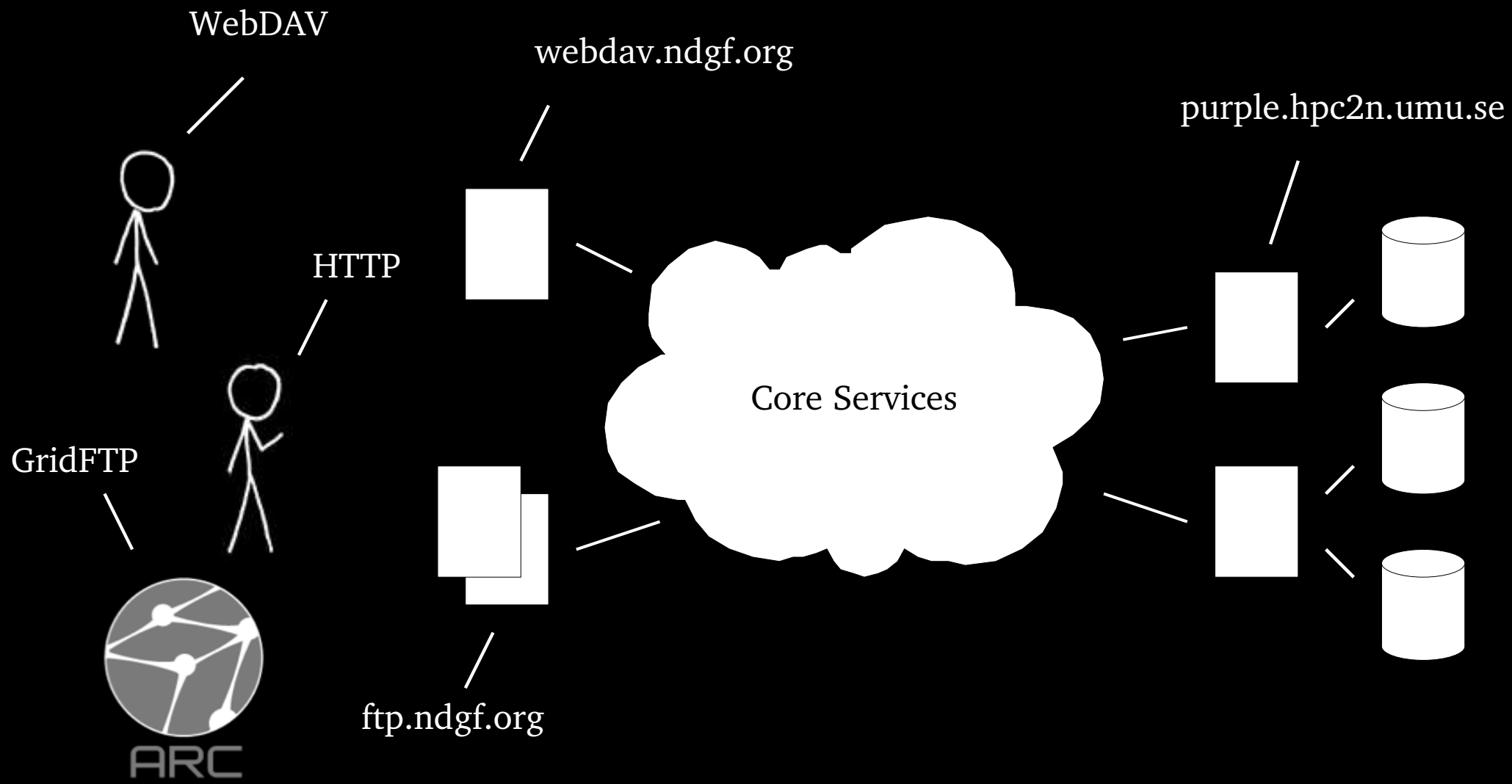

Users
Cache
jobs
mount
copy/link
session dir
/cache
GridFTP server
Grid Manager
job
Grid storage & catalogs
LFC
SRM
CE front-end
Worker nodes
download/upload


WebDAV
HTTP
GridFTP
ARC

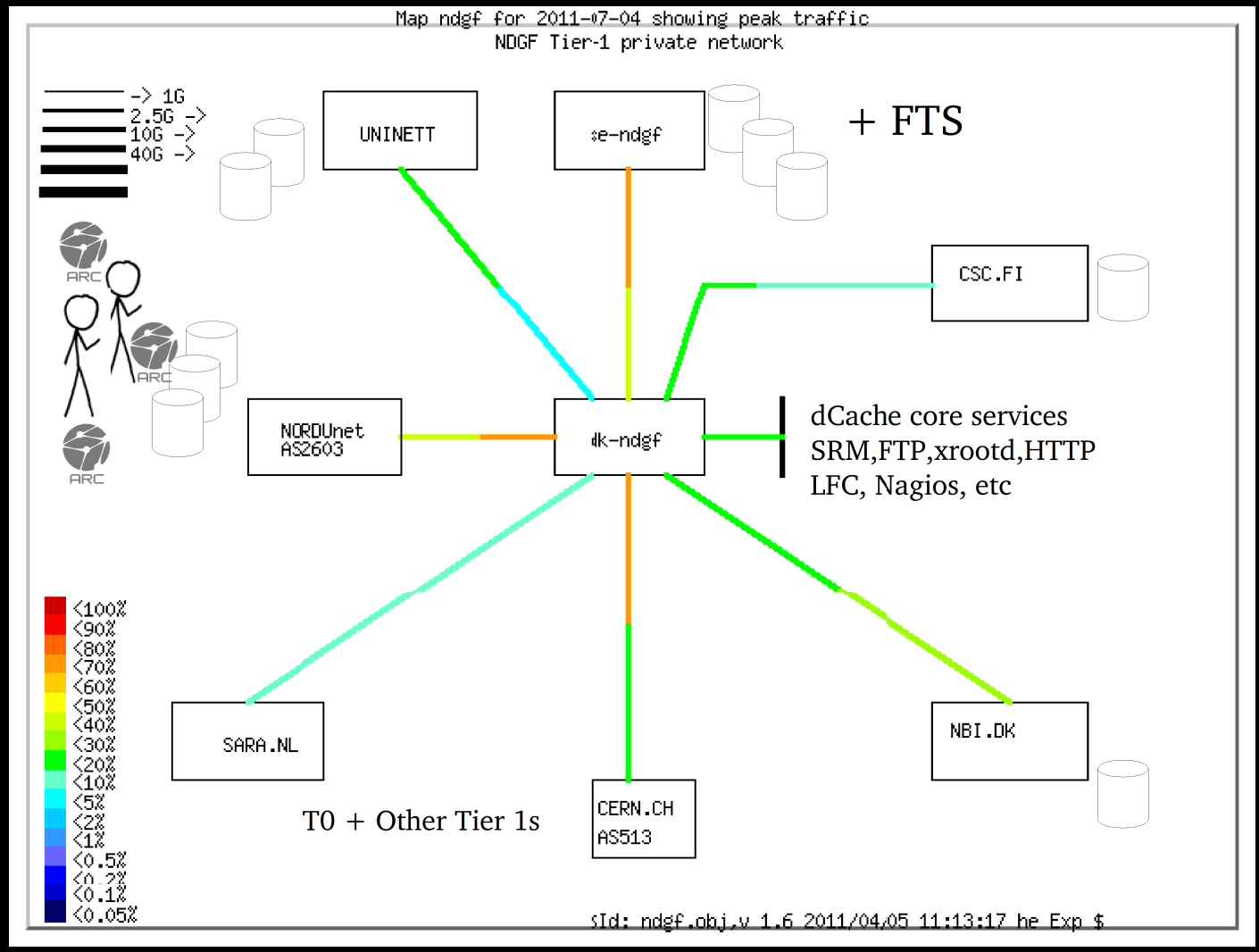

WebDAV
webdav.ndgf.org
HTTP
GridFTP
ftp.ndgf.org
ARC

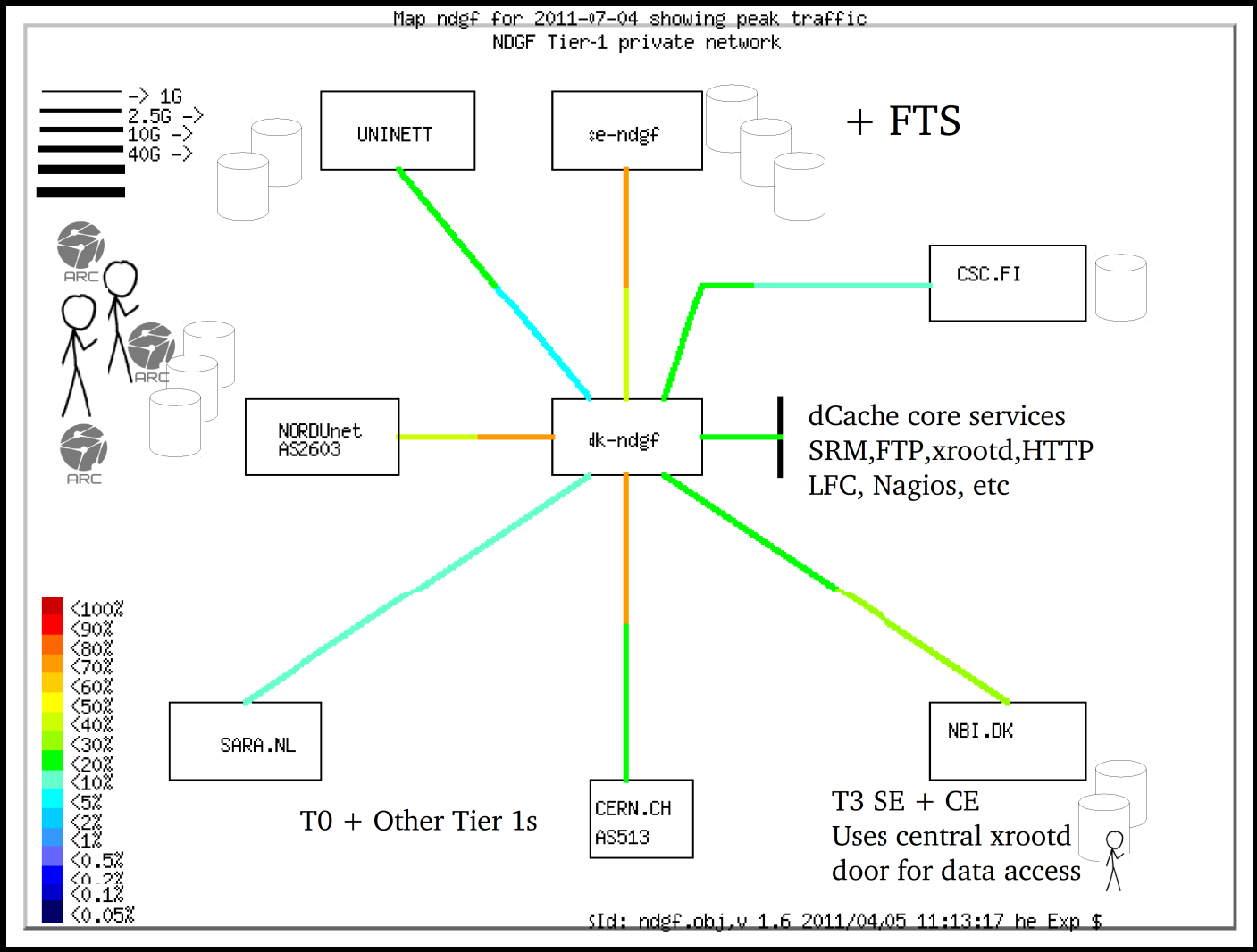


NDGF
NORDIC DATAGRID FACILITY



NDGF
NORDIC DATAGRID FACILITY


Map ndgf for 2011-07-04 showing peak traffic
NDGF Tier-1 private network
-> 1G
2.5G ->
10G ->
40G ->
UNINETT
se-ndgf
+ FTS
CSC.FI
NORDUnet
AS2603
dk-ndgf
dCache core services
SRM,FTP,xrootd,HTTP
LFC, Nagios, etc
<100%
<90%
<80%
<70%
<60%
<50%
<40%
<30%
<20%
<10%
<5%
<2%
<1%
<0.5%
<0.2%
<0.1%
<0.05%
SARA.NL
NBI.DK
T0 + Other Tier 1s
CERN.CH
AS513
$Id: ndgf.obj,v 1.6 2011/04/05 11:13:17 he Exp $


Map ndgf for 2011-07-04 showing peak traffic
NDGF Tier-1 private network
-> 1G
2.5G ->
10G ->
40G ->
UNINETT
se-ndgf
+ FTS
CSC.FI
NORDUnet
AS2603
dk-ndgf
dCache core services
SRM,FTP,xrootd,HTTP
LFC, Nagios, etc
<100%
<90%
<80%
<70%
<60%
<50%
<40%
<30%
<20%
<10%
<5%
<2%
<1%
<0.5%
<0.1%
<0.05%
SARA.NL
CERN.CH
AS513
NBI.DK
T0 + Other Tier 1s
T3 SE + CE
Uses central xrootd
door for data access
$Id: ndgf.obj,v 1.6 2011/04/05 11:13:17 he Exp $

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