

# Atelier 2 - VRE et plateformes d'analyse

---

Intervenants :

- Jean-François Piollé
- Erwan Bodéré

L'atelier s'est découpé en 2 parties, avec une 1ère partie autour d'une présentation sur ce qu'est une VRE, qui a laissé place à une 2ème partie d'échange et de discussion.

## Présentation

VRE pour Virtual Research Environment (ou Environnement Virtuel de Recherche en français) est à l'intersection de plusieurs concepts et besoins, proposant ainsi différents services:

- Accès aux données,
- Partage d'outils et de code,
- Traitement local VS distant
- ...

Au final on peut résumer ça en définissant une VRE sous la forme:

**VRE = Données + Infrastructure + Software + Communauté**

Plusieurs défis sont soulevés lorsque l'on parle de VRE:

- Rupture avec le modèle classique, ici on déporte les codes et traitements au plus près des données.
- Volume croissant de donnée
- ...

Mais cela implique aussi des contraintes sur la découverte et l'accès aux données:

- Référencement des données disponibles dans ODATIS
- Accès direct via la VRE
- Services de sélection et subsettings
- Accès réseau local ou accès cloud
- Transformation des données en format ARCO (Analysis Ready, Cloud Optimized).
- ...

Une VRE est construite autour de plusieurs outils et services, comme par exemple:

- Des environnement scientifiques préconfiguré (Pangeo), mais aussi extensibles par l'utilisateur.
- Des frontaux d'analyse interactive (Jupyter Notebook, R-studio, ..)
- Gestion et Partage de workflows scientifique (Galaxy)
- Des outils desktop (ex: QGIS)
- Un support aux utilisateurs
- ...

Enfin une rapide présentation des infrastructures ODATIS a été faite avec TREX (CNES) et Datarmor (Ifremer). Et un exemple concret a été présenté via la VRE du CNES dans le cadre du projet SWOT.

## Discussions

Lors des discussions, de nombreux sujets ont été abordés mais peuvent être regroupés en 3 grandes questions ou thématique.

1) Est ce que les VRE sont amenés à perdurer ? La questions sous-jacente était, est ce que je dois commencer à me former sur ces outils, et commencer à former mes étudiants ?

Les VRE sont amenés à perdurer, en effet plusieurs instituts et laboratoires commencent à mettre en place des plateformes d'analyse partagée telles que JupyterHub.

Concernant les étudiants, une grande majorité d'entre eux utilise déjà ces outils durant leur formations/stages.

Cependant, il a été soulevé une nécessité d'accompagnement pour les personnes n'étant pas familière avec ces outils. Ces échanges ont donc amené à la 2ème question.

2) Quel est le niveau de maturité des utilisateurs cibles vis-à-vis des VRE ?

Les discussions ont soulevé qu'il n'est pas immédiat de s'approprier ces outils, en effet une certaine barrière à l'entrée existe et peut refroidir certains utilisateurs à changer leur pratique.

A ceci 2 points de réflexions ont été amené:

- la nécessité de cacher au maximum aux utilisateurs tous les détails techniques qui pourrait les refroidir (ex: ne pas parler de conteneurisation, HPC, parallélisation, ...)
- la nécessité d'effectuer une transition et donc d'accompagner les utilisateurs au travers du support aux utilisateurs, mais aussi de ressources en ligne tel que des formations, des tutoriels, des documentations, ...

Concernant les défis soulevés par cette transition, 2 exemples sont ressortis:

- Un défi concernant les formats de données, avec une transition entre les formats de données usuelles (NetCDF, CSV, ..) vers les formats ARCO (Zarr, Parquet)
- Un défi concernant les langages de programmation, avec une transition entre certains langages propriétaires largement répandus tel que Matlab vers des langages libres tel que Python, R ou Julia.

Dans le cas de Matlab, certains participants expliquaient travailler avec Matlab depuis 20 ans en ayant développé de nombreux outils, et étant dans l'incapacité de réécrire ces outils dans des langages libres. A celà, plusieurs propositions de solutions ont été amené:

- Utiliser les services desktops proposés par les VRE pour installer Matlab ou bien Octave
- Utiliser un noyau Octave sur jupyter pour faire tourner des notebooks écrits en Matlab.
- Compiler le code Matlab en Java pour se détacher du langage propriétaire.

Finalement il a été conclu la nécessité d'inclure et d'impliquer au maximum la communauté dans le développement de ces outils, afin de faciliter la transition avec par exemple la mise en place d'un comité de pilotage utilisateur représentatif des cas d'applications et d'utilisations finaux.

### 3) Comment accéder à des données extérieures ?

Suite à la présentation, un participant a soulevé une question importante qui est de savoir comment accéder à des données qui ne sont pas nativement fournies par la VRE ?

Une 1ère solution, serait d'identifier les données externes dont le besoin est partagé par plusieurs utilisateurs et mettre à disposition ces données en faisant une copie de ces données sur la VRE.

Une 2ème solution serait de permettre aux utilisateurs d'ajouter leurs propres données.

Cependant la meilleur solution reste encore de limiter le nombre de copies de données et donc de récupérer les données directement à la source, ceci implique de travailler sur une optimisation des accès et partages des données entre plateformes.

A ceci un participant à soulevé qu'il est nécessaire de rendre ça transparent aux utilisateurs et aussi d'accompagner les producteurs de données à déposer leurs données sur les plateformes répondant à leurs besoins.