

OPENDATA POUR LA SIMULATION ET LES EXPÉRIMENTATIONS AU LEGI

Interopérabilité et pérennisation des données de la recherche :
comment FAIR En pratique ? Retours d'expérience

Gabriel Moreau, Antoine Mathieu, Cyrille Bonamy, Julien Chauchat, Joël Sommeria

Laboratory LEGI - CNRS / UGA / Grenoble-INP - France

27 novembre 2018 / Paris

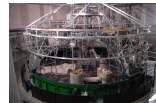
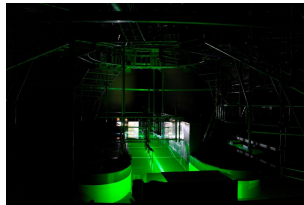
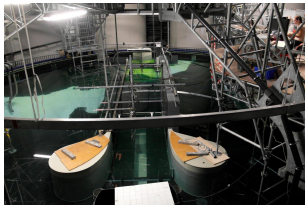


Niveaux pour le partage et la diffusion des données ouvertes

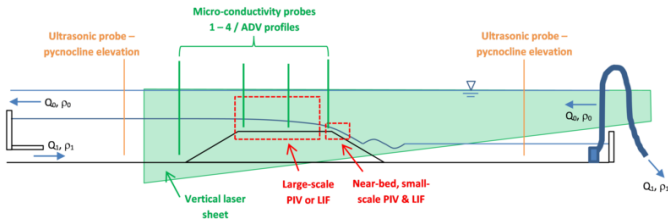
Un très bref aperçu

- ① Créer une archive de données et la publier simplement sur Zenodo
 - ② Créer un ensemble de données et le publier avec un accès distant OPeNDAP
 - ③ Partager et utiliser des scripts en ligne pour post-traiter à distance un ensemble de données (Jupyter Notebook + OPeNDAP)
 - ④ ...
- Nous allons **mettre l'accent** sur les **outils d'accès à distance** à un ensemble de données
 - Un **outil open source** projet-meta et des notebooks Jupyter ont été **développé** pour atteindre ces **objectifs**

Accès transnational à CORIOLIS - un exemple de jeux de données



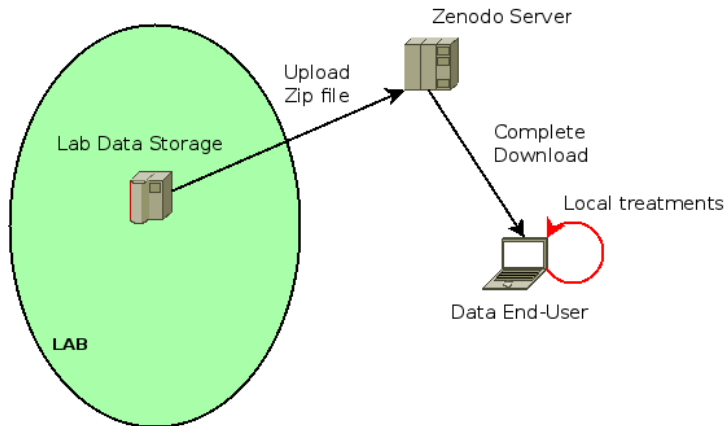
Estuarine Configuration:



Projet : 14CARR

- Année : 2014 - Meneur : Magda Carr
- Sujet : Mélange interne et dynamique en fond induit par des flux d'échange stratifiés restreints
- Équipe : 10 scient. (U. St Andrews, Heriott Watt U., Talinn U., U. Bergen, U. N. Norway, CNRS LEGI)
- Données : 1.4 To (principalement des fichiers XML, PNG et NetCDF)
- Instruments utilisés : (~10 synch.)
 - ▶ Optic. : PIV, LIF, Bathy laser scan
 - ▶ Acoust. : ADV, ADV, Limnileter

Niveau 1 : Zenodo



Pour

- Simple à utiliser
- ID unique par jeux de données (doi)

Contre

- Limité à 50 Go
- Téléchargement intégral de l'archive

Outils développés

- project-meta

Niveau 1 : Zenodo - Création d'un jeux de données libres

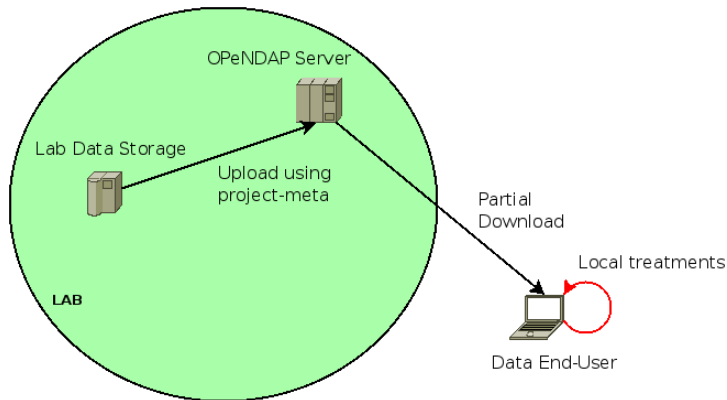
project-meta : boîte à outils en ligne de commande pour l'open-data

- Déclaration des métadonnées dans un fichier au format YAML
- Basée sur les spécifications **Dublin Core Metadata Initiative** (DCMI)
- Format descriptif simple composé de quinze propriétés relatives :
 - ▶ au contenu (titre, sujet, description, source, langue, relation, couverture)
 - ▶ à la propriété intellectuelle (créateur, contributeur, éditeur, gestion des droits)
 - ▶ à l'instanciation (date, type, format, identifiant de la ressource)
- Extension au Dublin Core pour déclarer les fichiers à mettre en données ouvertes
- Génération automatique à la racine de l'archive des fichiers : LICENSE.txt, README.txt, AUTHORS.txt, COPYRIGHT.txt...

Ligne de commande :

```
project-meta make-zip
```

Niveau 2 : OPeNDAP - Data Access Protocol



Pour

- Pas de limite en taille
- Téléchargement partiel possible (ex. sous-ensemble d'un fichier NetCDF)
- Encourage l'utilisation de format ouvert (NetCDF, HDF5... OPeNDAP backend)

Contre

- Pas d'ID unique (doi)
- Obligation d'avoir son propre serveur (pas de serveur publique)

Outils développés

- project-meta
- UVmat : compatibilité OPeNDAP

Niveau 2 : OPeNDAP - Vue web de dossier contenant des fichiers NetCDF



Contents of /coriolis/14CARR/ESTUARY3_160714/ESTUARY3b/Dalsa2.civ4.mproj

Name	Last Modified	Size	DAP Response Links	Dataset Viewers
0_EXE/	2015-06-20T23:01:38	-	- - - - - -	
0_LOG/	2015-06-20T23:01:41	-	- - - - - -	
0_OAR/	2015-06-20T23:01:40	-	- - - - - -	
0_XML/	2015-06-20T23:01:38	-	- - - - - -	
ing_1.nc	2015-06-20T23:01:43	157464	ddx dds das info html rdf	viewers
ing_10.nc	2015-06-20T23:01:46	157464	ddx dds das info html rdf	viewers
ing_100.nc	2015-06-20T23:02:22	157464	ddx dds das info html rdf	viewers
ing_101.nc	2015-06-20T23:02:22	157464	ddx dds das info html rdf	viewers
ing_102.nc	2015-06-20T23:02:22	157464	ddx dds das info html rdf	viewers
ing_103.nc	2015-06-20T23:02:23	157464	ddx dds das info html rdf	viewers
ing_104.nc	2015-06-20T23:02:23	157464	ddx dds das info html rdf	viewers
ing_105.nc	2015-06-20T23:02:24	157464	ddx dds das info html rdf	viewers
ing_106.nc	2015-06-20T23:02:24	157464	ddx dds das info html rdf	viewers
ing_107.nc	2015-06-20T23:02:24	157464	ddx dds das info html rdf	viewers
ing_108.nc	2015-06-20T23:02:25	157464	ddx dds das info html rdf	viewers
ing_109.nc	2015-06-20T23:02:25	157464	ddx dds das info html rdf	viewers
ing_11.nc	2015-06-20T23:01:47	157464	ddx dds das info html rdf	viewers
ing_110.nc	2015-06-20T23:02:25	157464	ddx dds das info html rdf	viewers

Niveau 2 : OPeNDAP - Vue web des métadonnées d'un fichier NetCDF

Dataset Information

Conventions: uvmat

nb_coord: 2.0000000000000000

nb_dim: 2.0000000000000000

Time: 10.149999999999998

Variables in this Dataset

coord_y: Array of 64 bit Reals [coord_y = 0..85]

coord_x: Array of 64 bit Reals [coord_x = 0..112]

U: Grid

Role: vector_x

U: Array of 64 bit Reals [coord_x = 0..112][coord_y = 0..85]

Role: vector_x

coord_x: Array of 64 bit Reals [coord_x = 0..112]

coord_y: Array of 64 bit Reals [coord_y = 0..85]

V: Grid

Role: vector_y

V: Array of 64 bit Reals [coord_x = 0..112][coord_y = 0..85]

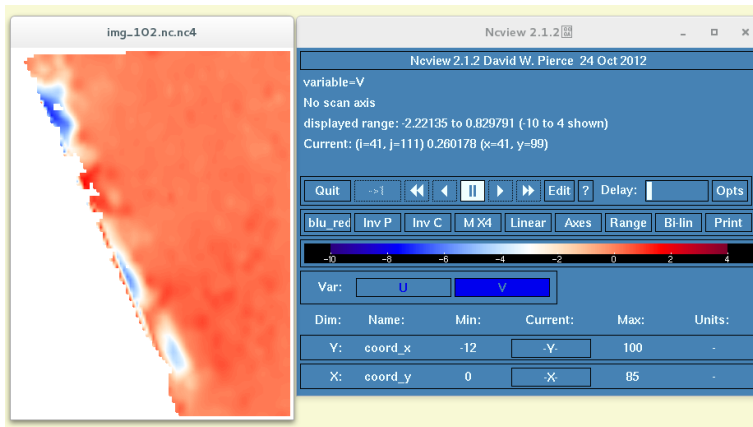
Role: vector_y

coord_x: Array of 64 bit Reals [coord_x = 0..112]

coord_y: Array of 64 bit Reals [coord_y = 0..85]

Niveau 2 : OPeNDAP - Affichage à distance de données (PIV) avec ncview

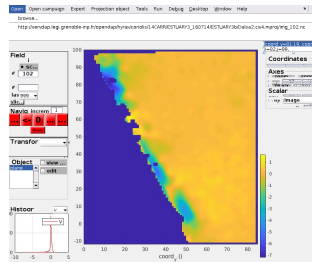
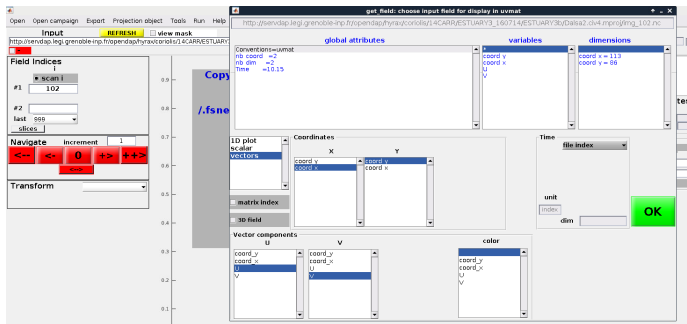
ncview - visualiseur basique NetCDF



Ligne de commande :

ncview http://servdap.legi.grenoble-inp.fr/opendap/hyrax/coriolis/14CARR/ESTUARY3_160714/ESTUARY3b/Dalsa2.civ4.mproj/img_102.nc.html

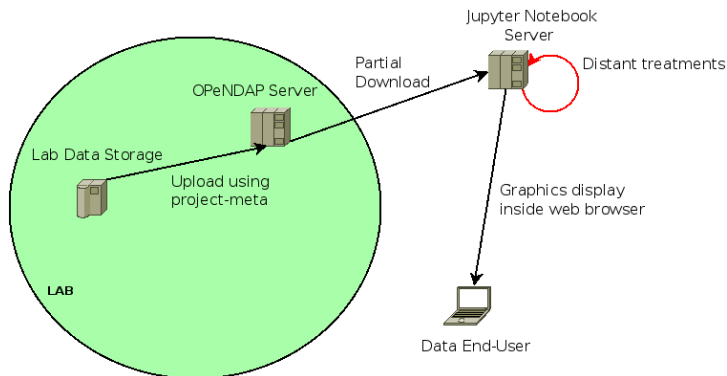
Niveau 2 : OPeNDAP - Visualisation distante (PIV) avec UVmat



UVmat : boîte à outils graphique libre Matlab pour la PIV

- Développeur principal **Joël Sommeria** (CNRS LEGI)
- **100% compatible OPeNDAP**
- Traitement local sur un petit sous ensemble de données distantes PIV sans téléchargement complet de l'archive !

Niveau 3 : Jupyter Notebook + OPeNDAP



Pour

- Pas de limite en taille, téléchargement partiel...
- Pas besoin d'installer de logiciel sur l'ordinateur de l'utilisateur final
- Utilise un langage ouvert pour le post-traitement (Python)
- Partage des scripts de post-traitement

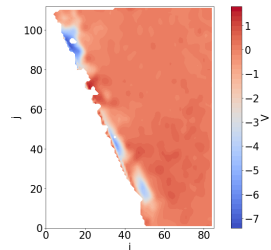
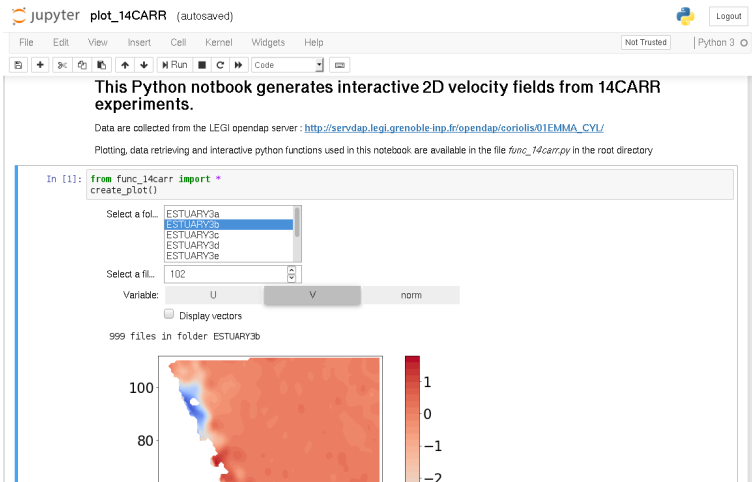
Contre

- Pas d'ID unique (doi)
- Obligation d'avoir son propre serveur (pas de serveur publique)

Outils développés

- De nombreux Python notebooks

Niveau 3 : Jupyter Notebook + OPeNDAP - Traitements distants



- Tout les post-traitements et les créations d'image sont effectuées depuis votre navigateur Web
- Résultat identique quel que soit les outils utilisés. . .

firefox <https://hub.mybinder.org/user/cyrillebonamy-notebook.opendata-sehy5hol/notebooks/plot.14CARR.ipynb>

Niveau 3 : Jupyter Notebook + OPeNDAP - Une vue rapide du code Python

```
13 def plot_func(variable, rootgrp, vector):
14     # Figure size
15     fig_xsize = 15
16     fig_ysize = 10
17     font_size = 25
18     fig = plt.figure(num=0, figsize=(fig_xsize, fig_ysize), dpi=60,
19                        facecolor='w', edgecolor='w')
20     plt.rcParams.update({'font.size': font_size})
21     plt.gca().set_aspect('equal')
22
23     #axis names
24     xaxis_name = 'i'
25     yaxis_name = 'j'
26     plt.xlabel(xaxis_name)
27     plt.ylabel(yaxis_name)
28     contourcolor = plt.cm.coolwarm
29     xlin = np.linspace(0, len(rootgrp.variables['coord_x'])-1, len(rootgrp.variables['coord_x']))
30     ylin = np.linspace(0, len(rootgrp.variables['coord_y'])-1, len(rootgrp.variables['coord_y']))
31     x, y = np.meshgrid(xlin, ylin)
32     if variable == 'norm':
33         norm = np.sqrt(rootgrp.variables['U'][:, :]**2 +
34                        rootgrp.variables['V'][:, :]**2)
35         mycontour = plt.contourf(norm, 50, cmap=contourcolor)
36         cbar = plt.colorbar(mycontour)
37         cbar.ax.set_ylabel('norm (m/s)')
38     else:
39         varplot = rootgrp.variables[variable][:, :]
40         mycontour = plt.contourf(varplot, 50, cmap=contourcolor)
41         cbar = plt.colorbar(mycontour)
42         cbar.ax.set_ylabel(rootgrp.variables[variable].name)
```

- Programmation Python dans le navigateur → assez simple
- Similaire à la programmation Matlab
- Même philosophie que les notebooks Mathematica
- Opérations sur les données et permet de les tracer (principalement 2D)

https://hub.mybinder.org/user/cyrillebonamy-notebook-opensdata-5u86dmd8/edit/func_14carr.py

Conclusion

project-meta = Gestion des données ouvertes

- Une même spécification YAML pour Zenodo et OPeNDAP
- Une grande liberté dans l'organisation des données
- Obligation de déclarer une licence, des auteurs...

Jupyter Notebook + OPeNDAP = Outils de visualisation en ligne

- Partager les données associées à votre publication
- Fournir des scripts de post-traitement pour créer des graphiques
- Rendre vos données plus lisibles par les autres et par vous même !
- Peut être utilisé et testé en tant qu'outil collaboratif avant la publication
- Plus besoin de télécharger des jeux de données complets

Merci de votre attention !

Cette présentation est sous : LICENCE ART LIBRE

<http://artlibre.org/>

