

Introduction to Programming and Computing for Scientists

Oxana Smirnova

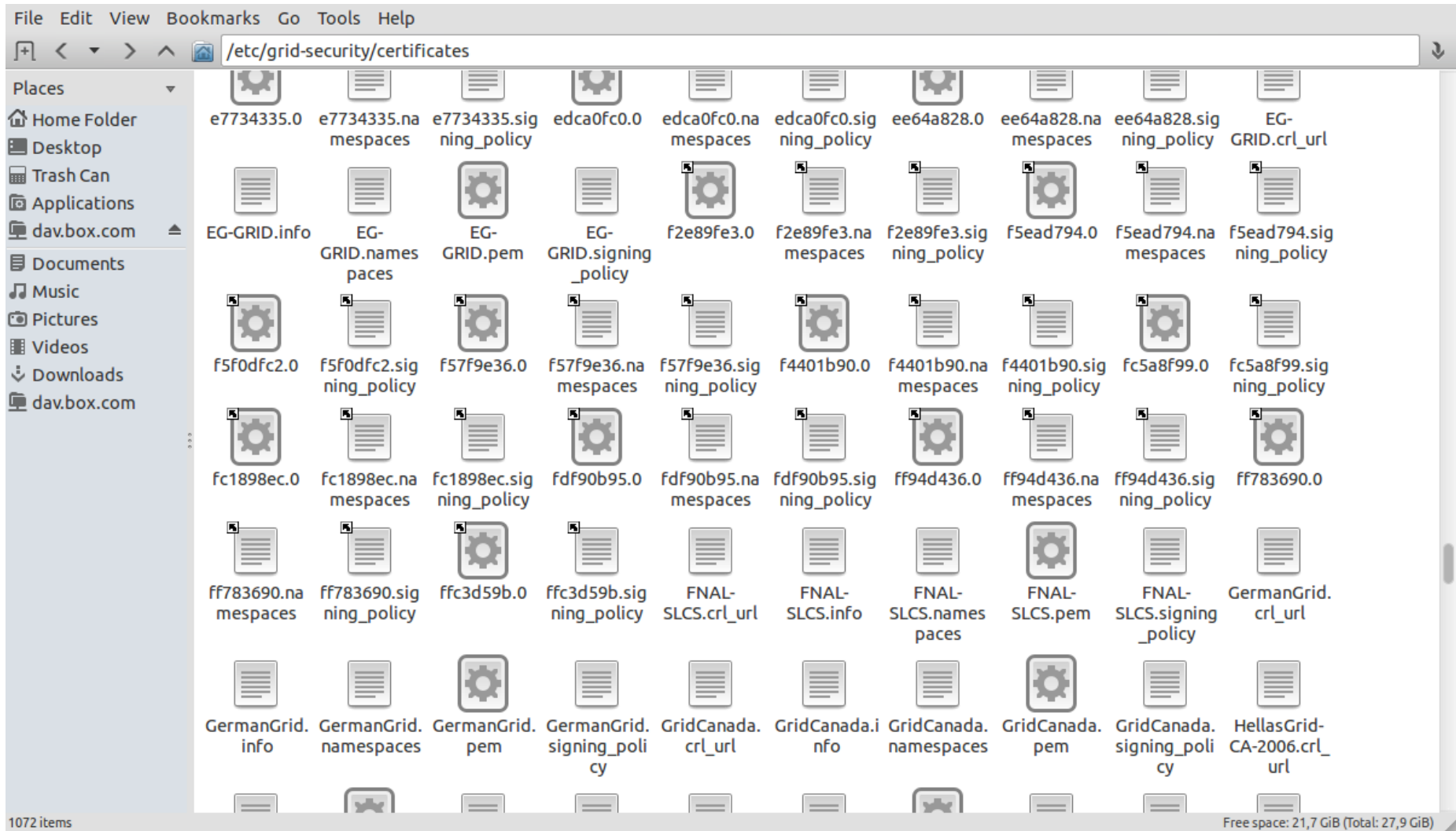
Lund University

Tutorial 4b: Grid certificates and jobs

Step 1: Install public Certificate Authority certificates

- Before doing anything on the Grid, you will need to obtain **IGTF Certificate Authorities certificates**
 - Packages are available from IGTF and some Grid repositories
 - The packages include Certificate Revocation Lists (CRLs)
 - Regular updates for CRL and IGTF packages must be in place
 - Usually happens automatically
- **Our virtual machines and cluster have them already:**
 - Start the course virtual machine
 - Inspect **`/etc/grid-security/certificates`**
 - Hint: use **`ls -al`**
- **Today we will work on the Iridium cluster only!**

/etc/grid-security/certificates has CA keys, CRLs and policies



Step 2: get your own keys and certificates

There are two main ways of storing personal certificates:

- Two files: private key and public certificate
 - Grid uses PEM encoding for keys and certificates (ASCII)
 - Standard file names: **userkey.pem** and **usercert.pem**
 - Note: public key is inside the CA-signed certificate **usercert.pem**
- Single file: PKCS#12 formatted certificate, containing private and public keys, as well as CA signature and CRL info
 - PKCS#12 certificate (**.p12**) is used mostly by browsers, but can also replace PEM files in some Grid tools
 - One can convert PKCS#12 file to PEM files and vice versa

- Private keys must not be copied over the network!
- Private keys can not be stored in public machines like ours in this class!
- Because of this, we will log in to the Iridium cluster and create the keys there

Log in to Iridium and launch a Web browser there:

- From any virtual machine or a Linux machine log in to Iridium:

```
ssh -X yourlogin@pptest-iridium.lunarc.lu.se
```

- Launch a Web browser (use Firefox, other browsers are less suitable):

```
firefox &
```

- Some useful commands – just a reminder:

<code>ls -al</code>	List all files in the current directory
<code>mkdir stuff</code>	Create a new directory called stuff inside current directory
<code>cd stuff</code>	Step down into the directory called stuff
<code>cp ~/course/file1 file2</code>	Copy a file from directory ~/course to the current directory
<code>geany &</code>	Open geany editor

How to get a certificate

The easy way:

Google for “Grid certifikat Lund” and find a link to “Certifikat | LDC” - “inloggningsportalen”

Log in using your university credentials

Follow the instructions (slide 7 onwards)

The certificate in PKCS#12 format will be stored in your browser certificate store

You can export the certificate into a file (.p12) and extract PEM files, if necessary

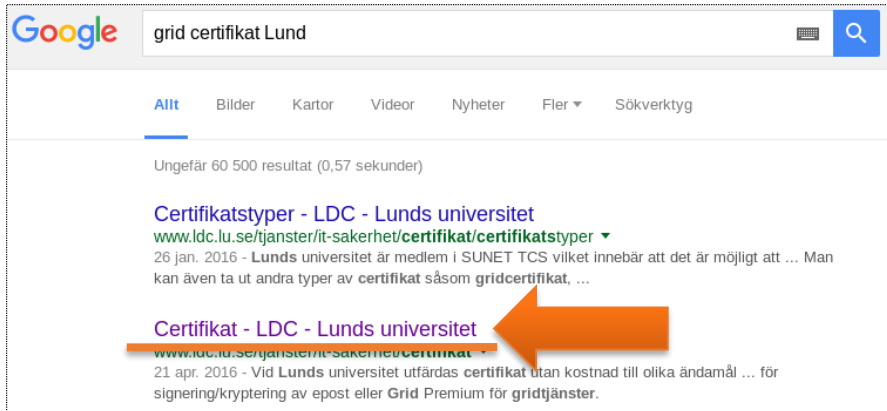
How does it work:

Browser creates private and public keys for you

Public key is sent to the CA, along with information that University provides about you

Signed public certificate is returned back to your browser, and merged with the private key to create a PKCS#12 certificate

Let's start: Google is your friend



Certifikat

Vid Lunds universitet utfärdas certifikat utan kostnad till olika ändamål men endast för tjänster, servrar och personal som tillhör myndigheten. Genom ett [samarbete via Géant](#) kan vi via [DigiCert](#) utfärda så kallade TCS certifikat. Det finns flera olika [certifikattyper](#) för olika ändamål, både för servrar, kodsignering och personliga för att signera och/eller kryptera epost.

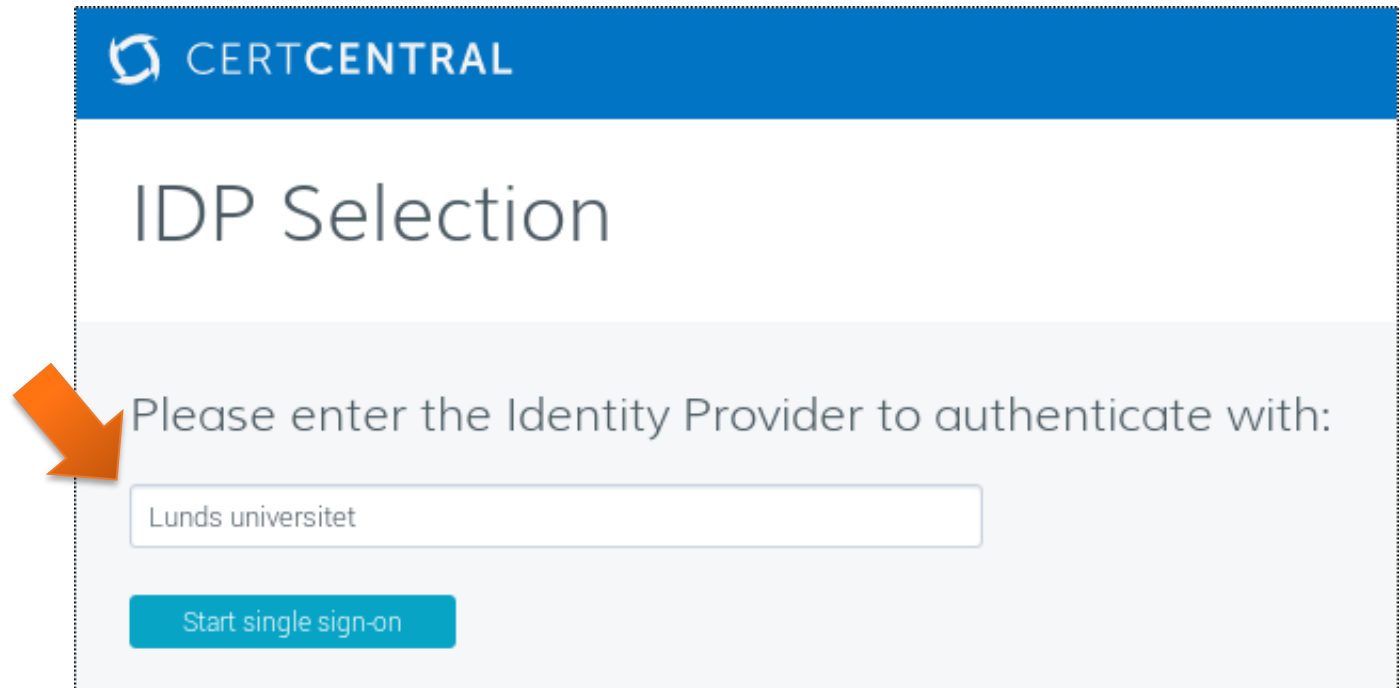
Snabbguide för att ta ut server- och kodsignerings-certifikat:

1. [Ansök om ett användarkonto](#) i DigiCerts system
2. Skapa ett certifikat och en CSR (Certificate Signing Request) utifrån det
3. [Logga in](#) och välj vilken [typ av certifikat](#) du skall ansöka om
4. Fyll i fälten på sidan och ansök om att få certifikatet signerat av DigiCert
5. När vi accepterat ansökan meddelar DigiCert via epost att certifikatet är signerat
6. Installera det signerade certifikatet

Snabbguide för att ta ut ett personligt certifikat:

1. Gå till [inloggningsportalen](#)
2. Skriv in "Lunds universitet" för att komma till vår inloggningssida. Logga in.
3. Välj vilken typ av certifikat du vill skapa. Använd Premium för

Login to the CertCentral Portal



CERTCENTRAL

IDP Selection

Please enter the Identity Provider to authenticate with:

Start single sign-on

- Type Lunds universitet in the window (NOT Lund University!)
- If asked whether to remember Lunds universitet as Identity Provider, feel free to answer “yes”

Enter your LU credentials



SVENSKA

LUND UNIVERSITY

Please enter your userid without "@lu.se" at the end.

Username:

Password:

LOGIN

This is Lund University Central Authentication Service (CAS). For privacy and security reasons, you should always log out and close all browser windows when you are done accessing services that require authentication. If you use Apple Mac OS X, you must also shut down the entire browser, not just the windows. If you use a public computer, it is especially important that you close all browser windows before you leave your computer.

Request “Grid Premium” certificate



Request a Certificate

Choose a product

Product:

CSR:
(optional)

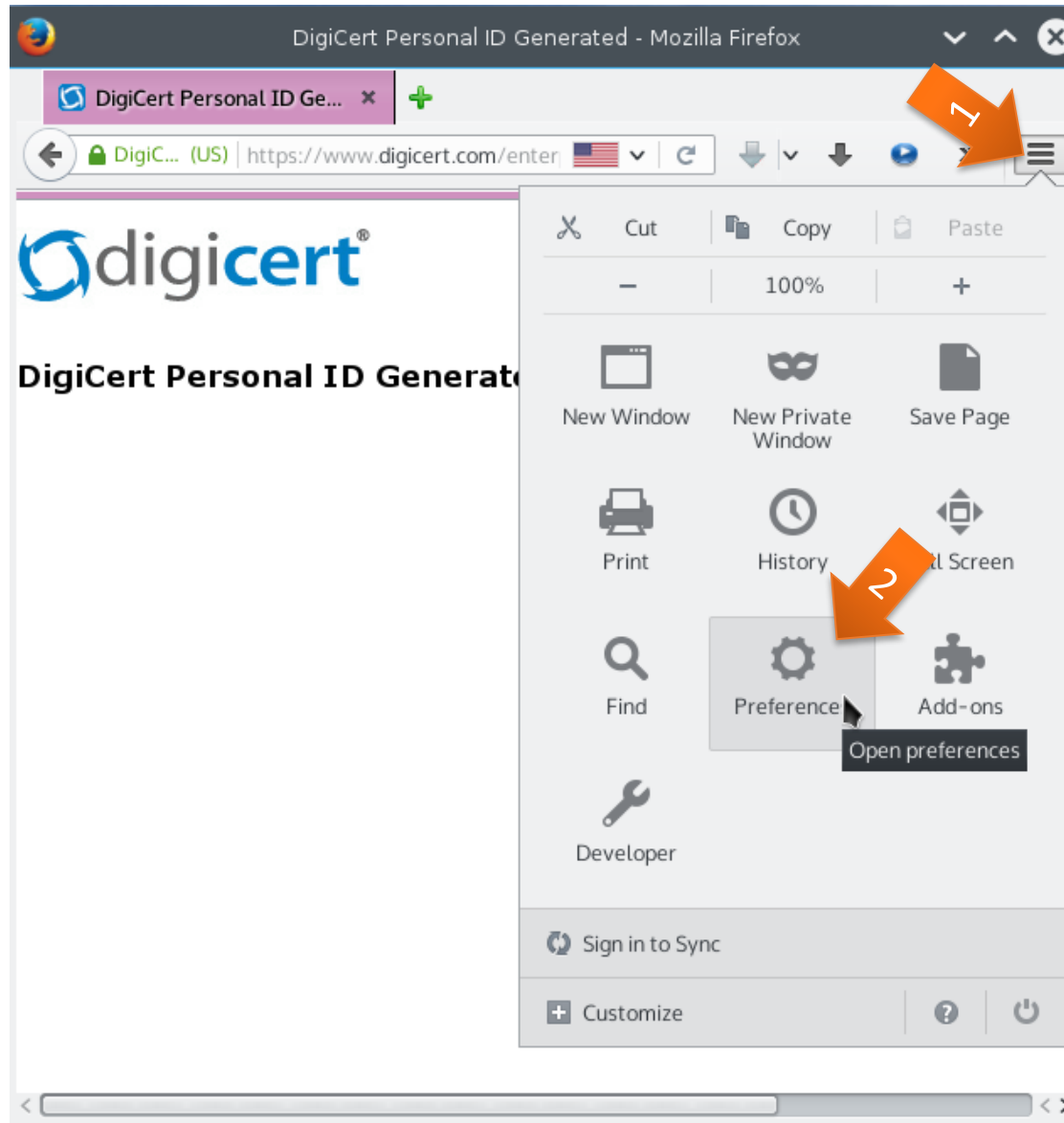
Common Name: Oxana Smirnova quar-osm@lu.se

Email: oxana.smirnova@hep.lu.se

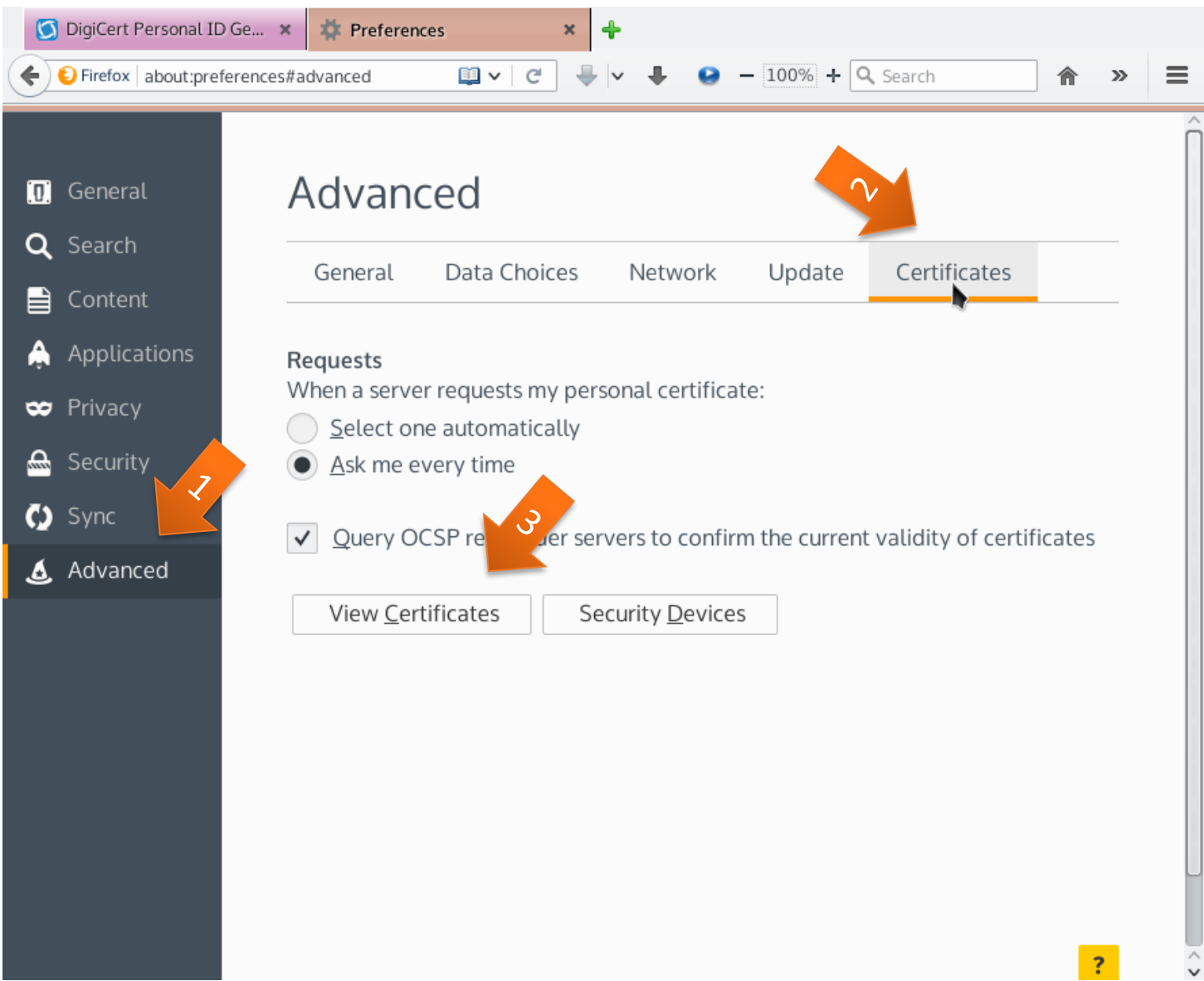
Organization: Lunds universitet

- Select “Grid Premium” Product
- If asked whether to trust the CA, feel free to tick “yes” everywhere

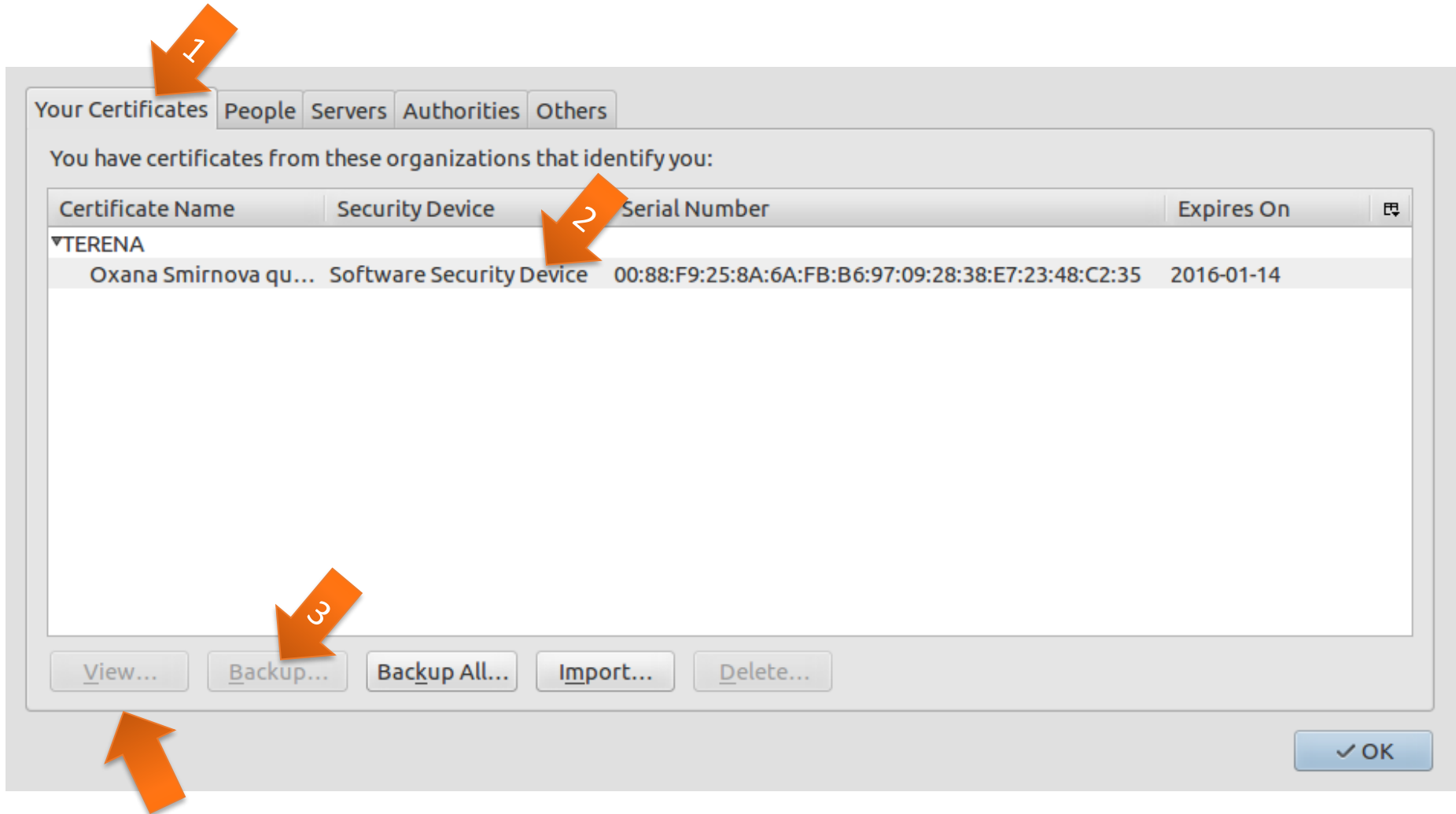
Backup (save) the certificate as a file



Go to Advanced – Certificates – View Certificates

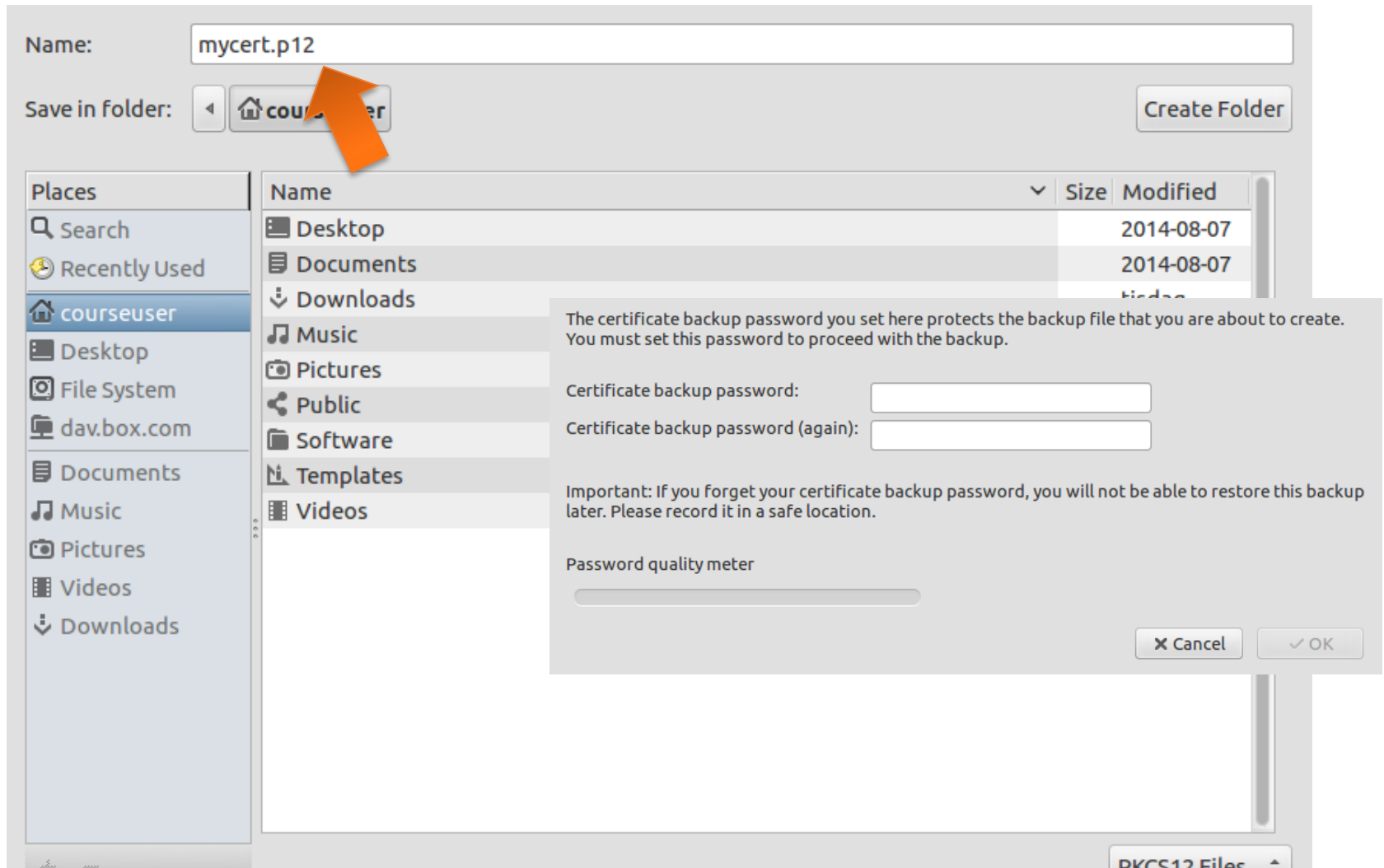


In Your Certificates, find yours, and use Backup to save it



Take some time to view the certificate content

Backup the certificate as a file with extension .p12



- Choose any password you like, make sure to memorise it!

Back to command line: extract private and public keys

- Create a hidden directory `~/ .globus`
 - This is the *default* location for Grid certificates
- Use `openssl` command to extract the keys inside `~/ .globus` :
 - Private key:

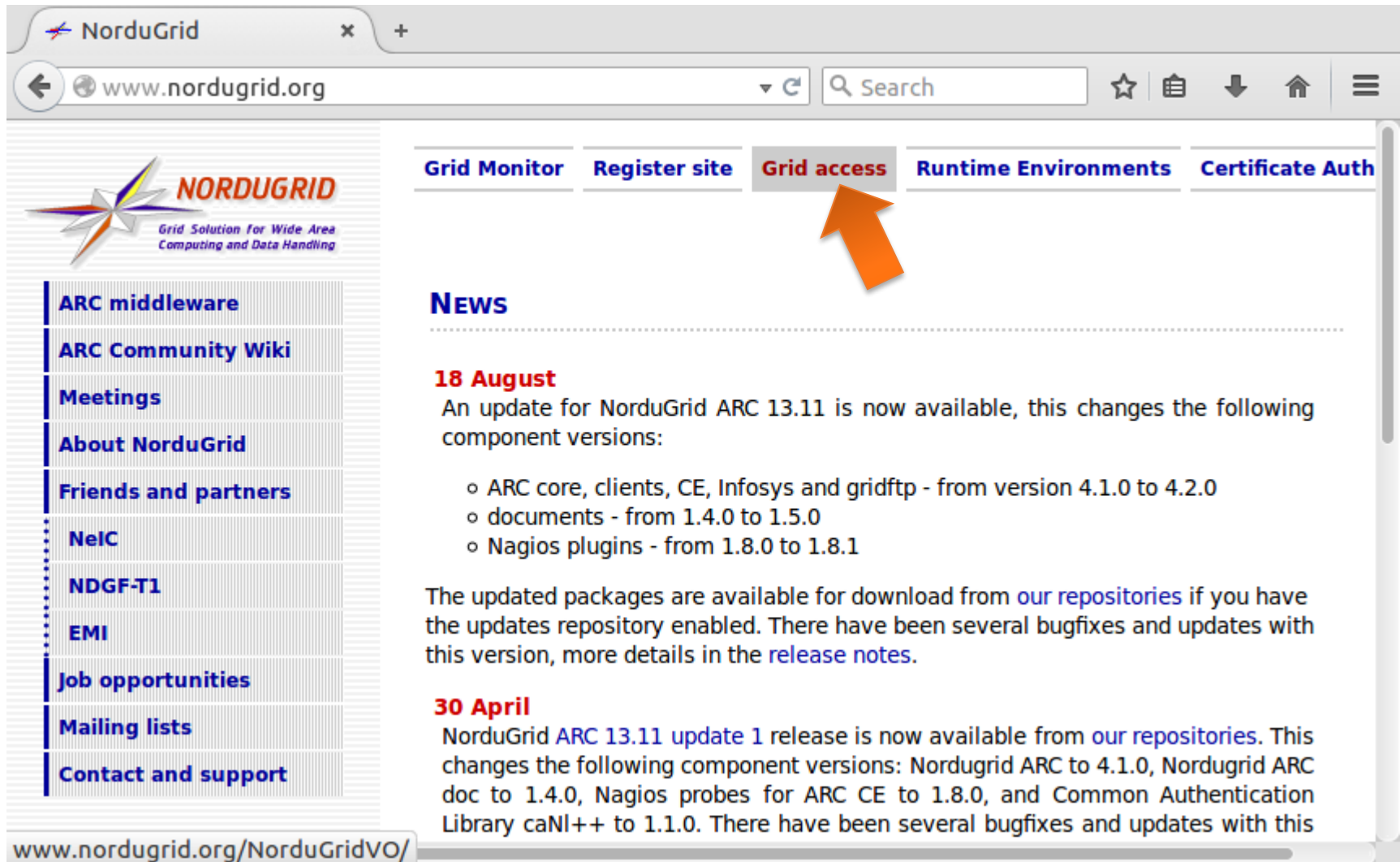
```
openssl pkcs12 -nocerts -in mycert.p12 -out userkey.pem
```
 - Public key:

```
openssl pkcs12 -clcerts -nokeys -in mycert.p12 -out usercert.pem
```
 - Hint: Google for “grid certificate howto” to find where to copy-and-paste from
 - “*Import Password*” is the one you used to backup `mycert.p12` from the browser (in slide 14)
 - “*pass phrase*” for PEM key is your own choice
 - You can use the same password in both cases
- Copy the files `userkey.pem` and `usercert.pem` into `~/ .globus`
 - Make sure that `userkey.pem` is readable only by you!
 - Hint: use `ls -al` and `chmod`

Summary of the steps:

```
oxana@gus:~ >
oxana@gus:~ >
oxana@gus:~ > cd .globus
oxana@gus:~/globus > ls -al terena-17.p12
-rw-r--r-- 1 oxana oxana 5761 Sep 17 21:26 terena-17.p12
oxana@gus:~/globus > openssl pkcs12 -nocerts -in terena-17.p12 -out userkey.pem
Enter Import Password:
MAC verified OK
Enter PEM pass phrase:
Verifying - Enter PEM pass phrase:
oxana@gus:~/globus > openssl pkcs12 -clcerts -nokeys -in terena-17.p12 -out usercert.pem
Enter Import Password:
MAC verified OK
oxana@gus:~/globus > ls -al userkey.pem usercert.pem
-rw-rw-r-- 1 oxana oxana 2280 Sep 27 21:50 usercert.pem
-rw-rw-r-- 1 oxana oxana 2019 Sep 27 21:49 userkey.pem
oxana@gus:~/globus > chmod 400 userkey.pem
oxana@gus:~/globus > ls -al userkey.pem usercert.pem
-rw-rw-r-- 1 oxana oxana 2280 Sep 27 21:50 usercert.pem
-r----- 1 oxana oxana 2019 Sep 27 21:49 userkey.pem
oxana@gus:~/globus > █
```


Step 3: Join a Virtual Organisation (Google for NorduGrid)



The screenshot shows a web browser window with the address bar displaying www.nordugrid.org. The page features a navigation menu with the following items: [Grid Monitor](#), [Register site](#), [Grid access](#) (highlighted with an orange arrow), [Runtime Environments](#), and [Certificate Auth](#). On the left side, there is a sidebar with a logo for NORDUGRID and a list of links: [ARC middleware](#), [ARC Community Wiki](#), [Meetings](#), [About NorduGrid](#), [Friends and partners](#), [NeIC](#), [NDGF-T1](#), [EMI](#), [Job opportunities](#), [Mailing lists](#), and [Contact and support](#). The main content area is titled **NEWS** and contains two news items:

18 August
An update for NorduGrid ARC 13.11 is now available, this changes the following component versions:

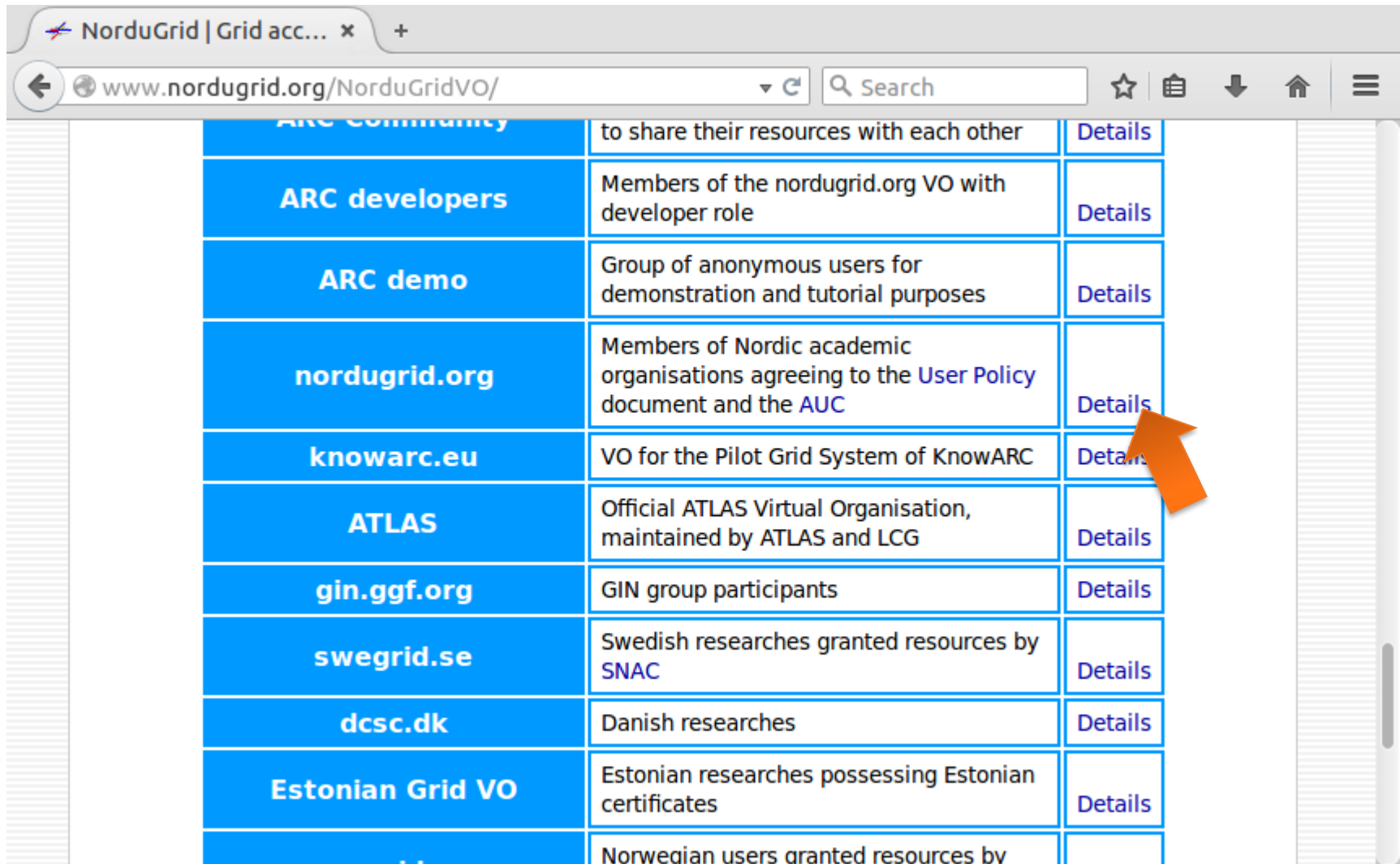
- ARC core, clients, CE, Infosys and gridftp - from version 4.1.0 to 4.2.0
- documents - from 1.4.0 to 1.5.0
- Nagios plugins - from 1.8.0 to 1.8.1

The updated packages are available for download from [our repositories](#) if you have the updates repository enabled. There have been several bugfixes and updates with this version, more details in the [release notes](#).

30 April
Nordugrid [ARC 13.11 update 1](#) release is now available from [our repositories](#). This changes the following component versions: Nordugrid ARC to 4.1.0, Nordugrid ARC doc to 1.4.0, Nagios probes for ARC CE to 1.8.0, and Common Authentication Library caNI++ to 1.1.0. There have been several bugfixes and updates with this

At the bottom of the page, the URL www.nordugrid.org/NorduGridVO/ is visible.

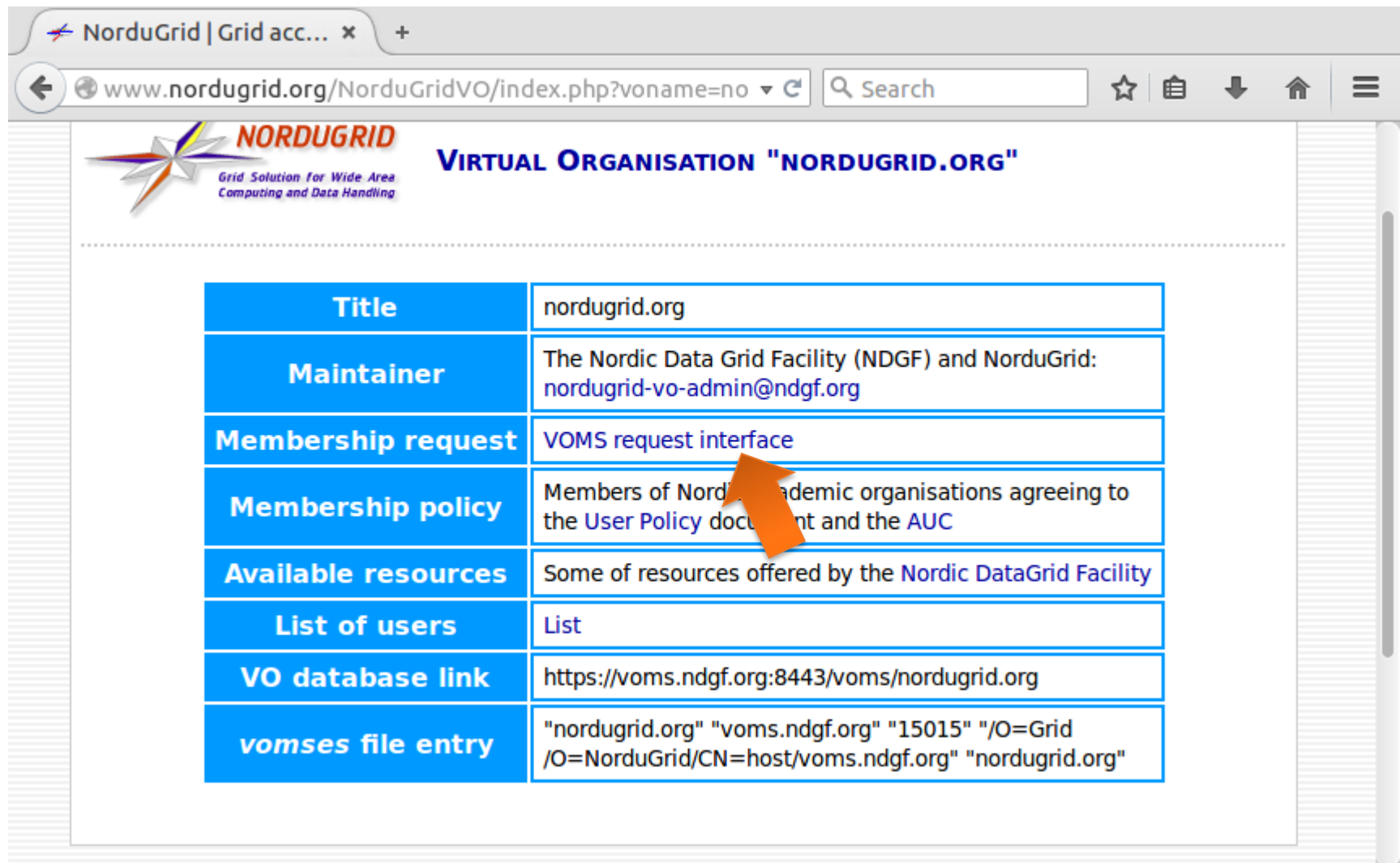
Find nordugrid.org and click Details



The screenshot shows a web browser window with the address bar displaying `www.nordugrid.org/NorduGridVO/`. The page content is a table listing various Virtual Organizations (VOs) with their descriptions and links to their details.

ARC community	to share their resources with each other	Details
ARC developers	Members of the nordugrid.org VO with developer role	Details
ARC demo	Group of anonymous users for demonstration and tutorial purposes	Details
nordugrid.org	Members of Nordic academic organisations agreeing to the User Policy document and the AUC	Details
knowarc.eu	VO for the Pilot Grid System of KnowARC	Details
ATLAS	Official ATLAS Virtual Organisation, maintained by ATLAS and LCG	Details
gin.ggf.org	GIN group participants	Details
swegrid.se	Swedish researches granted resources by SNAC	Details
dcsc.dk	Danish researches	Details
Estonian Grid VO	Estonian researches possessing Estonian certificates	Details
	Norwegian users granted resources by	

Click "VOMS request interface"



The screenshot shows a web browser window with the URL `www.nordugrid.org/NorduGridVO/index.php?voname=no`. The page header includes the NorduGrid logo and the text "VIRTUAL ORGANISATION 'NORDUGRID.ORG'". Below the header is a table with the following content:

Title	nordugrid.org
Maintainer	The Nordic Data Grid Facility (NDGF) and NorduGrid: nordugrid-vo-admin@ndgf.org
Membership request	VOMS request interface
Membership policy	Members of Nordic academic organisations agreeing to the User Policy document and the AUC
Available resources	Some of resources offered by the Nordic DataGrid Facility
List of users	List
VO database link	https://voms.ndgf.org:8443/voms/nordugrid.org
vomses file entry	"nordugrid.org" "voms.ndgf.org" "15015" "/O=Grid /O=NorduGrid/CN=host/voms.ndgf.org" "nordugrid.org"

The server requires your certificate:

This site has requested that you identify yourself with a certificate:
host/voms.ndgf.org (:8443)
Organization: "Grid"
Issued Under: "Grid"


Choose a certificate to present as identification:

Oxana Smirnova quar-osm@lu.se's TERENA ID [00:88:F9:25:8A:6A:FB:B6:97:09:28:38:E7:23:48:C2:35] ▾

Details of selected certificate:

Issued to: CN=Oxana Smirnova quar-osm@lu.se,O=Lunds Universitet,C=SE,DC=tcs,DC=terena,DC=org
Serial Number: 00:88:F9:25:8A:6A:FB:B6:97:09:28:38:E7:23:48:C2:35
Valid from 2014-12-14 01:00:00 to 2016-01-14 00:59:59
Certificate Key Usage: Signing,Key Encipherment
Email: oxana.smirnova@hep.lu.se
Issued by: CN=TERENA eScience Personal CA,O=TERENA,C=NL
Stored in: Software Security Device

Remember this decision



You have to establish trust with the server

Untrusted Connection x +

https://voms.ndgf.org:8443/voms/nordugrid.org

This Connection is Untrusted

You have asked Firefox to connect securely to **voms.ndgf.org**. This connection is secure.

Normally, when you try to connect securely, sites will prove they are going to the right place. However, this site's identity is not trusted.

What Should I Do?

If you usually connect to this site without problems, they may be impersonating the site, and you shouldn't continue.

[Get me out of here!](#)

- ▶ **Technical Details**
- ▼ **I Understand the Risks**

If you understand what's going on, you can tell Firefox to start trusting this site's identification. **Even if you trust the site, this error should mean that someone is tampering with your connection.**

Don't add an exception unless you know there's a good reason why this site doesn't use trusted identification.

[Add Exception...](#)

You are about to override how Firefox identifies this site. Legitimate banks, stores, and other public sites will not ask you to do this.

Server
Location: [Get Certificate](#)

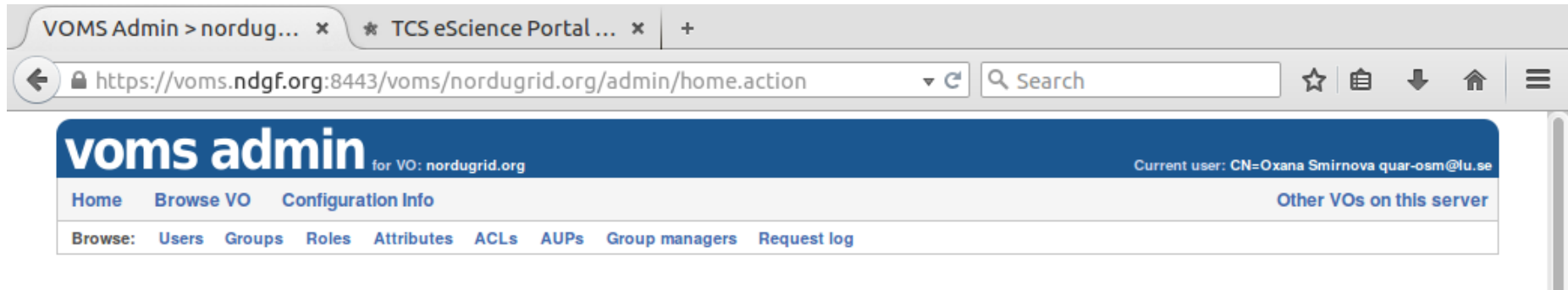
Certificate Status
This site attempts to identify itself with invalid information. [View...](#)

Unknown Identity
Certificate is not trusted, because it hasn't been verified by a recognized authority using a secure signature.

Permanently store this exception

[Confirm Security Exception](#) [X Cancel](#)

Fill in your details and request VO membership



- Check your e-mail: VOMS will ask to confirm the request
 - Give teacher a moment to approve the request
- Ask the teacher if something is unclear
- What happens if your certificate is issued by a non-trusted CA?

On to Grid: create a proxy!

```
File Edit Tabs Help
courseuser@Lubuntu-VirtualBox:~$ arcproxy
Enter pass phrase for private key:
Your identity: /DC=org/DC=terena/DC=tcs/C=SE/O=Lunds Universitet/CN=Oxana Smirnova quar-osm@lu.se
Proxy generation succeeded
Your proxy is valid until: 2014-12-15 12:32:44
courseuser@Lubuntu-VirtualBox:~$
```

- Simply type **arcproxy** and enter your Grid password (*PEM pass phrase* for the private key, as in slide 15)
- Type **arcproxy -I** to check information about this proxy:
 - Issuer (must be you, of course)
 - Your Distinguished Name (DN), a.k.a Identity
 - Validity period

What actually `arcproxy` does?

- A **new** private/public key pair is created for each proxy
 - When a proxy expires, a new one must be created to continue working
 - Default expiration time is 24 hours
- A proxy is then constructed of:
 1. Public certificate (with public key embedded)
 - Certificate contains modified owner's Distinguished Name (has "*proxy*" appended to the name)
 - Owner's DN:
`/C=UK/O=Grid/OU=CenterA/L=LabX/CN=john doe`
 - Proxy DN:
`/C=UK/O=Grid/OU=CenterA/L=LabX/CN=john doe/CN=proxy`
 - Certificate is signed by the proxy owner's **real** private key
 - Certificate contains validity period
 2. Private key
 3. Optionally, Attribute Certificates – extensions containing additional information

The tale of two proxies

- A user always has to create a proxy certificate **P1**
 - Technically, it can be sent to the server, but it is a security breach
- Any Grid server (e.g. a Computing Element) creates itself a delegated proxy **P2** for each user request:
 1. Server generates a **new** private/public key pair (yes, that's a 3rd one...)
 2. Server returns the generated public key as a certificate request to the user
 3. User's tool signs that public key and inserts user information (DN etc), thus generating a public certificate. It uses the private key of proxy P1 for performing signing operation.
 - It can also use the actual private key, but that will require entering password every time!
 4. User's tool sends the signed public certificate back to the server
 5. Server adds generated private key to that certificate and creates a delegated proxy **P2**

What's the use of VOMS

- A Grid user must become a member of a Virtual Organisation (VO)
 - *VOMS* is the most common VO management system
- A Grid cluster administrator gets the list of authorised users from the VOMS database
- VOMS can add extra VO information to your proxy, if necessary
 - For example, your VO role, group etc
 - You should use **arcproxy** with special command-line options to request such extra information to be added
 - We won't try it today

Summary of the proxies

- Luckily, all authentication and delegation procedures are a part of the protocol, you only need to create a proxy
- You have to create a proxy before every Grid activity
- Proxies expire quickly!
 - Resist temptation to create long-living proxy: this will undermine your security
- Proxies may have special extensions, specific to Virtual Organisations
- If you forget your Grid password (PEM pass phrase), and even the browser Import Password, you will have to request a new certificate

Workflow: Grid vs PC/cluster

PC/cluster

Log in via SSH

- Different logins on different machines

Familiarize with the environment

- OS, installed software, storage space, batch system, sysadmin etc

Customize the environment

- Pathes, environment variables, own software, scripts, data files etc

Prepare for batch submission

- Interactive execution of short jobs, optimization of batch scripts

Submit jobs to the batch system, check their status

- Different batch systems (or none) on different machines

Log out

Log in later to fetch the output

Grid

Create proxy

- One for all machines

Create a Grid job description document

- Generalization of batch scripts, plus input/output data location etc

Test a couple of jobs, fix job description

Submit jobs to the Grid, check their status

- Same commands for all machines

Watch output appearing in the desired location

- Or fetch it manually

Simplest Grid job submission

- Your Grid client should:

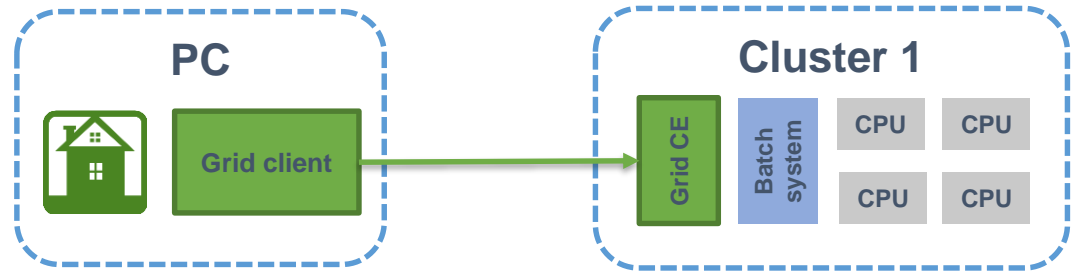
- Create a proxy:

- **arcproxy**

- Submit the job description document to the cluster:

- **arcsub -c arc-iridium.lunarc.lu.se hello_grid.xrsl**

- **arcsub** will refuse submission if the cluster does not meet job requirements



See the *ARC Clients manual* for info about all ARC client commands:
<http://www.nordugrid.org/documents/arc-ui.pdf>

- The CE on the cluster should:

- Check whether you are authorised
- Fetch input file (if requested)
- Convert job description to a batch script and start a batch job
- Upload output file (if requested)

Simplest Grid job description: hello_grid.xrsl

```
&( executable = "/bin/echo" )
  ( arguments = "hello grid" )
  ( stdout = "stdout_file" )
  ( stderr = "error_file" )
  ( cputime = "13" )
  ( gmlog = "grid_log" )
  ( jobname = "hello_grid" )
```

attribute

value

- Yes, this is yet another language:
XRSL – eXtended Resource Specification Language
 - File extension is **.xrsl**
- XRSL is not a standard language, but no standard exists
 - There are many other Grid languages and meta-languages
 - XRSL is an ARC extension of the original Grid language by Globus
 - It was actually modelled on the LDAP database query language
 - Is a list of attribute-value pairs

Main attributes of job description

Job attribute description	Attribute name (XRSL)	Example value
Main executable (binary or script)	executable	MyAnalysis.py
Arguments of the executable	arguments	-i input.dat -o output.dat
Input files	inputfiles	https://store.lu.se/physlab/2012/file1.dat
Output files	outputfiles	https://store.lu.se/physlab/2014/file1.dat
Standard input file	stdin	stdin.txt
Standard output file	stdout	stdout.txt
Standard error file	stderr	stderr.txt
Time (used by CPU)	cputime	1 hour
Memory (maximum needed, Mbytes)	memory	1000
Disk space (maximum needed, Mbytes)	disk	1000
Job name	jobname	My data analysis
Number of slots (cores) for the job	count	36

and many others: ARC job description language XRSL has 37 attributes, see <http://www.nordugrid.org/documents/xrsl.pdf>

Create and submit your hello_grid.xrsl

- Prepare job description for the “Hello Grid” task:
 - Use **geany** (or any text editor) to create a file **hello_grid.xrsl**
 - Use at least the following XRSL attributes: **executable**, **arguments**, **jobname**
 - Hint: copy the example from slide 30
- Submit your first Grid job to our Iridium cluster:
 - First, make sure that you have a valid proxy:
arcproxy -I
 - Use the **arcsub** command with explicit cluster selection:
arcsub -c arc-iridium.lunarc.lu.se hello_grid.xrsl
 - Find the returned **job ID** (a long string that looks like a URL)
 - Check the job’s status (copy-and-paste job ID):
arcstat <jobid>
 - Check what the job “session directory” looks like on the remote cluster:
arcls <jobid>
 - Check what does the job print out:
arccat <jobid>

Manipulate the jobs: kill, retrieve

- Submit a couple more jobs
 - You may want to change the job names in `hello_grid.xrsl`
 - Or you may even want to change what do the jobs produce
- Check the status of all your jobs:
`arcstat -a`
- Terminate some of them and check the status afterwards:
`arckill -k <jobid>`
`arcstat <jobid>`
 - `-k` here means “keep the job files”, otherwise they will be wiped out
- Retrieve job results (download job output):
`arcget -k <jobid>`
 - `-k` here has the same meaning as for `arckill`
- Find where the downloaded files are, and look what is there
 - Inspect the content of the `gmlog` sub-directory: it has files useful for error diagnostics and debugging

If you have some time left

- Find the hidden directory `~/.arc` and copy the file `/etc/arc/client.conf` into `~/.arc`:

```
cp /etc/arc/client.conf ~/.arc/.
```

- Open `client.conf` in Geany (or any other editor)
- Find blocks `[registry/index1]` , `[registry/index2]` etc and uncomment them and all their content
 - Save `client.conf` and quit the editor
- Try to submit `hello_grid.xrsl` to the entire Grid!

```
arcsb hello_grid.xrsl
```

(this is likely to fail, since nobody may authorize you yet, but at least you will see fascinating error messages...)