Grid data sync and any other business

Jérémie Dudouet

on behalf of the AGATA Analysis Working Group

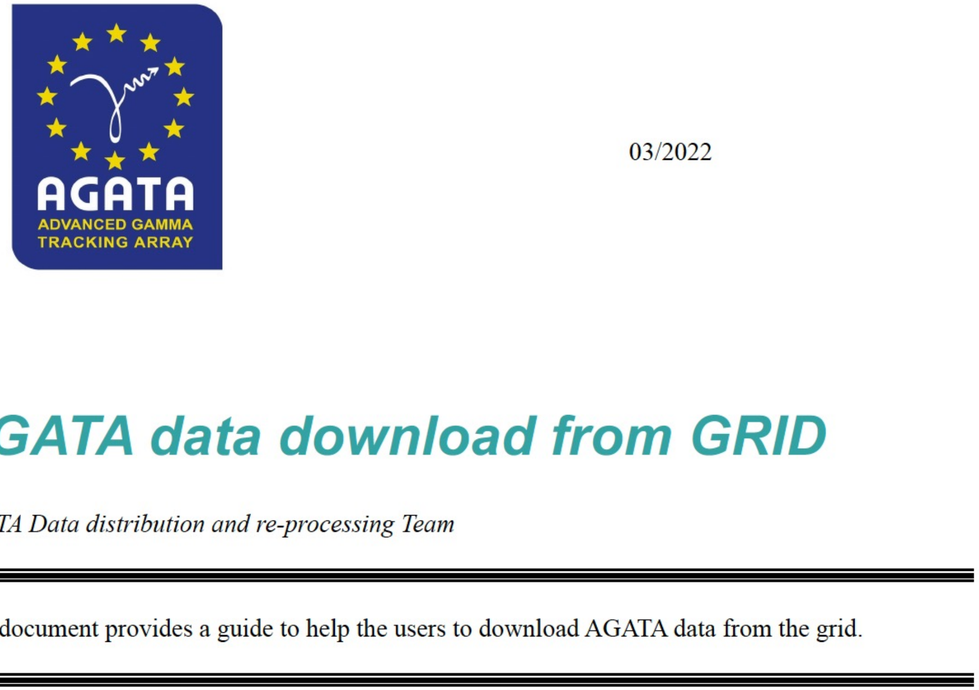
AGATA Analysis School: 15/09/2023, Legnaro

The grid script

After each experiment, the data are stored on the grid:

- On LYON CC center
- On Bologna center

Full documentation on:
https://atrium.in2p3.fr/f6ec4aaa-e918-4cdf-904a-8ca22d58b42b



The grid script

Prerequisites:

- A grid user interface with the **gfal2 python library** (to be installed by IT services)
  - ➢ provide the environment to interact with the Grid
- A Grid certificate
- Join the AGATA Virtual Organization
- The python3 script for AGATA data download:
  - ➢ Included in the AGATA software's packages “script”

The grid script

run_0018 — dudouet@lyoserv2: ~/GridTests_gfal2 — sshpass -p zzzzzzzz ssh -XY dudouet@lyoserv.in2p3.fr ▸ ssh — 150×29

```
[dudouet@lyoserv2:~/GridTests_gfal2$ ./GridDataSync.py

Usage: GridDataSync.py [options]

Browse and download AGATA data from the grid

Options:
  -h, --help          show this help message and exit
  --new_proxy         create a new proxy
  --proxy_status      print the proxy status
  --from_LYON         download data from CC Lyon (default)
  --from_CNAF         download data from Bologna
  --show_conf         show the current configuration (paths, patterns)
  --ls_dir            list the content of the given folder
  --input_dir=path    copy grid data from distant path
  --output_dir=path   copy grid data into local path
  --exc=patt          exclude patterns, separated by ":", will skip all files containing exc patterns (none to reset)
  --inc=patt          include patterns, separated by ":", will skip all files not containing inc patterns (none to reset)
                      (check https://regexone.com/references/python for python regexp format)
  --build_list        build the list of files to be downloaded (mandatory before start)
  --bring_online      move files from tape to disks (make the copy of files faster)
  --check_status      check the status of the files to be downloaded (locality, downloaded...)
  --verbose           increase the verbosity
  --start             launches the download of the files from the grid
  --force             force the download of offline files (much slower)
  --nochecksum        remove the checksum on each downloaded file
  --overwrite         Overwrite the already downloaded files
  --release           release all files from disk
dudouet@lyoserv2:~/GridTests_gfal2$
```

Step 1: create a new proxy (valid for 72h)

"./GridDataSync.py –new_proxy"

"./GridDataSync.py –proxy_status"

run_0018 — dudouet@lyoserv2: ~/GridTests_gfal2 — sshpass -p zzzzzzzz ssh -XY dudouet@lyoserv.in2p3.fr ▸ ssh — 150×15

```
[dudouet@lyoserv2:~/GridTests_gfal2$ ./GridDataSync.py --new_proxy

[Enter GRID pass phrase for this identity:
Contacting cclcgvomsli01.in2p3.fr:15007 [/O=GRID-FR/C=FR/O=CNRS/OU=CC-IN2P3/CN=cclcgvomsli01.in2p3.fr] "vo.agata.org"...
Remote VOMS server contacted succesfully.


WARNING: VOMS AC validation for VO vo.agata.org failed for the following reasons:
         LSC validation failed: LSC chain description does not match AA certificate chain embedded in the VOMS AC!
         AC signature verification failure: no valid VOMS server credential found.

Created proxy in /tmp/x509up_u2471.

Your proxy is valid until Sat Aug 19 09:47:22 CEST 2023
dudouet@lyoserv2:~/GridTests_gfal2$
```

run_0018 — dudouet@lyoserv2: ~/GridTests_gfal2 — sshpass -p zzzzzzzz ssh -XY dudouet@lyoserv.in2p3.fr ▸ ssh — 150×20

```
[dudouet@lyoserv2:~/GridTests_gfal2$ ./GridDataSync.py --proxy_status

subject   : /O=GRID-FR/C=FR/O=CNRS/OU=IPNL/CN=Jeremie Dudouet/CN=1742975744
issuer    : /O=GRID-FR/C=FR/O=CNRS/OU=IPNL/CN=Jeremie Dudouet
identity  : /O=GRID-FR/C=FR/O=CNRS/OU=IPNL/CN=Jeremie Dudouet
type      : RFC3820 compliant impersonation proxy
strength  : 2048
path      : /tmp/x509up_u2471
timeleft  : 71:58:50
key usage : Digital Signature, Non Repudiation, Key Encipherment, Data Encipherment, Key Agreement
=== VO vo.agata.org extension information ===
VO        : vo.agata.org
subject   : /O=GRID-FR/C=FR/O=CNRS/OU=IPNL/CN=Jeremie Dudouet
issuer    : /O=GRID-FR/C=FR/O=CNRS/OU=CC-IN2P3/CN=cclcgvomsli01.in2p3.fr
attribute : /vo.agata.org/Role=NULL/Capability=NULL
timeleft  : 71:58:49
uri       : cclcgvomsli01.in2p3.fr:15007


dudouet@lyoserv2:~/GridTests_gfal2$
```

Step 2: check the current configuration status

"./GridDataSync.py –show_conf"

- Data taken from Lyon CC
- Folder to download: e680 experiment
- Sets of include and exclude patterns to filter the data

Step 3: update the catalog

"./GridDataSync.py --build_list"

run_0018 — dudouet@lyoserv2: ~/GridTests_gfal2 — sshpass -p zzzzzzzz ssh -XY dudouet@lyoserv.in2p3.fr ▸ ssh — 154×45

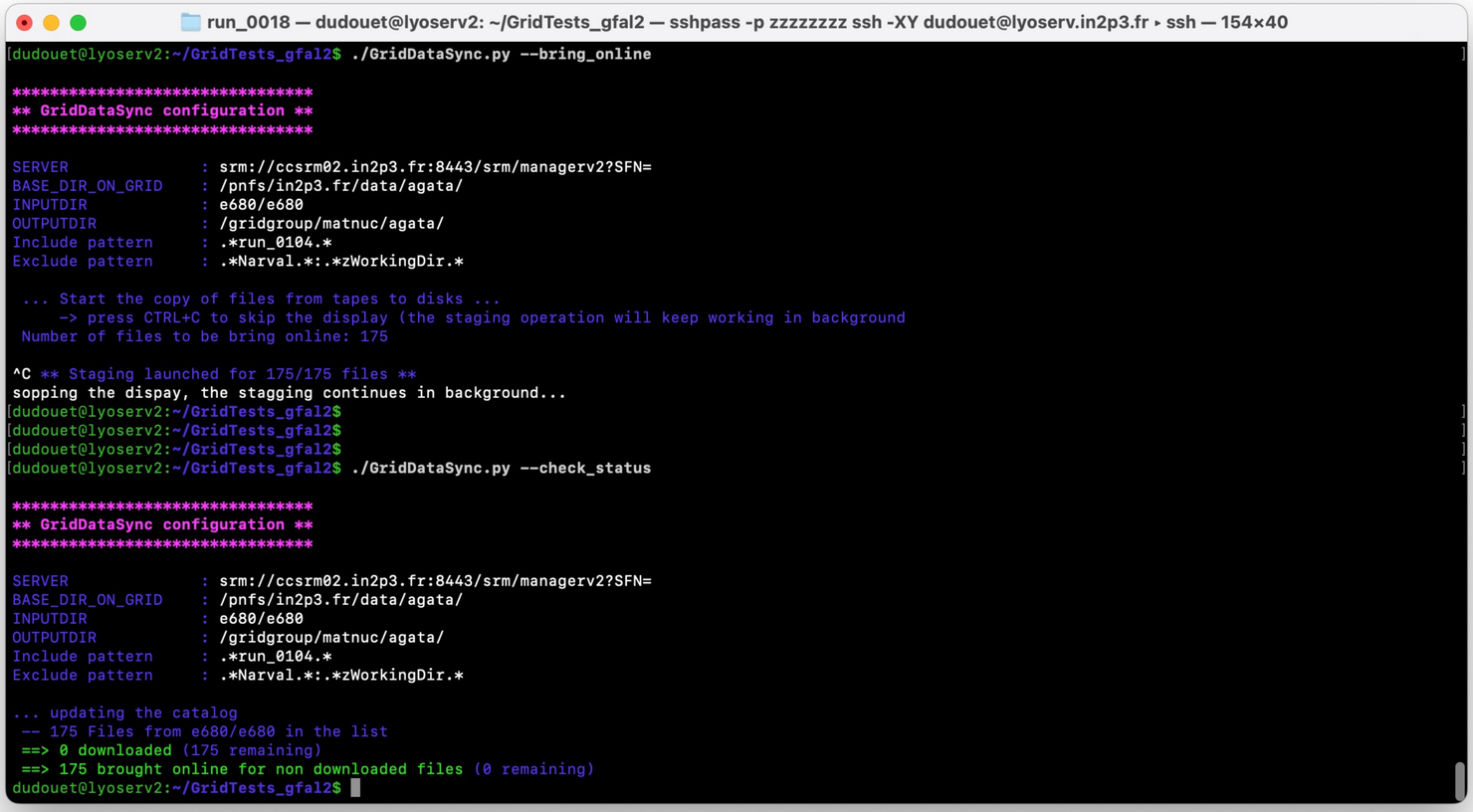
```
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12B/SRM_AGATA_event_mezzdata.cdat.0002
=> adding:  160.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12B/SRM_AGATA_event_mezzdata.cdat.0003
=> adding:  310.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12B/SRM_AGATA_psa_0000.adf
=> adding:   69.1 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12B/SRM_AGATA_small_files.tar
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_event_mezzdata.cdat.0000
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_event_mezzdata.cdat.0001
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_event_mezzdata.cdat.0002
=> adding:  640.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_event_mezzdata.cdat.0003
=> adding:  334.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_psa_0000.adf
=> adding:   69.1 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/12C/SRM_AGATA_small_files.tar
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_event_mezzdata.cdat.0000
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_event_mezzdata.cdat.0001
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_event_mezzdata.cdat.0002
=> adding:  302.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_event_mezzdata.cdat.0003
=> adding:  300.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_psa_0000.adf
=> adding:   69.1 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13A/SRM_AGATA_small_files.tar
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_event_mezzdata.cdat.0000
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_event_mezzdata.cdat.0001
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_event_mezzdata.cdat.0002
=> adding:  691.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_event_mezzdata.cdat.0003
=> adding:  328.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_psa_0000.adf
=> adding:   69.1 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13B/SRM_AGATA_small_files.tar
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_event_mezzdata.cdat.0000
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_event_mezzdata.cdat.0001
=> adding:    3.6 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_event_mezzdata.cdat.0002
=> adding:  336.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_event_mezzdata.cdat.0003
=> adding:  330.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_psa_0000.adf
=> adding:   69.1 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/13C/SRM_AGATA_small_files.tar
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0000.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0001.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0002.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0003.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0004.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0005.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0006.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0007.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0008.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0009.adf
=> adding:    2.0 GB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0010.adf
=> adding:  957.0 MB e680/e680/run_0104.dat.19-05-15_22h37m08s/Data/vamos/SRM_AGATA_vamos_0011.adf
=> adding:  330.0 kB e680/e680/run_0104.dat.19-05-15_22h37m08s/RCC_conf_files/SRM_AGATA_small_files.tar
=> adding:   40.0 kB e680/e680/run_0104.dat.19-05-15_22h37m08s/SRM_AGATA_small_files.tar
 -- 175 Files from e680/e680 added to the list of files to be downloaded
 --> Total files size:  287.7 GB
dudouet@lyoserv2:~/GridTests_gfal2$
```

Step 4: data staging

"./GridDataSync.py –bring_online"

"./GridDataSync.py –check_status"

- On the grid, data are stored on **tapes**
- To be downloaded, data need to be temporary move from **tapes to disks** (data stagging)



Step 5: data download

"./GridDataSync.py –start"

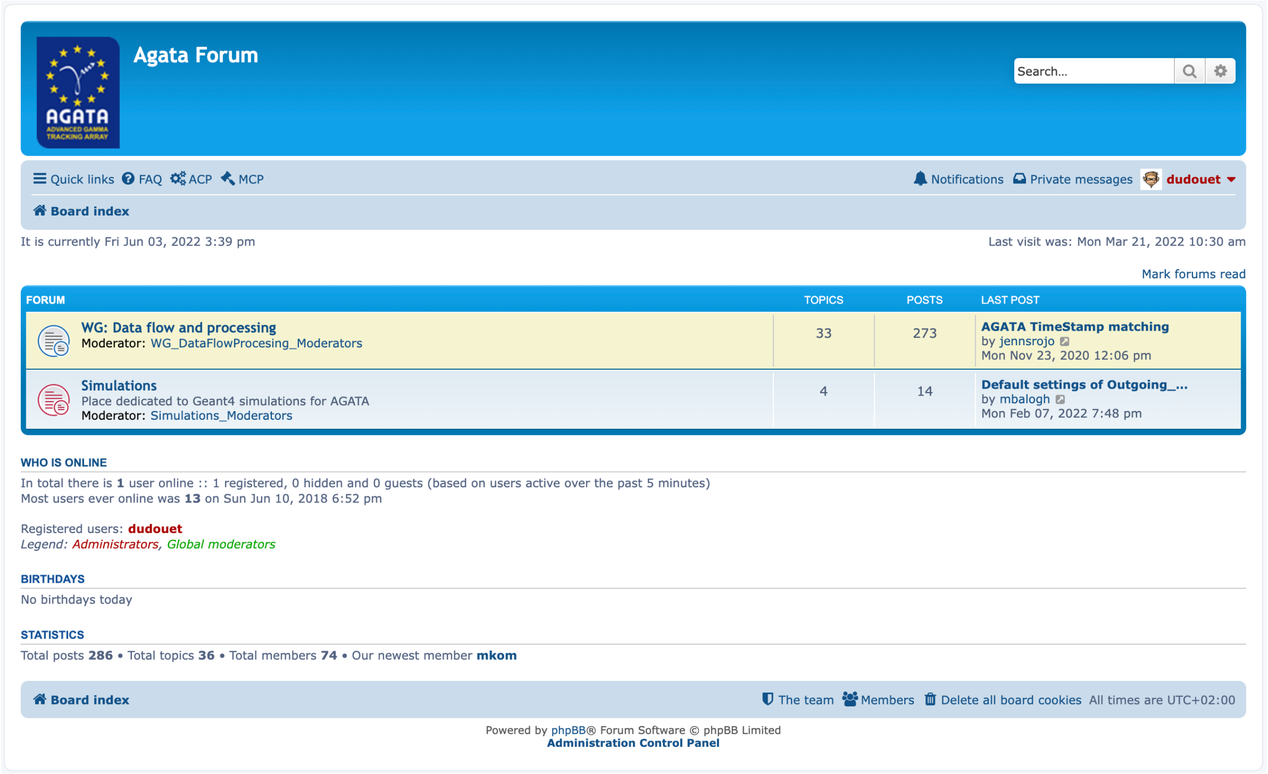
run_0018 — dudouet@lyoserv2: ~/GridTests_gfal2 — sshpass -p zzzzzzzz ssh -XY dudouet@lyoserv.in2p3.fr ▸ ssh — 154×19

```
[dudouet@lyoserv2:~/GridTests_gfal2$ ./GridDataSync.py --start

*********************************
** GridDataSync configuration **
*********************************

SERVER              : srm://ccsrm02.in2p3.fr:8443/srm/managerv2?SFN=
BASE_DIR_ON_GRID    : /pnfs/in2p3.fr/data/agata/
INPUTDIR            : e680/e680
OUTPUTDIR           : /gridgroup/matnuc/agata/
Include pattern     : .*run_0104.*
Exclude pattern     : .*Narval.*:.*zWorkingDir.*

... updating the catalog
 -- 175 Files from e680/e680 in the list
 ==> 0 downloaded (175 remaining)
 ==> 175 brought online for non downloaded files (0 remaining)
...starting to download the 175 requested files...
Copied files: 16/175, current:   80.0 kB, total:    1.2 MB/ 287.0 GB, rate= 304.1 kB/s
```

The AGATA forum

www.agata.org/forum

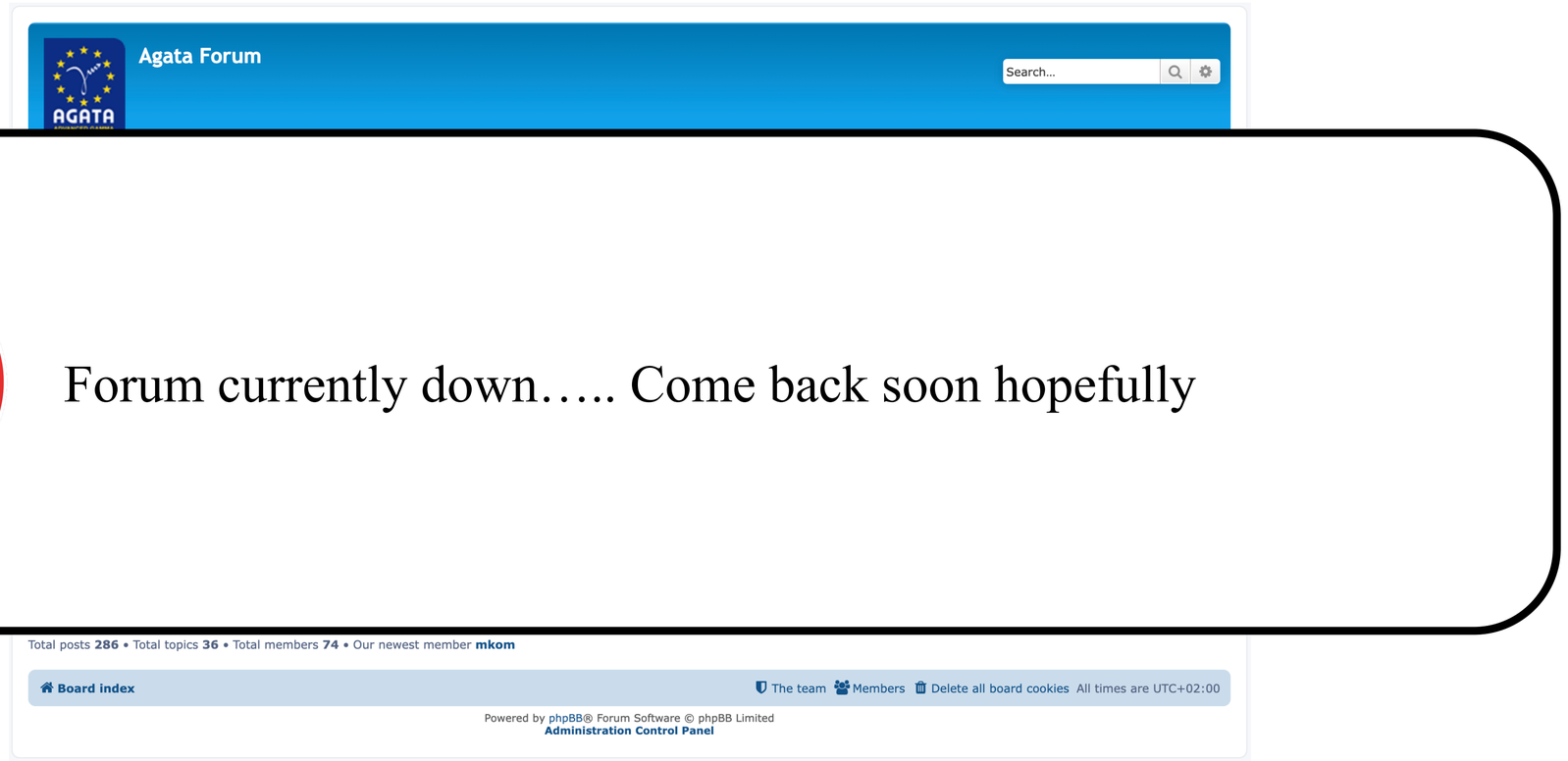
- The forum is dedicated place for getting any help in:
  - ➥ Software installation, Grid data download, Data analysis, G4 simulations…
  - ➥ Documentation download



The AGATA forum

www.agata.org/forum

- The forum is dedicated place for getting any help in:
  - ➥ Software installation, Grid data download, Data analysis, G4 simulations…
  - ➥ Documentation download



The TkN library

A C++ interface to nuclear databases

Jérémie Dudouet, IP2I Lyon, CNRS
Diego Gruyer, LPC Caen, CNRS

tkn.in2p3.fr
gitlab.in2p3.fr/tkn/tkn-lib

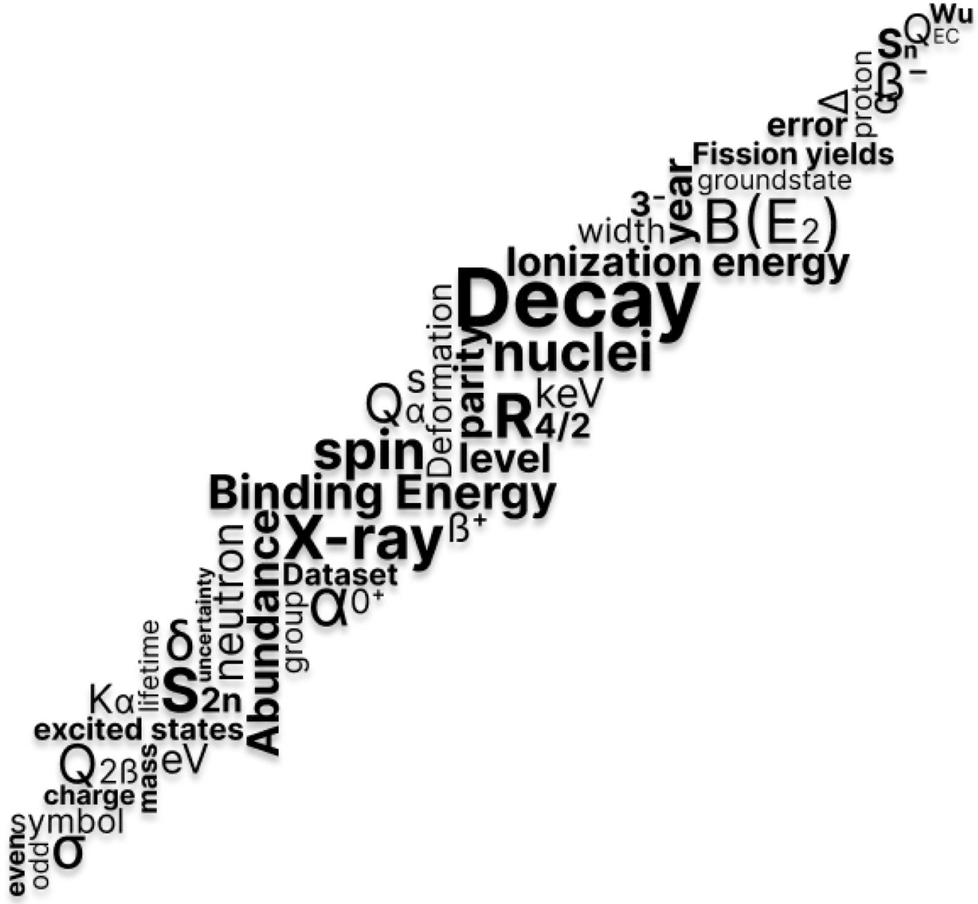
The TkN library

The nuclear physicist's playground

Many objects
- 118 chemical elements
- ~ 3000 nuclei
- ~ 200 000 excited states
- ~ 500 000 nuclear decays

Each with multiple properties
- energies
- lifetimes
- Q-values
- binding energies
- ...

➥ Millions of nuclear properties published



The TkN library

The nuclear physicist's nightmare

Data accessibility often non-trivial

- PDF file decoding
- web page parsing
- old database format decoding

➥ Each nuclear physicist has already created her/his own macros, physics case dependant

```
12C   L 0          0+              STABLE
12C 2 L ISPIN=0 $G=2.0010415963 45 (2002Be82)
12C 3 L XREF=ABDEGHIKMNOPTUWXZacdefijklmopqstuvwxyz123456789
12C   L 4439.82   212+              10.8E-3 EV6
12C 2 L %IT=100 $ ISPIN=0 $MOME2=6 3 (1983Ve01)
12C 3 L XREF=ABDEGHIKMNOPTUWXYacdefijklmopqstuvwxyz12346789
12C  cL E$From average of values given in (1967Ch19, 1967Ko14, 1971St22,
12C 2cL 1974Jo14, 1974No07, 2016Mu06).
12C 3cL The value is dominated by E|g=4438.91 keV {I31} in (1967Ch19).
12C  cL WIDTH$From average of (1958Ra14,1967Cr01,1968Ri16,1970Co09,1970St10).
12C   G 4438.94      100      E2
12C 2 G WIDTHG=10.8E-3 EV 6$BE2W=4.65 26
12C   L 7654.07   190+              9.3 EV    9
12C 2 L ISPIN=0 $ %IT=4.16E-2 $ %A AP 100
12C 3 L XREF=ABDEGHIJKMNOPTUVWXacdefijklnopqruvwxyz1234789
12C  cL E$See discussion in (1976No02). Note: E{-x}=7657.8 keV {I10} is
12C 2cL obtained from analysis of |g rays measured in (2016Mu06).
12C  cL WIDTH$Using |G{-|p}/|G=(6.7 {I6})|*10{+-6} (average of
12C 2cL 1972Ob01,1977Ro05,1977Al31) and |G{-E0}=|G{-|p}=(62.3 |meV {I20})
12C 3cL (see discussion in 2010Ch17,2011Vo16).
12C  cL $|G{-rad}/|G=(|G{-|g}+|G{-|p})/|G=(4.16 {I11})|*10{+-4}.
12C 2cL From 10{+4}|*|G{-rad}/|G=3.3 {I9} (1961Al23), 3.5 {I12} (1964Ha23),
12C 3cL 4.20 {I22} (1974Ch03), 4.4 {I2} (1975Da08), 4.15 {I34} (1975Ma34),
12C 4cL 4.09 {I27} (1976Ob03), 3.87 {I25} (1976Ma46).
12C 5cL The value from (1961Al23) has sometimes been miscopied as 3.4, but
12C 6cL it has no impact on the average. The value of (1975Da08) has been
12C 7cL corrected, as indicated in (1976Ob03). The value (2.82 {I29})|*10{+-4}
12C 8cL (1963Se23) is a statistical outlier; including this value yields
12C 9cL the average (3.99 {I18})|*10{+-4} that is the weighted average
12C acL using the external uncertainty. The value in (1990Aj01) did not
12C bcL use the corrected (1975Da08) value. In (2014Fr09), the
12C ccL value (4.19 {I10})|*10{+-4} is deduced by rounding the above
12C dcL values to the nearest tenth.
12C  cL $|G{-rad}=3.87 meV {I39} and |G{-E2}=|G|{-|g}=3.81 meV {I39}
12C  cL $Decay mechanisms were analyzed in (2017Sm03);
12C 2cL the decay is >99.92% via sequential |a-decay to {+8}Be{-g.s.}
12C 3cL and <0.047% via direct decay into 3|a-particles.
12C 4cL This is relevant for the astrophysical 3|a rate, via detailed balance.
12C 5cL  Also see
12C 6cL (2011Ra43, 2012Ma10, 2012Ki07, 2013Ra20, 2014It01, 2016Mo05, 2017De25)
12C   G 3213.79      100      E2
12C 2 G FL=4439.82
12C 3 G WIDTHG=3.81E-3 EV 39$BE2W=8.26 85
```

The TkN library

The nuclear physicist's nightmare

Data accessibility often non-trivial

- PDF file decoding
- web page parsing
- old database format decoding

➥ Each nuclear physicist has already created her/his own macros, physics case dependant

➥ Various software exist, with their own (partial) database implementation

The TkN library

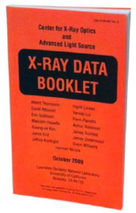
The TkN database

Data sources

- Chemical element properties
- X-ray data
- Isotopes properties
- Levels and decays properties

PubChem

NuDat 3.0



SQLite3 database

- directly **embedded** in the library
- **stable**, **cross-platform**, backward compatible (2050),
- widely used by many companies

AIRBUS

SQLite

➥ **Automatic monthly database update (stored on IP2I's servers)**

The TkN library

The TkN database : summary

Database content
- 3 559 isotopes
- 22 906 datasets
- 550 080 levels
- 703 109 decays

To be added in future TkN versions
- particle decays
- charge radii
- other properties proposed by the users
- python/julia interface

Reading performances
- 1rst full database read ~ 15s
- 2nd full database read ~ 0.02s
- Extracting randomly 10^8 levels (1 thread) ~ 250s
- Extracting randomly 10^8 levels (16 thread) ~ 30s

The TkN library

The TkN user interface

Dedicated c++ classes are provided in order to

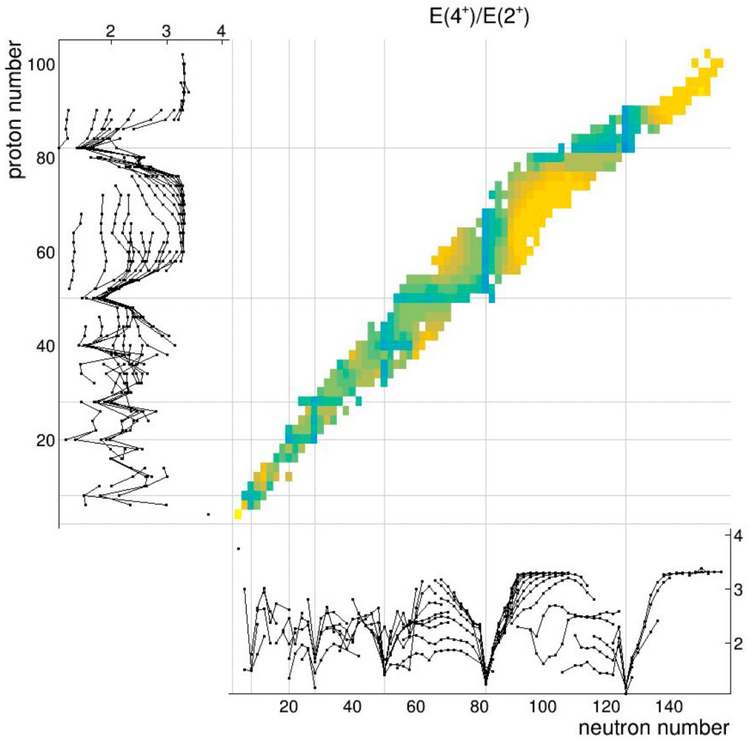
- browse the **database** (tkmanager)
- extract the desired **properties** (tknucleus, tklevel, tkdecay...)
- handle **uncertainties**, including asymetric unc, limit values... (tkmeasure)
- handle **units** and units conversion (tkunit_manager)

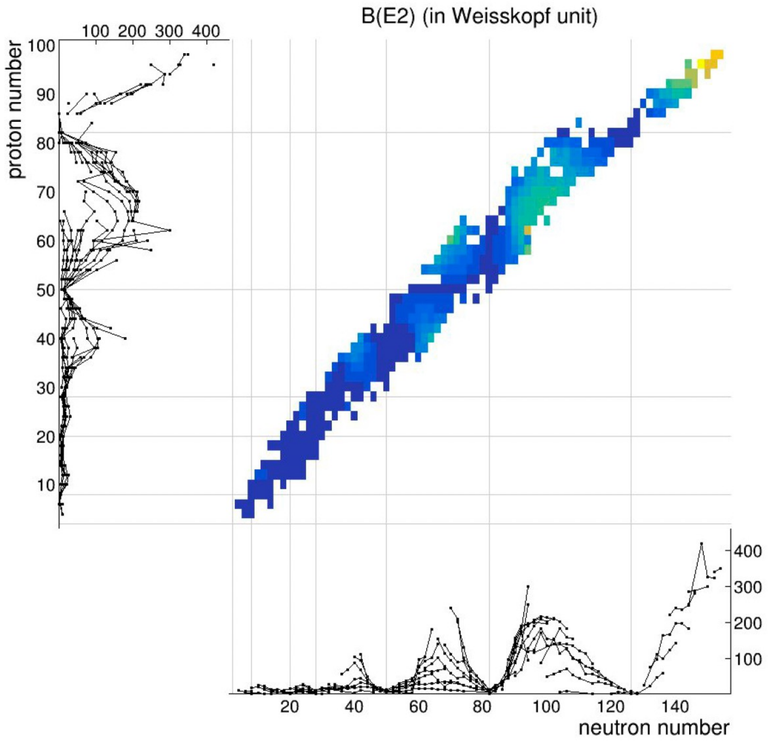
Dedicated tools are provided in order to:

- download/update the database (tkn-db-update)
- start an **interactive root** terminal with TkN linked (tkn-root)
- **link** TkN to an external project (cmake example, git submodule)

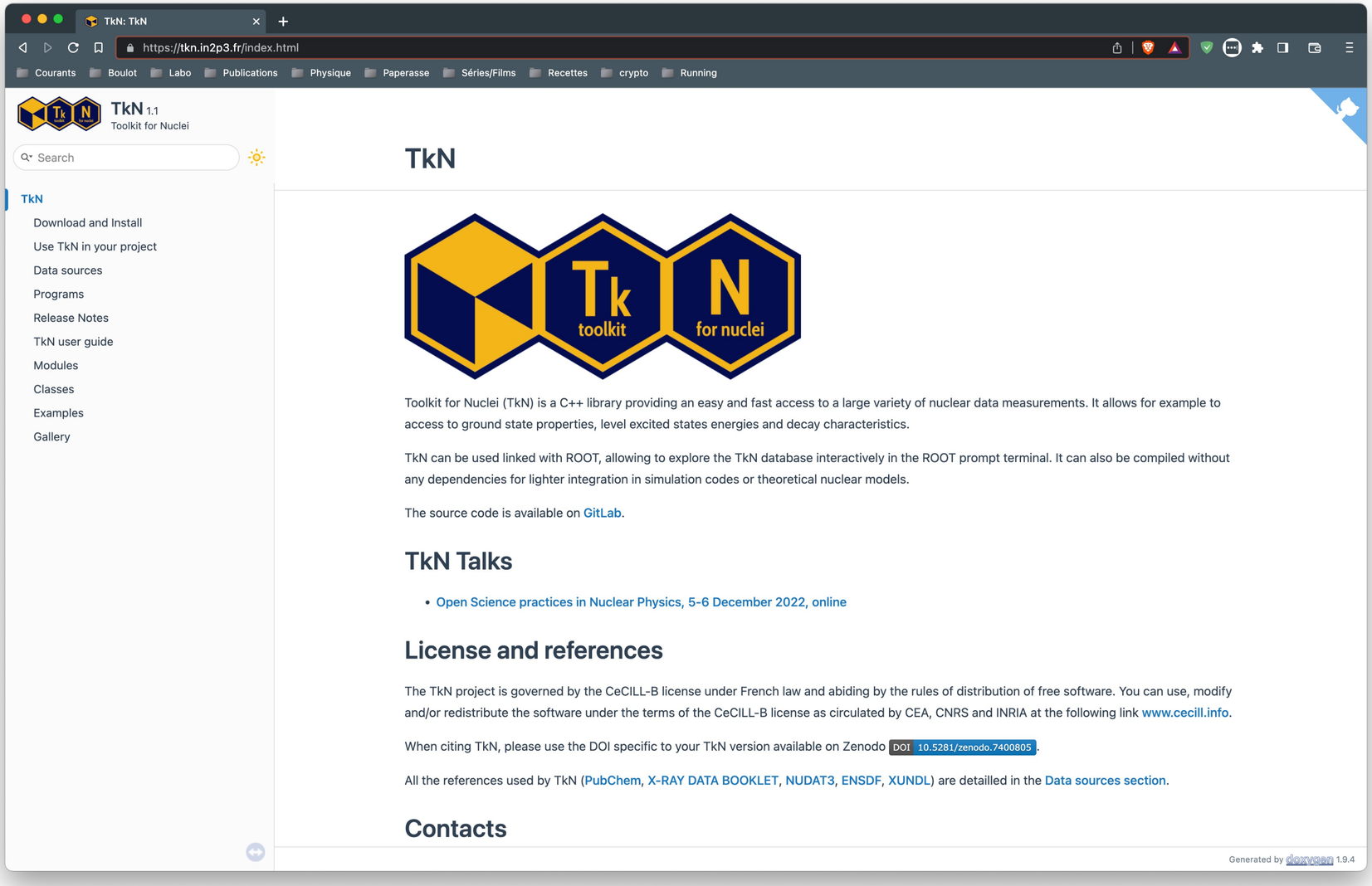
The TkN library

Some examples : level and decay properties

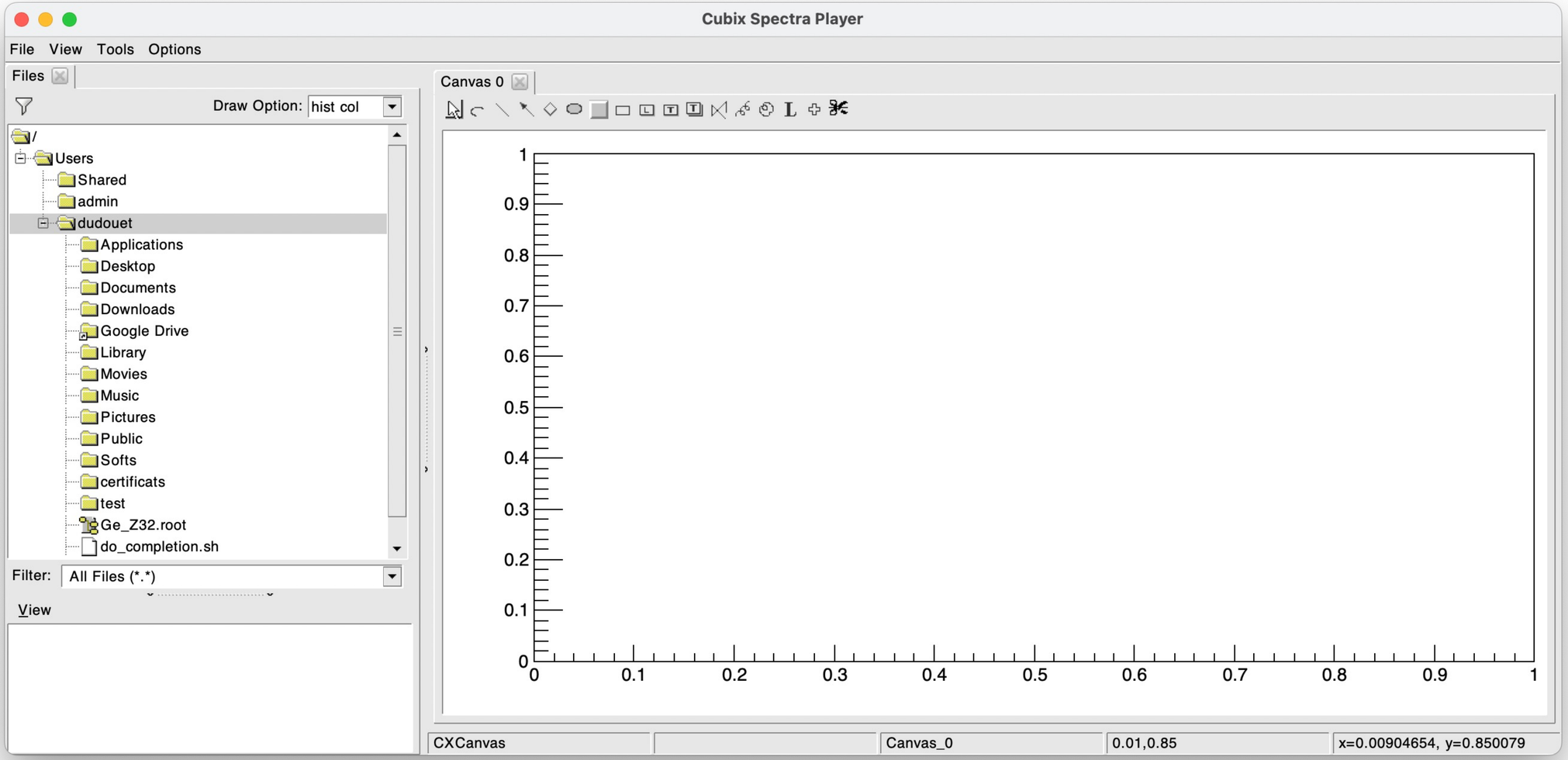




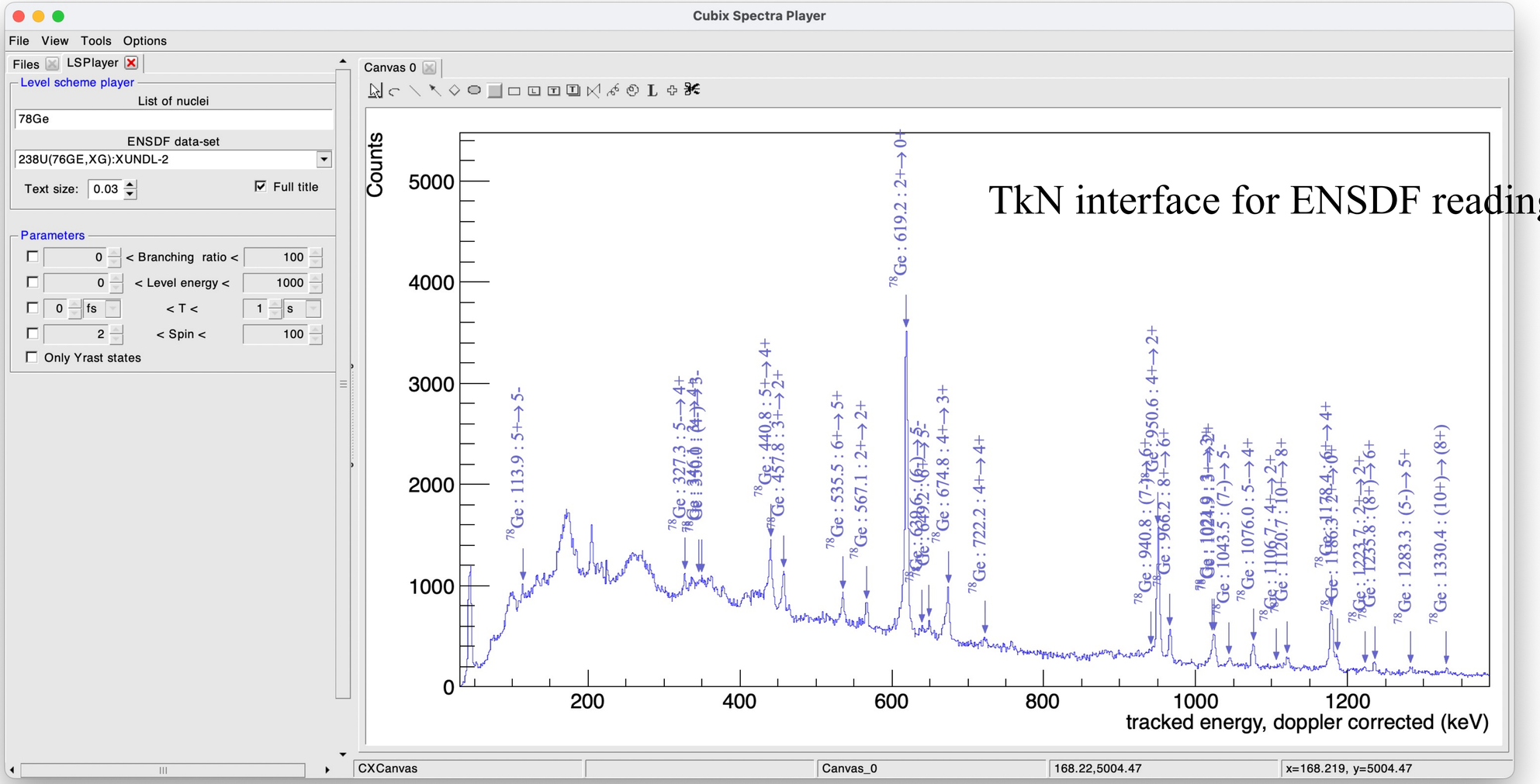
The TkN library: tkn.in2p3.fr



The Cubix software

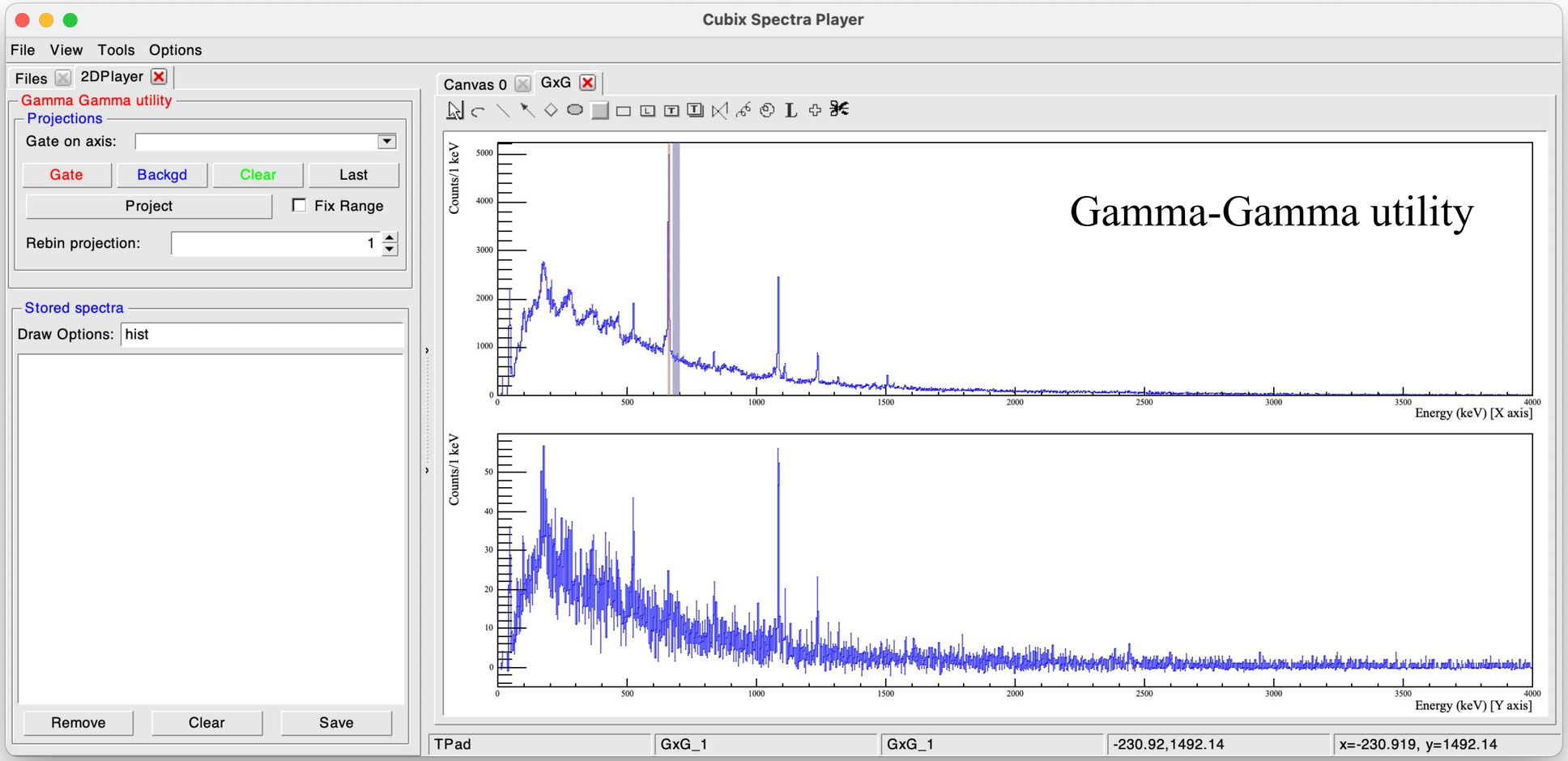


The Cubix software

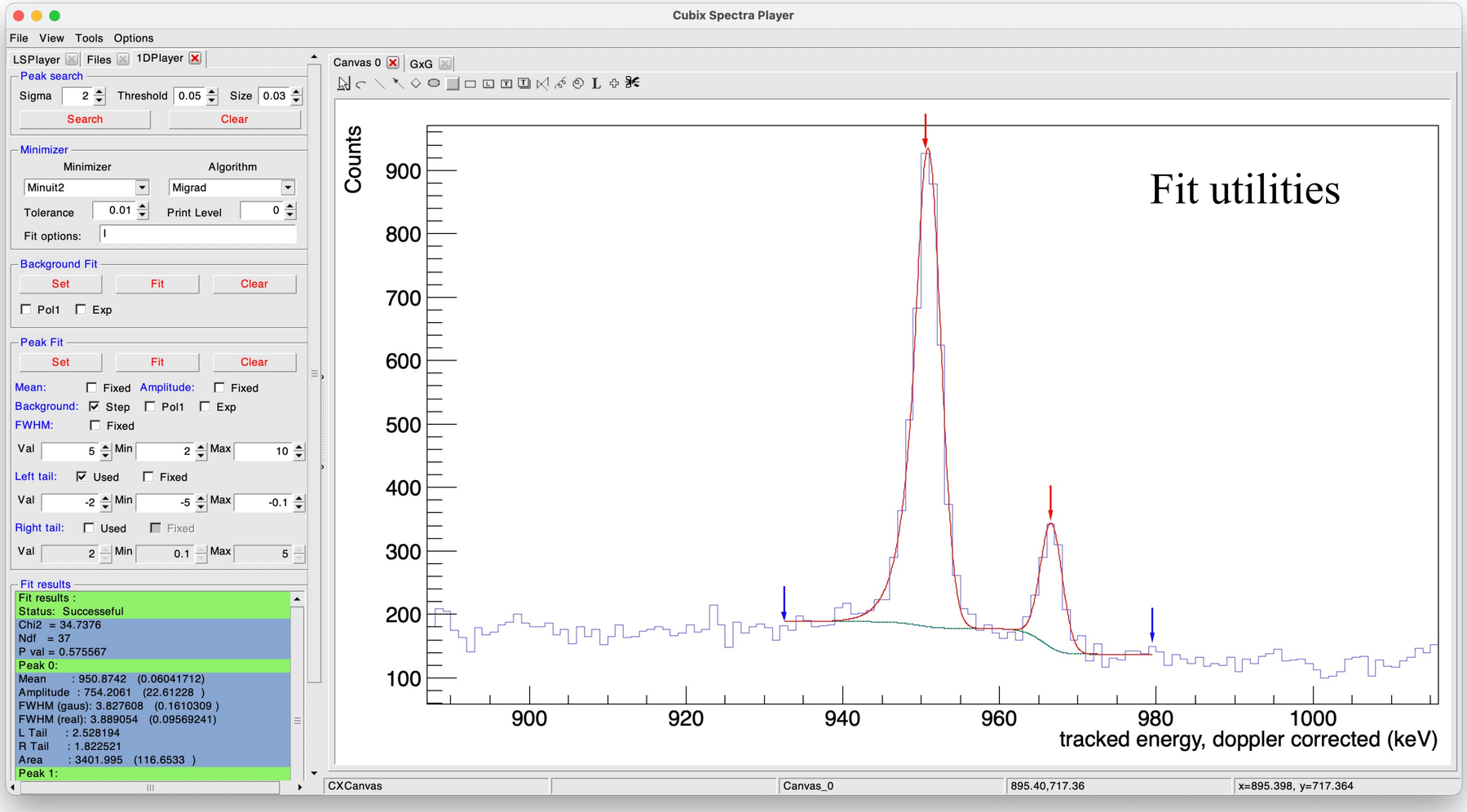


TkN interface for ENSDF reading

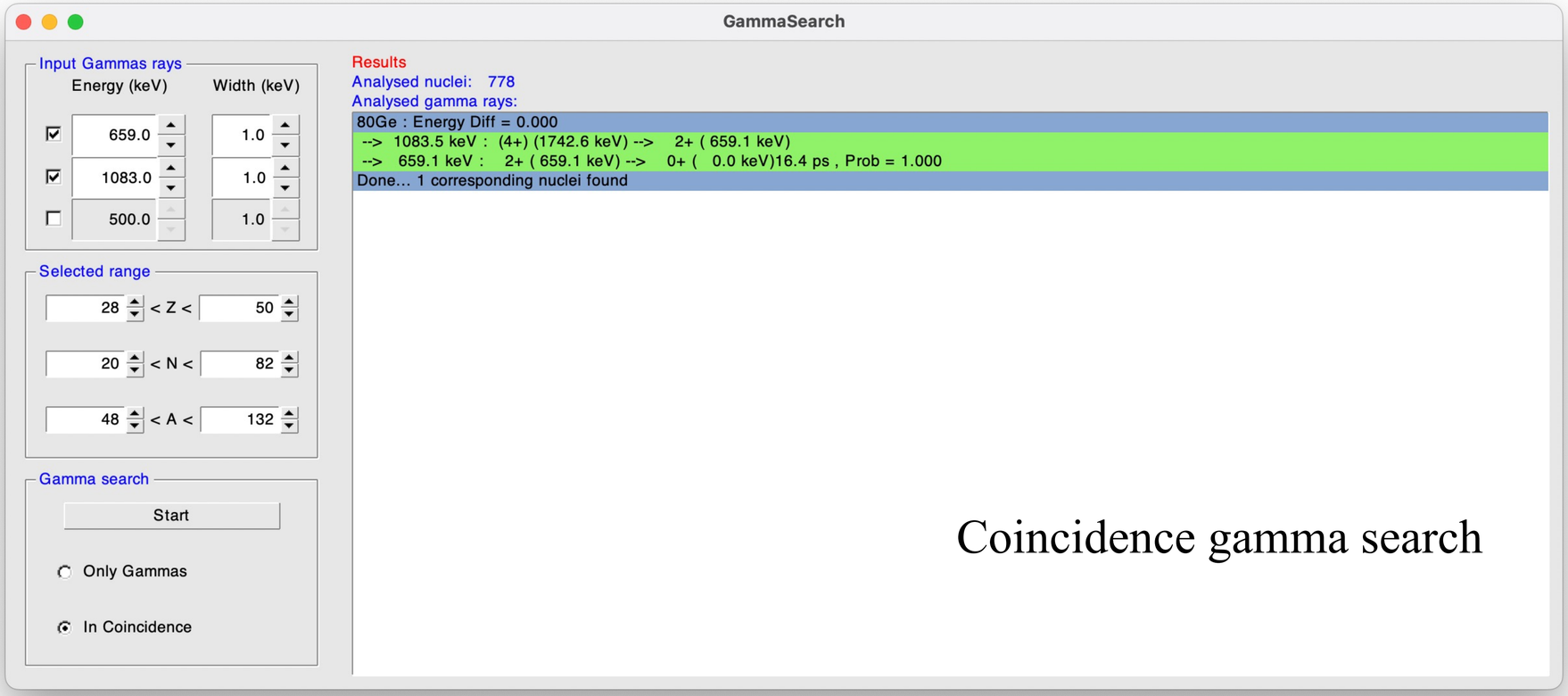
The Cubix software



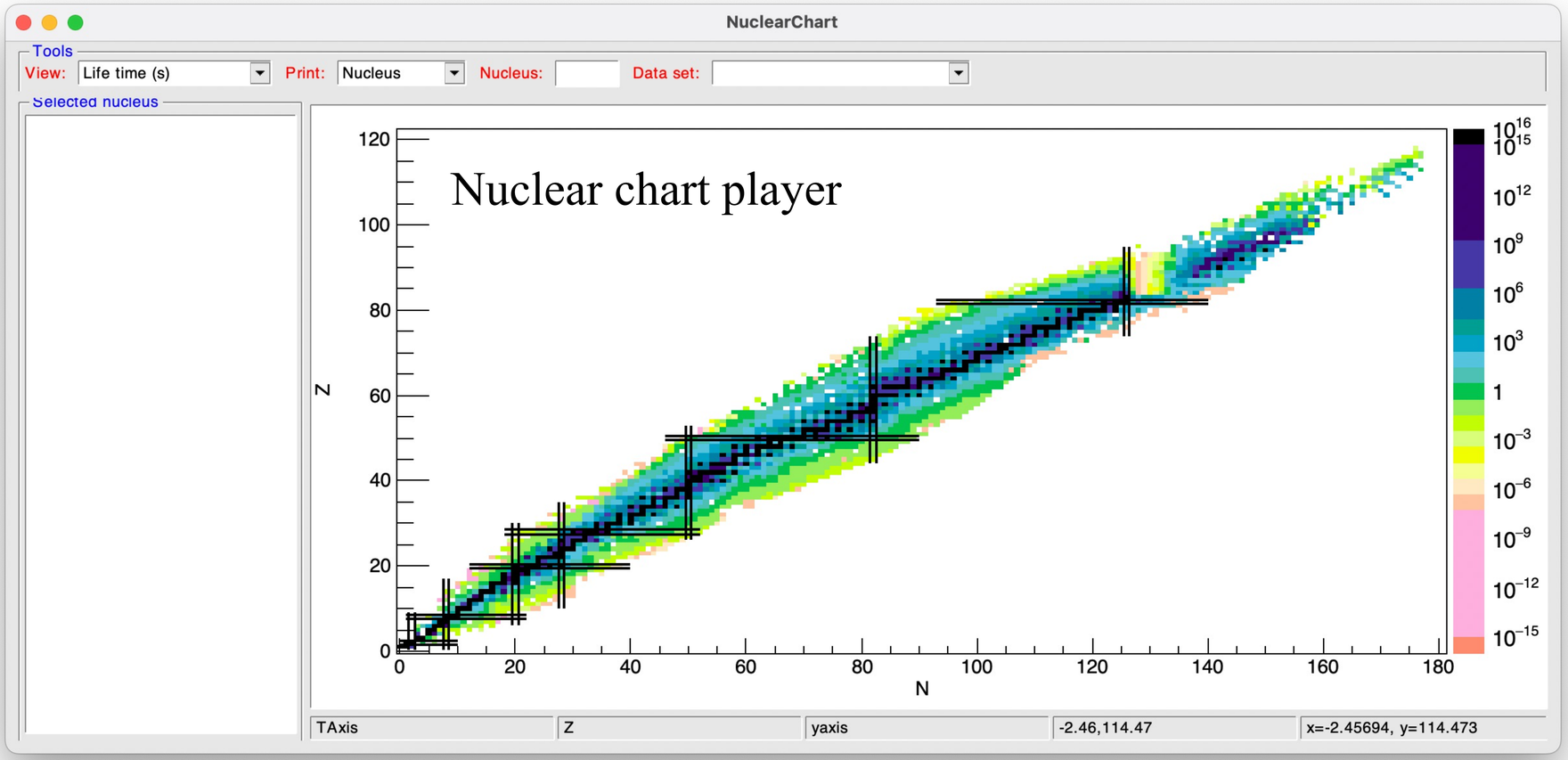
The Cubix software



The Cubix software

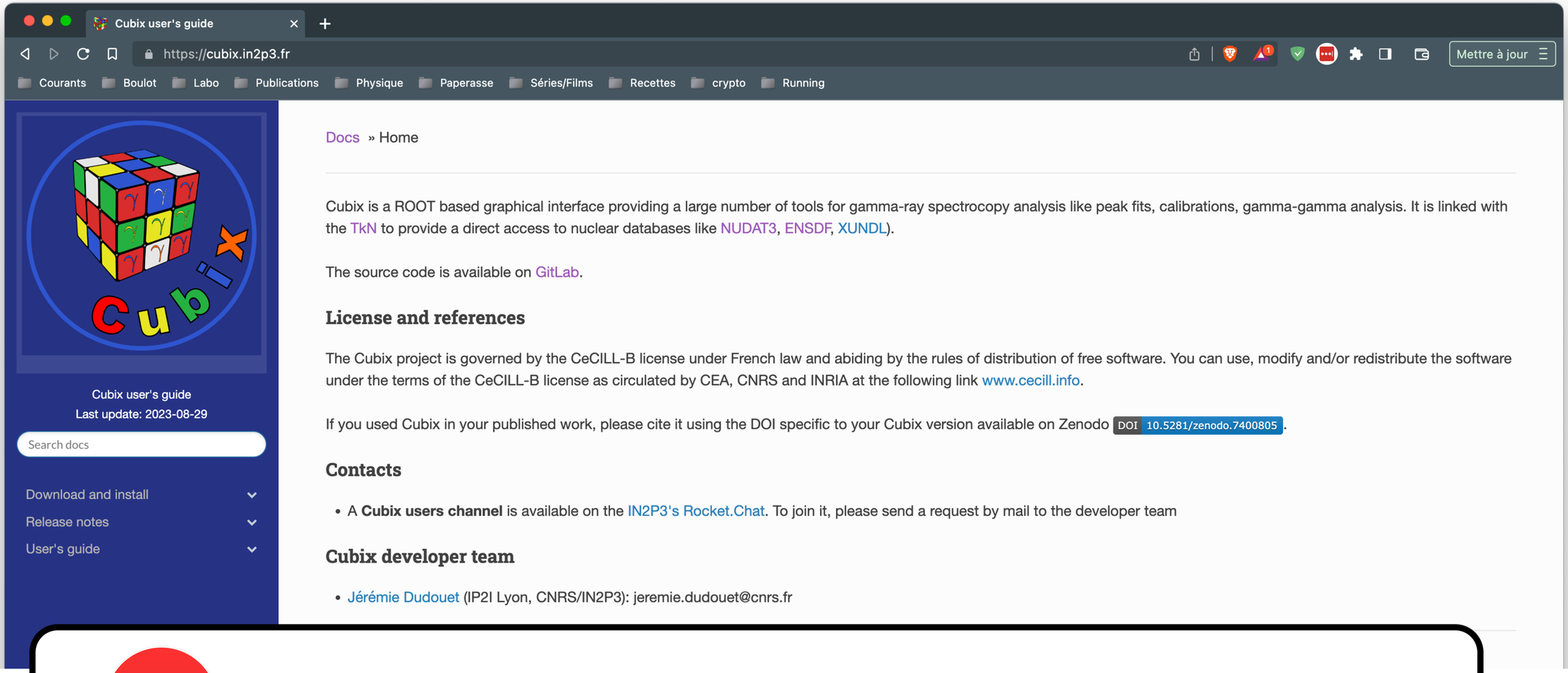


The Cubix software



The Cubix software

The Cubix software: cubix.in2p3.fr



((!)) Still under development, can show some unstabilities