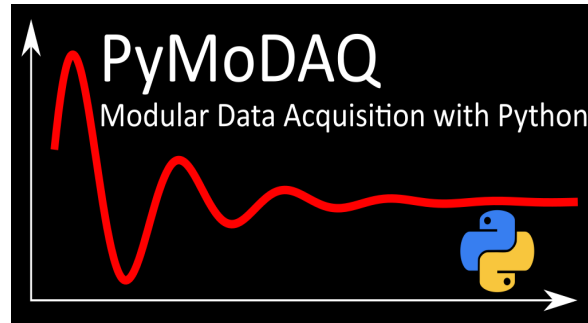




CENTRE
D'ÉLABORATION
DE MATÉRIAUX
ET D'ÉTUDES
STRUCTURALES



FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE

HOW DO YOU
OPEN A .XZBQ FILE?



INTEROPERABLE



REUSABLE

Sébastien Weber

PyMoDAQ - Acquisition Modulaire de données FAIR ?

Documentation: <http://pymodaq.cnrs.fr>

Publication: [Review of Scientific Instruments 92, 045104 \(2021\)](#)

Webinaire IN2P3
20-mars-25

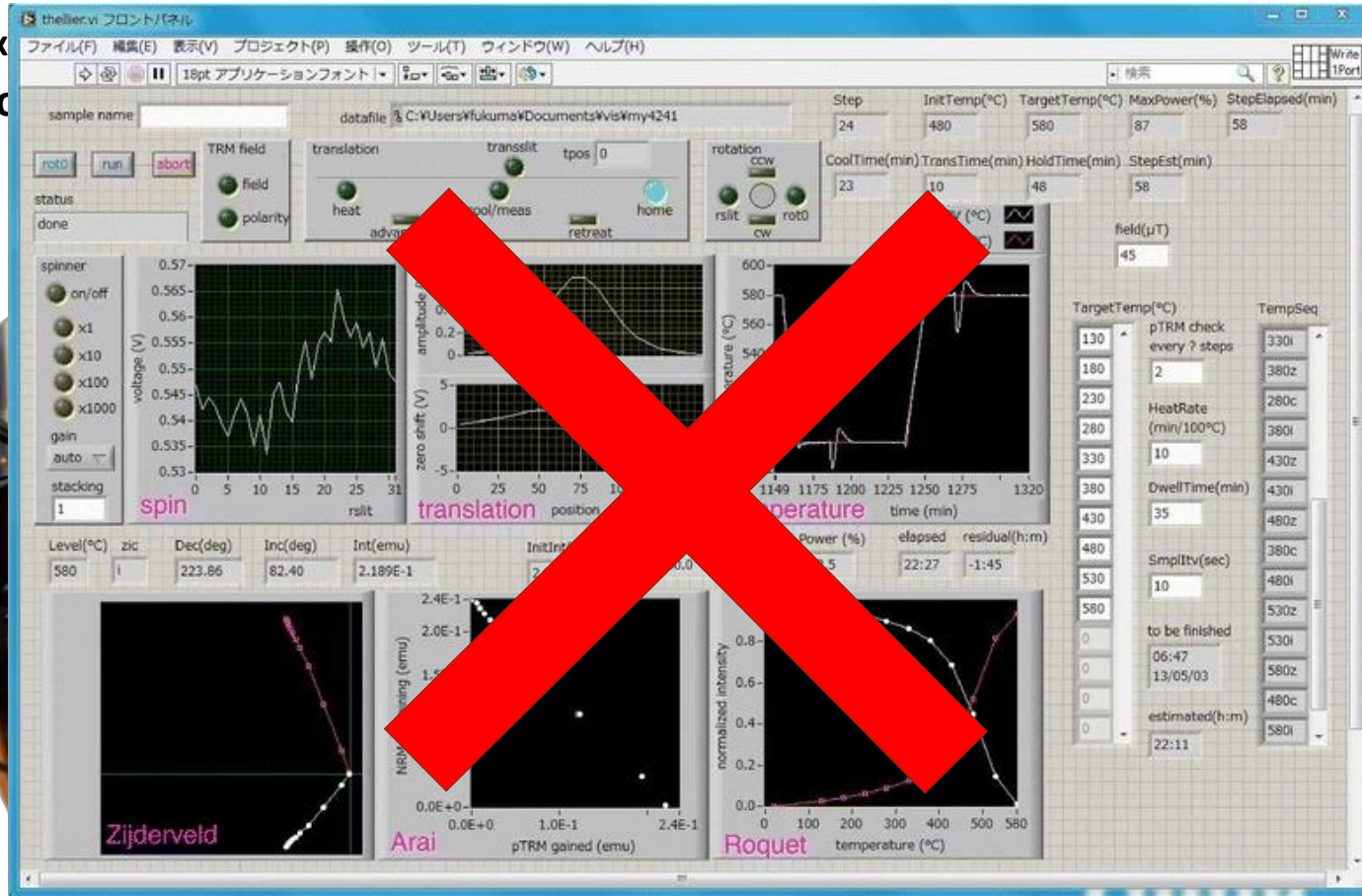
Sébastien Weber
Ingénieur de Recherche

CEMES-CNRS (UPR-8011)



2

L'acquisition des données et l'obsolescence des logiciels!



Un dispositif expérimental :

- Nombreux instruments
- Plusieurs réglages

=> Je développe une interface de contrôle pour mon besoin



Mais :

- Prend du temps
- Ajout de hardware !
- Obsolète !

Besoin d'une solution modulaire, pérenne et communautaire

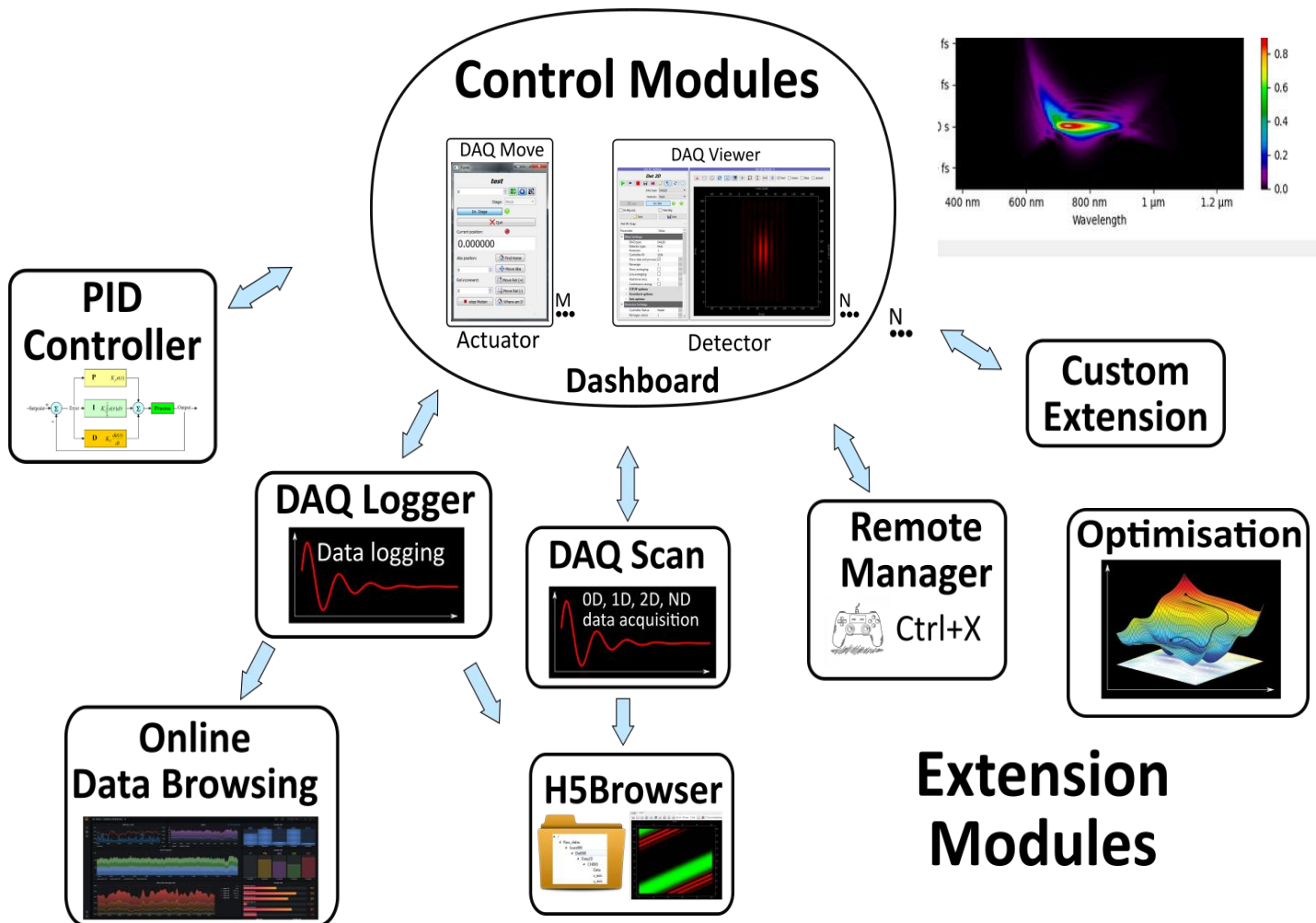


Tableau de bord :

- 1 setup = 1 Configuration
- 1 instrument = 1 module de contrôle générique
- Complètement modulable

Extensions spécifiques:

- Scan
- Logger
- Optimisation
- ...

Outils pour extensions à la demande

Système de plugins :

- Ajout d'instruments
- Ajout de fonctionnalités

4

Naissance et animation d'une communauté

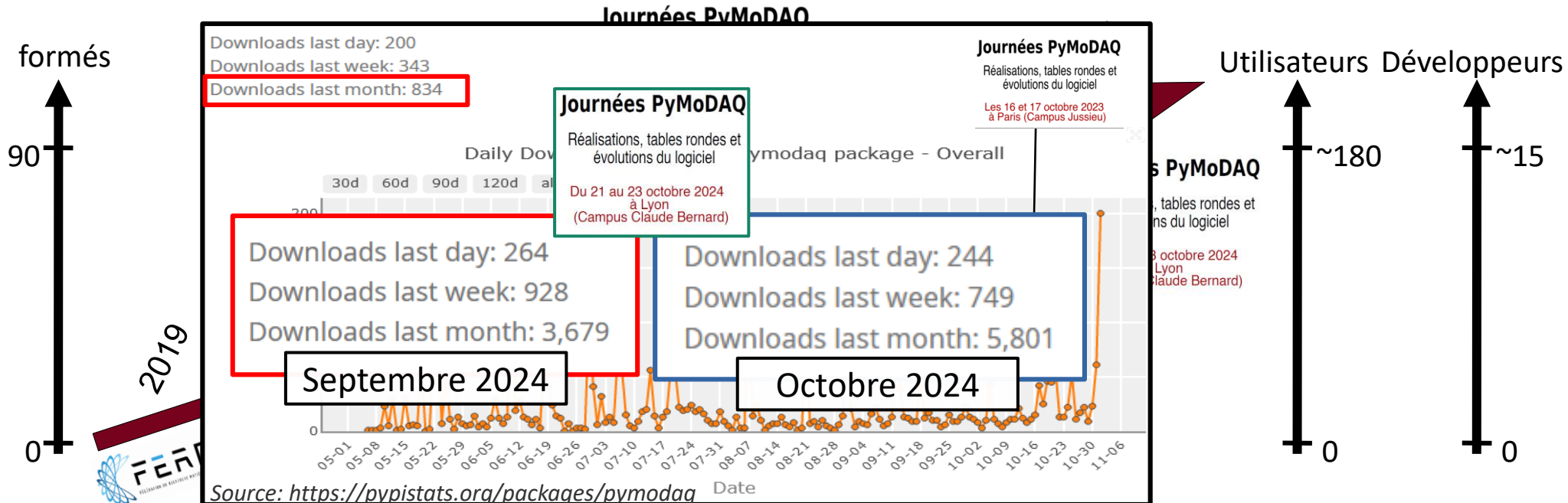


Première publication publique du code en 2018 :

- Adoption progressive
- Formations successives et tutos vidéo
- Journées thématiques

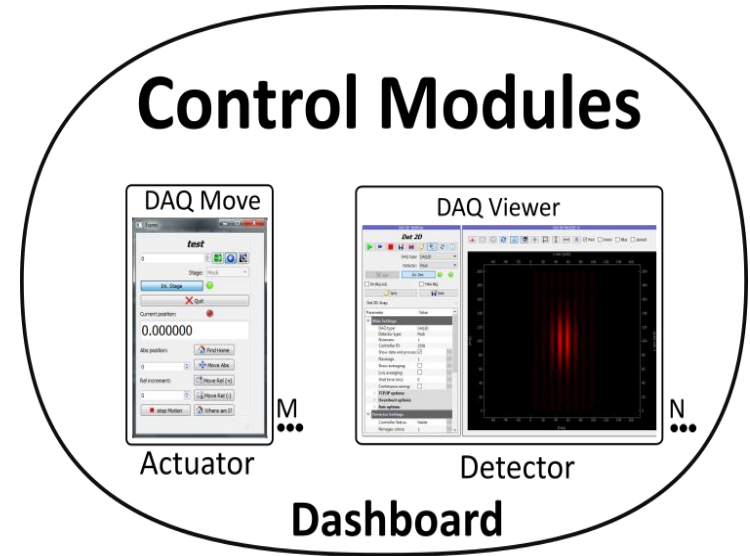
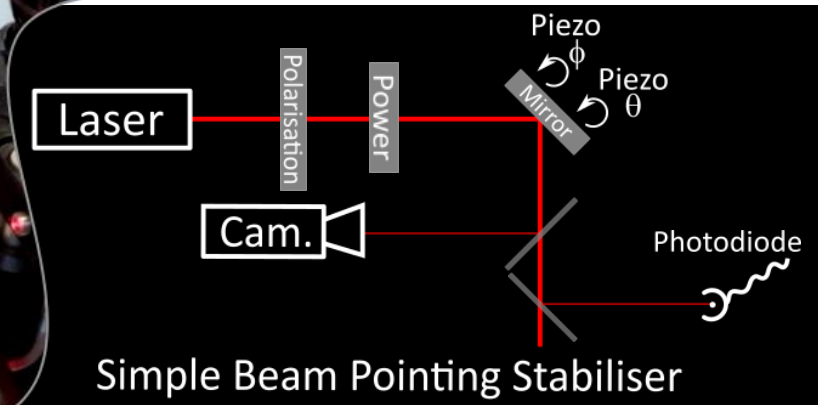
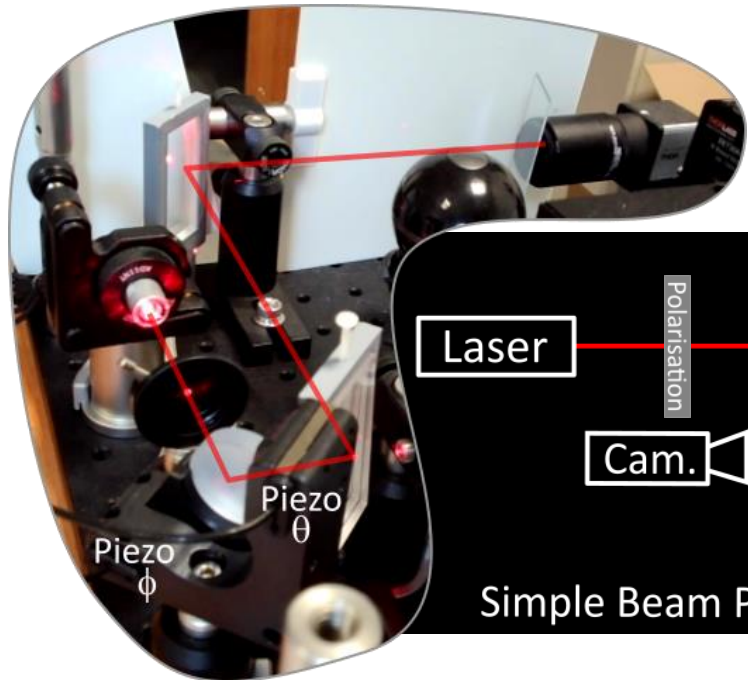
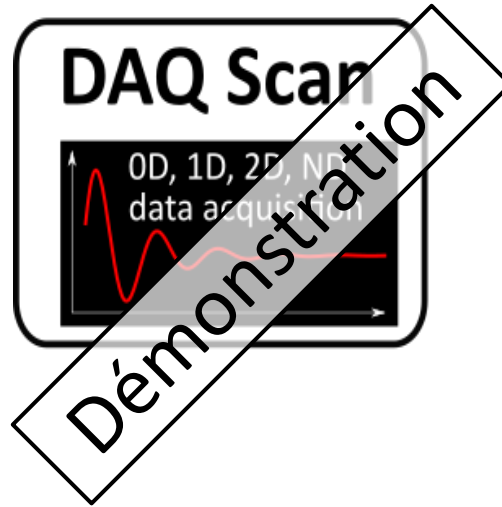
→ Création d'une communauté (90 formés, mailing liste 180 pers.)
Public : chercheurs et ingénieurs

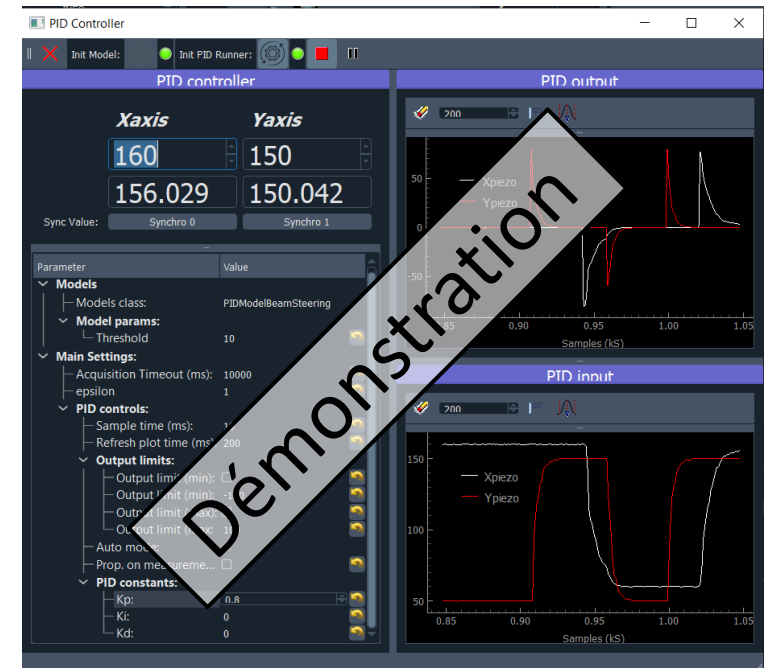
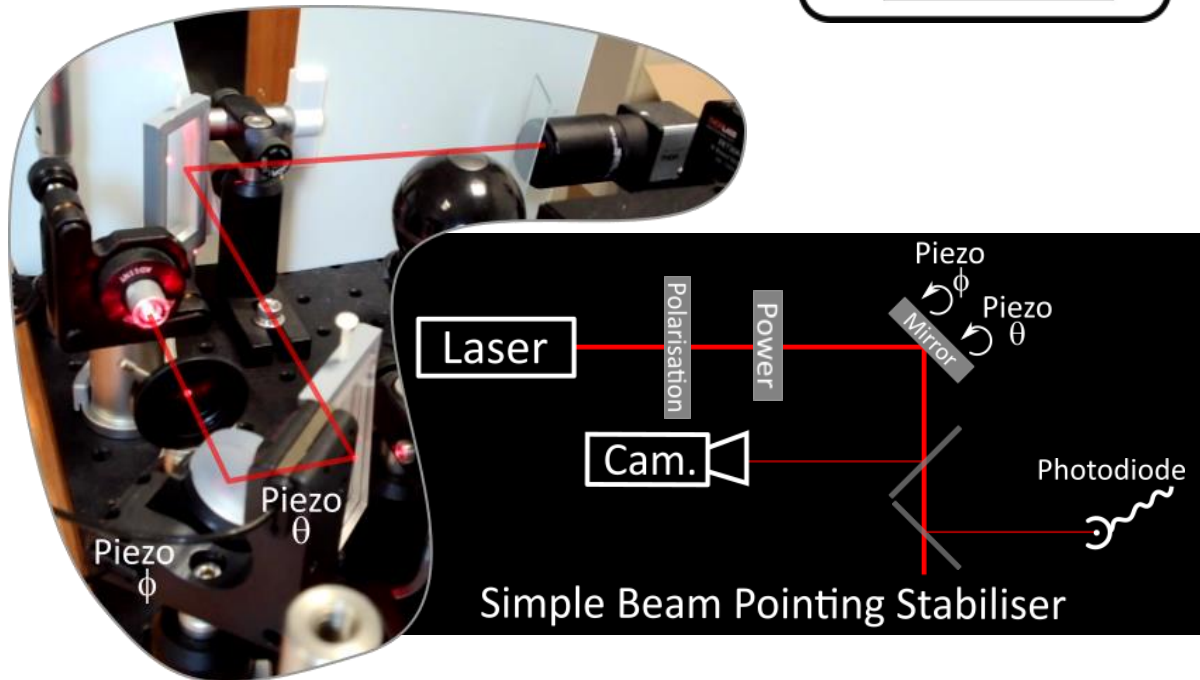
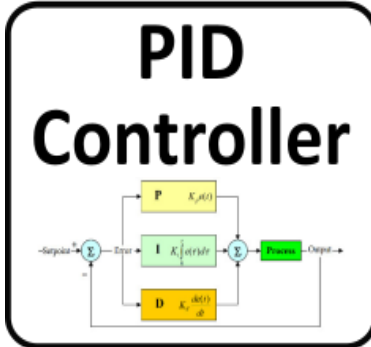
- *The use of PyMoDAQ has really accelerated our experimental development [...] 100% reliable*
- *[...] The updates of the modules, the training sessions [...] also reflects the vitality of the community*



5

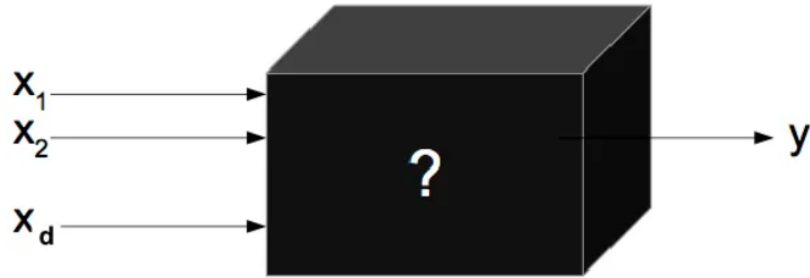
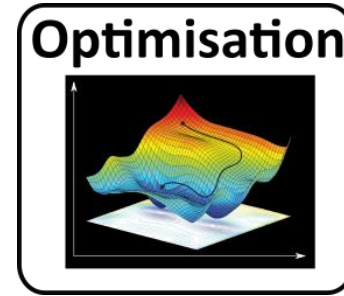
PyMoDAQ pour l'orchestration : Dashboard et DaqScan



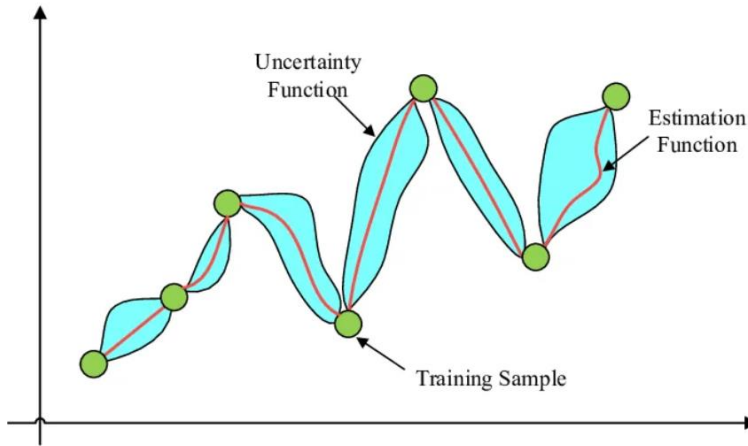


7

PyMoDAQ pour l'optimisation : Algorithme Bayésien



Black box. [Source](#)



Gaussian processes are specified by the estimation function and the uncertainty function. [Source](#)

Concept de « Black Box » :

- Plusieurs potentiels inputs...
- ...donne une valeur « signal »
- L'évaluation à un « coût »

=> *Quelle stratégie pour optimiser le signal ?*

- *Grid/Random sampling ?*
- *Bayesian Optimisation*

Bayesian Optimisation :

- Distribution de fonctions potentielles
- Probabilités sur ces fonctions...
- ...renseigne sur un bon optimum

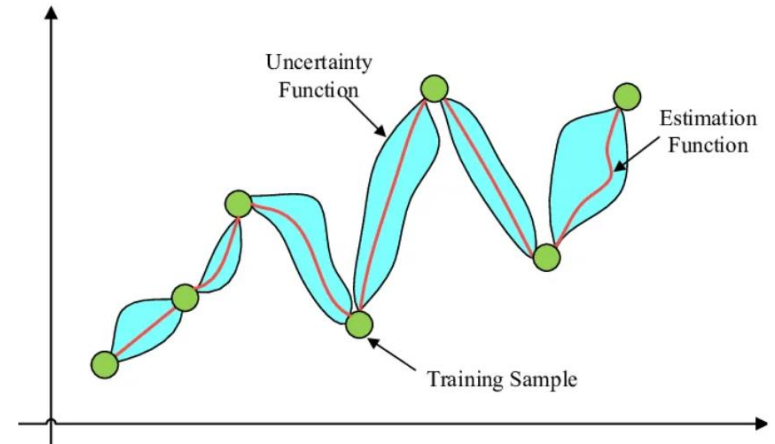
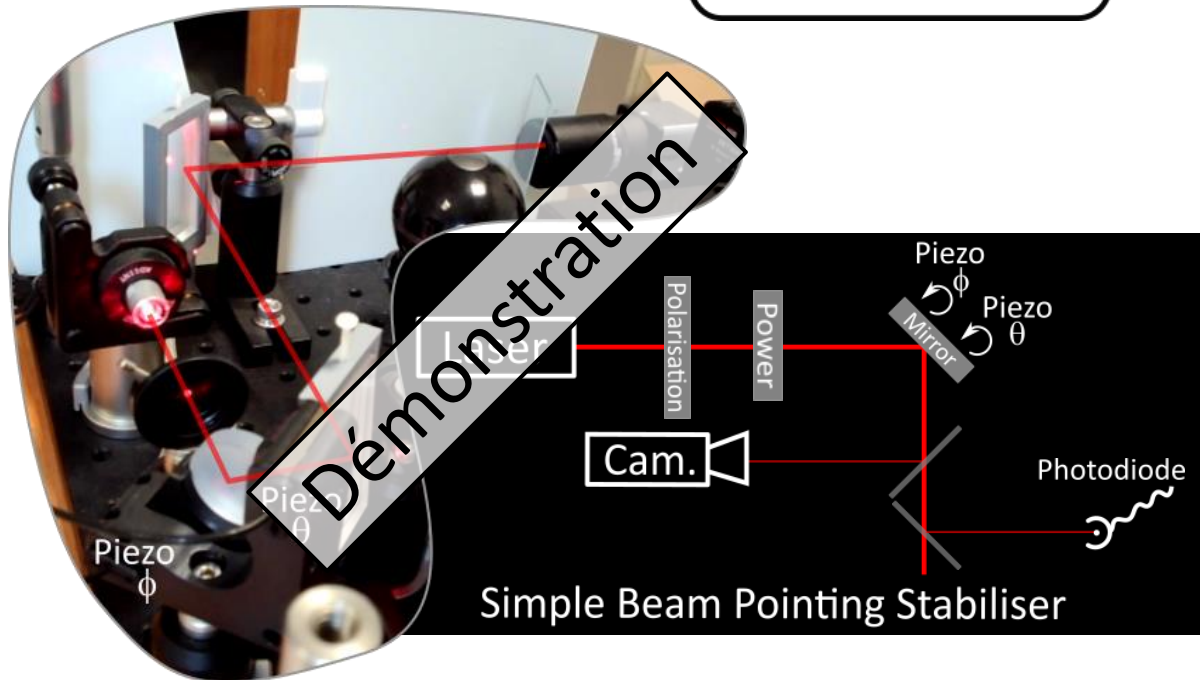
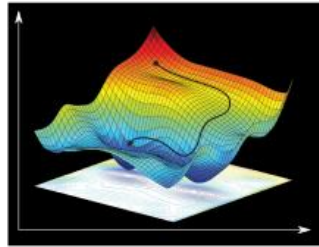
Avantages :

Algo intelligent qui apprend seul

PyMoDAQ pour l'optimisation : Algorithme Bayésien

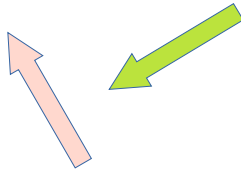
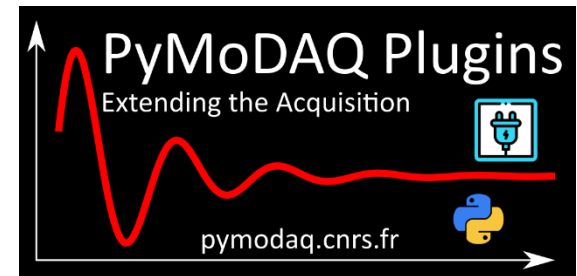
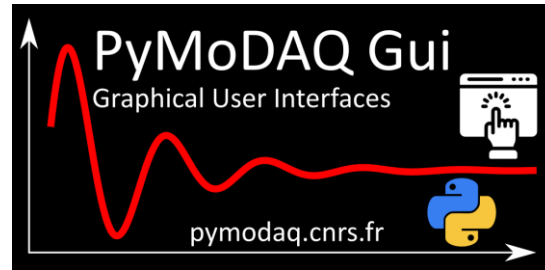
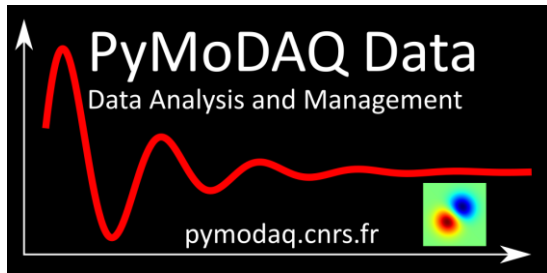
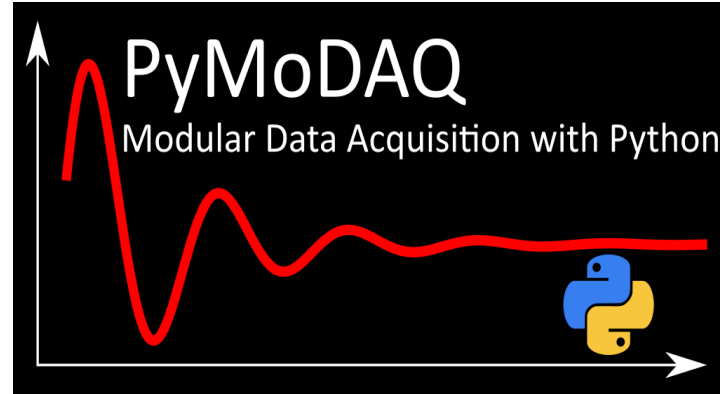
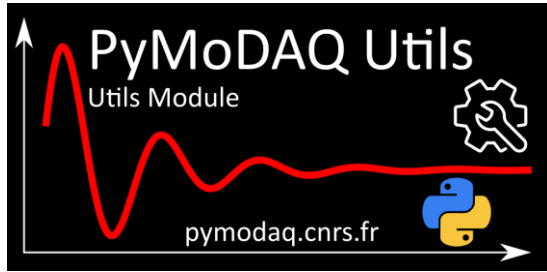
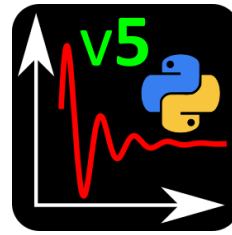


Optimisation



9

PyMoDAQ right now! Version 5 !





PyMoDAQ : Lauréat du programme OPEN 2024

- Financement d'un ingénieur CDD pour l'édition du code
- Soutien à l'organisation des Journées PyMoDAQ
- 6 Lauréats



INNOVATION

ICHEM – Criblage moléculaire

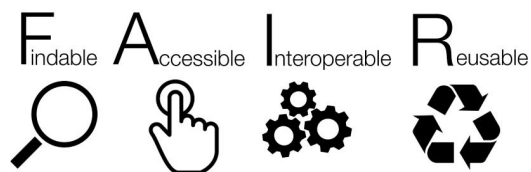
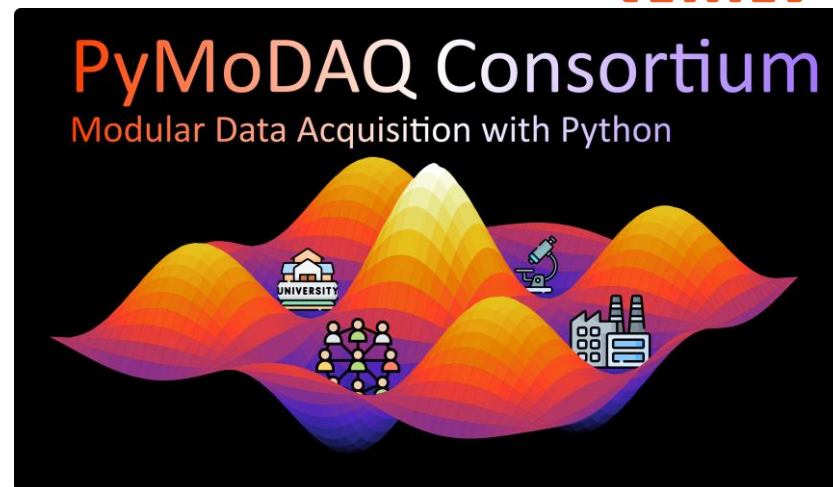
LabNBook – Support aux travaux pratiques

PyMoDAQ - Orchestration de dispositifs expérimentaux

aGrUM - Modélisation et analyse de modèles graphiques
probabilistes

TexMacsCloud - Création de documents mathématiques

RTK - Imagerie tomographique



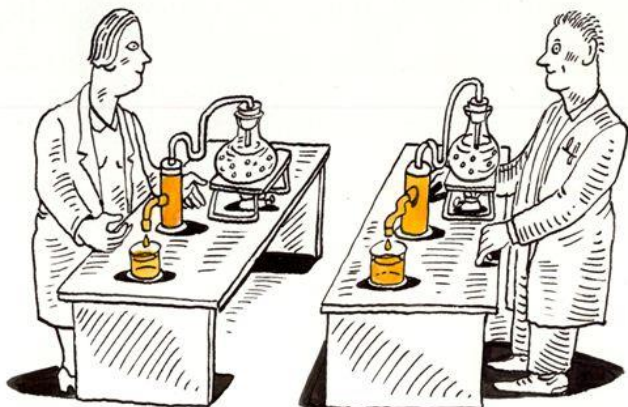
NOMAD

**Materials science data
managed and shared**

NOMAD lets you manage and share your materials science data in a way that makes it truly useful to you, your group, and the community. **Free and open source.**

Produire des données reproductibles?

11

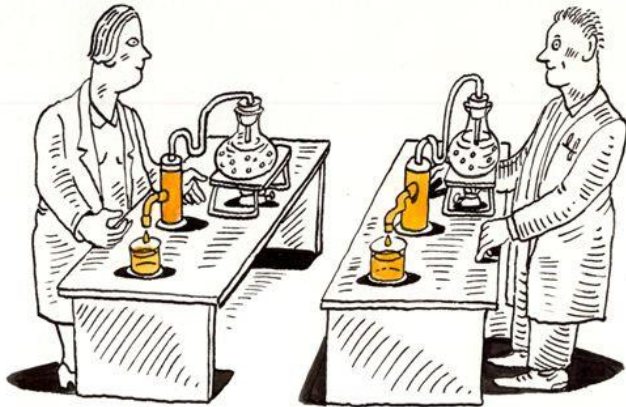


Reproducibility is *“obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.”*

For a given paper, reproducibility may be understood as *“the ability to re-create all of the figures and numbers [...] from the code and data provided by the authors”*

Replicability is *“obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data”*

For a given paper, replicability is then *“the ability to re-perform the experiments and computational analyses. . . and arrive at consistent results”*



Conditions nécessaires:



Disponibilité des données de bases: données brutes, code informatique, échantillons, ...



Disponibilité des méta-données: conditions expérimentales, paramètre des instruments (même ceux qu'on ne varie pas), protocoles, ...

Problématiques:

Faux positifs, Faux négatifs...

Fabrication des échantillons... sujet de recherche en lui-même

Données statistiques? Echantillons uniques, dispositifs expérimentaux uniques...

Absence de standardisation des techniques expérimentales et des données

Qu'est ce qui entrave la reproductibilité?

13



- Motivation de faire qqch de nouveau
- Nouveaux résultats plus faciles à publier
- Evaluations, financement et carrière: We Want Ground Breaking Research !!!

Ce qui conduit à une recherche non répliquable voire même non reproductible (même pour vous dans vos labo genre après 6 mois...) :

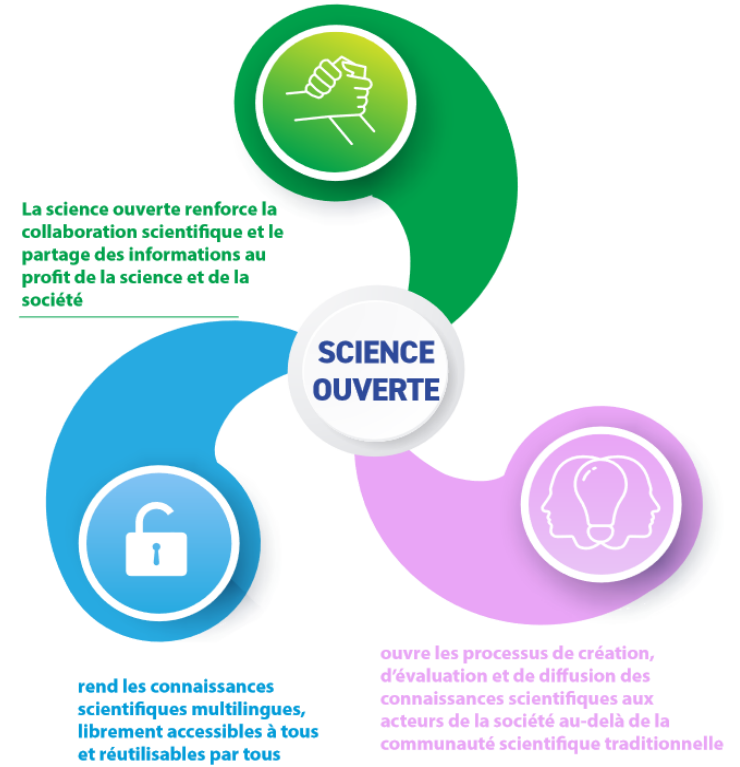
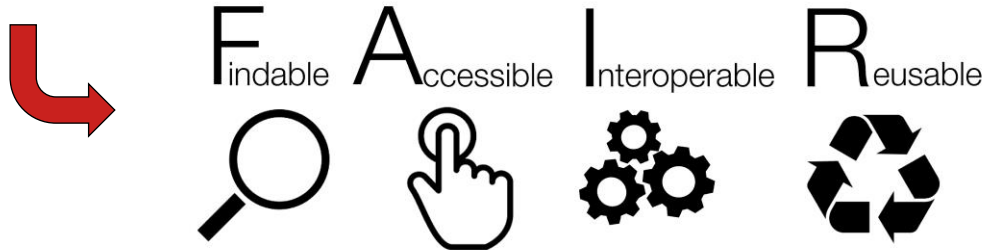
- Données brutes non disponibles ou non-accessibles ou non-lisibles
- Protocoles expérimentaux mal/peu décrits
- Logiciel d'analyse non-disponible ou obsolète
- Données publiques partielles
- Erreurs d'interprétation
- ...

Quoi faire et qui faire?



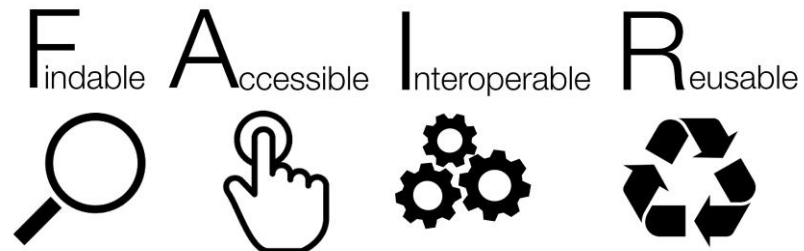
Recommandations à l'échelle:

- du chercheur
- des éditeurs et des journaux
- de l'université et organismes de recherche
- du gouvernement et des agences de financement



Recommandation de l'UNESCO sur une science ouverte

Données FAIR: c'est quoi, pourquoi et comment?



FAIR :

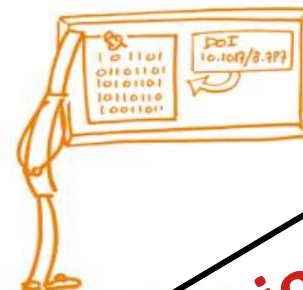
- Findable
 - ◆ Recherche rapide : thématique, mots clés
 - ◆ Identifiant unique
- Accessible
 - ◆ Accès simple
 - ◆ Contrôle d'accès
 - ◆ Utilisation de standards : identification...
- Interoperable
 - ◆ Format standardisé
 - ◆ Compatibilité avec plusieurs logiciels
- Reusable (Reproductible?!)
 - ◆ Bien documentée
 - ◆ Définition des méthodes (théorie/expérience)
 - ◆ Licence d'utilisation

FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE

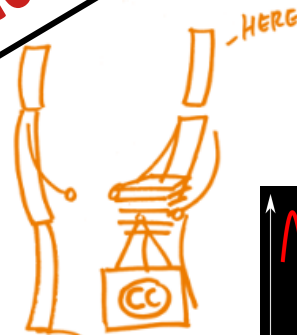
Entrepôt de données :
Quoi ?
Comment ?

Standardisation

HOW DO YOU OPEN A .XYZ FILE?

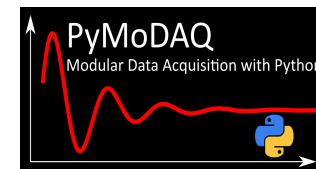


INTEROPERABLE

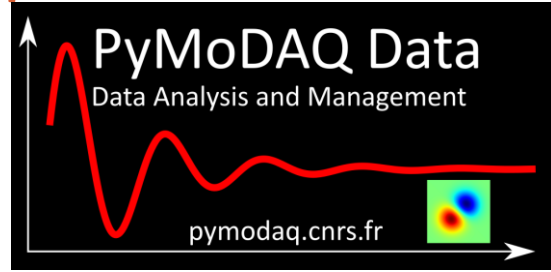


Reproductible

Logiciels?
Outils?



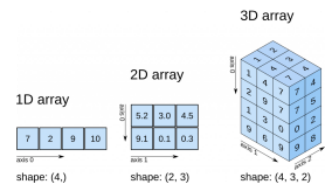
PyMoDAQ: déjà interopérable et réutilisable?



DataLowLevel: •name •timestamp

DataBase:

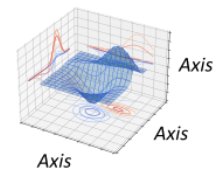
- Dimensionality: 0D, 1D, 2D, ND
- Source: Raw or calculated
- Distribution: Uniform or Spread
- Data: list of numpy arrays with same shape



DataWithAxes:

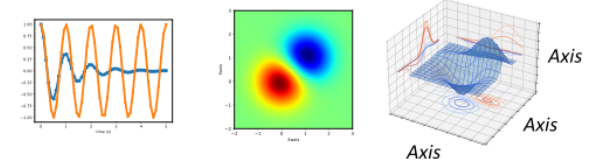
• axes: list of Axis objects

- DataRaw
- DataCalculated
- DataFromPlugins
- DataActuator



DataToExport:

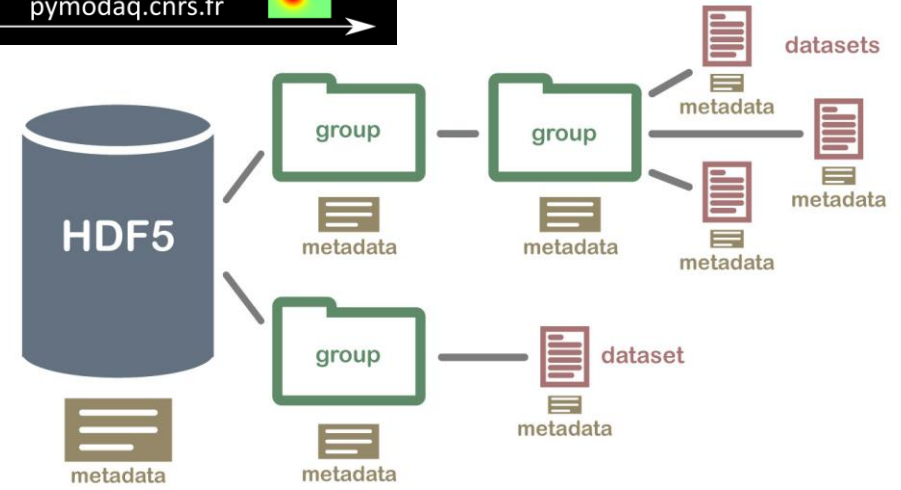
• Data: list of DataWithAxes



Axis:

•label
•units

- Data: numpy array
- index: with respect to its data shape



HDF5 : Hierarchical Data Format

- Extensible
- Interopérable



H5Browser

File ?

Node

- RawData
 - Scan012
 - Actuator000
 - Detector000
 - Data00
 - CH00
 - Data00
 - CH01
 - Data00
 - NavAxes
 - Axis00

Parameter

- daq_viewer_settings
 - Main Settings:
 - Detector Settings
 - Controller Status:
 - VISA:
 - Manufacturer:
 - Serial number:
 - Model:
 - Timeout (ms):
 - Configuration:
 - Reset:
 - Channels in separate viewer:
 - Channels:

Navigation

eur de données du noeud

Arborescence

Attributs du groupe "Detector000"

Parameter	Value
CLASS	GROUP
TITLE	Lodin
VERSION	1.0
settings	b'<All_settings type="group"><daq_viewer_se
type	detector

Représentation de l'attribut settings

FAIR DATA PRINCIPLES

AH!

FINDABLE

ACCESSIBLE

?

HERG

REUSABLE

- Mais stockage local... : pas trouvable
- Des métadonnées instruments mais variables...

ROI_00/MAG

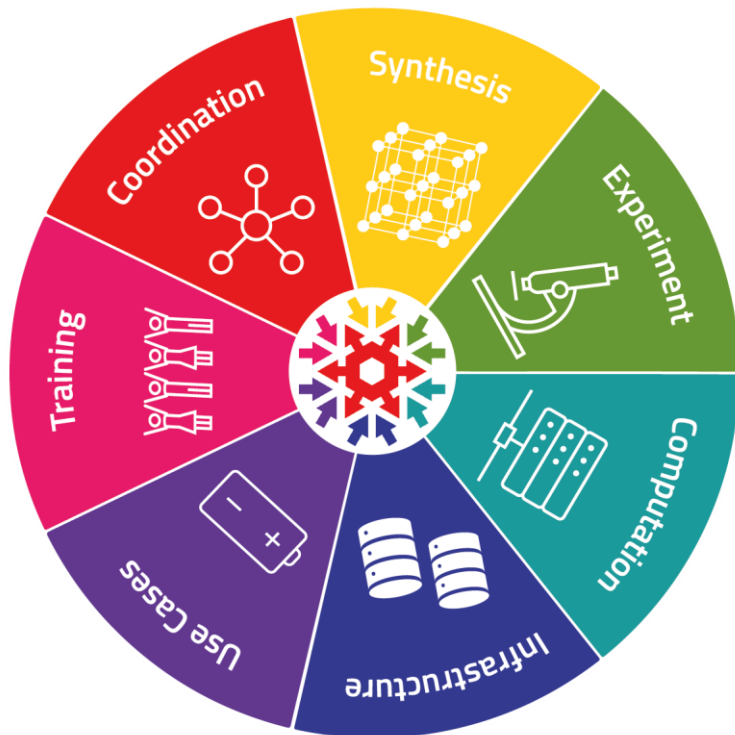
delay (pmm)

Average (x1e-09)

Samples (S)



NOMAD



FAIRmat

is building a FAIR data infrastructure for condensed-matter physics and the chemical physics of solids

Gestion, dépôt et partage des données

- Principes FAIR
- Focus sur la science des matériaux

Entrepôt possible : NOMAD



FAIR DATA PRINCIPLES

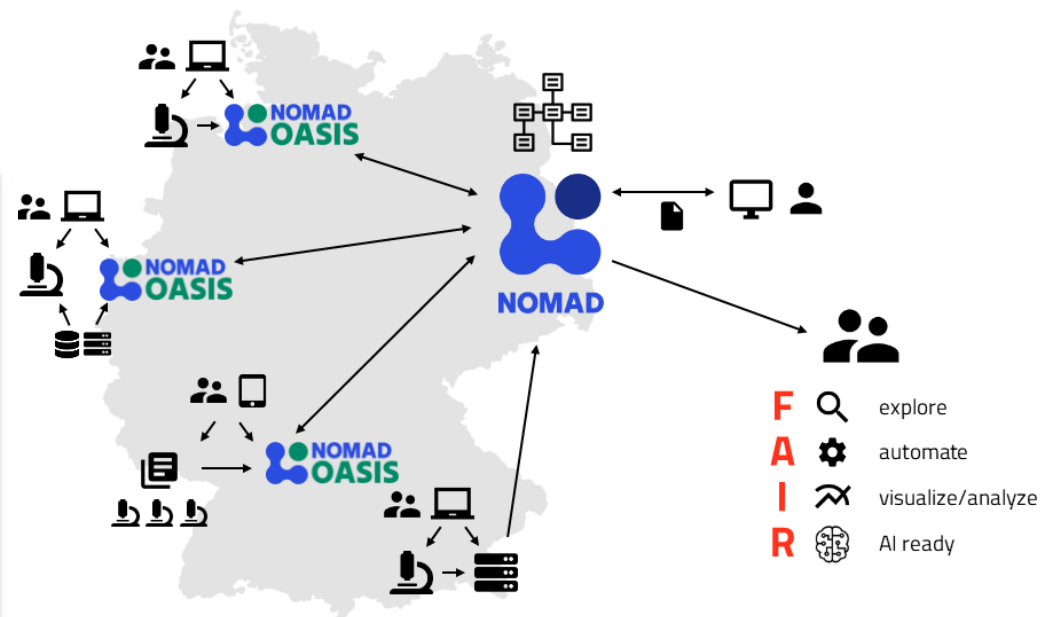
AH!



FINDABLE

The screenshots show the NOMAD web interface. The main search results page displays a list of simulation entries with columns for Name, Formula, Entry type, and Authors. The search criteria are set to 'only compositions that exclusively contain these atoms'. The interface also includes a periodic table and a sidebar with navigation options like 'HOME', 'METHODS', 'SCATTERING', and 'PERIODIC TABLE'.

Name	Formula	Entry type	Authors
H2T exciting simulation	H2S	exciting simulation	Andrés Galià, Stefan Goedecker
0T exciting simulation	0T	exciting simulation	Andrés Galià, Stefan Goedecker
F2T VASP simulation	F2O	VASP simulation	Raz Song, Christopher Chantrel et al.
F3T VASP simulation	F2O	VASP simulation	Raz Song, Christopher Chantrel et al.
W2T VASP simulation	W2O	VASP simulation	Raz Song, Christopher Chantrel et al.
T2 WENZ	Si	WENZ	Marius Schweiger
H3T WENZ	H3S	WENZ	Marius Schweiger
02T exciting simulation	02T	exciting simulation	Andrés Galià



F 🔍 explore
A ⚙️ automate
I 📊 visualize/analyze
R 🧠 AI ready

NOMAD

- Base de donnée
- Dépôts : DOI
- Plugin pour reconnaître les formats de fichier
- OASIS : serveur local

NOMAD : exemple simple de structuration des données et métadonnées



https://nomad-lab.eu/prod/v1/gui/user/uploads

PUBLISH EXPLORE ANALYZE ABOUT

Temporal Waveform

section TemporalWaveform

sub section data

QUANTITIES

author = Sébastien Weber

num points = 10

delta t = 0.00250

name = data

starting Time = 18/10/2024 10:41:57

method

ID

location

description

SUB SECTIONS

instruments

results

CH1

CH2

Waveform

section Waveform

sub section results

QUANTITIES

name = CH2

amplitude = 10 list

time = 10 list

REFERENCED BY closed

Time

quantity time

VALUE

0
$2.5 \cdot 10^{+12}$
$5 \cdot 10^{+12}$
$7.5 \cdot 10^{+12}$
$1 \cdot 10^{+13}$
$1.25 \cdot 10^{+13}$
$1.5 \cdot 10^{+13}$
$1.75 \cdot 10^{+13}$
$2 \cdot 10^{+13}$
$2.25 \cdot 10^{+13}$
(*)
fs

REFERENCED BY closed

Metadata

Instruments

TemporalWaveform

Author
Sébastien Weber

Num points
10

Delta t
0.0025

name
data

starting Time
18/10/2024 10:41:57

Method

Des données FAIR :

- Un auteur
- Reconnaissance d'un instrument (éventuel lien)
- Normalisation des données :
 - ◆ Deux channels (type Waveform)
 - ◆ Un axe temporel

18/10/2024 10:41:37

Last processing time
18/10/2024 10:41:57

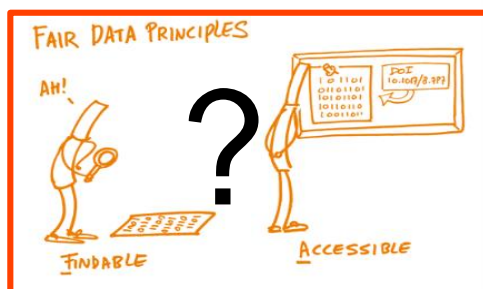
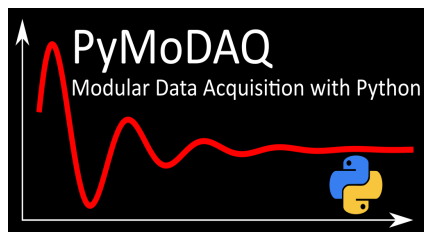


Des données FAIR, c'est :

- Une technique expérimentale
- Des ressources
- Des données



a free and open source electronic lab notebook



Started on 2023-09-22

Holography

Team Service Optique et Magnétisme

Category NOT SET

Status ● RUNNING

Visibility + Only owner
Groups

Can write + Only owner and admins

LINKED RESOURCES

- CAMERA Zelux
- LASER HeNe (HNL-S008L)
- SETUPS Holography

MAIN TEXT

Setup for phase shaping:

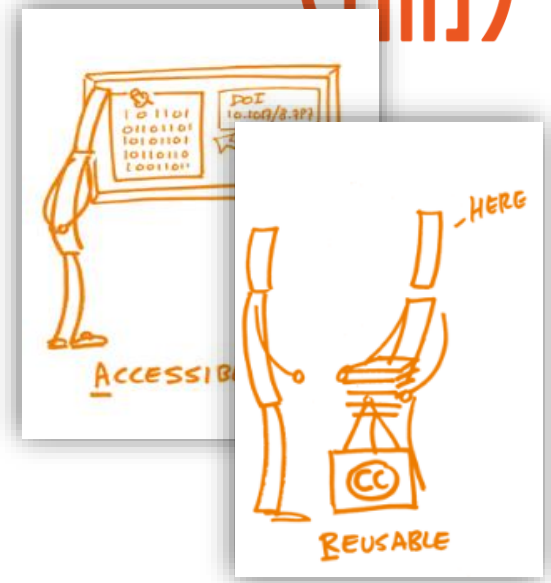
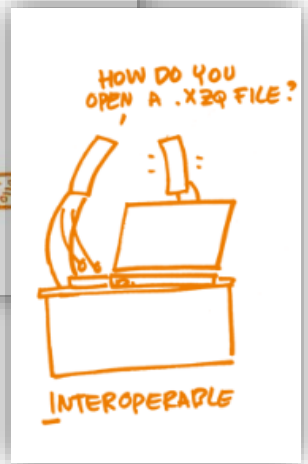
Figure 11: Experimental setup for phase modulation

From the documentation, the angles are referred with respect to the vertical axis (that is 0° blocks the horizontal polarization).
In my setup, using the HeNe laser from Thorlabs: [Laser] HeNe (HNL-S008L) so the polarization is vertical meaning $\theta_{\text{laser}}=0^\circ$

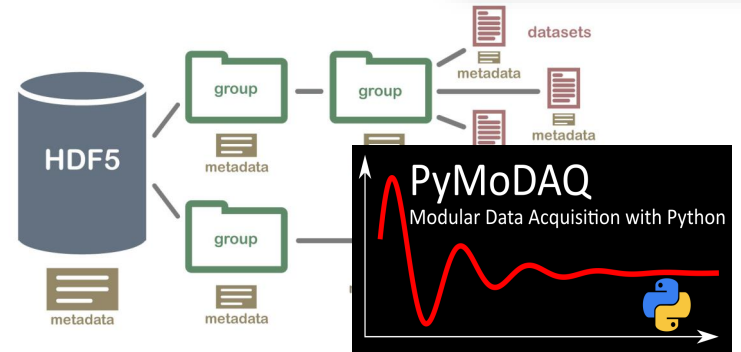
Le trio gagnant?



NOMAD




eLabFTW
a free and open source electronic lab notebook



25

Actions à venir!



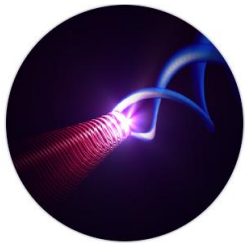
INNOVATION



CENTRE
D'ÉLABORATION
DE MATÉRIAUX
ET D'ÉTUDES
STRUCTURALES



quantumm



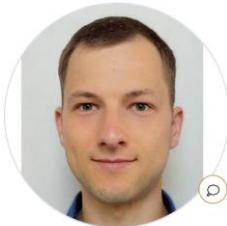
rgeneaux



Aurore Finco
aurore-finco - she/her



Jérémie Margueritat



Benedikt Burger



Paul Levasseur

PyMoDAQ
Modular Data Acquisition with Python

Actions 2025

	<p>Action Nationale de Formation</p> <p>Développer PyMoDAQ</p> <p>Du 23 au 26 juin 2025 à Toulouse Laboratoire CEMES</p>	<p>Journées PyMoDAQ</p> <p>Réalisations, tables rondes et évolutions du logiciel</p> <p>Du 8 au 10 octobre 2025 Université de Montpellier</p>
--	---	--