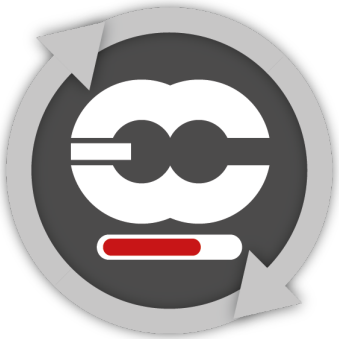


ECMWF Data Downloader Documentation

Release 0.8.4

Olivier Henry

Jun 08, 2018

CONTENTS:

CHAPTER

ONE

INTRODUCTION

The ECMWF Data Downloader is a software written in Python 3.5 and PyQt5, and designed with Qt Creator. Its purpose is to download data from ECMWF public datasets using the official Web API.

The idea behind EDD is to have a software with a minimal Graphical User Interface in order to prepare a query for data in all ECMWF public datasets, using the official web API. The idea comes from a discussion I had with my wife, a geophysist, and the use for the first time of the ECMWF web API with Python. Few years ago, it was still possible to use the ECMWF website to download data on a large period of time. Few months ago, I noticed that it was no longer possible: now the user can donwload data only for a month or for a year. That limitation is not present if one uses the official web API. After struggling for a certain amount of time with the web API and the preparation of a simple query, which worked on Linux but failed on Windows, I decided to write a software dedicated to this task, and which could work on Windows and Linux. Even if the software makes use of the official web API, the official ECMWF python library is not used, the query and the download of data rely on a homemade function and on the library `requests` to handle HTTPS protocol.

Note: EDD is still in beta state. Few queries can lead to errors. The user should use its account to check the status of the query and why it failed (http://apps.ecmwf.int/webmars/joblist/).

CHAPTER

TWO

INSTALLATION

The latest version of ECMWF Data Downloader (EDD) can be obtained from https://github.com/olivierpascalhenry/ECMWF-Data-Downloader. EDD comes in three different versions:

- Sources
- A stand-alone package for Windows (https://github.com/olivierpascalhenry/ECMWF-Data-Downloader/releases, .msi package)
- A stand-alone package for Linux (https://github.com/olivierpascalhenry/ECMWF-Data-Downloader/releases, .tar.gz package)

2.1 Sources

Use of EDD from sources requires the following packages:

- Python 3.5.4 or newer. Available at https://www.python.org/
- PyQt5 5.10 or newer. Available at https://www.riverbankcomputing.com/software/pyqt/download5
- hurry.filesize 0.9 or newer. Available at https://pypi.org/project/hurry.filesize/
- requests 2.18 or newer. Available at https://pypi.org/project/requests/
- selenium v3.11 or newer. Available at https://pypi.org/project/selenium/. Not mandatory to run EDD. Only to build a part of the database.

Then, to launch the software, download sources, uncompress the package, navigate into the new directory and run the following command (Windows):

```
PS User> python ecmwf_data_downloader.py
```

Or (Linux):

```
& python ecmwf_data_downloader.py
```

2.2 Installation of the stand-alone package

2.2.1 For Windows

The installation of the package for Windows is really simple. Just donwload the .msi file from its repository on GitHub, double click on it to launch the installation and follow the instructions on screen. Once the installation is done, you can launch the software from its icon on Windows desktop, or from its folder in Windows start menu.

To avoid issues with admin rights, EDD should be installed outside of Program Files folder: two files are created when it is launched for the first times, *ecmwf_downloader.ini* which contains all options (by default) of the software, and *ecmwf_downloader_log.out* which contains all log messages of the software. Installing EDD in Program Files could lead to the rejection of the creation of both files.

2.2.2 For Linux

The installation of the package for Linux is also really simple. Just donwload the .tar.gz file from its repository on GitHub, uncompress it somewhere in your HOME folder, navigate to the new folder and launch the executable by double-clicking on it or from the terminal.

2.3 Options

A file is creating when the software is executed for the first time. This file, called *ecmwf_downloader.ini* contains the different options of the software. It is possible to modify those options in a dedicated window, detailed in the following chapter. Here is a list of those options:

- *level*: the logging level (`DEBUG`, `INFO`, `WARNING`, `CRITICAL`, `ERROR`). In beta version, by default it is set on `DEBUG`. In final version, it will be set on `INFO`. It can be changed by the user from the Options window.
- *path*: where the logging file should be saved. For those who want to keep all their logging file at the same place. By default the option is empy, which means the logging file is created in the software folder. It can be changed by the user from the Options window. If the path doesn't exist, an error message is added to the log file and the path is reseted to the default path.
- *check_update*: when on `TRUE`, the software will check automatically for an update at each startup. `TRUE` by default, it can be change by the user from the Options window.
- *language*: set on english actually. Probably used in a future version, once translations are available. It can't be changed by the user at this time.
- *display_api_info*: when on `TRUE`, a window will be displayed at startup to give information about the ECMWF web API. `TRUE` by default, it can be change by the user from the Options window.
- *email*: the email registered in the ECMWF account. It is only used when using the ECMWF web API. It can be changed by the user from the Options window.
- *url*: the url of the ECMWF web API. By default it is alreay embedded in the software. It can be changed by the user from the Options window.
- *key*: the key provided by the user ECMWF account. Mandatory to connect to the ECMWF web API. It can be changed by the user from the Options window.
- *folder*: where files are stored once downloaded. It can be changed by the user from the Options window. EDD check the path at each startup. If the path doesn't exist, an error message is added to the log file and displayed to the user, and the path is reseted to the default path.

2.4 Log

A logging system is available in EDD. By default, the output file is available in the directory of EDD and the logging level has been set to DEBUG until a final release. Both options for logging level and logging location have been set in a config file, those options can be changed by the user in the Options window.

If issues are noticed, the logging file should be attached to the message when reporting the issue.

2.5 Update

An automatic update system is available for EDD, when installed from stand-alone packages. Once an update is available for EDD, the Update icon is enabled. If the user click on it, a warning window about the availability of an update will be displayed, asking him to click on **Update** to start the automatic update procedure.

With sources, the same icon informs about an update and proposes to the user to download the update. Then it is up to the user to install the update or not.

The checking of an update can be disabled in the Options window.

CHAPTER

THREE

DESCRIPTION

3.1 Important information

In the following tutorial, all pictures and all commands have been captured on a Windows 10 system.

3.2 Exploring ECMWF Data Downloader

The simplest way to start working with EDD (ECMWF Data Downloader) is to run it by double-clicking on the executable once the software has been downloaded and installed.

EDD is easy to use and self explaining by the presence of information buttons (once a button is clicked, a popup appears with text to explain the purpose of the area actually used) and tool tips. It is composed of different windows, designed to display information to the user and accept interactions.

3.2.1 Current GUI limitations

Actually, the GUI has few limitations based on the fact that the software doesn't include any database or any links to the ECMWF database. The interactions betweens the different objects of the main window and how the tabs behave between them is purely based on the assumptions of the developer. Thus it could lead to the rejection of queries by the ECMWF web API because queries can contain errors. In that case, a window which handle raw inputs has been included for expert users.

3.2.2 The main window

The main window of the GUI (cf. Fig. 3.1) is what the user sees once he launches the software.

The main window is composed of three parts, from up to down:

- the first one (from 1 to 7) is a toolbar containing 7 icons and giving access to different kind of functions and windows:
    1. the Exit icon is used to exit the software.
    2. the Save icon is used to save a query in an XML file.
    3. the Open icon is used to open an XML file and load a query written inside.
    4. the Expert icon displays a window where a user can prepare a query based on the keywords used by ECMWF for the web API.
    5. the About icon displays a window whith information about EDD and the changelog.
    6. the Options icon displays a window containing all options of EDD. The user can change all options from here.
    7. the Update icon is only enabled when an update for EDD is available. By clicking on it, a user will launch the automatic update procedure.

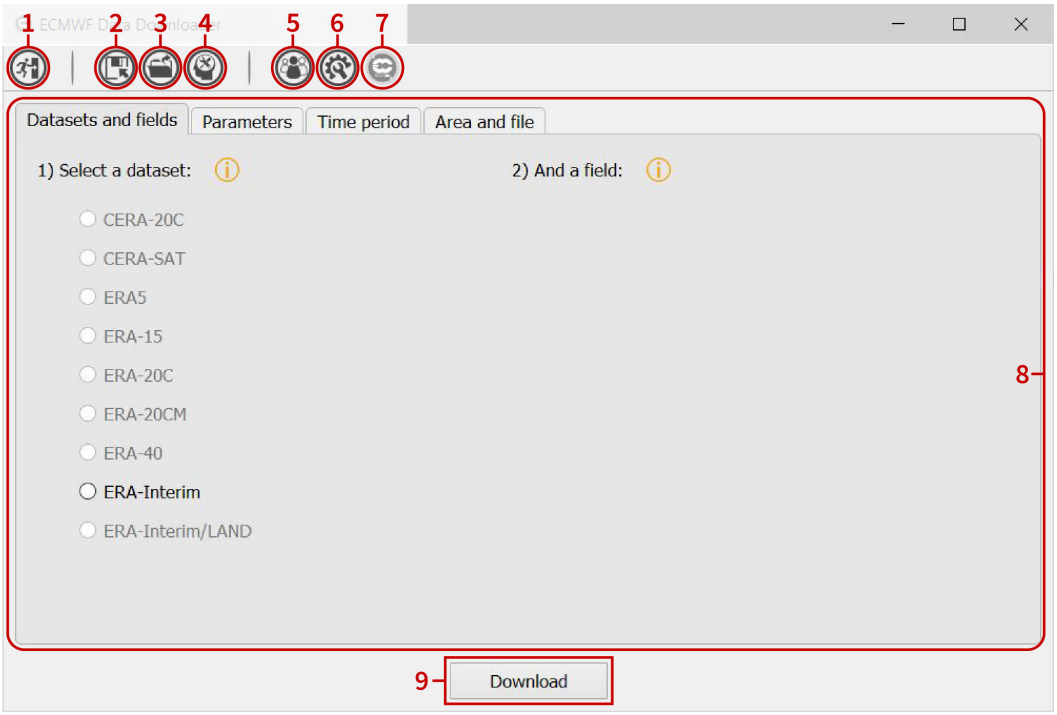


Fig. 3.1: The main window in EDD.

- the second one (8) is a central widget whose purpose is to let the user to prepare a query based on its choices in the GUI.
- the last one (9) is a simple button, used to validate the choices of the user and to send the query to ECMWF web API.

3.2.3 The central widget

The central widget is composed of four tabs. The first tab (cf. Fig. 3.2), called Datasets and fields, is dedicated to the dataset and the class/field in the dataset. Actually, only ERA Interim is included in the software. The user has to select a dataset to populate, automatically, the field part.

The second tab (cf. Fig. 3.3), called Parameters and populated once the dataset and the field have been selected, displays the list of parameters (or variables) extracted from the dataset and the field chosen by the user. It is possible to select one or multiple parameters. If another dataset and/or another field are selected by the user, the parameters list will be reseted.

The third tab (cf. Fig. 3.4), called Time period and enabled once one or more parameters have been selected, displays the different times, steps and time periods that the user can select. Depending on the parameter, the dataset and the field, the times and steps can change and can be available or not for the user. Concerning the time period, it is accessible from two different ways. By selecting a departure date and an arrival date. Or by selecting one or more months/years in a table. The first way is only accessible to daily parameters.

The fourth tab (cf. Fig. 3.5), called Area and file, is enabled once one or more parameters have been selected. It displays the different options to change the resolution and the area of the dataset and field. By default, options are already selected for the area and resolution, based on default option stored on the ECMWF website. From the same tab, it is also possible to indicate a file name and the format for the file. If the user doesn't inform the file name and the file format, defaults are used, output and Grib respectively.

3.2.4 The download button

Once the user has made its choices in the different tabs, it is the right time to click on **Download** (cf. Fig. 3.1, 9). Before sending the query to the ECMWF web API, the software executes a checking function to be sure that all

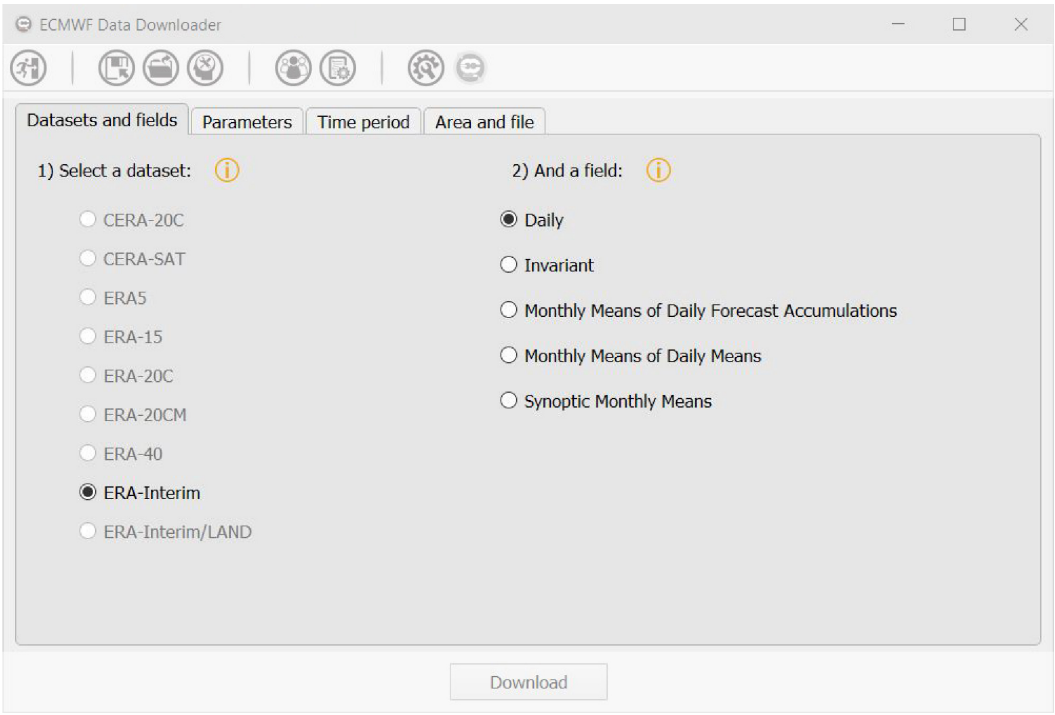


Fig. 3.2: The first tab in EDD, with a dataset and a field selected.

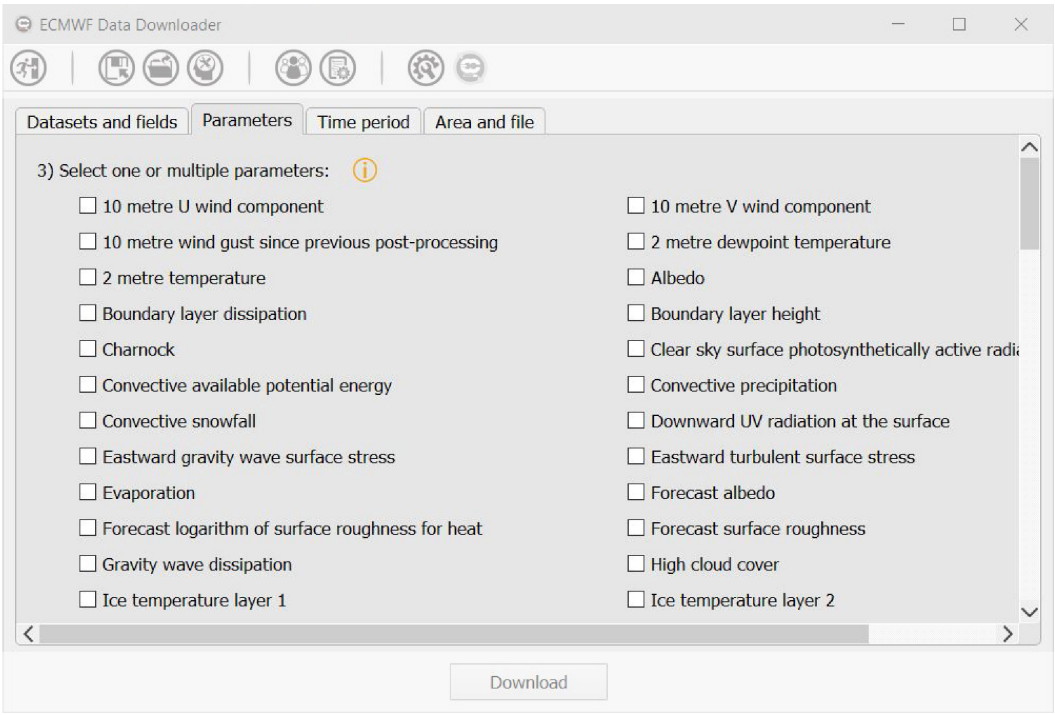


Fig. 3.3: The second tab in EDD.

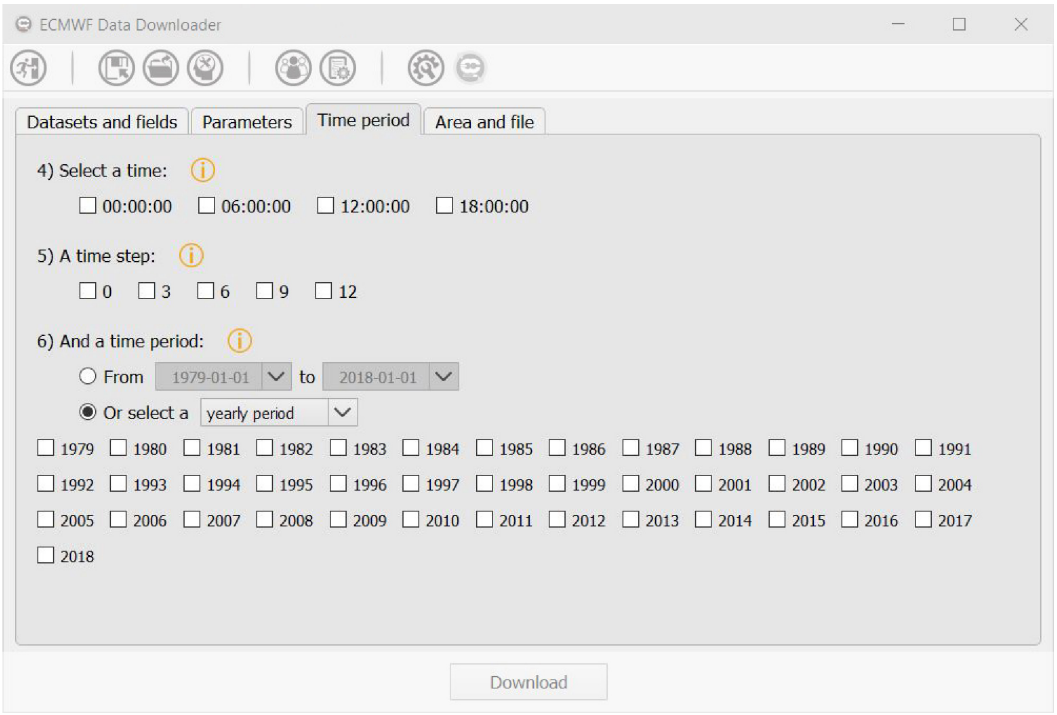


Fig. 3.4: The third tab in EDD, with a yearly time period selected.

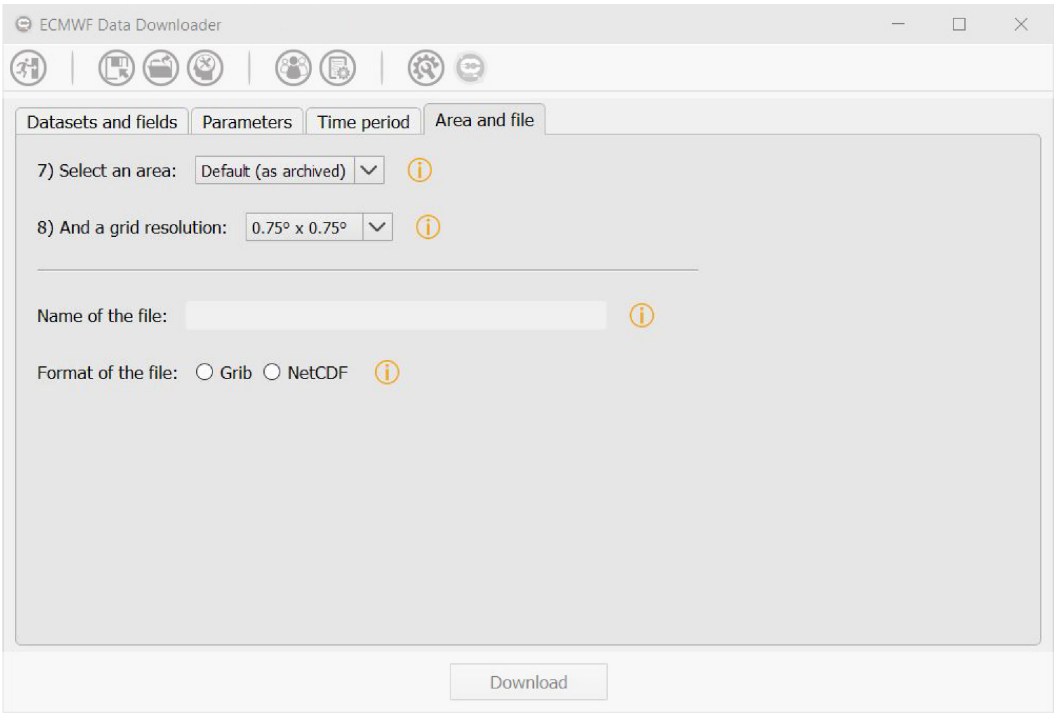


Fig. 3.5: The third tab in EDD, with default options.

mandatory fields have not been forgotten. If an error is discovered, a warning window will appear to inform the user that items in red should be reviewed. If the checking of all fields is successfull, the request is prepared by the software to comply with the ECMWF web API, and send right away. At the same time, a new window appear and displays messages from the ECMWF web API (cf. Fig. 3.6, left), beginning with the connection status and so on. Once the request has been accepted, validated and processed, the download begins (cf. Fig. 3.6, right). Depending on the number and complexity of the different requests processed by ECMWF servers, the user request can stay on **queued** a certain amount of time. It is not possible to cancel a request from EDD if it is queued, even if it is possible to close the window. In that case a warning window is displayed (cf. Fig. 3.7) and inform the user what he must do. Actually, the cancellation of a queued request can only be done from the user account. When the download has started, it is possible to cancel the download from EDD.

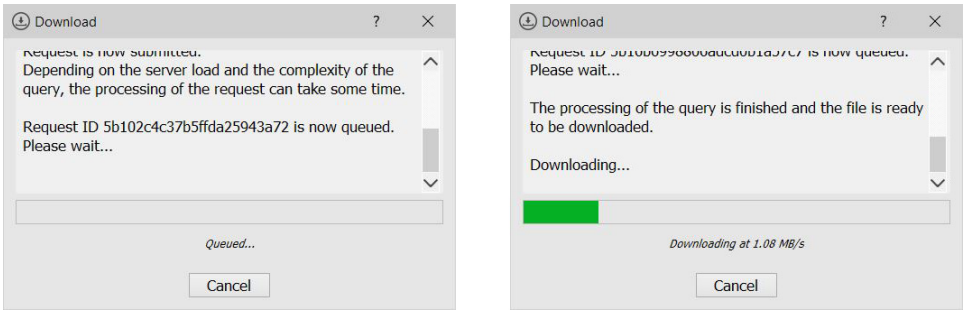


Fig. 3.6: The download window, left: the user's request queued on ECMWF servers ; right: a file is actually downloaded.

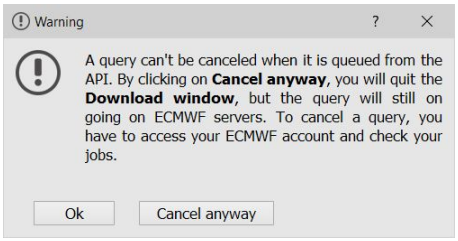


Fig. 3.7: The cancel window, warning the user about the implications of cancelling a request.

3.2.5 The expert window

The Graphical User Interface can be a limitation for those who wants to send a complex query in ECMWF datasets (actually covered or not by EDD). In that case, a kind of expert mode has been included. It is accessible by clicking on the expert icon (cf. Fig. 3.1, 4). A window (cf. Fig. 3.8) appears and gives the possibility to the user to prepare a query based on ECMWF keywords. 15 keywords are available, and if more keywords are needed, the user can enter them in the *other keywords* field in the last tab. Info buttons are here to help the user to understand each keyword, with a link to the complete explanation on the ECMWF website.

There is no mandatory keywords, the user is absolutely free to let all keywords which are not needed for its query empty. Once the query is ready, the user has to click on **Submit** to send the request to ECMWF servers. The download window is then displayed and the ECMWF web API sends back information about the validation of the request.

3.2.6 The about window

The about window (cf. Fig. 3.9), accessible by clicking on the about icon (cf. Fig. 3.1, 5), displays information about EDD and the changelog available in the *documentation* folder.

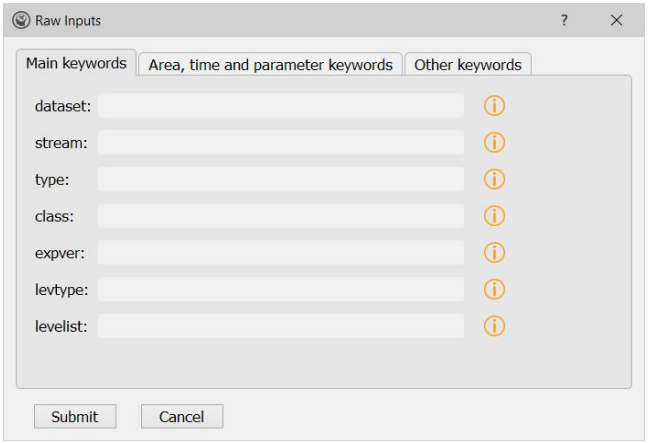


Fig. 3.8: The expert window, where the user can prepare its query based on ECMWF keywords.

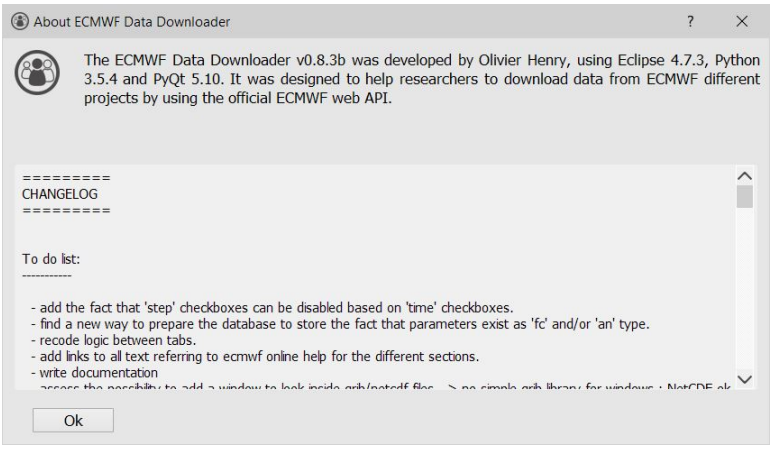


Fig. 3.9: The about window, displaying information about EDD and the changelog.

3.2.7 The options window

During the first startup, EDD creates an .ini file where all options are stored. A window (cf. Fig. 3.10) is accessible to the user to change those options through an icon in the main window (cf. Fig. 3.1, 6). Here are the available options:

- logging level (*level* in the .ini file): a specific level can be chose in the combo box.
- Path of the logging file (*path* in the .ini file): the path where to save the log file. If the path doesn't exist, an error message is added to the log file and the path is reseted to the default path.
- API URL (*url* in the .ini file): the URL of the API should be entered here (already embedded by default)
- User key (*key* in the .ini file): a personal key, provided by ECMWF, is required to use the web API. For more information and to obtain your key, please click on the following link:https://software.ecmwf.int/wiki/display/WEBAPI/Access+ECMWF+Public+Datasets#AccessECMWFPublicDatasets-key.
- User email (*email* in the .ini file): the user email address is required to use the web API.
- ECMWF file folder (*folder* in the .ini file): a folder where to save the file/data downloaded on ECMWF servers. EDD check the path at each startup. If the path doesn't exist, an error message is added to the log file and displayed to the user, and the path is reseted to the default path.
- Display ECMWF web API information at startup (*display_api_info* in the .ini file): If checked, ECMWF Data Downloader displays information about the ECMWF web API at startup.
- Check ECMWF Data Downloader updates on GitHub (*check_update* in the .ini file): allows EDD to check for an update online.

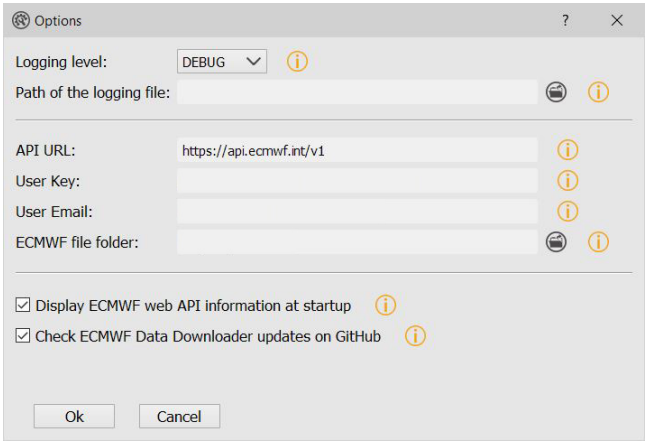


Fig. 3.10: The options window, where the user can change EDD options.

3.2.8 The update system

A simple update system has been included in EDD. How it works is based on the Operating System of the user. If EDD detects a new update in the Release tab of its repository on GitHub, the update icon (cf. Fig. 3.1, 7) is enabled. Once the user click on it, a window (cf. Fig. 3.11) appears and displays information. It's up to the user to click on **Update** or **Download** to launch the update procedure.

Depending on the Operating System, the update procedure is different:

- on Windows, once the user clicks on **Update**, EDD will download the update (an .msi file), close itself and execute the update file.
- on Linux, once the user clicks on **Update**, EDD will download the update (a .tar.gz file), copy the update script in the OS temp folder, close itself and execute the update script.

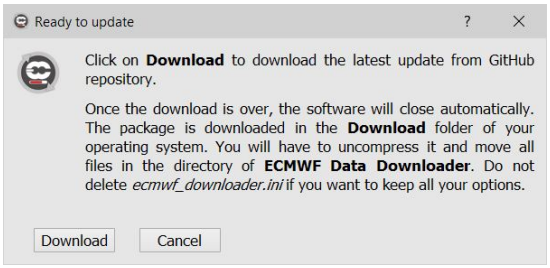


Fig. 3.11: The update window, sources are detected by EDD and the update system proposes a .zip update.

- from sources, once the user clicks on **Download**, EDD will download the update (a .zip file). The user has to uncompress and move all files to the EDD folder.

3.2.9 The download system

To download data from ECMWF servers, EDD uses a homemade function to connect to the ECMWF web API. Once the user validates its query by clicking on the **Download** button, all fields are inspected by a function to check that all mandatory fields have been well filled in and/or selected. If the result is satisfying, all fields are then sent to another function to translate the user inputs into ECMWF keywords. The query is a dictionary constituted of all those pairs keyword/value. Finally the sending of the query is done by embedding the query dictionary in a request method POST. If received by ECMWF servers, the query is queued a certain amount of time and an answer is sent back by the web API, containing the status code of the answer, the ID of the request and other information like the *retry* parameter. The *retry* parameter is used by EDD to interrogates the ECMWF servers about the status of the query. If the query is still queued or validated and being processed, the answer will be the same, with the same status code (202). If the query is not validated (because a keyword is not recognized, for example), the answer will change with an error message and a status code above 400. Once the query is accepted, validated and processed, the anwser contains a status code equal to 303 and a new link to download the requested file. Following that answer, EDD downloads the file in the directory specified by the user.

The cancelation of a query through EDD is only possible during a download, a request method DELETE is sent to the ECMWF server with the link to the file to tell to the servers that the file can be deleted. The cancelation of a query when it is queued on ECMWF servers is not possible through EDD. Even if the process in EDD can be canceled and closed, the user has to visit its account and the webpage dedicated to its queries (http://apps.ecmwf.int/webmars/joblist/) to cancel a query.

CHAPTER

FOUR

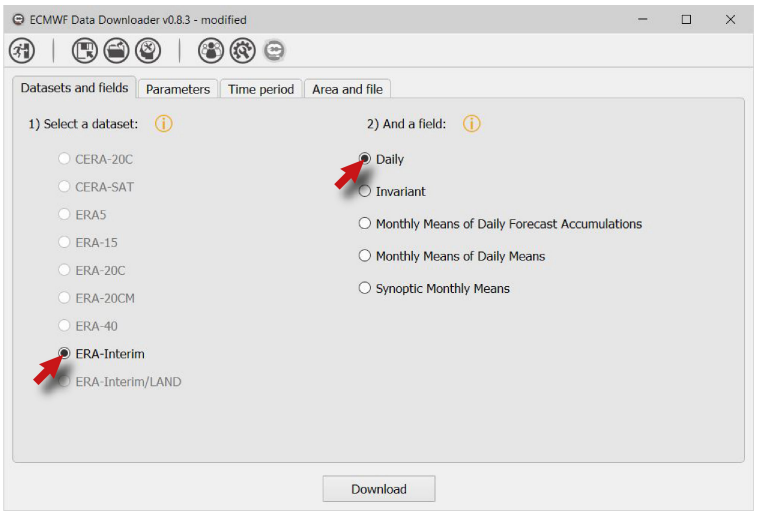
TUTORIAL

4.1 Important information

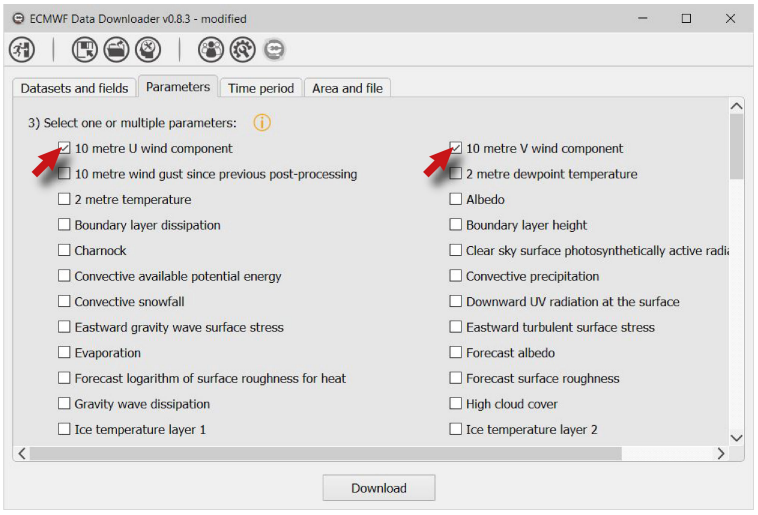
The tutorial proposed here is produced using the ECMWF Data Downloader 0.8.3 for Windows. It shouldn't differ to what the user can expect from the Linux version or the sources.

4.2 How to download data ? (from the GUI)

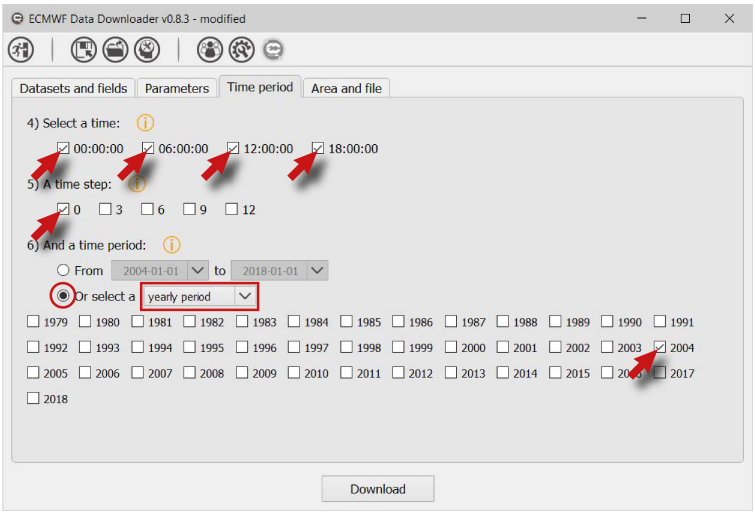
- First a dataset has to be selected and then a field.



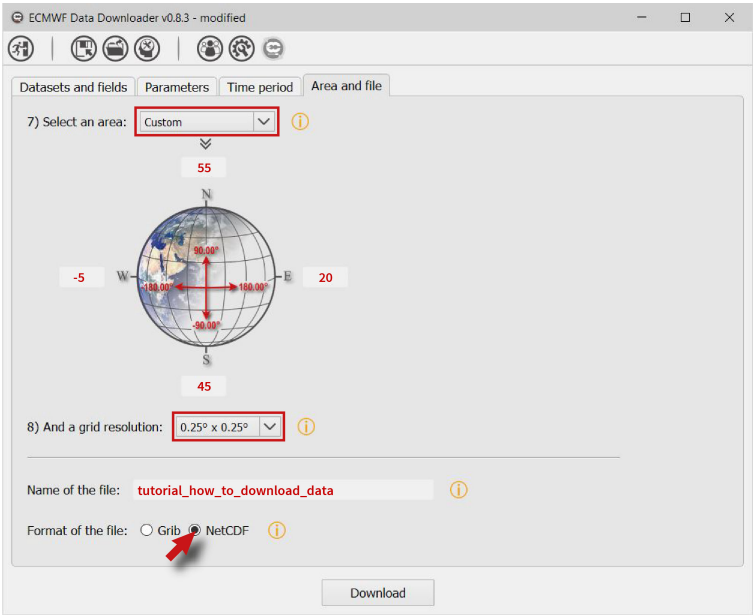
- Select one or more parameters.



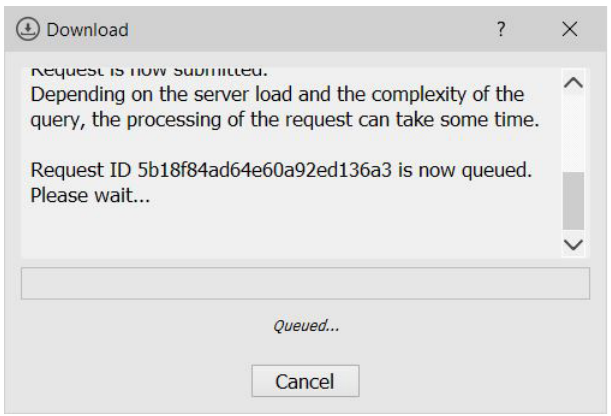
- And a time (or more), a step (or more) and a period.

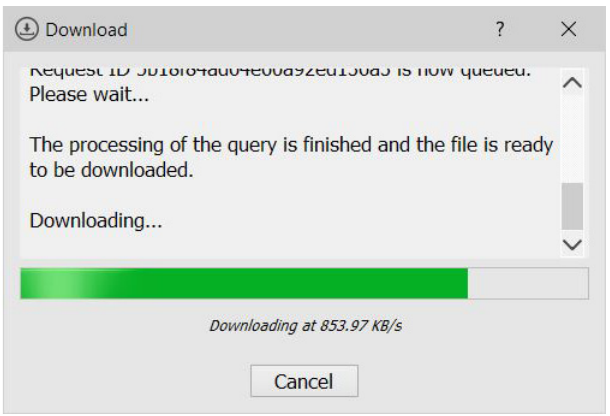


- Choose an area and a resolution. And finally, choose a name for the file and a file format.

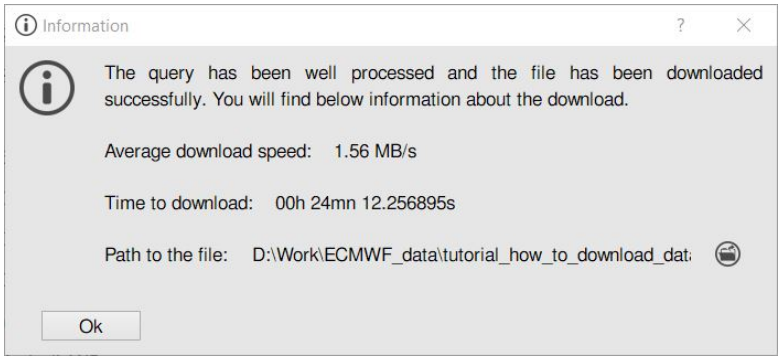


- Click on **Download** to send the query. If the query is well received by ECMWF servers, it will be queued. Once the file is ready, EDD downloads it.





- Finally, EDD displays a window at the end of the process to confirm that the file have been well downloaded.


Information
The query has been well processed and the file has been downloaded successfully. You will find below information about the download.
Average download speed: 1.56 MB/s
Time to download: 00h 24mn 12.256895s
Path to the file: D:\Work\ECMWF_data\tutorial_how_to_download_dat
Ok