Prototype DIRAC Web Portal for EISCAT data — Step by step instructions

Carl-Fredrik Enell, EISCAT Scientific Association

June 7, 2019

1 Introduction and history

EGI (`http://www.egi.eu`) is a federated e-Infrastructure set up to provide advanced computing services for research and innovation. The Horizon 2020 EINFRA-1-2014 project EGI-Engage supported a number of competence centres (CCs) for research infrastructures. In this framework, the CC for EISCAT_3D started to develop a data search and analysis portal for the future users of EISCAT_3D.

Several services developed by EGI, as well as the related Horizon 2010 projects Indigo DataCloud and EUDAT `http://www.eudat.eu`, are now part of the European Open Science Cloud Hub (EOSC-Hub), `http://www.eosc-hub.eu`, a currently running project funded by the Horizon 2020 research and innovation programme under grant agreement No 777536.

Thus, the development initiated in the EGI-Engage CC is continued in the EOSC-Hub CC for EISCAT_3D. Partners in this project are EISCAT HQ, Sweden, CNRS-IN2P3, France, and CSC in Finland.

1.1 DIRAC

The EISCAT_3D competence centre aims at developing the EISCAT_3D user portal, which is to be the primary interface through which users will browse, download and analyse EISCAT_3D data. We chose to work closely with the developers of the Distributed Infrastructure with Remote Agent Control (DIRAC) project, `http://diracgrid.org`, which was originally developed for distributing data from the LHCb project at CERN. DIRAC is an interware, mostly written in Python, that provides command-line, web and API interfaces to grid computing and storage resources. The major task of the EISCAT_3D CC so far has been to implement a prototype DIRAC portal for legacy EISCAT data, i.e. lag profile data (time-averaged spectral data in the autocorrelation domain). These data are stored in a Matlab version 4 compatible format at EISCAT HQ and indexed in a MySQL database.

Detailed user and administrator instructions for DIRAC can be found at `https://dirac.readthedocs.io/en/latest/UserGuide/index.html`. It is highly recommended to read the user instructions carefully to obtain a good understanding of what DIRAC can do.

1.2 Layout of the EISCAT DIRAC service

The prototype EISCAT DIRAC portal comprises the required functions for user authentication, data search, and job submission to computing resources at CSC.

Virtual organisation membership service (VOMS) User access to the EISCAT DIRAC portal is managed through membership in groups in the EGI virtual organisation (VO) **eiscat.se**.

The VOMS used by EGI is called Perun and runs at Cesnet, accessible to VOMS managers through a web GUI at `https://perun.cesnet.cz`

Storage element Access to storage in DIRAC is handled by the storage element (SE) server. This is a Python program and this project uses an instance that runs on server at EISCAT Headquarters. It reads and serves data files from a top level directory, where presently data directories from 1981–1983, 2007 (IPY) and 2011–2019 are linked from the main storage server.

File catalogue This is a database of files available on the storage element, and their metadata. It runs on EGI resources at Cyfronet (`http://www.cyfronet.krakow.pl`) and allows users to list and search the data with a command-line client and the file catalogue application in the web portal.

The file catalogue is updated by a Python script that runs monthly (from cron) on the DIRAC server at EISCAT Headquarters. This script reads metadata from the MySQL database at EISCAT, as available at `https://www.eiscat.se/schedule/schedule.cgi` by selecting Archived data.

Web interface A web interface is the main point of access for users. It also runs at Cyfronet and is accessible at `https://dirac.egi.eu:9443/DIRAC/`.

This web interface presents the user with a GUI environment (Fig. 1) which has applications for file catalogue search, job submission, and more.

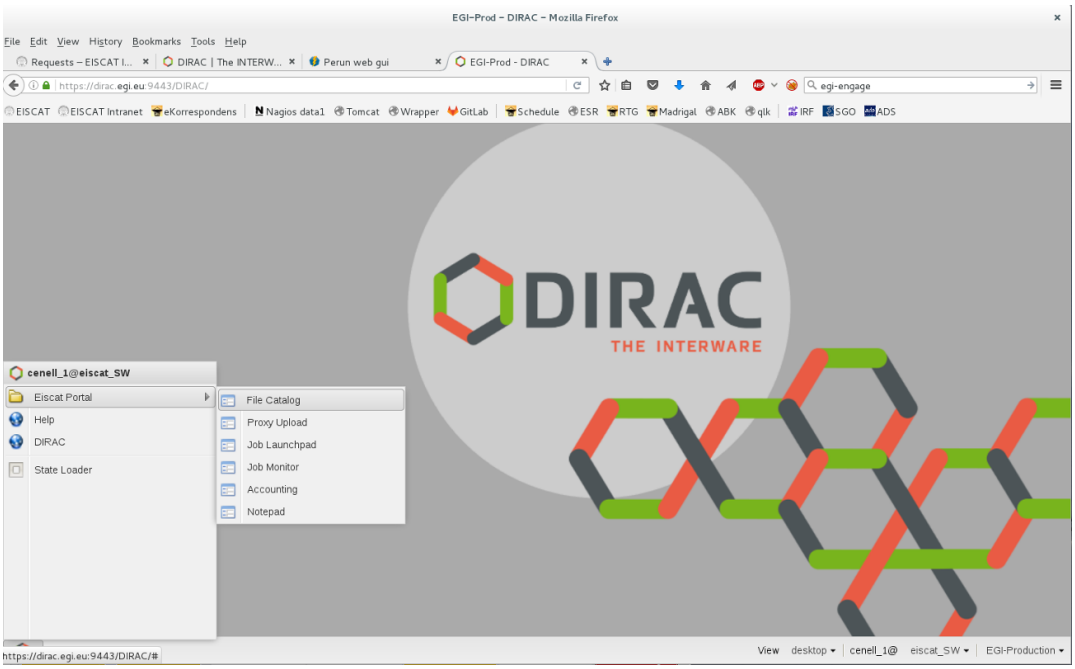

Figure 1: The DIRAC web GUI

2 Accessing the EISCAT_3D DIRAC portal

Access to data in DIRAC requires user authentication and access authorisation through membership in one ore more groups in the Virtual Organisation (VO) eiscat.se.

For user authentication two options are implemented: X.509 certificate authentication and authentication with the EGI Checkin service. Certificate authentication is the most mature system. However, this has turned out to be unknown to most EISCAT users, even those who in principle have access to certificate authorities, and thus a detailed description follows here. For instructions on EGI Checkin authentication, skip to Sect. 2.2 below.

2.1 Certificate authentication

2.1.1 Get a certificate

Skip this section if you already have a Grid Premium type X.509 certificate. In this case it should be installed in your web browser and exportable from there.

1. Log in to your certificate authority. At least for EISCAT users in the Nordic region, this is usually Digicert. Browse to `https://www.digicert.com/sso/` using a compliant browser. Firefox, MS Internet Explorer and Safari should all work, but Google Chrome and MS Edge were not compatible when last tried (this may change).

   On the first page shown you will have to type in the name of your identity provider, i.e. your university or institute. Fig. 2 shows what this looks like for the author, whose identity provider is the Swedish Institute of Space Physics. This will redirect you to a login page where you should be able to log in with the user credentials of your organisation. In many cases this will be the same name and password that you use for your university email and internal web pages.

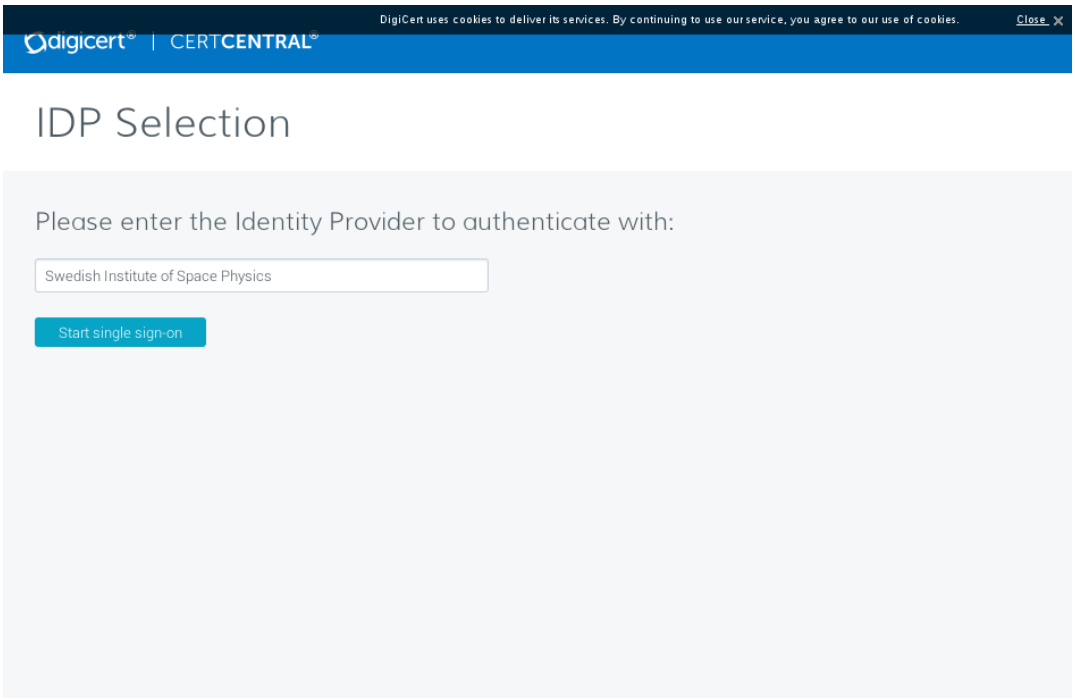

Figure 2: Providing the Identity Provider information to Digicert. Type the name of your provider (such as University of X). This figure shows the identity provider of the author, which is the Swedish Institute of Space Physics.

2. Request a Grid Premium certificate by selecting this in the Product menu and clicking *Request certificate*, as in Fig. 3. Depending on your browser, the certificate may be

installed automatically as part of this step. If not, follow the next step below.

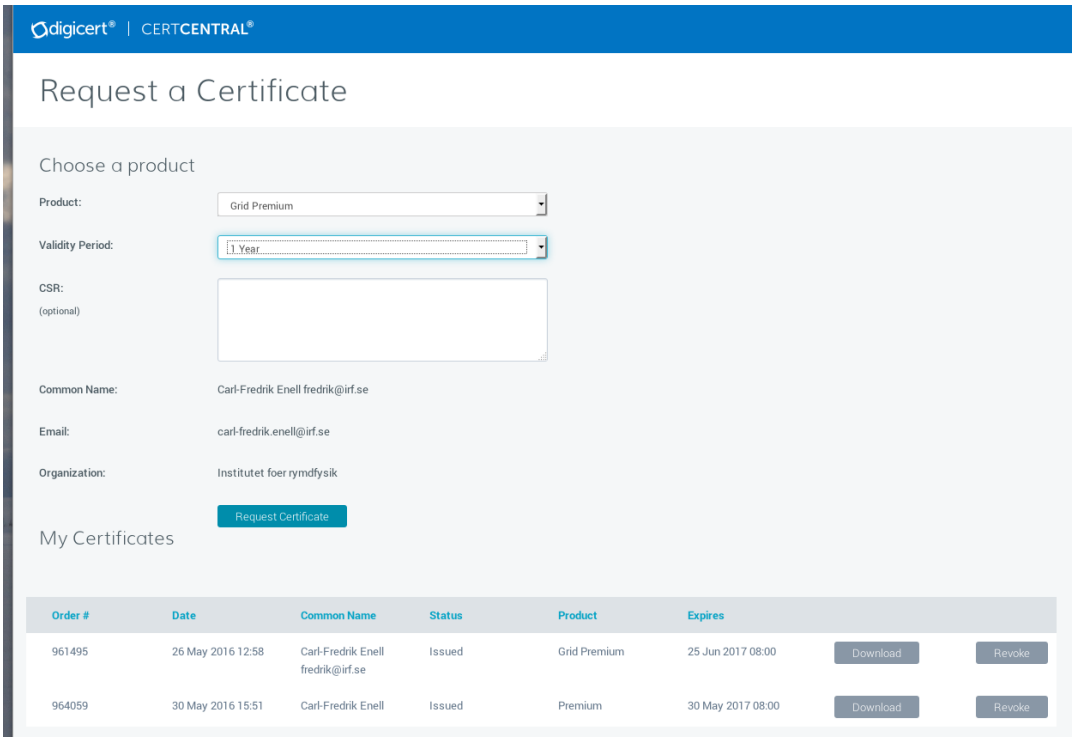

Figure 3: Requesting a Grid Premium certificate from Digicert.

3. If the Request procedure did not install the certificate automatically in your browser, it can be downloaded by looking up the new certificate in the list *My certificates* and clicking the download button. This will give you a zip archive with your personal certificate together with a few other authority files. The certificate file (called something like `yourname.crt`) can then be imported.

   Fig. 4 shows the details of the author's certificate after a successful import into Firefox.

2.1.2 Register Virtual Organisation membership

1. Register to the **eiscat.se** VO through the *Perun* service at `https://perun.metacentrum.cz/cert/registrar/?vo=eiscat.se`. You will have to enter required information and then wait for approval. Fig. shows a completed successful registration.

2. Ask a VO manager (e.g. the author) to add your VO user to the access groups that you are entitled to (usually your EISCAT associate country and common programme data). Fig. 7 shows the author editing access groups, and a similar procedure must be performed for your account.

2.1.3 Proxy upload

1. Upload your certificate to the DIRAC proxy, which is the gateway that allows you to access the DIRAC grid services. This is easiest using the DIRAC command line tools. They are Python and UNIX (bash) shell scripts. If you cannot install them, the GUI also provides a proxy upload applet.

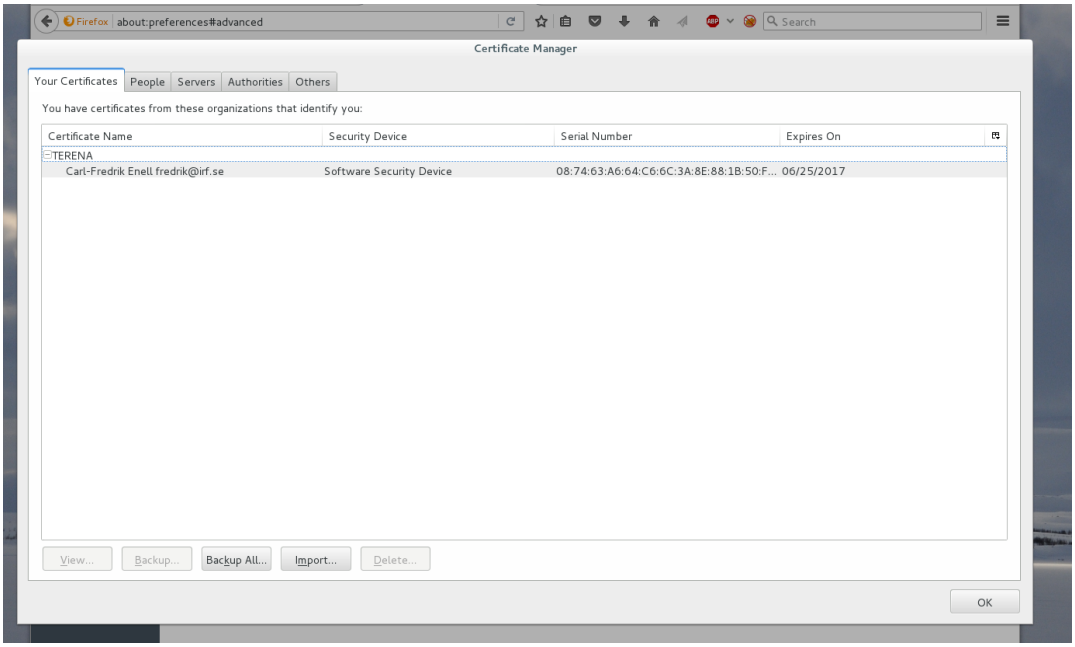

Figure 4: The certificate manager in Firefox (Edit → Preferences → Advanced → Certificates) after successfully importing a certificate from Digicert.

- Export your certificate in p12 format. The export function of your web browser should work. This will ask you to set a password for encryption.
- Option 1: Use the web GUI (see Fig. 6) to upload your certificate to the proxy.
  - Select the Proxy upload applet in the web GUI
  - Browse to your p12 format certificate
  - Type in the password of the p12 certificate
  - Click Upload
- Option 2: Use the DIRAC CLI to install the certificate for the DIRAC services and upload to the proxy. Refer to `http://diracgrid.org` for instructions on how to install the DIRAC client or run it in a Docker container.

  ```
  dirac-cert-convert.sh <YOUR_CERTIFICATE>.p12
  ```

- Upload the certificate and initialise the DIRAC proxy:

  ```
  dirac-proxy-init -M -U -g eiscat_<GROUP>
  ```

  for example:

  ```
  dirac-proxy-init -M -U -g eiscat_FI
  ```

  You will be asked for the password of your certificate (the one you set in the export step above). You can then check whether the upload succeeded like this:

  ```
  dirac-proxy-get-uploaded-info
  ```

  and you should then see something like

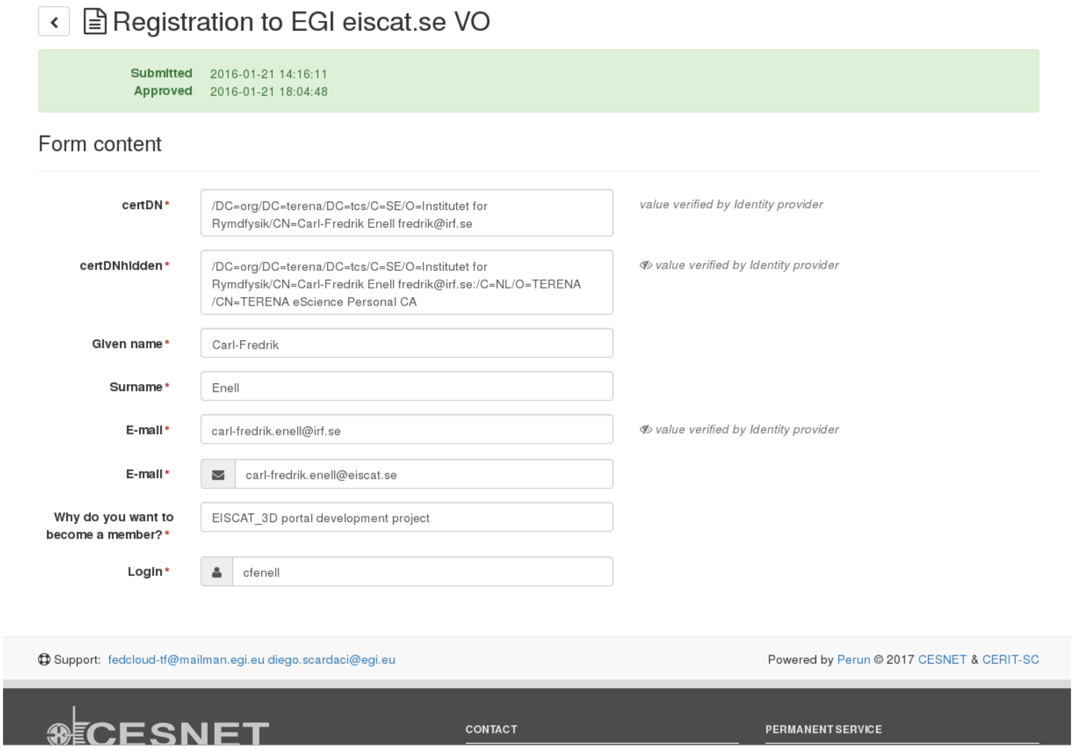

Figure 5: Registration to the EGI VO **eiscat.se** through Perun.

```
Checking for DNs /DC=org/DC=terena/DC=tcs/C=SE/O=Institutet foer rymdfysik/CN=Carl-Fredrik Enell fredrik@irf.se
--------------------------------------------------------------------------------------------------------------------------------------------------------------------
| UserName | UserDN                                                                                        | UserGroup    | ExpirationTime      | PersistentFlag |
--------------------------------------------------------------------------------------------------------------------------------------------
| cenell_1 | /DC=org/DC=terena/DC=tcs/C=SE/O=Institutet foer rymdfysik/CN=Carl-Fredrik Enell fredrik@irf.se | eiscat_SW    | 2017-06-25 11:54:59 | True           |
--------------------------------------------------------------------------------------------------------------------------------------------------------------------
```

To connect to the web GUI, browse to `https://dirac.egi.eu:9443/DIRAC/`. The browser may ask you to use the installed certificate. If not, click the label "Visitor" at the bottom left and choose certificate authentication.

2.2 EGI Checkin authentication

It has turned out that the project needs an optional authentication mechanism to account for users who do not have access to certificates or work outside Europe. EOSC-Hub offers three options, originating from EGI (Checkin), EUDAT (B2ACCESS) and Indigo (IAM) respectively. These can all work with several identity providers, and we chose EGI Checkin for the DIRAC portal. This type of authentication has now been implemented.

To authenticate with EGI Checkin, follow these simple steps:

1. Connect to the EISCAT DIRAC portal at `https://dirac.egi.eu:9443/DIRAC/`. If you don't have a certificate, the login selector in the lower right corner will show "Visitor".

2. Select the EGI Checkin login option, as shown in Fig. 9. Checkin will allow you to select an identity provider, which handles the actual login. Options include EduGAIN and social media accounts.

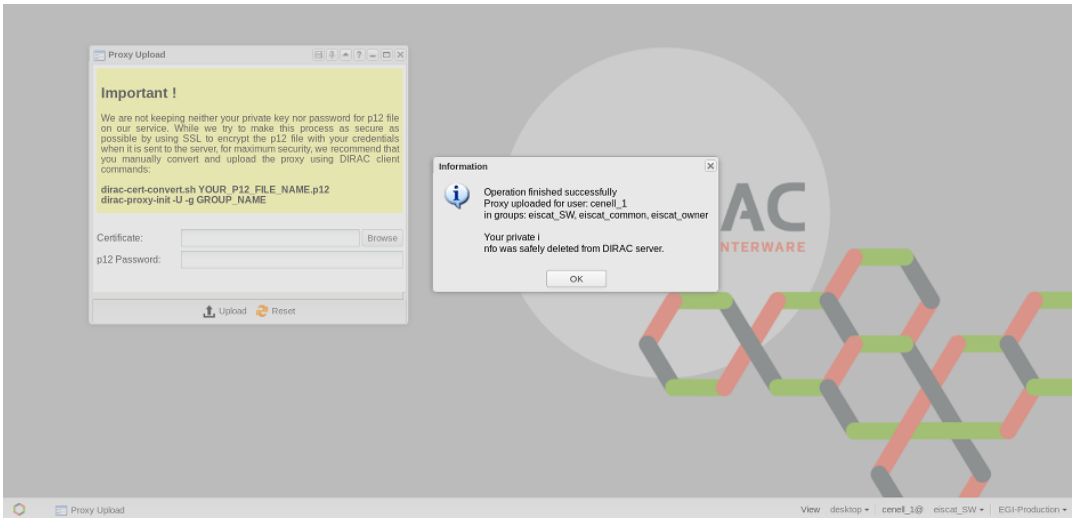

Figure 6: Proxy upload through the web GUI

3. The identity provider will send the credentials back to Checkin, and you will have to accept this.

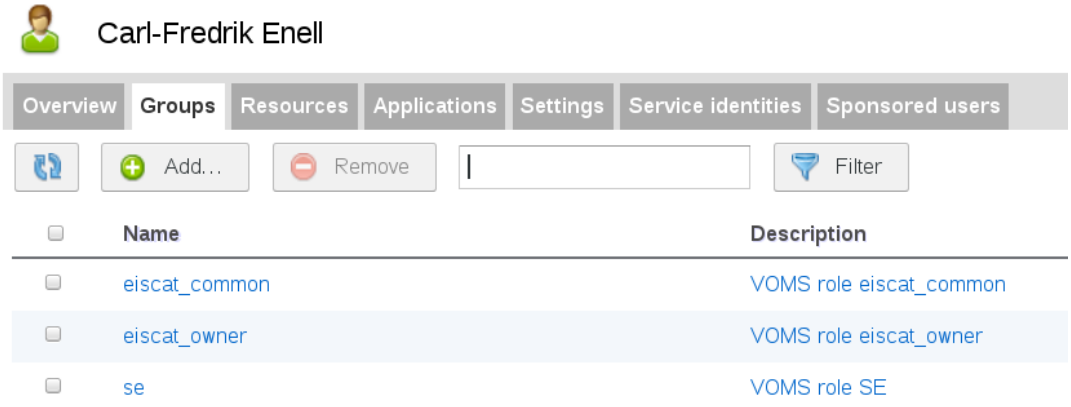

Figure 7: The Groups tab of the Perun VOMS GUI, showing access groups of the author. This registration will be handled by one of the VO managers so as a normal user you will not see this. It is important that you are a member of the appropriate VOMS group(s), however.

Figure 8: Selecting the EGI Checkin login option in the DIRAC Web GUI.

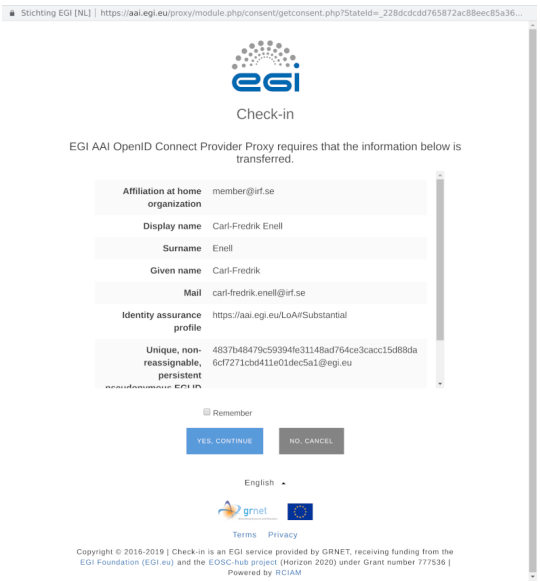

Figure 9: After successfully logging in to the identity provider, EGI Checkin will require permission to transfer the required information. Click Accept on this page.

3 Searching EISCAT data

Once authenticated you can start using the web GUI. The text "Visitor" should now have changed to "EGI Production", and selectors for user name(s) and granted VOMS access groups will appear, as shown in Fig. 10.

Figure 10: Successfully logged in to the DIRAC portal. The lower right text will change from "Visitor" to "EGI Production" and the user can select access groups in which membership has been granted.

The point of start in the portal is the file catalogue GUI. To access it:

1. Connect and authenticate as described
2. Select your VOMS group, such as `eiscat_SE`
3. Go to the Windows style menu at the bottom left and browse to the file catalogue GUI.

3.1 Basic search

- The file catalogue GUI will look much like any file browser. The search will, however, start from a top level directory that you select by right-clicking on a directory in the listing (top right pane) and then clicking `Set as starting path` (See Fig. 11). EISCAT data are in the directory hierarchy `eiscat.se/archive/<year>/`.
- After setting the starting path, additional search criteria can be added using the GUI in the top left pane. See Fig. 12.
- Click Submit (below the bottom left pane).

The search will list any found files in the bottom right pane. After finding the desired files, you can select files for download or processing in the bottom right pane.

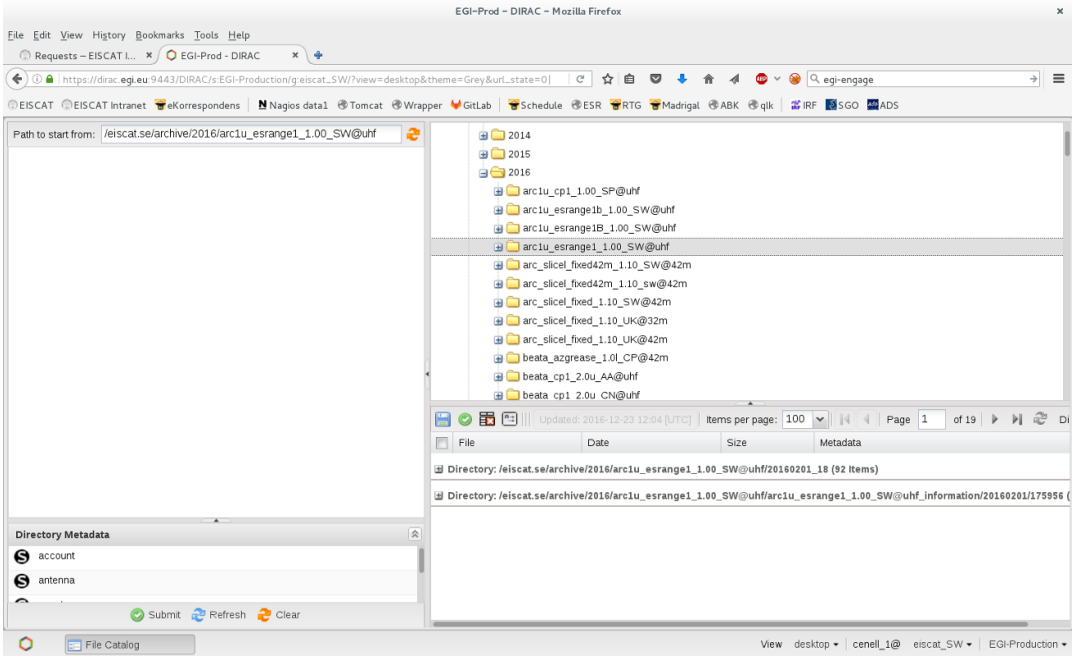

Figure 11: Selecting a top level directory in the file catalogue GUI

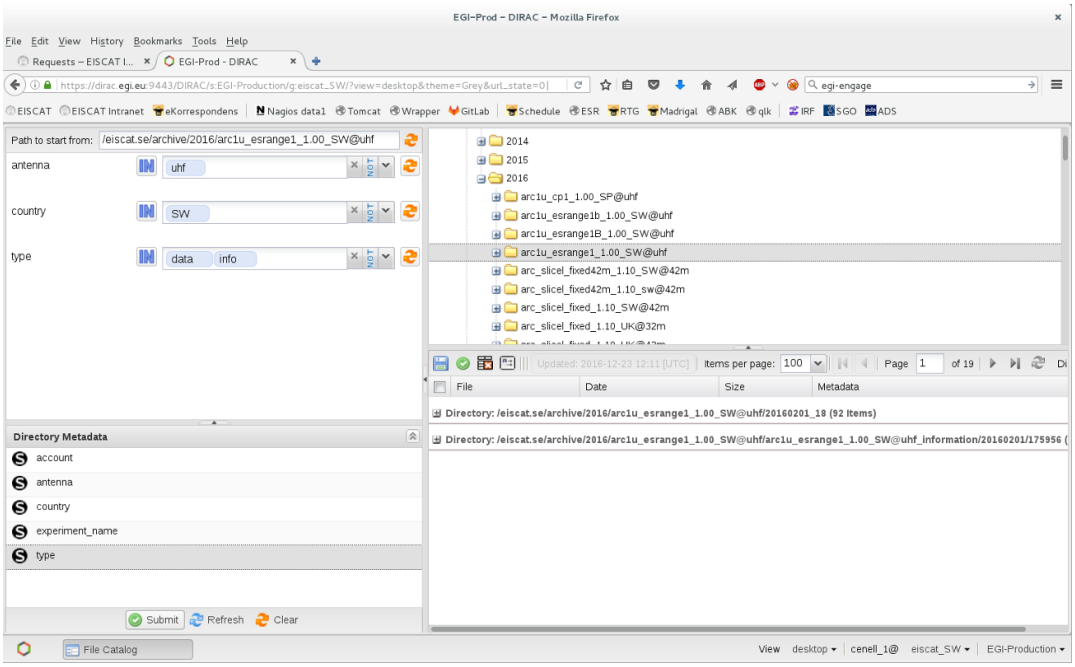

Figure 12: Adding metadata search criteria. In this particular case, the top level directory contains only SW UHF data, so the filters shown as an example on the top left will have no effect. After selecting the search criteria, click the Submit button

4 Downloading EISCAT data

Files selected in the bottom right pane can be downloaded as a ZIP archive by clicking on the diskette icon, as shown in Fig. 13.

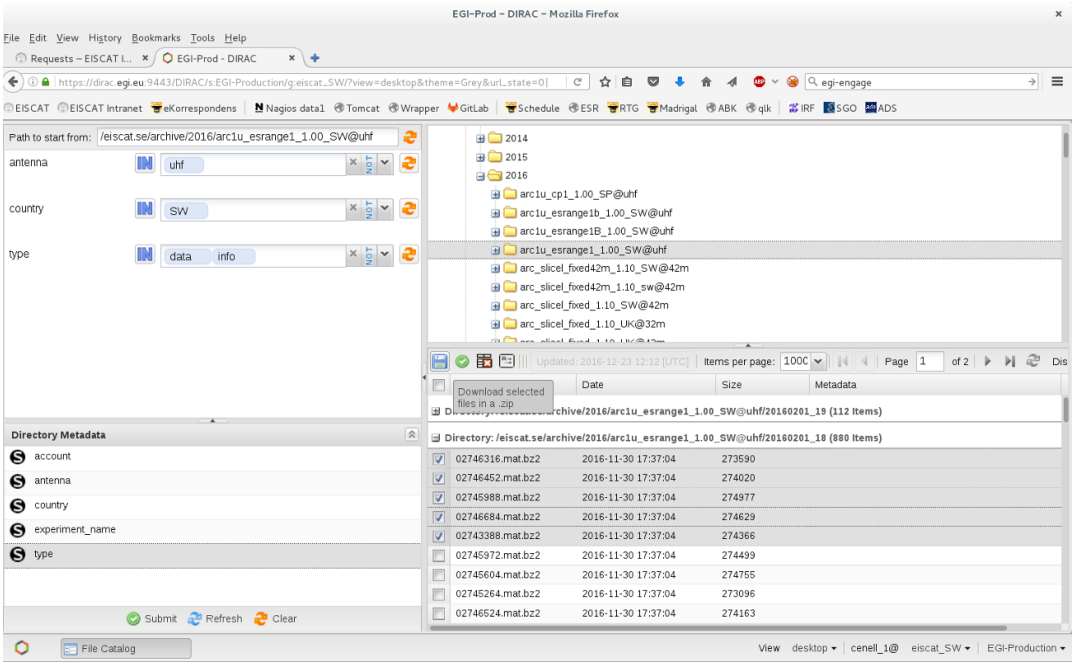

Figure 13: Selecting files to download.

5 Submitting user analysis jobs

5.1 Using the Job Launchpad GUI

The GUI can also submit processing jobs. This procedure will essentially

1. start a pilot job on the compute resources
2. the pilot job downloads and runs the specified software with the selected files as input.

At this stage the standard EISCAT Real Time Graph (RTG) has been implemented and will plot the content of the selected data files (by running the RTG script on the open source Matlab-compatible software Octave).

The Job Launchpad is a separate application in the GUI, but it can also be accessed from the file catalogue as follows:

1. Select files in the same way as for download
2. Click the green job launchpad icon above the file list
3. The job launchpad GUI should open. Right-click on the `EISCAT png maker` directory icon and select `Apply to the selected parameters`. See Fig. 14. The field `Executable` should now show the path to `webtg4dirac`.

4. Check the other parameters as well and click Submit at the bottom of the job launchpad window.

5. Go to the main menu and browse to the job monitor (Fig. 15). Make sure that your user name is preselected and click Submit to see the status of the job.

6. Once the job is finished, the output will be in the Sandbox. Right-click on the job line to see the menu as in Fig. 16.

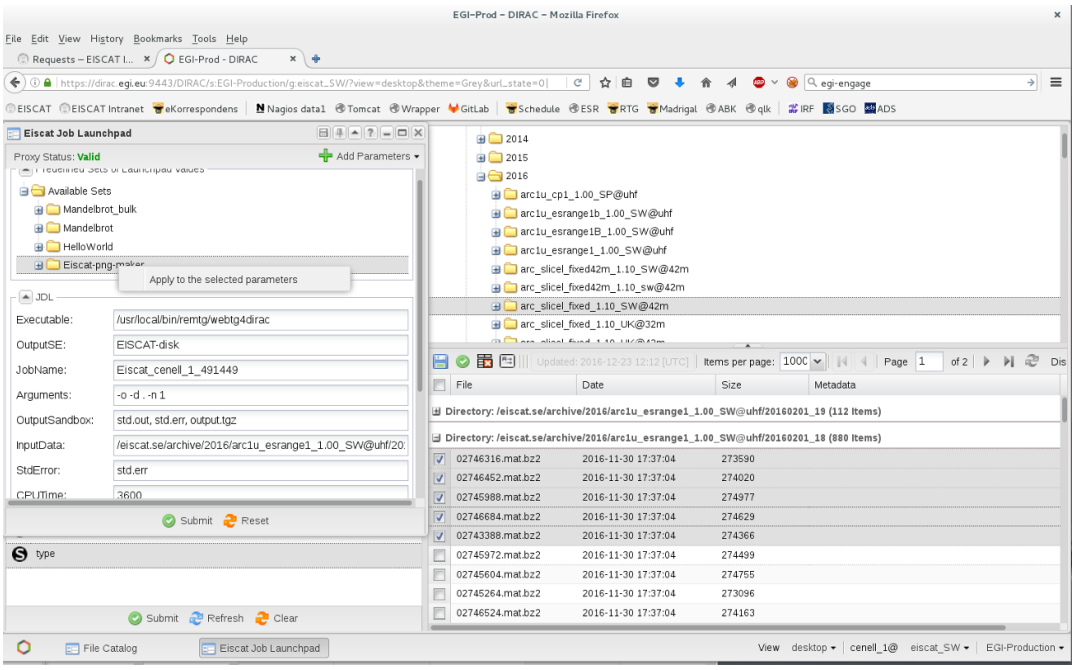

Figure 14: Submitting a job that will plot the selected EISCAT data files with RTG.

5.2 Using the CLI

Jobs can also be submitted from the DIRAC command line. For this, a job description file has to be written and submitted. A typical job description file has this format, known as JDL (Job Description Language):

```
[
  Executable = "run_rtg_docker.sh";
  Arguments = "";
  JobName = "my_job_name_string";
  Site = "Cloud.CSC.fi";
  CPUTime = 86400;
  InputSandbox = {"run_rtg_docker.sh",
        "LFN:/eiscat.se/archive/yyyy/experiment/file1",
        "LFN:/eiscat.se/archive/yyyy/experiment/file2",
        ...
       }
  OutputSandbox = {"output/*"};
]
```

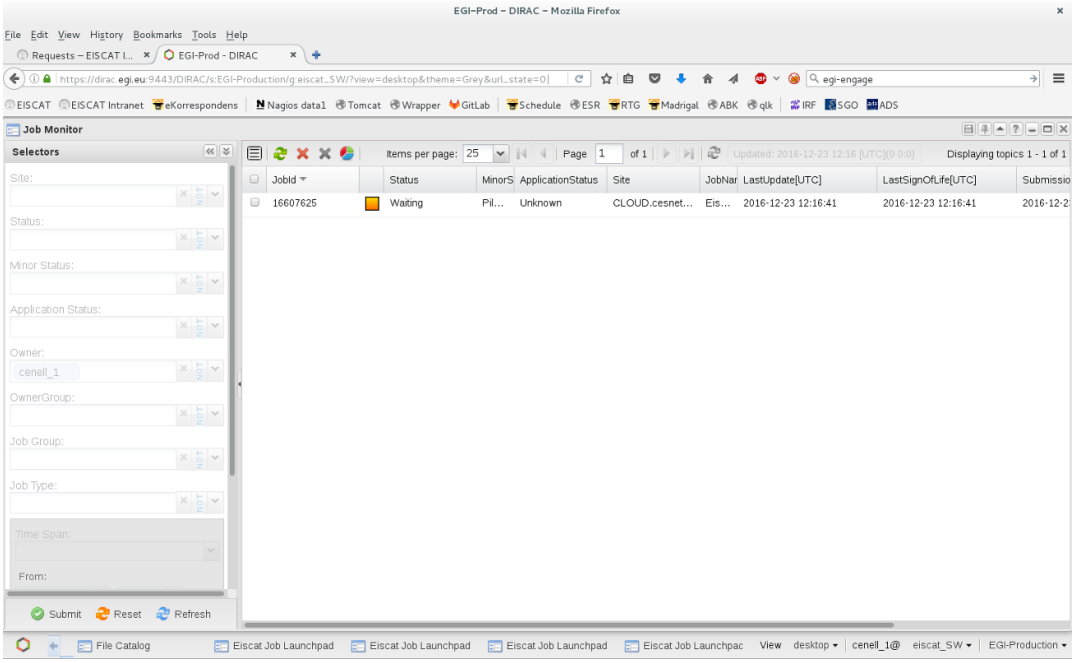

Figure 15: The DIRAC job monitor showing the status of the submitted job.

The principle of job submission through the CLI is thus:

Edit and save a JDL file for your analysis case, e.g. `my_job.jdl`.

Submit the job: `dirac-wms-job-submit my_job.jdl`

For further instructions, please refer to the instructions at `https://dirac.readthedocs.io/en/latest/UserGuide/index.html`

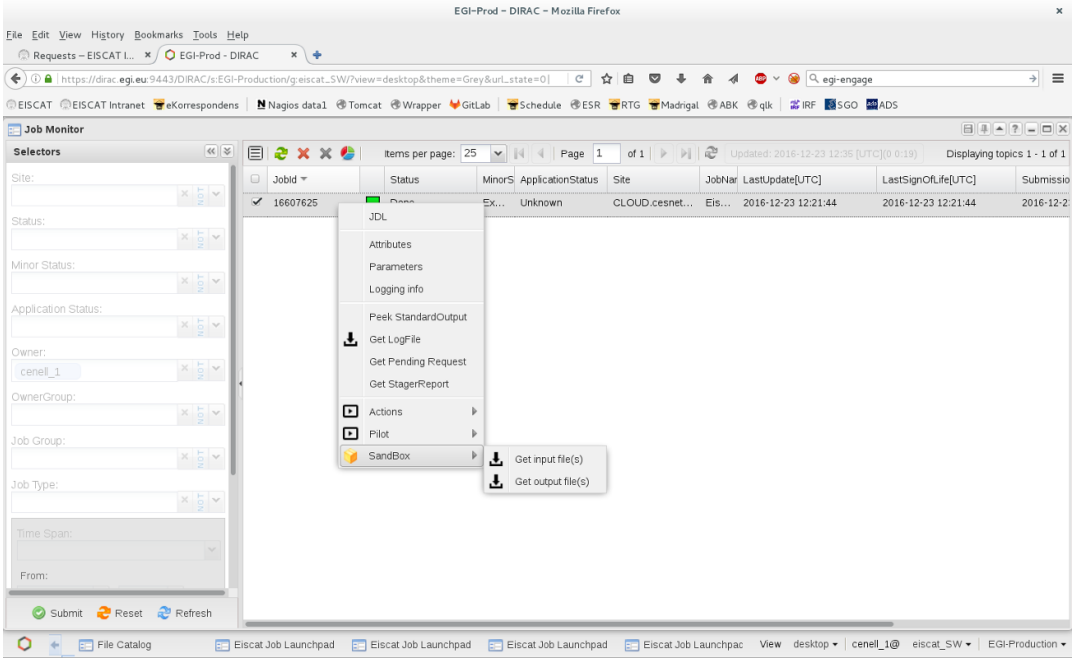

Figure 16: The sandbox contains the output files for download.

6 Feedback

Feedback and bug reports through the EGI RT tracker system will be appreciated. EISCAT portal reports should be submitted through `https://rt.egi.eu/rt/Search/Results.html?Query=Queue%20%3D%20%27dirac4egi-eiscat3d-requirements%27%20AND%20(Status%20%3D%20%27new%27%20OR%20Status%20%3D%20%27open%27%20OR%20Status%20%3D%20%27accepted%27%20OR%20Status%20%3D%20%27developed%27%20OR%20Status%20%3D%20%27stalled%27%20OR%20Status%20%3D%20%27feedback%27)`.