LEC6 FLEX_EXTRACT: Overview of ECMWF data extraction

FLEXPART Trainings Course 2019

Anne Philipp (anne.philipp[at]univie.ac.at)

(with main contribution of Leopold Haimberger)

July 2^{nd} 2019

universität wien imgw Institut für Meteorologie und Geophysik

Flex_extract is a Software which …

- … will be available in version 7.1 on *https://www.flexpart.eu/flex_extract/*
- Tar for today:

flex_extract_v7.1.tar.gz

- … provides the meteorological input for FLEXPART
- … is mainly handled by a Python program with some additional wrapping shell scripts (operational and on demand)

  -> uses ECMWF software/libraries and the MARS archive

- … consists a FORTRAN program for numerical calculations

  -> transform ECMWF data to fit FLEXPART grid

The task of flex_extract is to …

- … retrieve meteorological fields from ECMWF MARS archive (T,q,u,v,…)
- … calculate instantaneous surface flux values from

the „accumulated“ fluxes stored in MARS (LSP/CP/SSHF/EWSS/NSSS/SSR)

- … calculate hybrid coordinate vertical velocity (if necessary: from

horizontal wind field)

- … also retrieve additional 2D fields (e.g. cloud cover/vegetation cover/soil temperature etc.)
- … check if all fields have been retrieved
- … transfer data from/to local host
- … prepare final FLEXPART input files

	Parameters
Model level	U, V, T, Q, ETADOT, (CIWC+CLWC=QC)
Surface level	LNSP, MSL, 10u, 10v, 2T, 2D, SD, TCC, SR, Z, SDOR, LSM
Flux data	LSP, CP, SSHF, EWSS, NSSS

Changes in v7.1

New features

- first set of UNIT tests
- first set of regression tests
- structured documentation with Sphinx
- ~~local retrieval via CDS API for ERA 5 data~~ (Canceled by ECMWF)
- simplified installation process
- disaggregation of precipitation with a new algorithm
- ensemble retrievals

Changes

- upgraded to Python3
- applied PEP8 style guide
- use of genshi templates
- modularization of python source code
- upgrade from grib_api to ecCodes
- completely revised/refactored python section
- restructured program directories

User modes

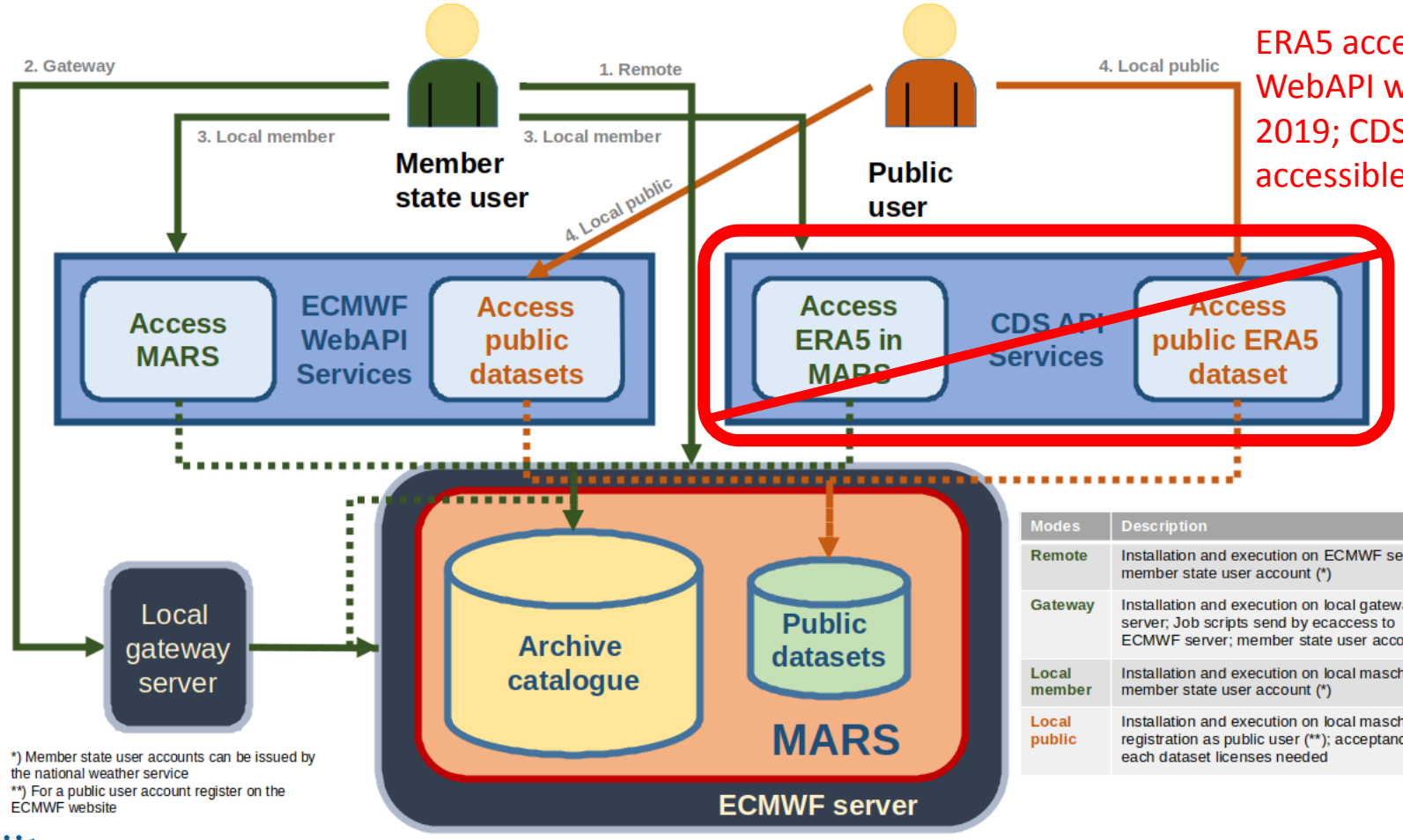
Member state user

- work directly on ECMWF servers or through a local Gateway server
- theoretically, no limits
- registration through computing representative

Public user

- work locally through Web API‘s
- access only to public datasets (here: CERA-20C, ERA-Interim, and ERA5 in the future)
- registration through web mask

Application modes

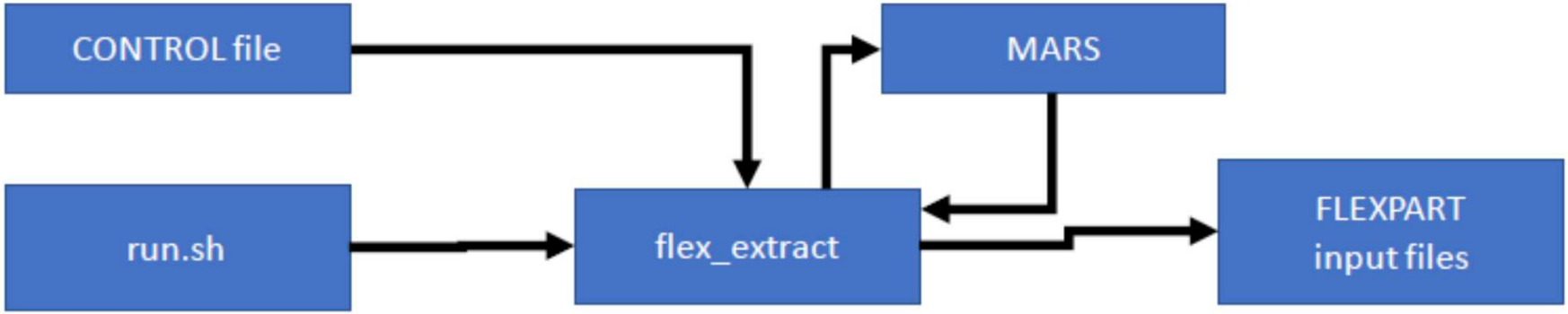


Modes	Description
Remote	Installation and execution on ECMWF server; member state user account (*)
Gateway	Installation and execution on local gateway server; Job scripts send by ecaccess to ECMWF server; member state user account (*)
Local member	Installation and execution on local maschine; member state user account (*)
Local public	Installation and execution on local maschine; registration as public user (**); acceptance of each dataset licenses needed

ERA5 access through ECMWF WebAPI was cancelled in Mar 2019; CDSAPI not yet fully accessible

General processing of flex_extract

Our problem:
Extract and convert data from the MARS archive and transfer them to local storage, where it can be accessed by FLEXPART software.



Input data

Input data

- CONTROL files
- [ECMWF user credential file]
- Templates

Controlling files

- setup.sh
  - compilejob.ksh
  - Fortran Makefile
- run.sh
  - job.ksh

Installation procedures

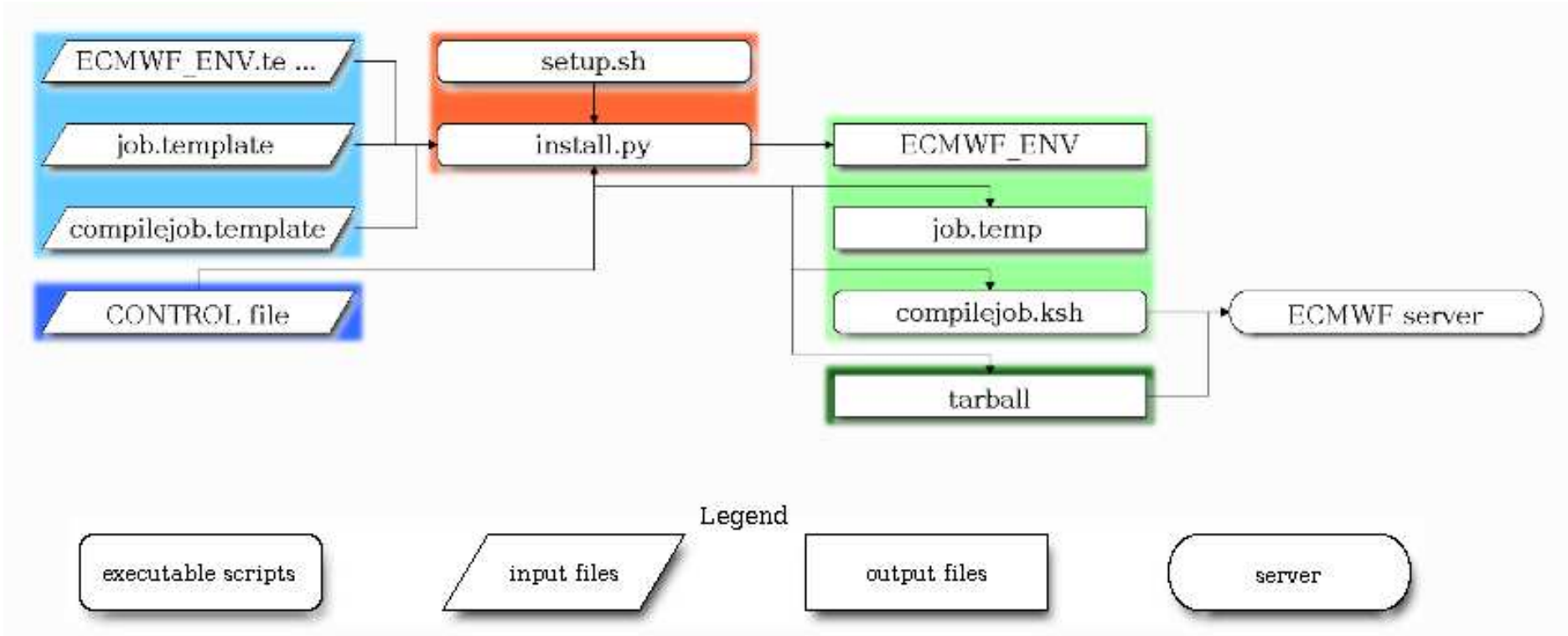
Everything is described in detail in the online documentation!

- Accessible through the tarball or the git repository https://www.flexpart.eu/browser/flex_extract.git (dev branch)

- Will be made available online at the community website as soon as version 7.1 is officially released!

https://www.flexpart.eu/wiki/FpInputMetEcmwf

General setup process



Detailed installation instructions in online documentation!

Example: extraction in local mode

CONTROL_EA5*

```
START_DATE
DTIME 6
TYPE AN AN AN AN
TIME 00 06 12 18
STEP 00 00 00 00
ACCTYPE FC
ACCTIME 06/18
ACCMAXSTEP 12
CLASS EA
STREAM OPER
GRID 0.28125
LEFT 0.
LOWER 0.
UPPER 2.25
RIGHT 2.25
LEVELIST 135/to/137
RESOL 799
ETA 1
PREFIX EA
```

* Minimum number of parameters to be set

./run_local.sh

flex_extract

run_local.sh (user specification)

```
# ------------------------------------------------------------
# AVAILABLE COMMANDLINE ARGUMENTS TO SET
#
# THE USER HAS TO SPECIFY THESE PARAMETERS
#
QUEUE='',
START_DATE=20180108
END_DATE=None
DATE_CHUNK=None
JOB_CHUNK=None
BASETIME=None
STEP=None
LEVELIST=None
AREA=None
INPUTDIR='./workspace',
OUTPUTDIR=None
FLEXPARTDIR=None
PP_ID=None
JOB_TEMPLATE='',
CONTROLFILE='CONTROL_EA5',
DEBUG=1
REQUEST=2
PUBLIC=0
# ------------------------------------------------------------
```

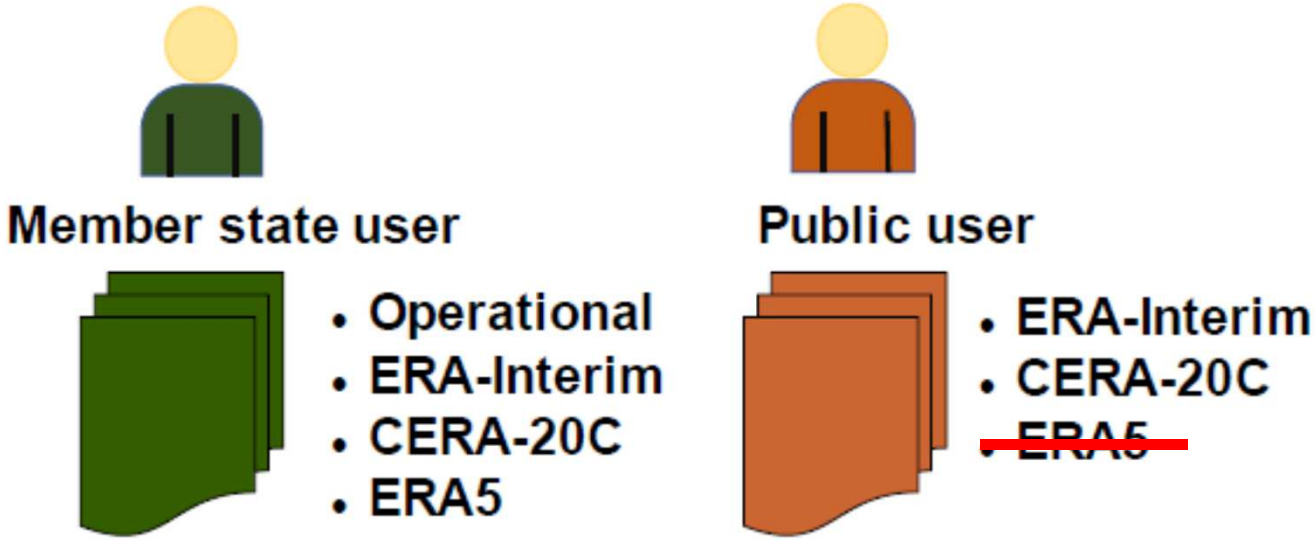
Red box: resulting files as input for FLEXPART;
the rest are temporary files

```
ANOG__ML.20180108.16098.16099.grb  FCOG_acc_SL.20180107.16098.16099.grb  flux2018010818  fort.11  fort.21
ANOG__SL.20180108.16098.16099.grb  flux2018010700                       flux2018010900  fort.12  fort.22
ANSH__SL.20180108.16098.16099.grb  flux2018010706                       flux2018010906  fort.13  fort.4
date_time_stepRange.idx            flux2018010712                       flux2018010912  fort.15  mars_requests.csv
EA18010800                         flux2018010718                       flux2018010918  fort.16  OG_OROLSM__SL.20180108.16098.16099.grb
EA18010806                         flux2018010800                       flux2018011000  fort.17  VERTICAL.EC
EA18010812                         flux2018010806                       flux2018011006  fort.18
EA18010818                         flux2018010812                       fort.10         fort.19
```

CONTROL file

- File name format: CONTROL_<Dataset>[.optionalIndications]
- Example files are listed in: flex_extract_v7.1/run/control/
- NEW: Comments with „#“
- NEW: Only parameters which should overwrite default values must be set
- The file „CONTROL.documentation“ contains a list of all parameters and their default values
- Detailed desciption of parameters in online documentation

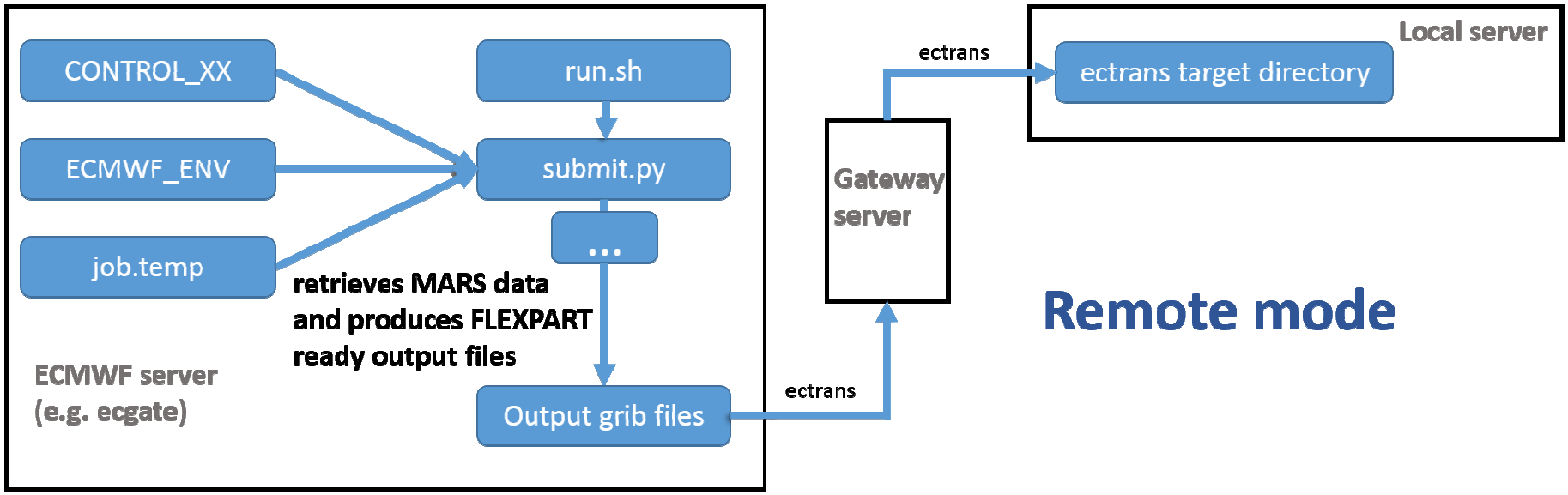
CONTROL file



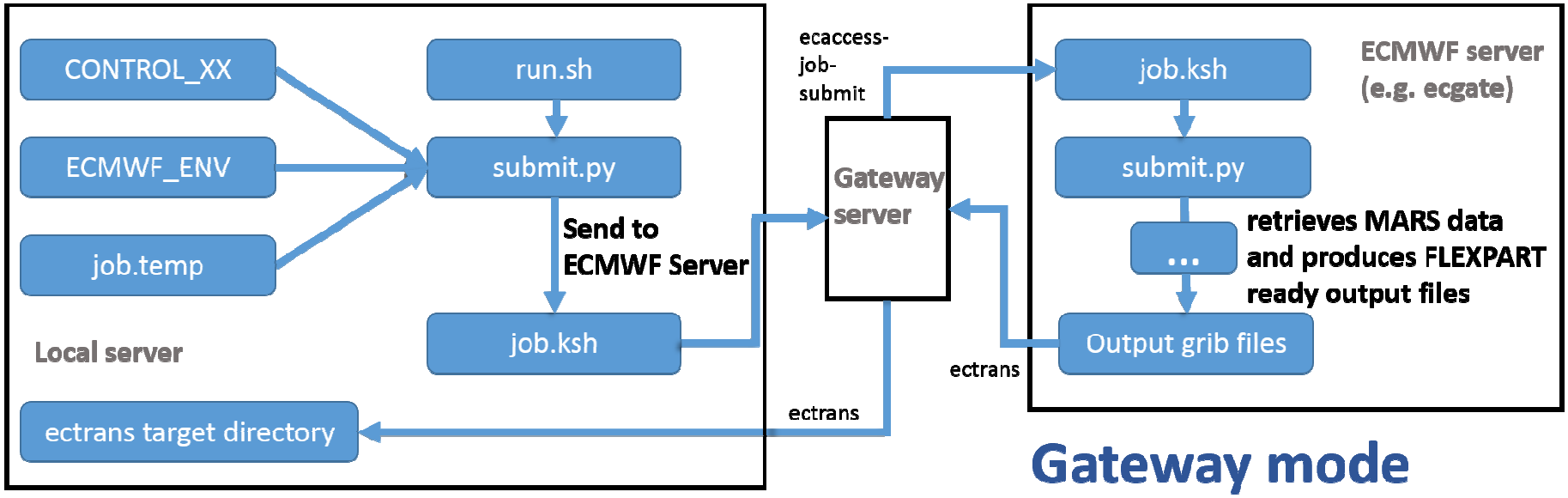
CONTROL file

	Operational	ERA-Interim	ERA5	CERA-20C
Period	12/1985 - ongoing	01/1979-12/2018	01/1979-12/2018	09/1901-12/2010
Streams	oper/elda/enfo	oper	oper/enda	enda
Types	AN/FC/4V/(PF(1992-2019)/CF(1992-2019)/CV (2006-2016) – only for enfo)	AN/FC/4V	AN/FC/4V	AN/FC
FC base time	0/12 UTC	0/12 UTC	06/18 UTC	18 UTC
Max. time resolution	1-hourly AN/FC mix	3-hourly AN/FC mix (6-hourly AN for public user)	1-hourly AN	1-hourly AN
Highest resolution	~0.1°	0.75° (80km)	0.28125° (31km)	~1.25° (125km)
Levels	137 (starting from 25/06/2013; less levels before, see docu)	60	137	91
Ensemble members	11 (enda); 26 (elda, upto 50 with sythesized doubles); 50 (enfo, since 1996)	-	10 (3-hourly)	10 (3-hourly)
Eta- coordinate vertical velocity	yes	No, only reduced gaussian grid	yes	yes

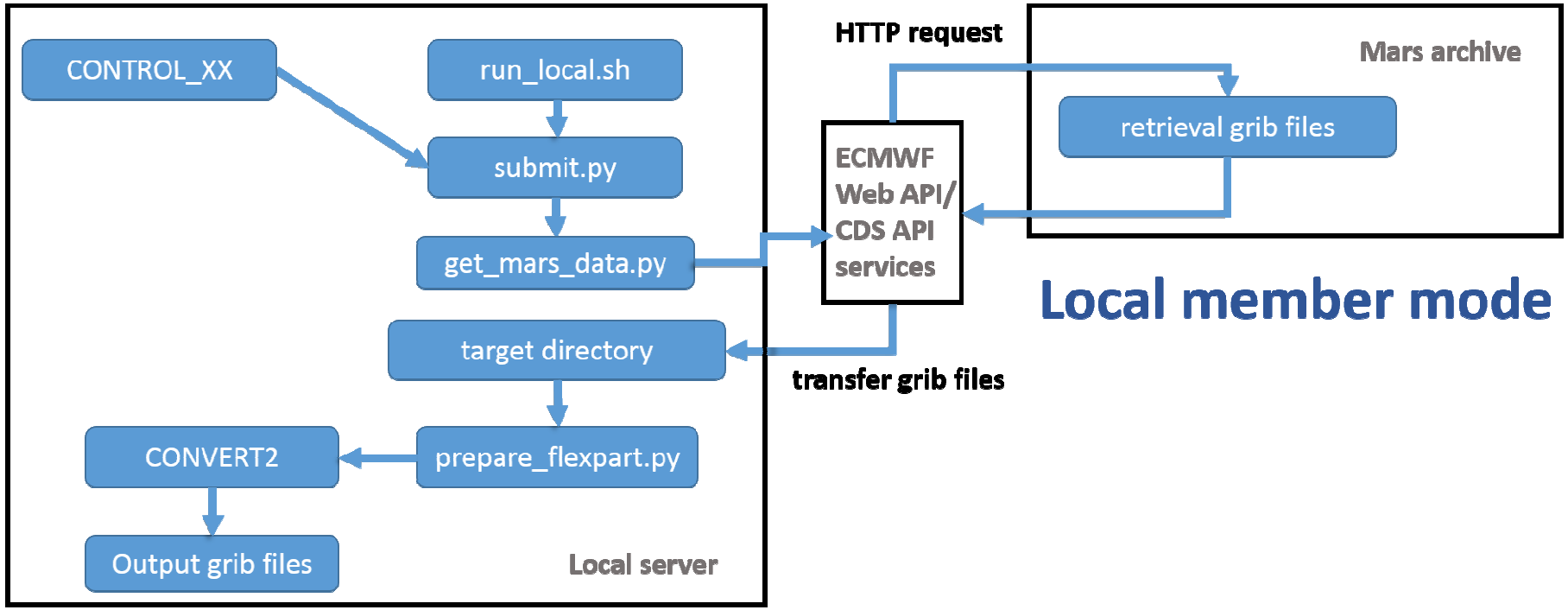
Remote mode


CONTROL_XX
ECMWF_ENV
job.temp
run.sh
submit.py
...
retrieves MARS data
and produces FLEXPART
ready output files
ECMWF server
(e.g. ecgate)
Output grib files
ectrans
Gateway
server
ectrans
Local server
ectrans target directory
Remote mode

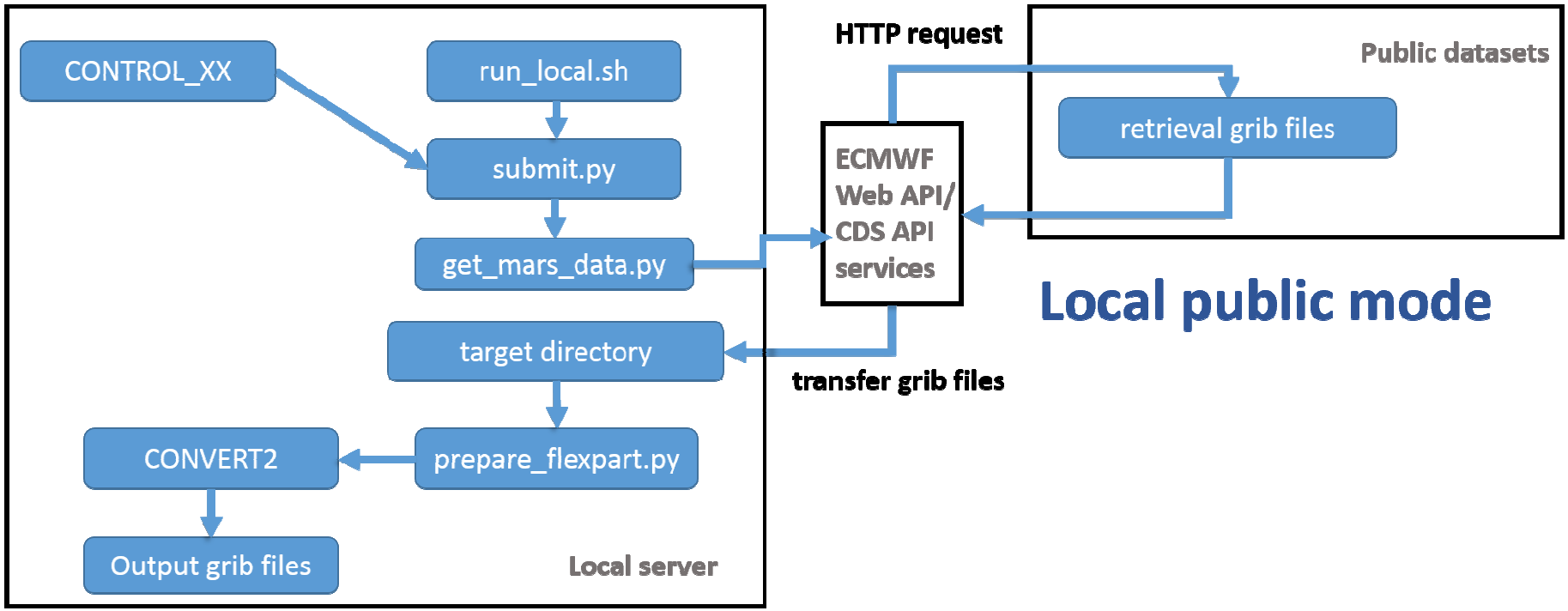
Gateway mode



Local member mode



Local public mode



Hybrid coordinate

When is hybrid coordinate vertical velocity calculated from horizontal wind field?

- In general use ETA=1 whenever possible -> minimize computation time !!!
- When setting ETA=0 or ETADIFF=1 or OMEGADIFF=1
  - etadot is calculated through vertical integration of the continuity equation.
  - That takes a lot of time and memory for high spectral resolution. For resolutions>511 it fails on ecgate due to memory constraints
  - Calculation more time consuming but also more accurate if GAUSS=1
- ETA=0 is recommended only if etadot is not available in MARS, e.g.
  - for operational fields before September 2008
  - for ERA-Interim or ERA-40 data

Flex_extract structure

Knowledge of directory structure and the files tell a lot about what's going on –> thus extremely useful for debugging

Directories

- documentation
- for_developers
- run
- source
- templates
- test

Scripts and Files

- setup.sh
- README.md
- LICENSE.md
- CODE_OF_CONDUCT.md

Files on local host

In the „flex_extract_v7.1“ directory is a „run“ directory which contains all necessary files for running a job:

- **Directories in run**:
  - **control** – contains the CONTROL files, some examples are already there
  - **jobscripts** – contains the job scripts which are sent to the ECMWF server
  - **workspace** – contains the output of retrieved data in the local application mode
    - Final FLEXPART input files (mostly known as „EN“ files)
    - Other temporary files created during execution of scripts:
      - MARS output (*.grb)
      - Control files for conversion software (fort.*, index file, flux*, …)
- **run.sh** – *wrapping Shell script which starts flex_extract with the command line options*

Files on $SCRATCH at ecgate

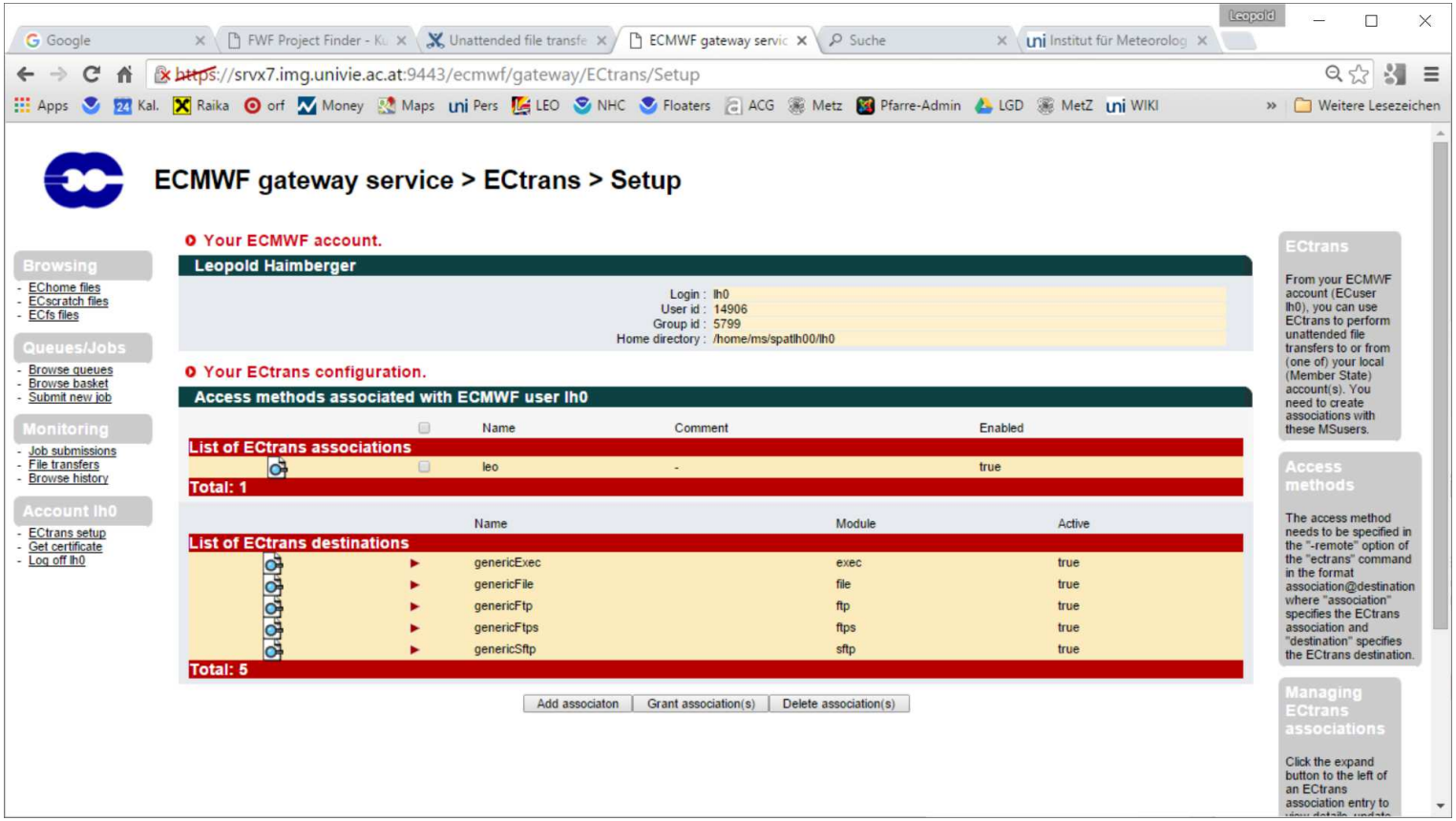
- Each on demand script has its own output directory
  - named python?????? # ? is the process number
  - use „ls –rthl“ to see which one is the newest
- flex_ecmwf.??????.out
  - That‘s the output from the executed Shell script # ? is the job ID
  - That output is **not** sent via email!
- python??????/*
  - CONTROL file
  - Log file called „prot“
  - Output directory „work“
    - Final FLEXPART input files (mostly known as „EN“ files)
    - Other temporary files created during execution of scripts:
      - MARS output (*.grb)
      - Control files for conversion software (fort.*, index file, flux*, …)

Managing Destinations

- „Destinations“ like username@genericSftp are aliases used by ectrans on ecgate to send data to places on local servers
- They need to be specified when using update_script.ksh
- They are set at the ecaccess pages http://ecaccess.ecmwf.int under „ectrans setup“
  - There one has to enter a
    - Transfer method (GenericFile or GenericSftp)
    - Destination server
    - Destination directory
    - User name & password

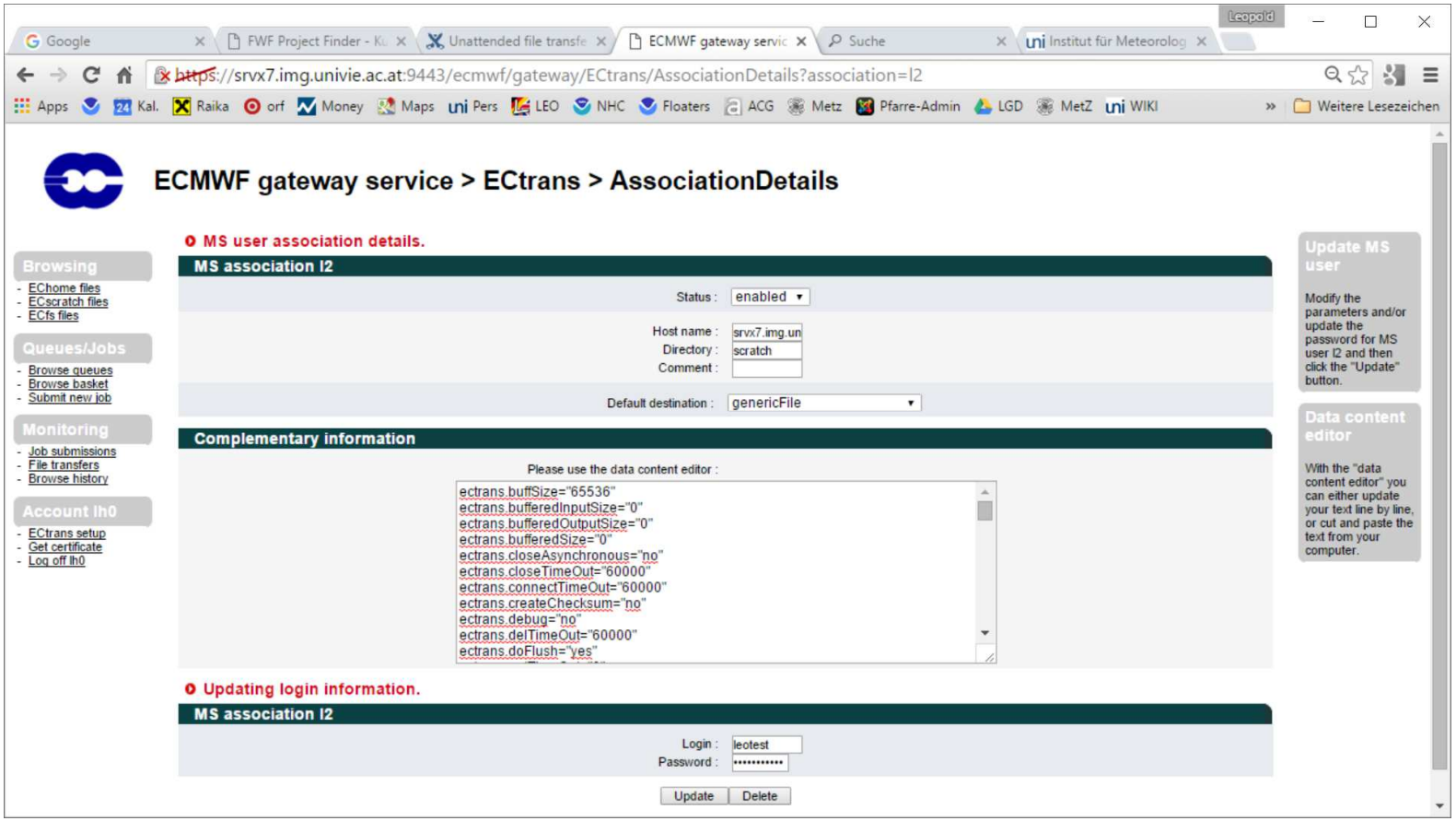
ECTRANS setup via gateway web interface

- Go to https://ecaccess.ecmwf.int:9443/ , then to Ectrans setup



ECTRANS setup via gateway web interface

- Go to https://ecaccess.ecmwf.int:9443/ , then to Ectrans setup



FLEXPART patch

- The current FLEXPART version cannot read all surface fields if encoded in GRIB2.
- With a patch FLEXPART can read GRIB2 encoded upper air and ALL surface fields
- Two fields (convective precipitation and snow depth) require rescaling because of change of units
- Input fields numerically equal

Troubleshooting on ecgate

- If you think the job runs for too long, you may:
  - Check your email for a log file
    - After job ends you get either flex.ecgb.????? or ERROR! flex.ecgb.????? with useful debug information.
  - use command **ecaccess-job-list**
    - Job should be in state EXEC or DONE or INIT or WAIT
    - If in state STOP it failed and you should have received an email
  - If the job succeeded, did the file transfer to local gateway server work?
    - Do you find files in the directory set in the destination?
    - If not -- was parameter ECTRANS in CONTROL file set to 1?
    - enter **ecaccess-ectrans-list** to see if transfers failed
      - If they failed check ectrans log and ectrans setup on ecaccess web interface

Troubleshooting on ecgate

- If all of these checks do not help
  - Run the job locally via the ECMWF WebAPI or CDS API to see if the job fails there as well. A misconfigured CONTROL file will likely cause trouble there as well.
  - Log on to ecgate
    - **cd $SCRATCH**
    - **ls –rthl** lists the latest log file and the latest python directory
    - The log file **flex_ecmwf.??????.out** may contain information about a failure
    - In the **python?????** Directory you can find the CONTROL file used and also a job log. In the **work** subdirectory you can find the temporary files, which may give you a clue what happened.

Troubleshooting on local host

flex_extract does not create a log file in the local mode

-> output is directly shown in your terminal window!

Support

- **FLEXPART's community website and ticket system**: https://flexpart.eu
- **flex_extract information:** https://www.flexpart.eu/wiki/FpInputMetEcmwf
- **Git repository:** https://www.flexpart.eu/browser/flex_extract.git
- **Mailing list:** flexpart[at]lists.univie.ac.at

Literature

- FLEXPART community website: https://www.flexpart.eu
- Online documentation of flex_extract:
  - for now use the offline website and open the following file in a browser: flex_extract_v7.1/documentation/html/Ecmwf/ec-links.html
  - will be reachable online soon from here: https://www.flexpart.eu/wiki/FpInputMetEcmwf
- User documentation at ECMWF: https://confluence.ecmwf.int/display/UDOC