Hands-on session with Globus 5

GridKa School 2011

7. September 2011, Karlsruhe Germany

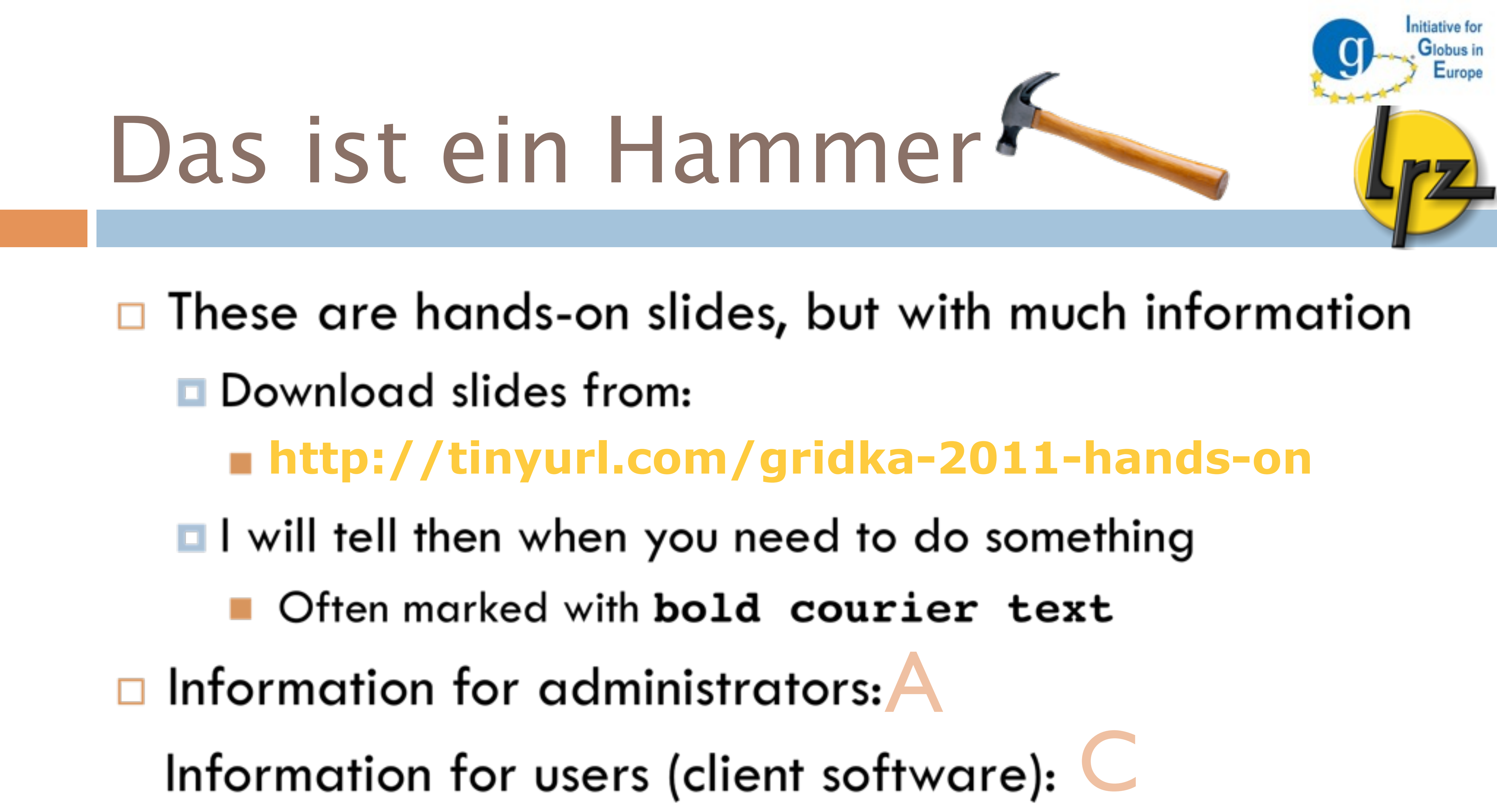
Florian Zrenner (zrenner@lrz.de)
Adrian Colesa (adrian.colesa@cs.utcluj.ro)

Slides from Jarno Laitinen

Overview of the hands-on session

Goal: to be able to act as a Globus admin/user

- Authentication and authorisation
  - Certificates
  - Authorisation file
- Globus components: configuration and usage
  - Interactive access
  - Data transfer
  - GlobusOnline – file transfer service
  - Job submission
  - MyProxy proxy storage service



Das ist ein Hammer

- □ These are hands-on slides, but with much information
  - Download slides from:
    - **http://tinyurl.com/gridka-2011-hands-on**
  - I will tell then when you need to do something
    - Often marked with **`bold courier text`**
- □ Information for administrators: A

  Information for users (client software): C

Questions: Who..
- might install Globus in future (not just use it)?
- is familiar with Globus, but expects to hear about GT5?

A Installation: Overview

- □ Where to download Globus
- □ How to install it from source
- □ How to use IGE repository

A

Installation: where to find GT5?

- GT 5.0.4 download available at http://globus.org.eu/toolkit
  - Documentation, Downloads and Support
- Source available
  - Builds on Ubuntu, Apple OS X, RedHat, Fedora Core, Debian, SuSE, FreeBSD, and Solaris
- IGE Releases
  - Repositories for Fedora, Red Hat, Debian and Ubuntu
  - rpms and debs

Setup Today

- Every attendee has it own instance
- Facts
  - Ubuntu 1 ECU (EC2 Compute Unit)
    0.6 GB RAM
  - globus compiled /usr/globus-src/
    - but not installed
  - Host certificates (/root/.)
  - User certificates (/home/ubuntu/.globus/.)

Login to your hands-on machine

- Ready to login
  - Windows? Download PuTTY:

http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
(http://bit.ly/1kyS98)

- Login to your personal host:

```
ssh -l ubuntu <your personal host>
```

Add IGE Repository

- Browse to http://www.ige-project.eu/releases/downloads
- Add IGE repo key to the system
  - `wget -q http://repo-deb.ige-project.eu/DEB-GPG-KEY-IGE.asc -O- | sudo apt-key add -`
- Download IGE repo data (as root: **sudo su -**)

```
cd /etc/apt/sources.list.d
wget http://tinyurl.com/ige-list
mv ige-list ige.list
```

- Reload repository
  - `apt-get update`

Install Software From IGE Repository

- Find a list of available software on:
  - http://www.ige-project.eu/guide/component-installation-guide

  - `apt-get install gsissh-client`
  - `apt-get install globus-gass-copy-progs`
  - `apt-get install globus-gram-client-tools`

Already installed: `globus-proxy-utils`

A Info: Download and compilation

- Create user "`globus`" and group "`globus`"

```
useradd -m globus -G globus
```

- Installation directory for globus

```
mkdir /opt/globus
chown globus:globus /opt/globus/
```

- Install "`make`" environment
  - On ubuntu meta-package „build-essential“ works
  - Package **libssl-dev** is needed to build globus

Info: Download and compilation

- As user **`globus`** (**`su - globus`**)
  - Download of Globus 5.0.4 sources
  - Extract:
    - `tar xjvf gt5.0.4-all-source-installer.tar.bz2`
  - In extracted directory run `configure`
    - `cd gt5.0.4-all-source-installer`
    - `./configure --prefix=/opt/globus`
  - Compile: `make`
  - Install: `make install`
- It is also possible to install only some GT tools
  - **`cd /usr/globus-src/gt5.0.4-all-source-installer`** (as user globus) **`make install gsi-myproxy`**

AUTHENTICATION & AUTHORISATION

A

Globus environment variables

```
export GLOBUS_LOCATION=/opt/globus
export PATH=$GLOBUS_LOCATION/bin:$PATH
export LD_LIBRARY_PATH=$GLOBUS_LOCATION/lib:$LD_LIBRARY_PATH
export GLOBUS_TCP_PORT_RANGE=20000,25000
export GLOBUS_USAGE_OPTOUT=1
export GLOBUS_HOSTNAME=<set hostname here>
```

(csh / tcsh: `setenv GLOBUS_LOCATION /opt/globus`)

Globus environment variables

- Globus environment should be loaded automatically
- **`as root (sudo su -):`**
  - **`cd /etc/profile.d/`**
  - **`wget http://tinyurl.com/grid-env-sh`**
  - **`mv grid-env-sh grid-env.sh`**
  - log-out from your hands-on machine / log-in again
  - Test: **`echo $GLOBUS_LOCATION`**

A

CA certificates

- To authenticate certificates the **Certificate Authority (CA) files** are needed.
  - Globus requires
    - <hash>.0 and
    - `<hash>.signing_policy` files.
      - The unique <hash> is a digest of the **subject name** of the CA
- CA files can be found e.g. via search-by-country functionality on http://www.eugridpma.org/

EUGridPMA files via repository

- Install the EUGridPMA PGP key for apt

  -
    ```
    wget -q -O - https://dist.eugridpma.info/distribution/igtf/current/GPG-KEY-EUGridPMA-RPM-3 | apt-key add -
    ```
- Add the repo to your sources.list file (for APT)

  -
    ```
    deb http://dist.eugridpma.info/distribution/igtf/current igtf accredited
    ```
- Populate the cache and install the meta-package

  -
    ```
    apt-get update
    ```
- Install one or more of the Profiles you want to accept

  -
    ```
    apt-get install ca-policy-igtf-classic
    ```

Installation of CA certificates

- Installation directory:
  `/etc/grid-security/certificates`

- For this workshop a special package (including the simpleCA CA files) was created
  - As **root** (**sudo su –**): Download and unpack CA certificates:

```
mkdir /etc/grid-security/certificates
cd /etc/grid-security/certificates/
wget http://tinyurl.com/workshopcacert
tar zxvf workshopcacert
```

A Certificate revocation list

- Each CA maintains a file of revoked certificates.
- `<hash>.crl_url` in certificates directory point to URL to download `<hash>.r0` files.

- There is a tool to update the files:
  - http://dist.eugridpma.info/distribution/util/fetch-crl/
- If not up-to-date an authentication may fail

  → Add fetch-crl as a cron tab

- Globus command for CA check:
  `grid-cert-diagnostics`

A Installation of Host Certificates

- For this workshop host certificats for your machines were
  - already requested
  - signed by our workshop SimpleCA
  - copied to YourHandsOnMachine:/root/.
    - secret `hostkey.pem` and
    - signed `hostcert.pem` (Make sure it is only readable by your user)
- Copy host certificate in place. As **root** **(sudo su –)**
    - **`mv /root/host* /etc/grid-security/`**
- Check your host certificate
  - **`cat /etc/grid-security/hostcert.pem`**

User certificate

- As user ubuntu (su ubuntu): `grid-cert-info`

  (equals: `openssl x509 -in $HOME/.globus/usercert.pem -text -noout`)
- Create proxy

  As user ubuntu: `grid-proxy-init`

  (Password: IGE2011)
- To view information about the generated proxy (e.g. DN, validity time):

  `grid-proxy-info`

Certificate security issues

- The proxy file is readable only by your account
- Default location: `/tmp/x509up_${UID}`
- By default valid for 12 hours (`-valid <h:m>`)
- For security reasons you can delete your proxy on the machine when you do not need it anymore
  - `grid-proxy-destroy`

MyProxy

Client part

Store Credentials On MyProxy

```
myproxy-init -l <your last name> -s myproxy.lrz.de
```

- It will prompt for the passphrase of your private key (It will not use your existing proxy credentials)
- Will prompt twice for new passphrase to protect your uploaded credential on the MyProxy server
  - **Don´t** use the same passphrase as for your private key

MyProxy Tools

- To view status of the proxy **at** MyProxy server:

```
myproxy-info -l <your last name> -s myproxy.lrz.de
```

- To remove the proxy from MyProxy server:

```
myproxy-destroy -l <username> -s <server>
```

Retrieve proxy certificate

- To destroy local credential
  - **`grid-proxy-destroy`**
- To view your proxy status at the client machine:

```
grid-proxy-info
```

- To retrieve proxy from MyProxy:

```
myproxy-logon -l <your last name> -s myproxy.lrz.de
```

  - `-t <lifetime>` of proxy in hours (by default 12 h)
    - This cannot be greater that what was set with `-t` in `myproxy-init`
  - **`grid-proxy-info`**

MyProxy Tips (myproxy-init)

- Default MyProxy server can be set with environment variable: `MYPROXY_SERVER`

```
export MYPROXY_SERVER=<set myproxy host here>
```

- Credential lifetime on myproxy: **`-c <hours>`** (default one week=168h)

- Proxy lifetime of from MyProxy retrieved proxies:
  -t <hours> (default: 12 h)

AUTHORISATION

Who can use the service?

A grid-mapfile

- Check you certificate's Distinguished Name (DN):

```
grid-cert-info -subject
```

- Map your DN to user account (ubuntu):
- A root (sudo su -):

  -
    ```
    $GLOBUS_LOCATION/sbin/grid-mapfile-add-entry \
    -dn "YOUR DN" -ln ubuntu
    ```
  - (verify with `cat /etc/grid-security/grid-mapfile`)

- Info: To delete an entry:
  ```
  grid-mapfile-delete-entry -dn "<Distinguished Name>" -ln <user>
  ```
- To check for duplicate DNs and if mapped linux accounts exists:

```
$GLOBUS_LOCATION/sbin/grid-mapfile-check-consistency
```

Additional information (1)

- For testing and internal purpose Globus provides SimpleCA to act as a Certificate Authority.

- `$GLOBUS_LOCATION/setup/globus/setup-simple-ca` script can install CA files to any directory with `-dir` option. See more options with `-usage`.

- The script will create a tar.gz packet to be distributed on the machines where needed.
  - To create a certificate sign request for this CA

- As CA: to sign the certificate request:
  - **grid-ca-sign -in usercert_request.pem -out usercert.pem**

- See also SimpleCA Admin Guide: http://bit.ly/cDdC8q

Additional information (2)

- **Time** settings of client and server must be within 5 minutes tolerance (otherwise the authentication can fail)
  - e.g. as root: `ntpdate ntp1.lrz.de`

- Host certificate DN must have the fully qualified domain name (FQDN)
  - hostname –f
- If the host certificate does not match FQDN the client needs to specify the DN in Globus command parameter

INTERACTIVE ACCCESS via GSI-OpenSSH

GSI-SSH: sshd configuration init.d script

A

□ Globus provides init.d script for gsissh daemon

□ As **root** on hands-on machine:

```
cp $GLOBUS_LOCATION/sbin/SXXsshd \
   /etc/init.d/gsisshd

chmod 744 /etc/init.d/gsisshd
```

GSI-SSH: sshd_config and ssh_config

- In `sshd_config` (server) and in `ssh_config` (client)

```
cd  $GLOBUS_LOCATION/etc/ssh/
```

- As **root** change port to „2222“:

`Port 2222` (no comment mark #!)

- You can disable protocols which you do not need

```
Protocol 2
RSAAuthentication no
PubkeyAuthentication no
PasswordAuthentication no
ChallengeResponseAuthentication no
```

GSI-SSH: Log Separation

- To differentiate from normal `sshd` in log file `/var/log/messages`:

  -
    ```
    mv $GLOBUS_LOCATION/sbin/sshd \
    $GLOBUS_LOCATION/sbin/gsisshd
    ```
  - In `/etc/init.d/gsisshd` correct „`Description`“ row and „`SSHD`“ row to:

    ```
    # Description: Start the sshd daemon
    SSHD=${sbindir}/gsisshd
    ```

- Disable the usage statistic collection by adding:

  ```
  GLOBUS_USAGE_OPTOUT=1
  ```

GSI-SSH: configuration A additional information

- "If compiled with **PAM support** (`--with-pam`) set `"UsePAM yes"` in `$GLOBUS_LOCATION/etc/ssh/sshd_config` after installation. "

- If compiled with **TCP wrapper** edit `/etc/hosts.allow` e.g. `2222:ALL:ALLOW`

- Privilege separation method: See the required steps:
  http://grid.ncsa.illinois.edu/ssh/admin.html#privsep

A

GSI-SSH: /etc/services and start-up

- Edit `/etc/services` e.g. for `netstat -tap`:

```
gsisshd 2222/tcp
```

- To start it now (as root):

```
service gsisshd start
```

- To start gsissh during the boot:

```
update-rc.d gsisshd defaults
```

GSI-SSH

Client Part

GSI-SSH: gsissh client

- Usage of command line client:
  - Syntax: `gsissh [-p <port>] host`
    - Use full host name
    - Debug: `-v` or `-vv`
    - By default it uses the **port** set in
      - `$GLOBUS_LOCATION/etc/ssh/`**`ssh`**`_config`
- Login as first local account found in grid-mapfile

GSI-SSH: gsissh client

- As your user ubuntu:
  - `grid-proxy-init` (if not yet done)
  - `grid-proxy-info`
  - `gsissh -p 2222 localhost`
  - `exit`
  - `gsissh gt5-ige.drg.lrz.de` (account name!)
  - `grid-proxy-info`
  - From there hop to TU Dortmund cluster
    - `gsissh -p 2222 udo-gt01.grid.tu-dortmund.de` (check the account name there!)
  - Create a 10MB file there
    - `dd if=/dev/zero bs=1024 count=10000 of=10MB`

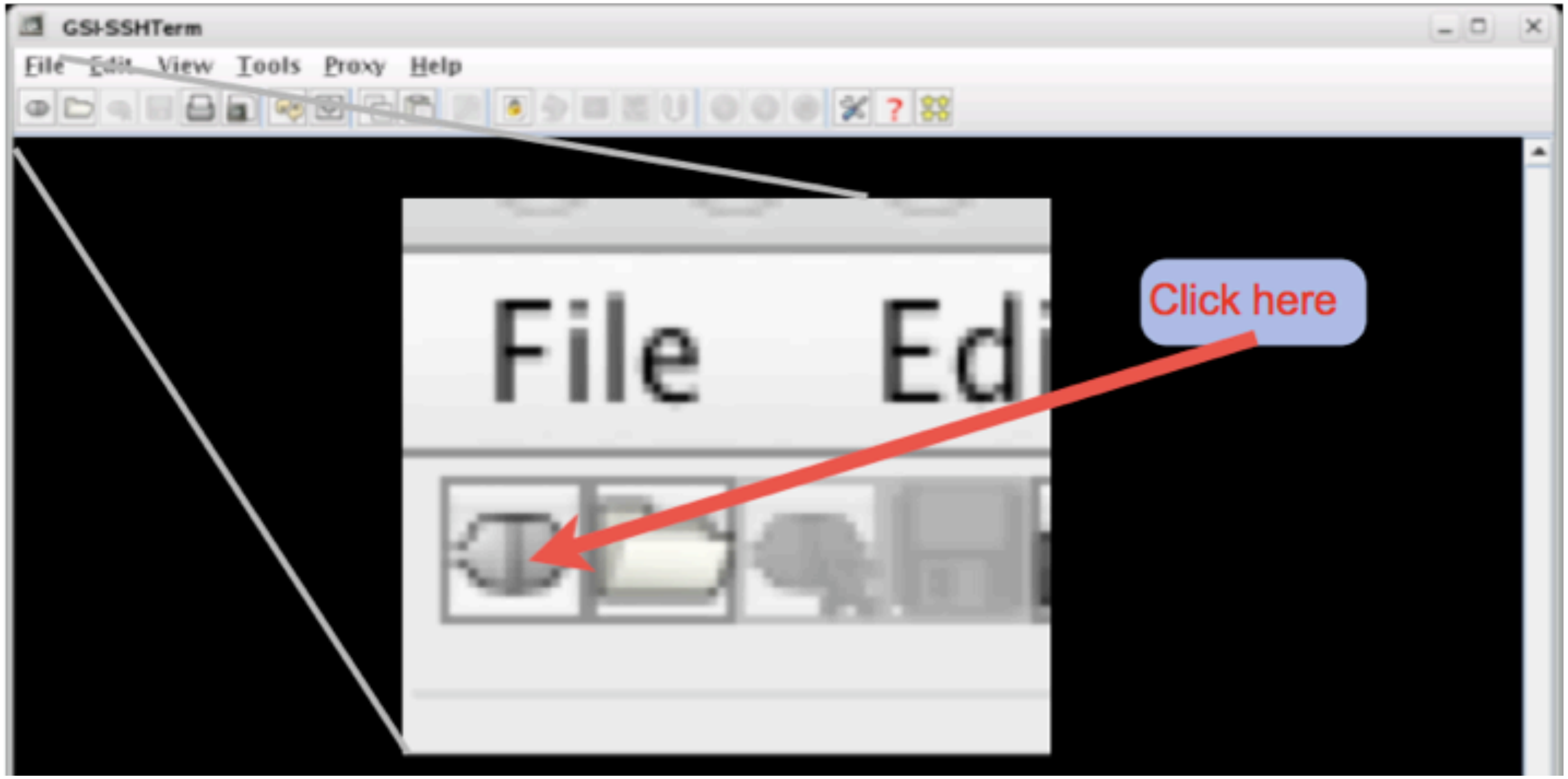
Java Webstart GSISSH-Term

- Java terminal client.
- 3rd party software (not from Globus)

- On your local operating system:
  Open Java Webstart GSISSH-Term
  - **Surf to** **http://tinyurl.com/gsissh-term**
    - There appear two "digital signature cannot verified" windows which you have to accept.
    - It is IGE´s version which takes care of CA certificates and supports using Safari browser certificate.

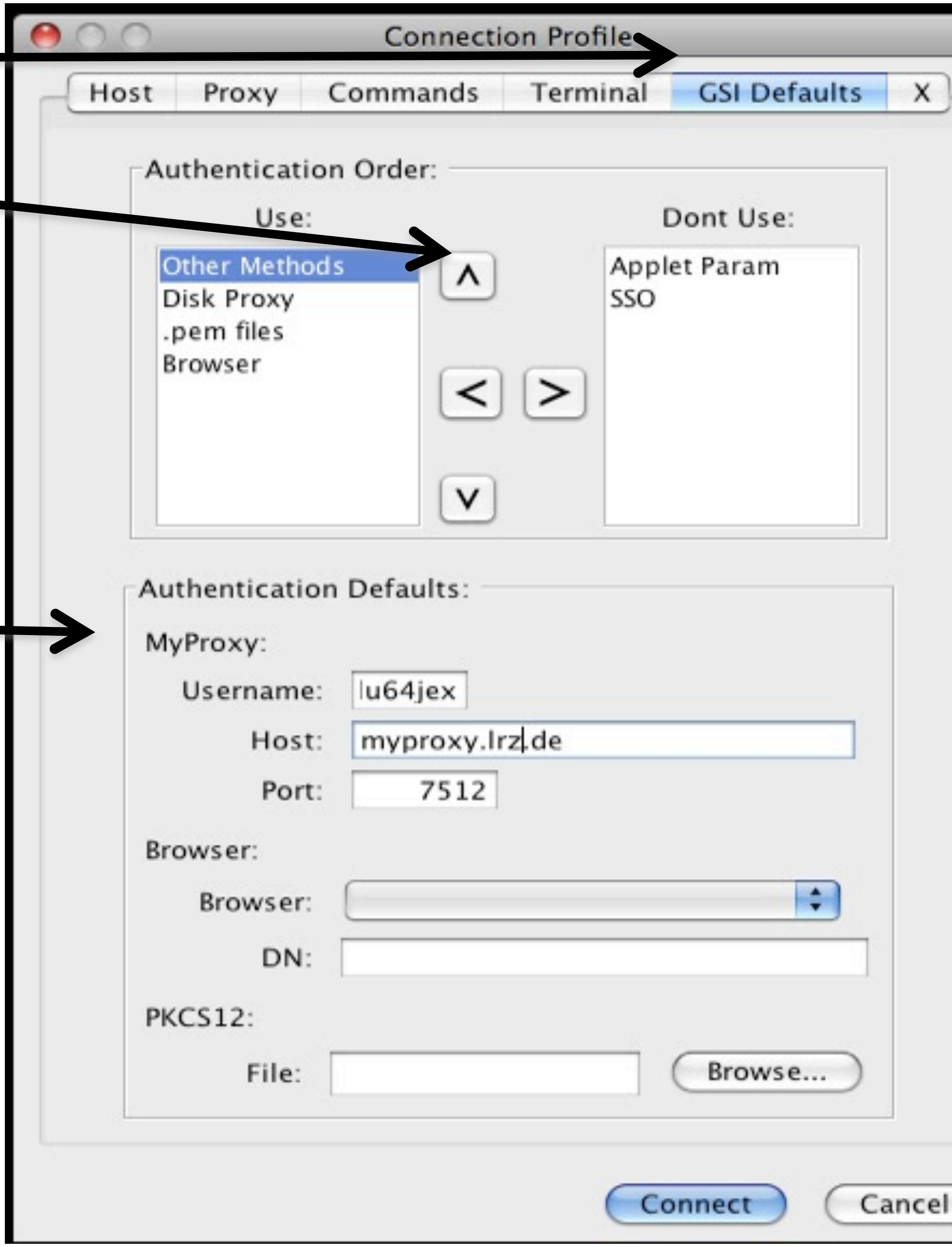
GSISSH-Term: start-up

- Login to IGE grid
  - Open a login window (see picture).
  - Connect to **gt5-ige.drg.lrz.de**
  - Click „Advanced“



Using Myproxy with GSI-SSH TERM

- Tab „Gsi-Defaults“
- Put "Other Methods" all to the top!
- Enter MyProxy information



Authentication Order:

Use:	Dont Use:
Other Methods	Applet Param
Disk Proxy	SSO
.pem files	
Browser	

Authentication Defaults:

MyProxy:

Username: lu64jex

Host: myproxy.lrz.de

Port: 7512

Browser:

Browser:

DN:

PKCS12:

File:

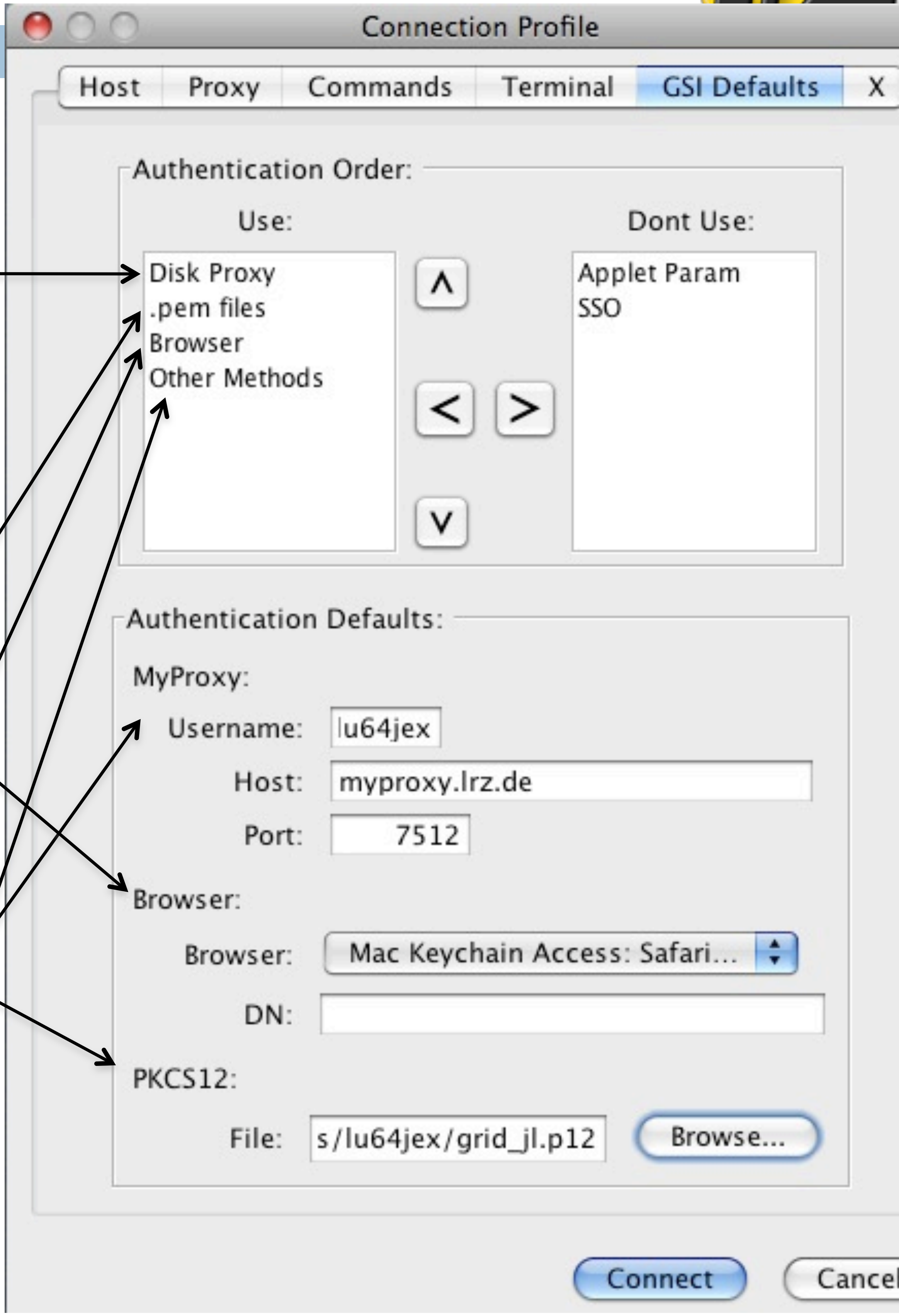
Information: GSISSH–Term Authentication methods

A proxy can exists already due:
- previous GSISSH-Term session
- Globus grid-proxy-init
- Short Lived Credential Service

- .pem files in .globus directory

- A web browser certificate

Other Methods:
- MyProxy service
- PKCS12 (.p12) file



Authentication Order:

Use:	Dont Use:
Disk Proxy	Applet Param
.pem files	SSO
Browser	
Other Methods	

Authentication Defaults:

MyProxy:
Username: lu64jex
Host: myproxy.lrz.de
Port: 7512

Browser:
Browser: Mac Keychain Access: Safari...
DN:

PKCS12:
File: s/lu64jex/grid_jl.p12

DATA TRANSFER with GridFTP

GridFTP: Overview

- Administration
  - Start-up script (xinetd)
  - Firewall issues

- Client
  - Globus `globus-url-copy`

A GridFTP with xinetd

- xinetd template:

```
instances    = 100
socket_type = stream
bind         = <your host IP>
env         += GLOBUS_TCP_PORT_RANGE=20000,25000
env         += GLOBUS_HOSTNAME=<YOUR HOSTNAME>
server       = /opt/globus/sbin/globus-gridftp-server
server_args = -i -l /opt/globus/var/gridftp.log \
               –d error,warn,info
port          = 2811
```

GridFTP: xinet.d script

- As **root** on hands-on machine:

```
apt-get install xinetd
cd /etc/xinetd.d
wget http://tinyurl.com/gsiftp
```

- Uncomment and correct:
  - The hostname for `GLOBUS_HOSTNAME` (find out with **`hostname –f`**)
    - `vi gsiftp`

A

GridFTP: firewall (1)

- **Control process** port is by default: 2811

- **Data port range**
  - Varies often from a hundred to some thousands
  - Port range should be set as your grid partner sites have.
  - The needed amount depends on the estimated amount of the clients.

A

GridFTP: firewall (2)

- Client set port range for outgoing firewall

-
  ```
  export GLOBUS_SOURCE_PORT_RANGE=20000,25000
  ```
  - (already done via grid-env.sh)

- Data port range is also used by the Globus job submission service for file transfer.

A GridFTP: Service start-up

- □ Start-up GridFTP server
- □ As **root**, `service xinetd reload`

GridFTP

Client Part

GridFTP: globus-url-copy

- Copy file from remote to local (**as ubuntu**) – check if you have valid proxy with **`grid-proxy-info`**

```
echo "some text" > mydata

globus-url-copy -vb file:///$PWD/mydata   gsiftp://
gridway.fdi.ucm.es/~/gassGlobusonline.data
```

  - Source: **GridFTP** server:
    `gsiftp://``host<:port>/path/file`
  - Target: **local** machine (no gridftp server):
    `file:///``path/file`

- ~ can be used to refer to home directory
- Paths must be **absolute**.

GridFTP: globus-url-copy switches

- More verbose output: -vb
- Copy files from subdirectories: -r
- Create destination directories if needed: -cd
- http://www.globus.org/toolkit/docs/5.0/5.0.2/data/gridftp/user/#gridftpUser
  (http://bit.ly/cNpSBk)

GridFTP: globus-url-copy performance options

- Optimal value depends on TCP settings of kernel, latency, bottlenecks. Just try now with e.g.
  - **Parallel streams**: `-p 4`
  - **TCP buffer** size: `-tcp-bs 4m`
  - **Concurrent FTP connections:** `-cc 2`

- If multiple data nodes are available following might help:
  - `-stripe`
  - `-sbs 0` (so called partitioned block size)

GridFTP: extra I/O: Netlogger

- □ Netlogger can provide some useful information of bottlenecks.
- □ Must be enabled in globus compilation step: --enable-netlogger
- □ Example output of `globus-url-copy` with '-nlb' switch:

```
Total instantaneous throughput:
   disk read     = 2278.8 Mbits/s
   disk write    = 1381.0 Mbits/s
   net read      = 664.3 Mbits/s
   net write     = 288.8 Mbits/s
Bottleneck: Unknown
```

A

GridFTP: O/S settings

- Bottleneck can be too low TCP buffer settings in operating system.
- New Linux kernel versions (>=2.6.17) should tune itself.
- E.g. in Linux root can adjust them with `sysctl -w` command or put them permanently into

  `/etc/sysctl.conf`

GridFTP: Mode E

- Can be more efficient than stream mode
  - Mode E: Out of order reception of data
  - Multiple Path
- `-p` <*number*>
- Data sending server establishes data channel
  - Data port range must be open on target server (firewall!)

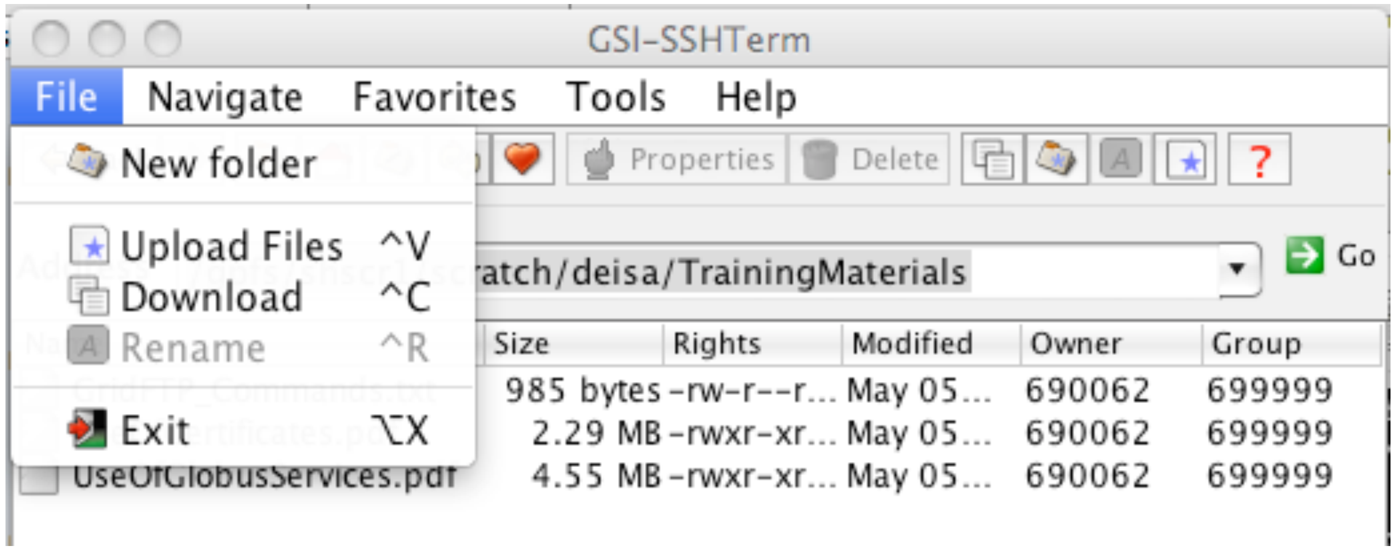
-
  ```
  time globus-url-copy -cc 10 -p 4 -vb -r gsiftp://
  gt5-ige.drg.lrz.de:2812/opt/dci_file/ gsiftp://udo-
  gt01.grid.tu-dortmund.de/~/
  ```

GridFTP: gsiscp client

- Globus provides also simple `gsiscp` client, which syntax is familiar from normal scp command:
  - `gsiscp –P 2222 source.txt <host>:target.txt`
  - `gsiscp –P 2222 <host>:source.txt target.txt`
  - 3rd party transfer did not seem to work

GridFTP: GSISSH-Term

- The Java Webstart tool has a file transfer GUI
- Requires GridFTP server on that server as well.
- Allows to upload and download files from/to your PC



atch/deisa/TrainingMaterials

	Size	Rights	Modified	Owner	Group
	985 bytes	-rw-r--r...	May 05...	690062	699999
	2.29 MB	-rwxr-xr...	May 05...	690062	699999
UseOfGlobusServices.pdf	4.55 MB	-rwxr-xr...	May 05...	690062	699999

Globusonline Identity

- Create account on webpage
  - www.globusonline.org
- Extract your public certificate (as user ubuntu)
  - cat ~/.globus/usercert.pem | grep –A 20 BEGIN
  - Select and copy to clipboard
- Add your DN to Globusonline
  - On Globusonline webpage goto “My Profile“ –> “Manage Identities“ –> “Add X.509 Certificate“
  - Paste your X.509 public certificate into the dialog box

Globusonline Endpoint Management

- Gsissh to CLI (as user ubuntu)
  - `gsissh cli.globusonline.org -p 22`
- Create testbed endpoints (while logged on to CLI !)
  - `endpoint-add -p gridway.fdi.ucm.es ucm`
  - `endpoint-add -p udo-gt01.grid.tu-dortmund.de udo`
  - `endpoint-add -p gt1.epcc.ed.ac.uk epcc`
  - `endpoint-add -p gt5-ige.drg.lrz.de:2812 lrz`
- Check your endpoints
  - `endpoint-list`

Globusonline CLI Data transfer

- List files on endpoints (while logged on to CLI !)
  - `ls -g udo/~/`
  - `ls -g epcc/~/`
- Use „scp“ to transfer files (blocking)
  - `scp -g udo:~/10MB epcc:~/10MB`
- Delegate proxy for endpoint “ucm“ to globusonline
  - `endpoint-activate -g ucm`
- Use „transfer“ to do non blocking transfer

  -
    ```
    transfer -g <enter>
     udo/~/10MB_0 ucm/~/10MB_0
     udo/~/10MB_1 ucm/~/10MB_1
    <Ctrl-D>
    ```
  - Returns <contact ID>
  - `status <contact ID>`

Globusonline Web Data transfer

- Transfer files via web application
  - Activate lrz via myproxy
- Create GlobusConnect Endpoint
  - It´s a gridftp server on your local computer
  - Windows / MAC / linux
- Transfer local file to endpoint udo
- Email notification was sent

JOB SUBMISSION via GRAM5

GRAM5: Overview

- Administration:
  - Start-up script
  - Configuration

- Client:
  - `globus-job-run`
  - `globusrun`
  - a batch job (non-blocking)
  - a batch scheduling system jobs
  - GRAM5 job scripts (RSL)

GRAM5: configuration xinet.d script example

```
service gsigatekeeper
{
  socket_type  = stream
  protocol     = tcp
  wait         = no
  user         = root
  env          += GLOBUS_TCP_PORT_RANGE=20000,25000
  server       = /opt/globus/sbin/globus-gatekeeper
  server_args  = -conf /opt/globus/etc/globus-gatekeeper.conf
  # bind       =  <optional if one interface. Otherwise set here IP address>
  port         = 2119
}
```

A

GRAM5: Configuration xinet.d

- As **root**:

```
cd /etc/xinet.d/

wget http://tinyurl.com/mygsigatekeeper

mv mygsigatekeeper gsigatekeeper

chmod 744 gsigatekeeper
```

Configuration globus-gatekeeper.conf

- globus-gatekeeper.conf was referred in start-up script:

```
-x509_cert_dir /etc/grid-security/certificates
-x509_user_cert /etc/grid-security/hostcert.pem
-x509_user_key /etc/grid-security/hostkey.pem
-gridmap /etc/grid-security/grid-mapfile
-home /opt/globus
-e libexec
-logfile var/globus-gatekeeper.log
-port 2119
-grid_services etc/grid-services
-inetd
```

Configuration default job manager

- Symbolic link points to the default job manager

```
lrwxrwxrwx 1 globus globus-g  jobmanager -> jobmanager-fork
-rw-r--r-- 1 globus globus-g  jobmanager-fork
-rw-r--r-- 1 globus globus-g  jobmanager-sge
```

```
 ll $GLOBUS_LOCATION/etc/grid-services
```

GRAM5: configuration etc/globus-fork.conf et al.

- The `$GLOBUS_LOCATION/etc` directory contains LRMS configuration files e.g. `globus-fork.conf`, `globus-sge.conf` – referring to respective log files

- Fork's log file is in `$GLOBUS_LOCATION/var/` with following permissions (622)

- There is also `globus-gatekeeper.log` (600)

GRAM5: configuration globus-job-manager.conf (1)

- □ Correct hostname (hostname –f) in /opt/globus/etc/globus-job-manager.conf

• $GLOBUS_LOCATION/etc/globus-job-manager.conf

```
-home "/opt/globus"
-globus-gatekeeper-host <your hostname>
-globus-gatekeeper-port 2119
-globus-host-cputype x86_64
-globus-host-manufacturer unknown
-globus-host-osname Linux
-globus-host-osversion 2.6.34-12-desktop
```

A

GRAM5: configuration globus-job-manager.conf (2)

```
-globus-toolkit-version 5.0.2
-stdio-log "$(HOME)"
-log-levels 'FATAL|ERROR'
-state-file-dir /opt/globus/tmp/gram_job_state
-globus-tcp-port-range 20000,25000
-stdio-log "$(HOME)"
-disable-usagestats
-log-levels 'ALL|FATAL|ERROR'
-state-file-dir /opt/globus/tmp/gram_job_state
```

□ It is not recommend state-file-dir to be on a shared file system.

A Firewall configuration and /etc services

- In `globus-job-manager.conf` can be set **gridftp data port range** (e.g. 20000,25000).

- In `globus-gatekeeper.conf` and in xinet.d script is set **gatekeeper port** (default 2119).

- As **root** in `/etc/services` set:

```
gsigatekeeper 2119/tcp
```

GRAM5: Start

- □ Start gatekeeper:

```
service xinetd reload
```

GRAM5

Client Part

GRAM5: Hints for client

- For logs see your home directory (`ls -lart gram*`)

- See also in `$HOME/.globus/job/`

- If your job seems to get stuck try to kill your job-manager processes
  - `killall globus-job-manager`

- Gatekeeper log
  - `$GLOBUS_LOCATION/var/gatekeeper.log`
  - It might be readable by administrators only

globus-job-run blocking submission

- With `globus-job-run` it is simple to submit a job
  - **Blocking** command:
    it does not release the shell until the job finishes

- Example (As user **ubuntu**):

```
globus-job-run localhost /bin/hostname
```

- Gsissh to gt5-ige.drg.lrz.de and run

```
globus-job-run <your ec2 name> /bin/hostname
```

- It is possible to give various parameters e.g. directing standard output or error.
  See `-help` or user guide http://bit.ly/c8FYK0

globus-job-run blocking submission

-
  ```
  globus-job-run udo-gt01.grid.tu-dortmund.de/jobmanager-pbs /bin/hostname
  ```

-
  ```
  globus-job-run gt5-ige.drg.lrz.de/jobmanager-sge /bin/hostname
  ```

Globus-job-submit non-blocking submission

- `globus-job-submit`
  - returns to shell right after the submission and prints **job_contact_string (https://...)**
  - **non-blocking**

- `globus-job-status` ***<job_contact_string>***
- `globus-job-get-output` ***<job_contact_string>***
- `globus-job-clean` ***<job_contact_string>***

Globus-job-submit non-blocking submission

- From your ec2 hands-on machine (as ubuntu):
- ~>globus-job-submit gt5-ige.drg.lrz.de /bin/sleep 60
  - https://gt5-ige.drg.lrz.de24383/16145785939916773883 1/2666570055213425/
- ~> globus-job-status https://gt5-ige.drg.lrz.de:24383/161....
  - DONE
- ~>globus-job-get-output https://gt5-ige.drg.lrz.de:38/....
  - Mon Jul 4 18:40:43 CEST 2011
- ~> globus-job-clean https://gt5-ige.drg.lrz.de:24383/161....
  - WARNING: Cleaning a job means:
    - Kill the job if it still running, and
    - Remove the cached output on the remote resource
    Are you sure you want to cleanup the job now (Y/N) ? Y
    Cleanup successful.

Globusrun and RSL (1)

- `globusrun` command is the most suitable for real "production" jobs
- It takes as a parameter a script written in **Globus Resource Specification Language (RSL)** vs. command line parameters as used on last slides

- RSL script can be passed:
  - from a command-line (in " ")

```
~> globusrun -s -r gt5-ige.drg.lrz.de "&(executable=/bin/date)"
   Mon Jul  4 18:40:43 CEST 2011
```

  - in an RSL file

GRAM5: globusrun and RSL (2)

- The simplest RSL script is specifying the executable:

```
&(executable=/bin/date)
```

  - Please store this line to a file `job.rsl`

- The `&` is needed only on the first row.

- All rows are surrounded in ().

GRAM5: globusrun command line parameters

- Submission which streams (`-s`) standard output and error to the display

```
globusrun -s -r gt5-ige.drg.lrz.de -f job.rsl
```

```
Mon Jul  4 18:40:43 CEST 2011
```

- For complete list of possible attributes see http://bit.ly/d6cQbL

GRAM5: globusrun non-blocking operation (1)

- With **-b** option non-blocking command is sent and a contact string is then returned.
- Edit `sleep.rsl`:

```
&(executable=/bin/sleep)
(arguments=1000)
```

- Run:

```
globusrun -b -r udo-gt01.grid.tu-dortmund.de/
jobmanager-pbs -f sleep.rsl
```

GRAM5: globusrun non-blocking operation (2)

- Status query:
  - `globusrun -status <job_contact_string>`
    - Possible job statuses:
      - ACTIVE
      - FAILED
      - SUSPENDED
      - DONE
      - UNSUBMITTED
      - STAGE_IN
      - STAGE_OUT
      - UNKNOW JOB STATE

- Cancelling the job:
  - `globusrun -k <job_contact_string>`

GRAM5: RSL

- Some useful RSL attributes:

```
& (rsl_substitution = (DIR "/tmp/my_dir") )
(environment = (MSG 'Hello'))
(stderr = $(DIR)/stderr.txt)
(stdout = $(DIR)/stdout.txt)
(executable=/usr/bin/env)
```

- Variable set in OS `environment` is not accessable in the RSL script.

GRAM5: File staging (1)

- Possible staging steps in a job are:
  - File **stage in**: files from client to GRAM5 server
  - File **stage out**: files from GRAM5 server to client
  - File **clean-up**: remove files on GRAM5 server

- Internal or external GridFTP can be used.
- To use internal file transfer mechanism (GASS) uses predefined variable

GRAM5: File Staging Gridftp Example

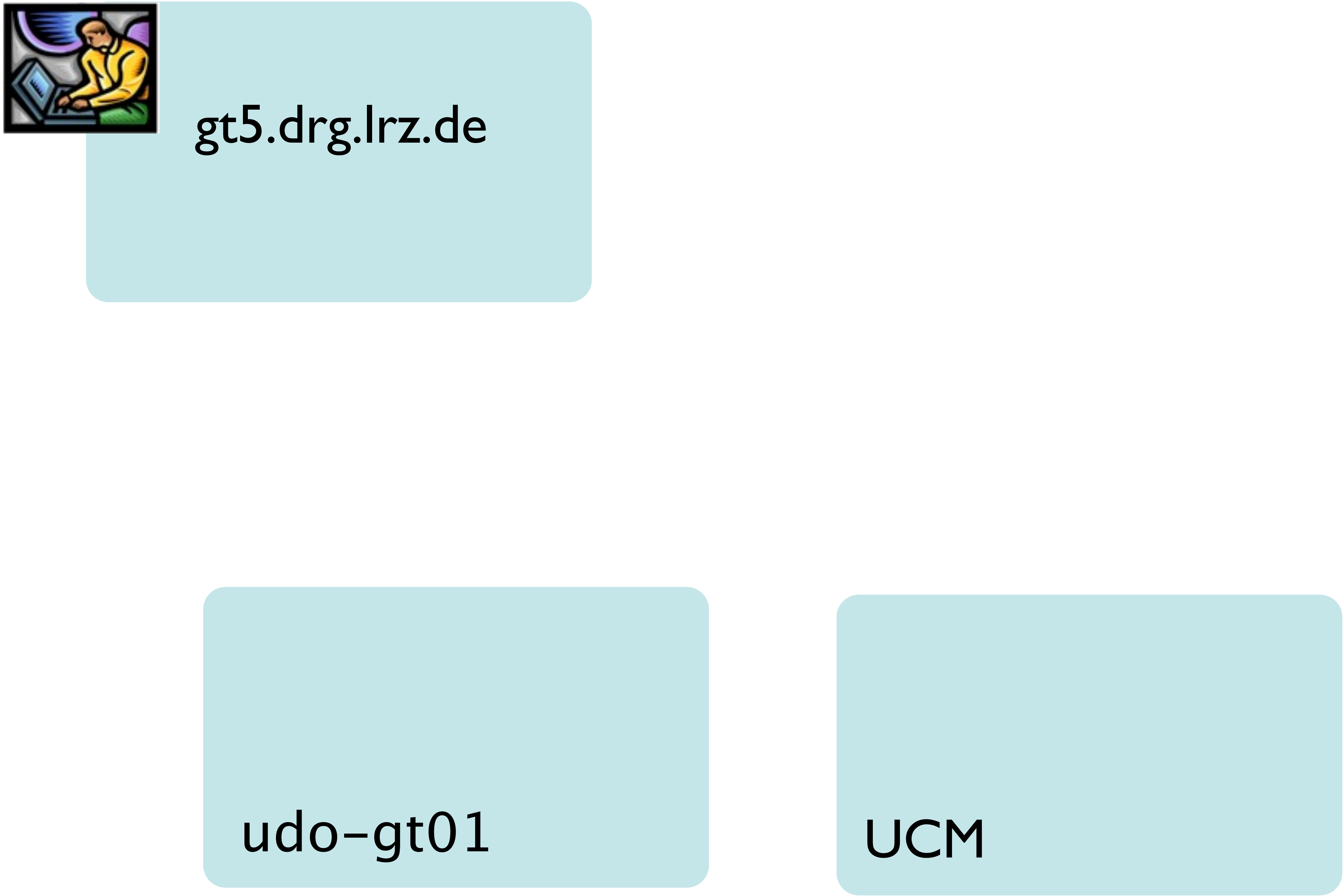
```
& (rsl_substitution = (GRIDFTP_SERVER gsiftp://<GFTPHOST>))
  (executable=/bin/cat)
  (arguments=$(HOME)/input_file)
  (stdout=stdout.txt)
  (stderr=stderr.txt)
  (file_stage_in = ($(GRIDFTP_SERVER)/$(HOME)/input_file $(HOME)/input_file))
  (file_stage_out = (stderr.txt $(GRIDFTP_SERVER)/$(HOME)/stderr.txt)
                    (stdout.txt $(GRIDFTP_SERVER)/$(HOME)/stdout.txt))
  (file_clean_up = $(HOME)/input_file)
```

- Obsolete?

GRAM5: File staging GASS example (1)

```
& (executable=/bin/sh)
(rsl_substitution = (username ige_user001))  (* Fix username! *)
(arguments= globusonline.sh)
(stdout=stdout.txt)
(stderr=stderr.txt)
(file_stage_in =
($(GLOBUSRUN_GASS_URL)/opt/home/$(username)/globusonline.sh globusonline.sh))
(file_stage_out =
(stderr.txt $(GLOBUSRUN_GASS_URL)/opt/home/$(username)/stderr.txt)
(stdout.txt $(GLOBUSRUN_GASS_URL)/opt/home/$(username)/stdout.txt))
(file_clean_up=stdout.txt)
(file_clean_up=stderr.txt)
```

GRAM5: File staging
GASS example (1)

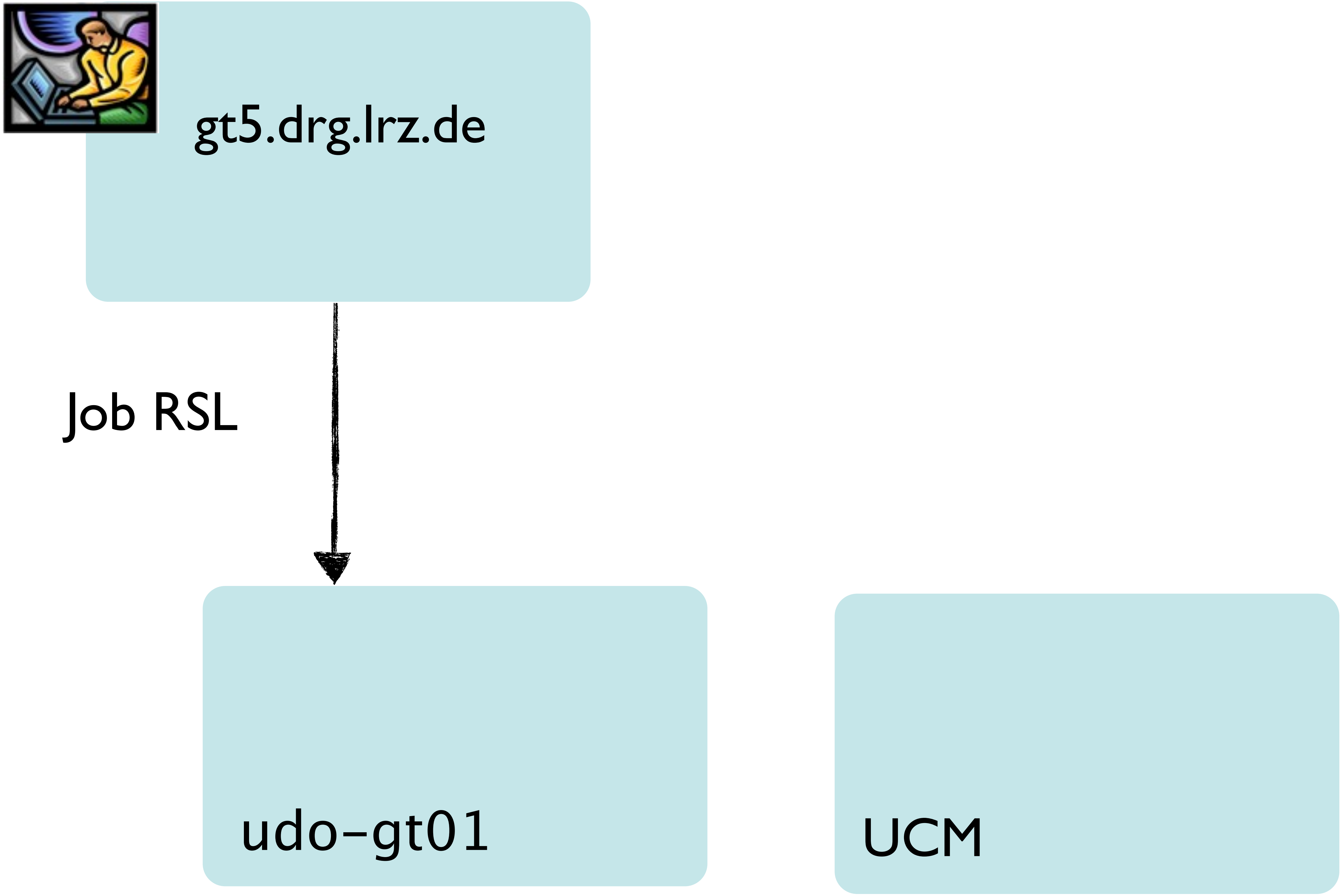


gt5.drg.lrz.de

udo-gt01

UCM

GRAM5: File staging GASS example (1)



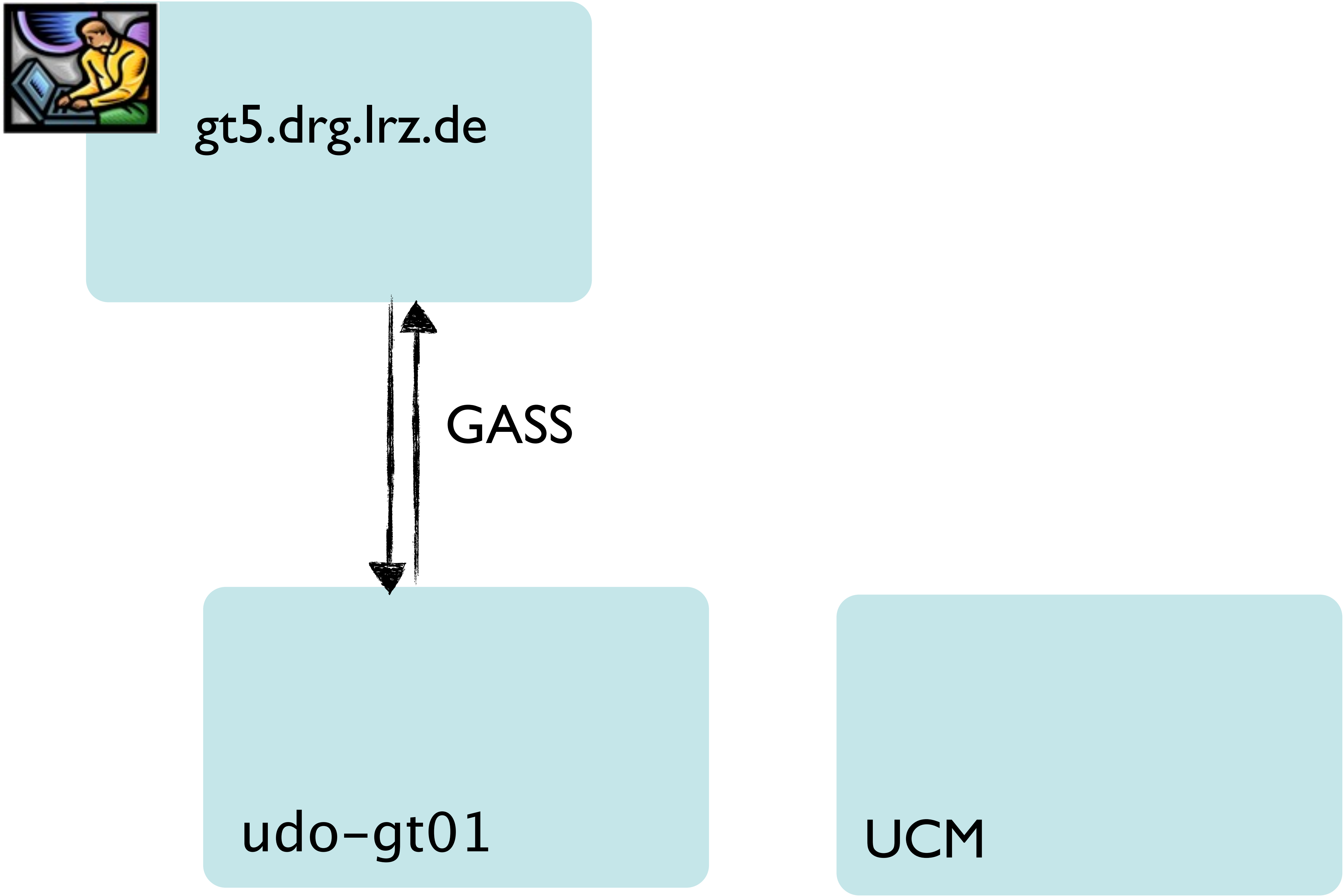
gt5.drg.lrz.de

Job RSL

udo-gt01

UCM

GRAM5: File staging GASS example (1)



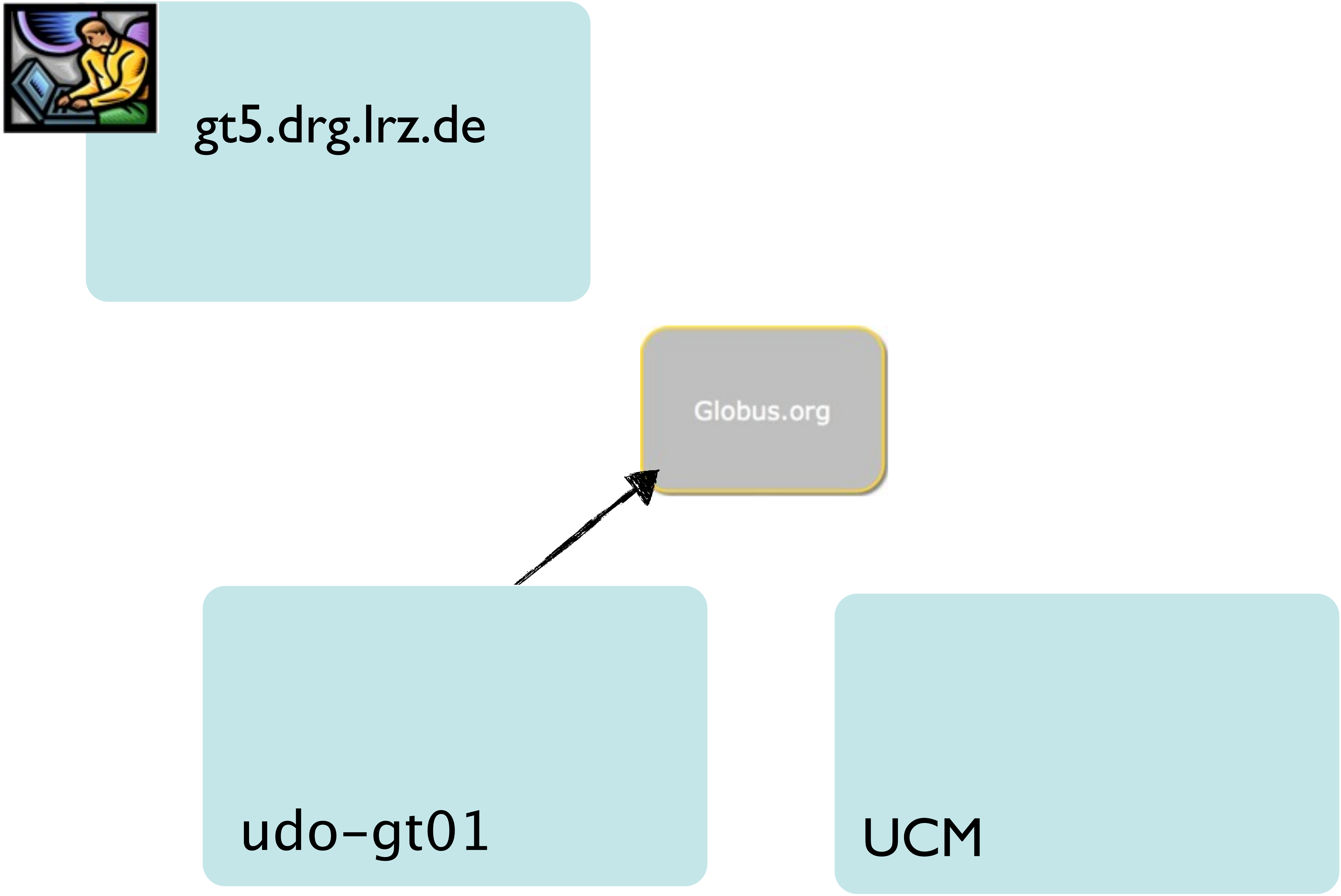
gt5.drg.lrz.de

GASS

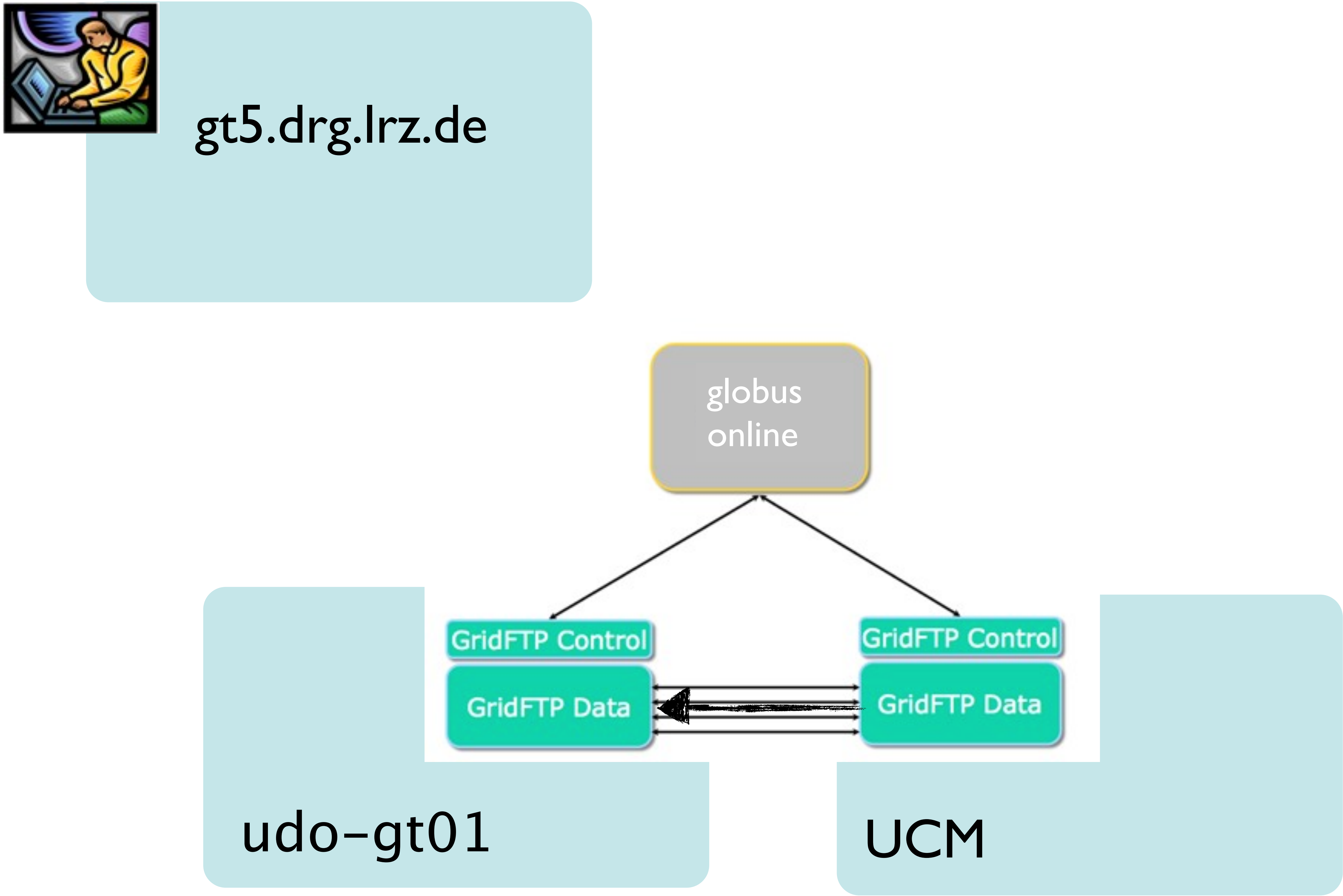
udo-gt01

UCM

GRAM5: File staging GASS example (1)



GRAM5: File staging GASS example (1)



gt5.drg.lrz.de

globus online

GridFTP Control

GridFTP Data

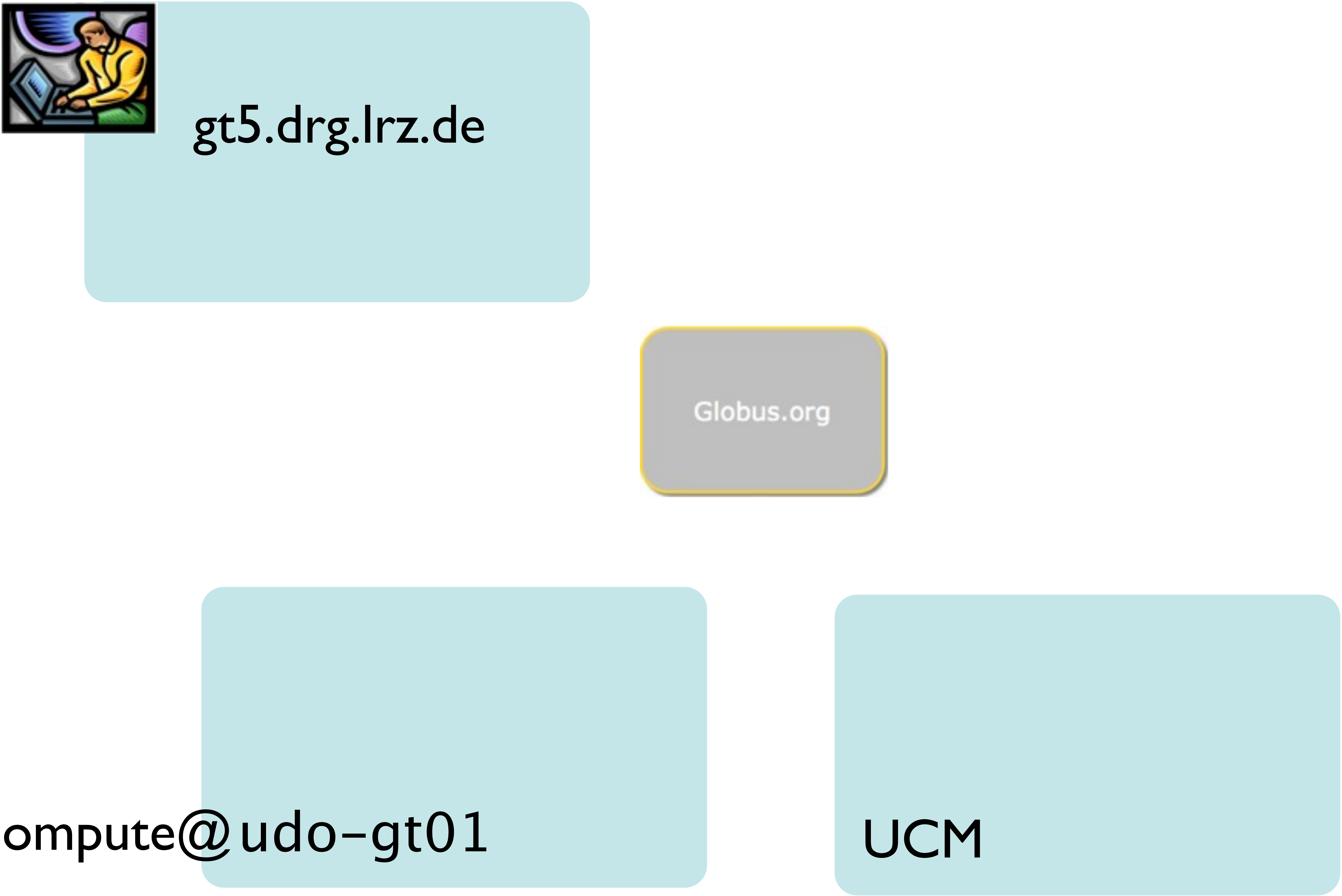
GridFTP Control

GridFTP Data

udo-gt01

UCM

GRAM5: File staging GASS example (1)



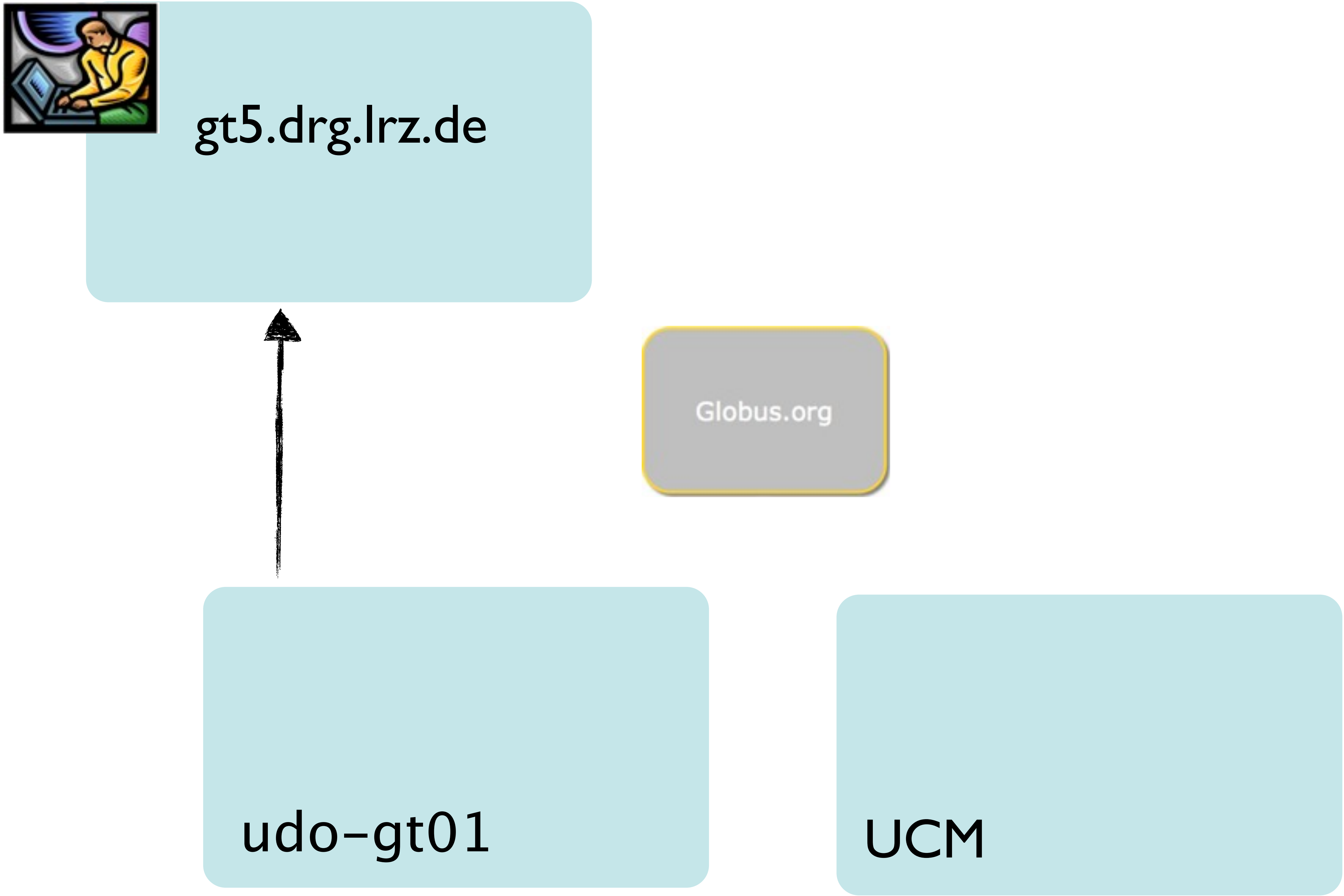
gt5.drg.lrz.de

Globus.org

compute@udo-gt01

UCM

GRAM5: File staging GASS example (1)



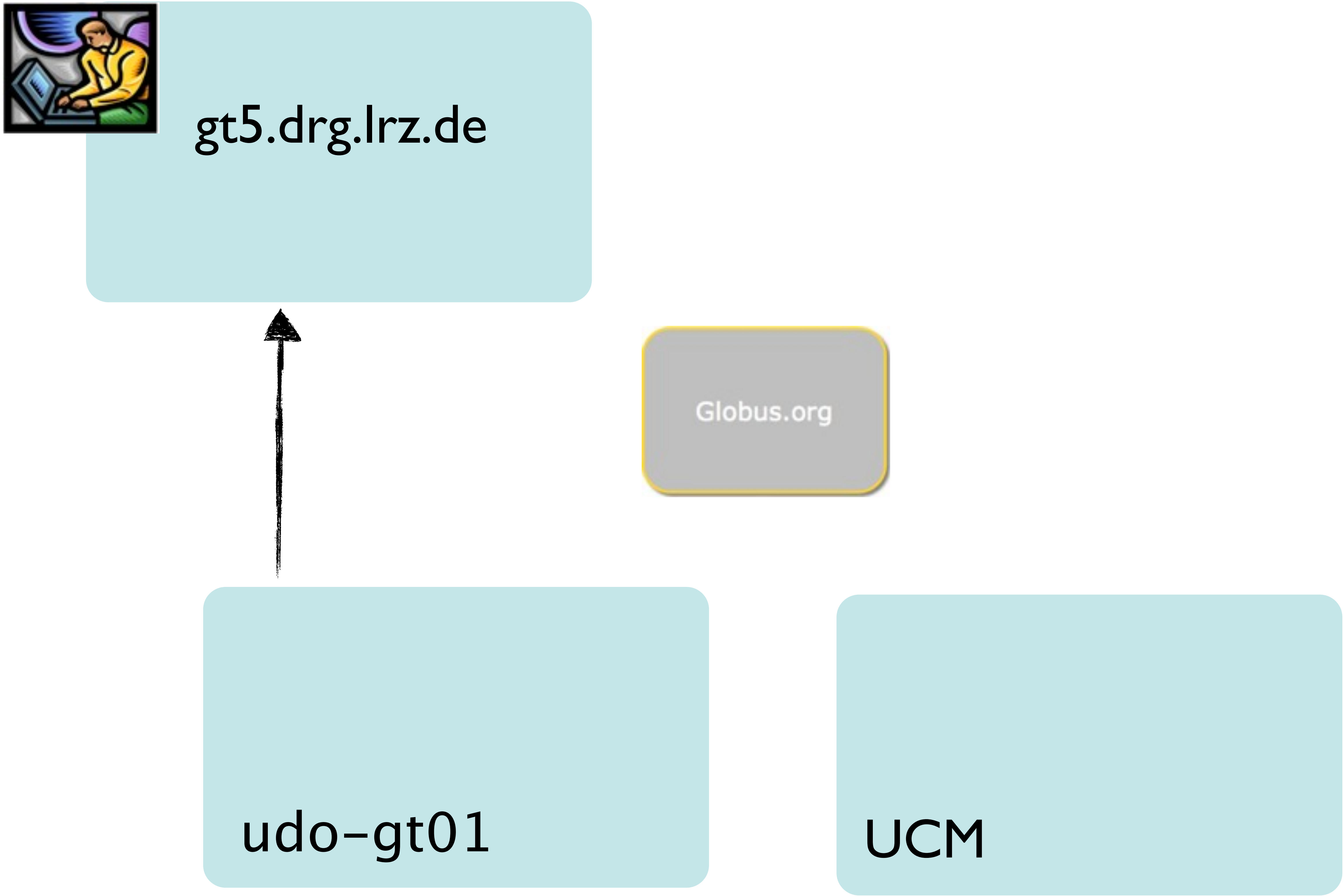
gt5.drg.lrz.de

Globus.org

udo-gt01

UCM

GRAM5: File staging GASS example (1)



gt5.drg.lrz.de

Globus.org

udo-gt01

UCM

GRAM5: File staging GASS example (2)

- Log in to ucm and create in home dir data file

  -
    ```
    gsissh gridway.fdi.ucm.es
    ```

  -
    ```
    echo “something“ > gassGlobusonline.data
    ```
- Log on to LRZ

  -
    ```
    gsissh gt5-ige.drg.lrz.de
    ```
- Download RSL script

  -
    ```
    wget http://tinyurl.com/globusonline-rsl
    ```
- Download globusonline execute script

  -
    ```
    wget http://tinyurl.com/globusonline-sh
    ```
- Submit this job

  -
    ```
    globusrun -D -s -r udo-gt01.grid.tu-dortmund.de -f globusonline.rsl
    ```

GRAM5: Advanced RSL: Proxy renewal operation & dbg

- By default proxy certificate lives 12 hours
- If proxy expires and need to get results of the job:
  - `grid-proxy-init`

  -
    ```
    globusrun -r <host> \ "&
    (restart=<job_contact_string>)"
    ```

PROXY STORAGE SERVICE

MyProxy

MyProxy: proxy storage service

- Administration
  - Security remarks
  - Configuration of the service (MyProxy and system)
  - Logging and start-up

- Client
  - Storing proxy with Globus commands
  - Fetching the stored proxy with Globus command

3rd party tools:

  - Firefox plug-in
  - Java Webstart application

MyProxy: Security remarks

- Root can access the proxies on the server.
- A dedicated machine for this service only with restricted access and strict firewall
- Port 7512 (by default) must be open for Internet
- Subscribe yourself to the security alert mailing lists:

  http://grid.ncsa.illinois.edu/myproxy/security/

MyProxy: configuration compilation, init.d script

A

- If only MyProxy needs to be installed:

```
make gsi-myproxy
```

- You find **init.d** and xinet.d start-up scripts fro

```
 cd $GLOBUS_LOCATION/share/myproxy/
cp etc.init.d.myproxy /etc/init.d/myproxy
chmod u+x /etc/init.d/myproxy
```

  - **Set** (at least) `GLOBUS_LOCATION`
  - Activate the init.d script for the boot:
    - `chkconfig -a myproxy`

MyProxy: configuration removing expired proxies

- To delete invalid credentials:

```
cp $GLOBUS_LOCATION/share/myproxy/myproxy.cron \
/etc/cron.hourly
chown globus /etc/cron.hourly/myproxy.cron
chmod u+x /etc/cron.hourly/myproxy.cron
```

- Correct **`$GLOBUS_LOCATION`**

- Put it to run on `globus` user's crontab (**`crontab -e`**):

```
59 * * * * /etc/cron.hourly/myproxy.cron > /dev/null
```

MyProxy configuration: myproxy-server.config file (1)

□ Template file is located in

```
cp $GLOBUS_LOCATION/share/myproxy/myproxy-server.config \
  $GLOBUS_LOCATION/etc/
```

□ It is a long file with many commented settings.

□ You can add there:

```
disable_usage_stats  "true"
```

MyProxy configuration: myproxy-server.config file (2)

Uncomment the following settings from the template:

```
accepted_credentials       "*"
authorized_retrievers      "*"
default_retrievers         "*"
authorized_renewers        "*"
default_renewers           "none"
authorized_key_retrievers  "*"
default_key_retrievers     "none"
trusted_retrievers "*"
default_trusted_retrievers "none"
```

MyProxy: start-up and logging

- Add into `/etc/services` row

```
mypróxy-server  7512/tcp    # Myproxy server
```

- To start either init.d

```
 service myproxy start
```

- The logs are in `/var/log/messages`
- More verbose messages with `-d` and `-v` parameters

Appendix

Grid Acronym App

□ Stefan Freitag´s (from TU Dortmund) Android App

□ Let you search for >500 Grid related Acronyms

□ Available on Android market: https://market.android.com/search?q=GridAcro

A&A: Certificate conversion

- Just for your information
- To create .pem files from .p12 file:

```
openssl pkcs12 -clcerts -nokeys -in usercert.p12 -out usercert.pem
openssl pkcs12 -nocerts -in usercert.p12 -out userkey.pem
chmod 0400 userkey.pem && chmod 0600 usercert.pem
```

- Browsers typically need a .p12 file. To create it from .pem files:

```
openssl pkcs12 -export -inkey userkey.pem -out \
usercert.p12 -name "Firstname Lastname” -in usercert.pem
```

A&A: Get Certificate information

- To view e.g. validity of `usercert.pem` file:

  -
    ```
    openssl x509 -in $HOME/.globus/usercert.pem \
     -text -noout
    ```
  - OR just run **`grid-cert-info`** Globus command.

- For p12:
  - Again, you can use a Globus command: `grid-cert-info`
  - OR with `openssl` even temporary pem file is needed
    - `openssl pkcs12 -in cert.p12 -out temp.pem` (asks passwords)
    - `openssl x509 -in temp.pem -noout -enddate`
    - `rm temp.pem`

./configure (1)

- There are good to know switches for `./configure`

- Batch scheduling system (BSS) support: PBS (Torque), Condor, LSF and SGE e.g.

```
--enable-wsgram-pbs
```

- TCP wrappers mechanism for gsisshd:

```
 --with-gsiopensshargs="--with-tcp-wrappers"
```

A

./configure (2)

- Batch Scheduling System support must be compiled separately:
  - `make gram5-`**`pbs`** (or `gram5-`**`sge`**`/gram5-`**`lsf`**`/ gram5-`**`condor`**)
- Optional features for GridFTP:
  - `make `**`udt`**
  - `make globus-xio-extra-drivers`

A&A: .pem, .p12 conversion

- PEM files:
  - `$HOME/.globus/usercert.pem` (public certificate part)
  - `$HOME/.globus/userkey.pem` (private key)
    - `chmod 400 userkey.pem`
  - Non-default place or name can be set with the environmental variables: `$X509_USER_CERT` and `$X509_USER_KEY`
- Instead of the `.pem` files a `.p12` file can be used:
  - `$HOME/.globus/usercred.p12`
    - `chmod 400 usercred.p12`
- In Windows put the files into: `%HOMEPATH%\.globus`
  - To create `.globus` start `cmd` program and run

```
mkdir %HOMEPATH%\.globus
```

A

GridFTP: O/S settings (2)

- Example settings from `/etc/sysctl.conf`:

```
net.ipv4.tcp_rmem = 4096 2097152 8388608
net.ipv4.tcp_wmem = 4096 2097152 8388608
net.core.rmem_default = 2097152
net.core.wmem_default = 2097152
net.core.rmem_max = 8388608
net.core.wmem_max = 8388608
net.core.netdev_max_backlog = 2000
```

A

GridFTP: extra I/O: UDT

- Instead TCP transfer also UDP based UDT protocol is available.
- Might be useful in case of high latencies.

- Needs to be compiled (`make udt`) and configured

  (`-dc-whitelist udt,gsi,tcp` in xinet.d script).
- `globus-url-copy` has `-udt` parameter

GridFTP: reliability options

- Client can save status to a file to recover from some failures

```
#!/bin/sh
STATEF=statusfile.txt;
while [ ! -e $STATEF -o -s $STATEF ]; do
globus-url-copy -restart -rst-timeout 10 -vb -dumpfile $STATEF \
gsiftp://srchost/srcdirpath/ gsiftp://dsthost/dstdirpath/;
done;
```

  - Dumpfile contains untransferred URLs during the transfer
  - File will be emptied if transfer succeeds

GRAM5: MPI job example (1)

- MPI job should be submitted to batch scheduling systems.
  Since it will take time so `-b` option is used:

```
globusrun -s -b -r udo-gt01.grid.tu-dortmund.de/
  jobmanager-pbs -f mpigt5.rsl
```

GRAM5: MPI job example (2)

RSL script gt5mpi.rsl (1)

```
&(executable=$(HOME)/mpi_test)
(job_type=mpi)
(count=2)
(max_wall_time=20)
(max_cpu_time=10)
(max_memory=10)
(stdout=stdout.txt)
(stderr=stderr.txt)
```

□ `count:` Number of the MPI processes.

□ `job_type` must be set to `mpi`.

□ `max_memory:` needed memory (in MBs).

□ `max_wall_time:` Execution time (in minutes).

GRAM5: Auditing & accounting

- Globus does not support **accounting**.
  - Batch scheduling system can provide accounting information.
  - Tip: add a specific string to job name attribute in BSS Perl file to identify Globus jobs

- Globus provides some **auditing** database functionality.
  - You can look at http://bit.ly/cCVCpK