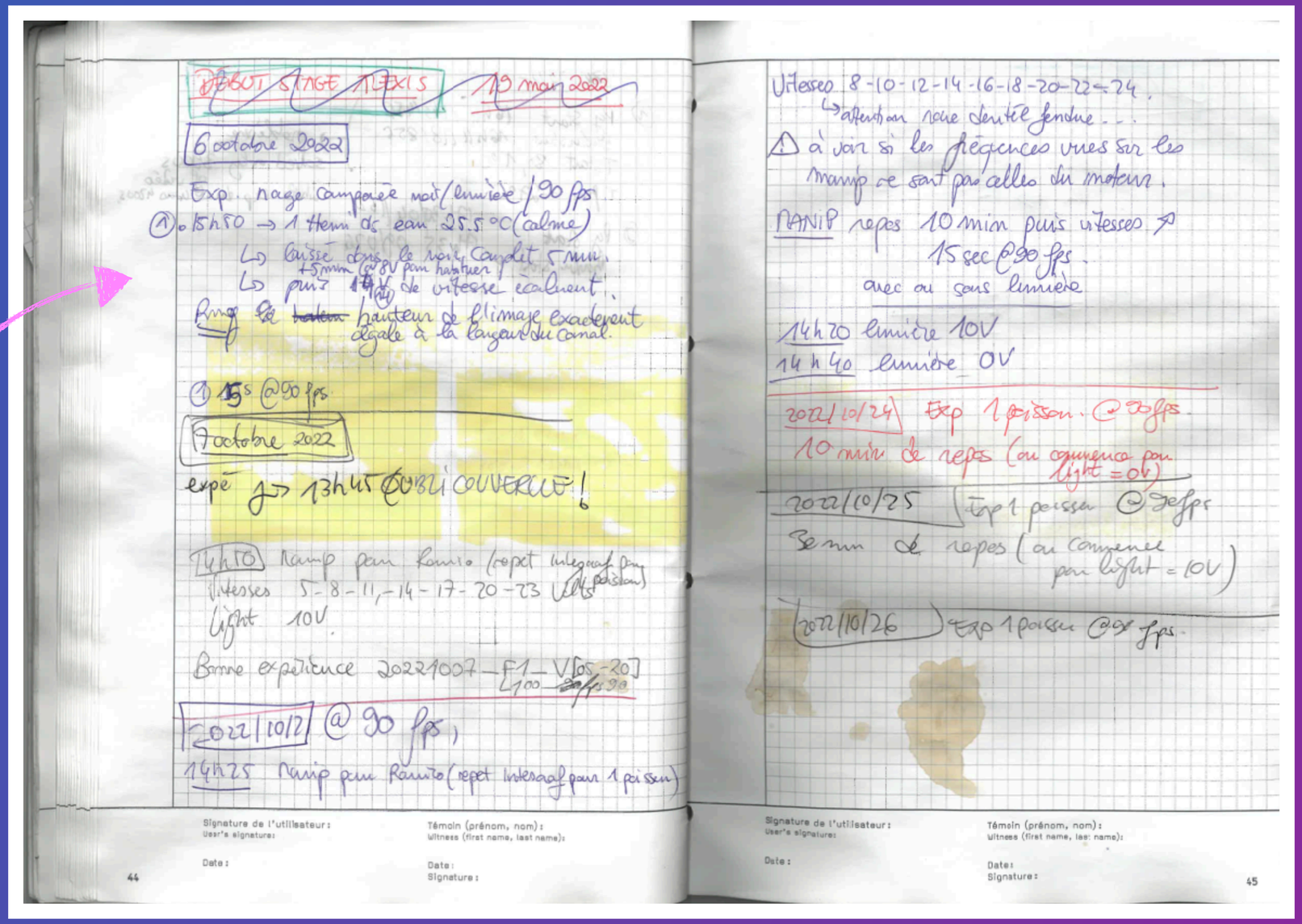
An introduction to XARRAY & NETCDF

A NICE WAY TO SHARE, STORE, LOAD AND PROCESS YOUR DATA

WHY WOULD YOU NEED THIS ?

- if you do **that**

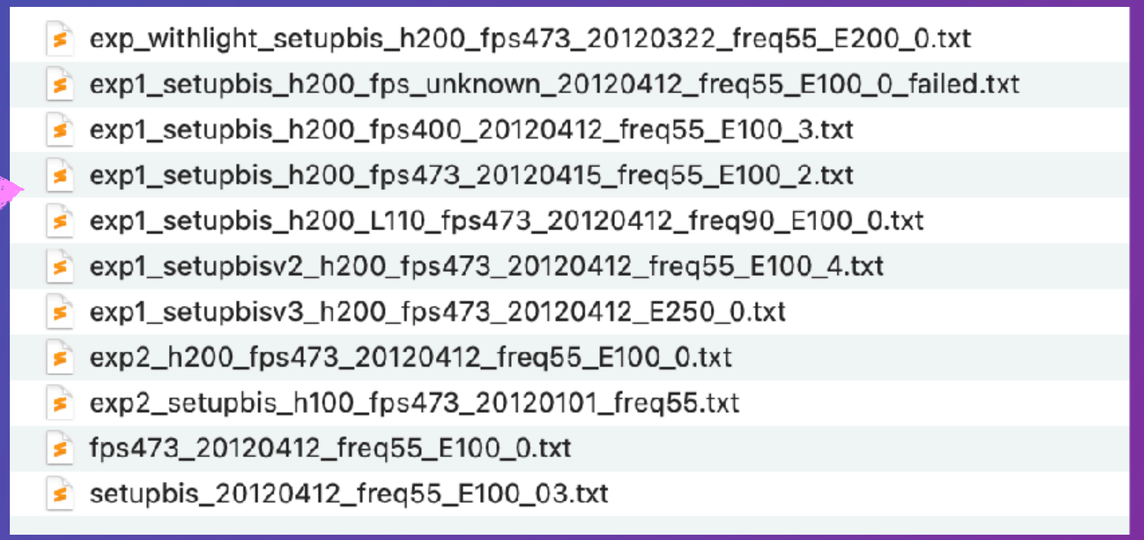
(but still always lose your metadata)



WHY WOULD YOU NEED THIS ?

- or if you do **this**

(and you cannot remember what it means)



WHY WOULD YOU NEED THIS ?

- or if you cannot **remember** what your code does after 1 week away from the lab

WHY WOULD YOU NEED THIS ?

and also :

- if you want to **share** your data (*Open-source publication, collaboration, …*)
- if you work with **large** data arrays
- if you work in **geoscience**
- if you have **non-aligned data** (data with different number of points)
- …

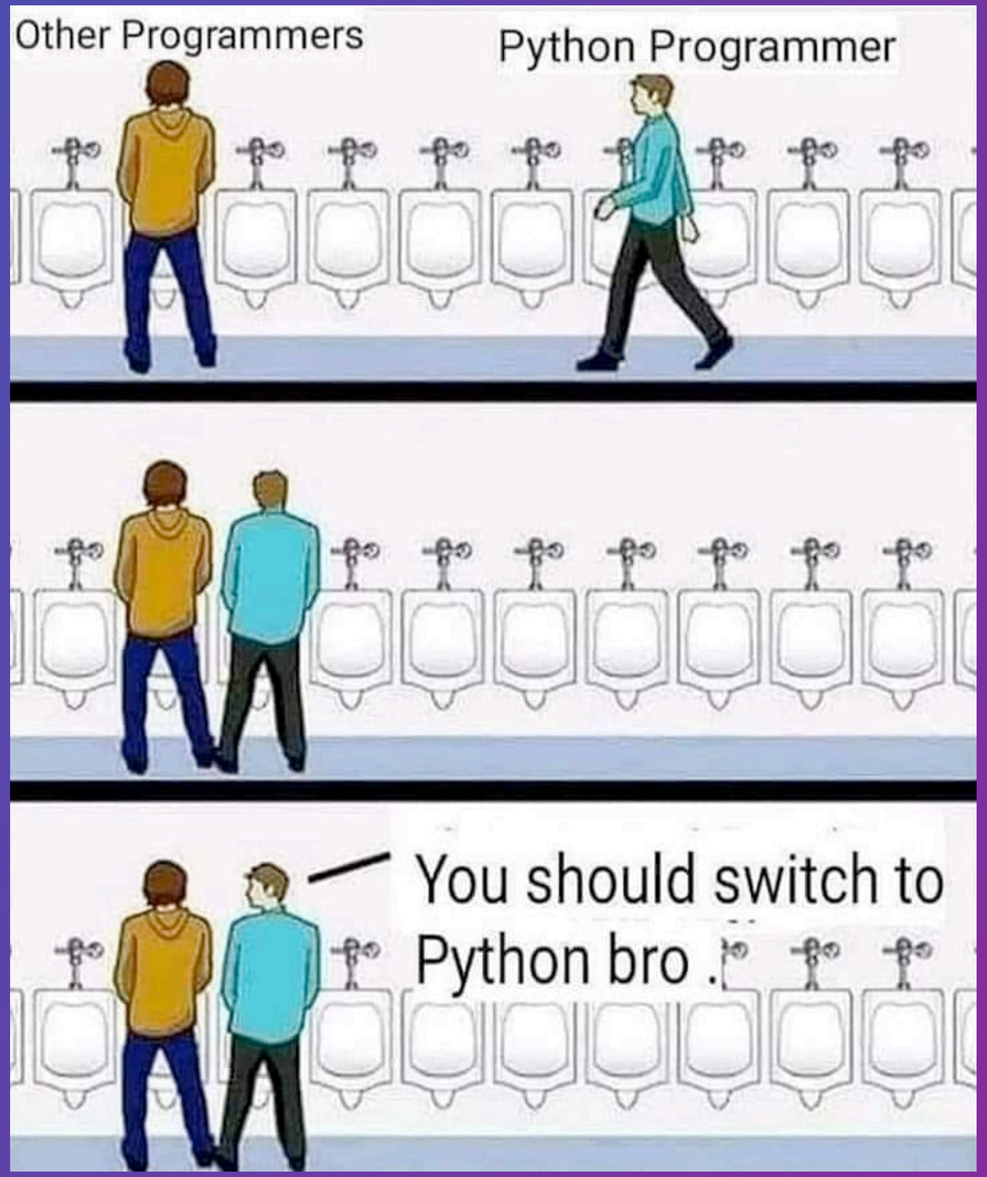
Then, You might be interested !

WHAT I'LL TRY TO DO

- Present what `NetCDF` and `Xarray` are
- **Code** an example with you —*PIV data processing*
- **Convince** you that it can make your life easier (and people you share data with)

DISCLAIMER

- I will code in Python, but it works in MATLAB too



DISCLAIMER 2

- This is my solution, not the best solution (I did not do a full « state of the art »)

NetCDF

NETWORK **C**OMMON **D**ATA **F**ORM

NETCDF

- A **file format** (.nc)
  - open-source
  - « self-describing » (= contains header with metadata)
  - binary (not human-readable, but can be opened with some software [*eg. Panoply*])
  - supports compression

NETCDF

- **Good points**
  - easy to share & use
  - widespread, esp. in geosciences (IPCC, NASA, …)
- **Bad points**
  - Objectively, none

XARRAY

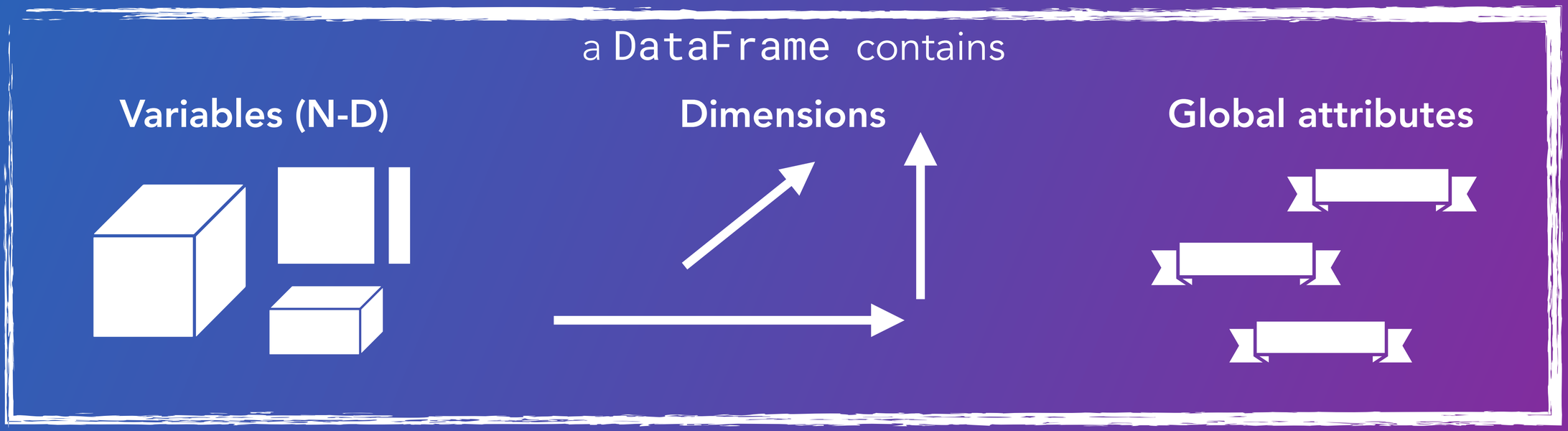
A PYTHON MODULE

XARRAY

- A **Python module**
  - **open-source**.
  - build on top of `pandas` and `numpy`
  - aim : **read**, **write** and **access** NetCDF files

XARRAY - PRACTICALLY

- Load data arrays into `DataFrame` (N-dimensional tensors)
- `DataFrame` have labels, are easy to call and to do operations on



XARRAY - PRACTICALLY

- **Short term payoff**
  - You write less code
  - You manipulate multiple dimensions easily
- **Long term payoff**
  - Everyone (including you) can understand your code, even months after writing it

XARRAY - USE CASES EXAMPLES

- **PIV, height maps** > *Juan, Tristan, Gatien, Samantha…*
- **Contours tracking** (bubble, drops, root, fish, …) > *Antoine, Maud, Aliénor, Manon, Alice*
- **Particle Tracking** (Lagrangian) > *Lars, Jeanne, Renaud*
- Experiments with lot of different « **specimens** » (cells, dragonfly, …) > *Chloé, Joseph, Camille*
- **Multiple** acquisition **rates** (force sensors, PIV, …) > *Gauthier*
- Anything with a **large number of points** > *Everyone*

> You could try Xarray !

AN EXEMPLE IS BETTER THAN 1000 WORDS