



La reproductibilité des calculs coûteux

Konrad Hinsén

Centre de Biophysique
Moléculaire
CNRS Orléans

- Simulation de macromolécules biologiques
- Machines parallèles : 10 à 100 processeurs
- Quelques jours voir semaines par calcul
- Quelques GOs de sortie

Centre de calcul



- Ressources coûteuses, accès réglementé
- Architecture unique
- Accès réseau très limité
- Restrictions sur l'installation de logiciels

- Ressources coûteuses, accès réglementé
- Architecture unique
- Accès réseau très limité
- Restrictions sur l'installation de logiciels

Refaire ses propres calculs: **pénible**

Refaire les calculs de quelqu'un d'autre: **impossible**

Pourquoi refaire un calcul?

Pour **vérifier** la reproductibilité d'un calcul, il faut le **refaire**.

Pourquoi refaire un calcul?

Pour **vérifier** la reproductibilité d'un calcul, il faut le **refaire**.

Pour **avoir confiance** dans la reproductibilité d'un calcul, il suffit que les outils et méthodes utilisés soient **fiables**.

Pourquoi refaire un calcul?

Pour **vérifier** la reproductibilité d'un calcul, il faut le **refaire**.

Pour **avoir confiance** dans la reproductibilité d'un calcul, il suffit que les outils et méthodes utilisés soient **fiables**.

Alors... **construisons** des outils fiables qui garantissent la reproductibilité!

Pourquoi refaire un calcul?

Pour **vérifier** la reproductibilité d'un calcul, il faut le **refaire**.

Pour **avoir confiance** dans la reproductibilité d'un calcul, il suffit que les outils et méthodes utilisés soient **fiables**.

Alors... **construisons** des outils fiables qui garantissent la reproductibilité!

Faisable à cause du **déterminisme** du calcul.

Vérification: qu'est-ce qui a été fait?

- le code
- les données
- l'environnement
- toute intervention humaine

Validation: explorer la robustesse

- un autre compilateur
- un autre ordinateur
- une nouvelle version d'une bibliothèque

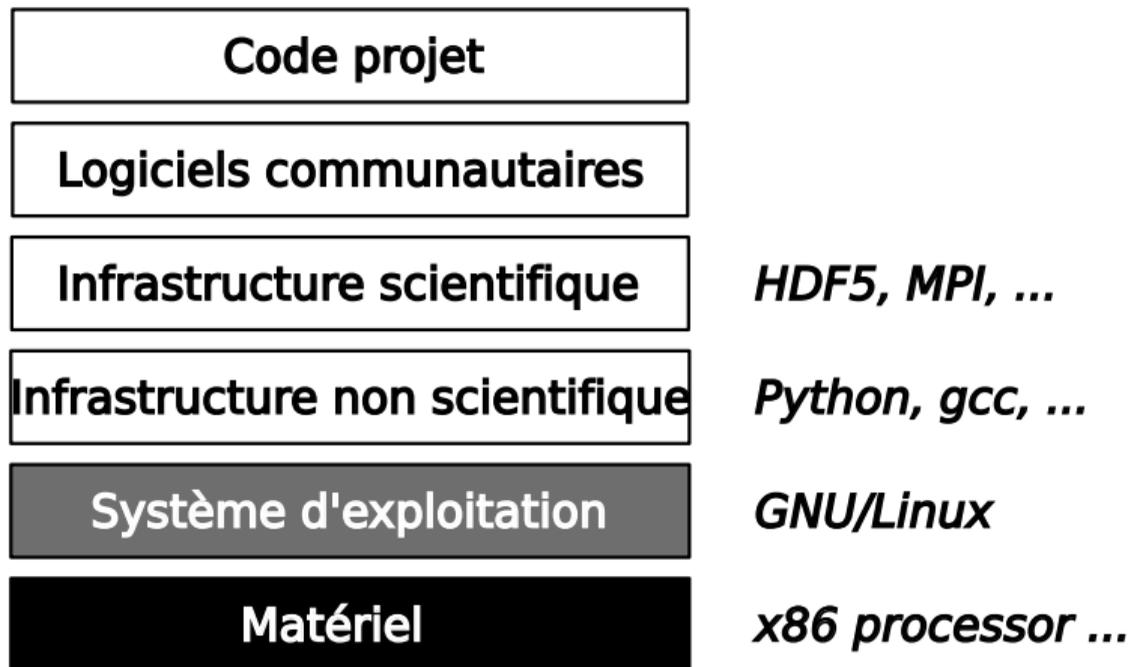
Projet de recherche démarré en 2010

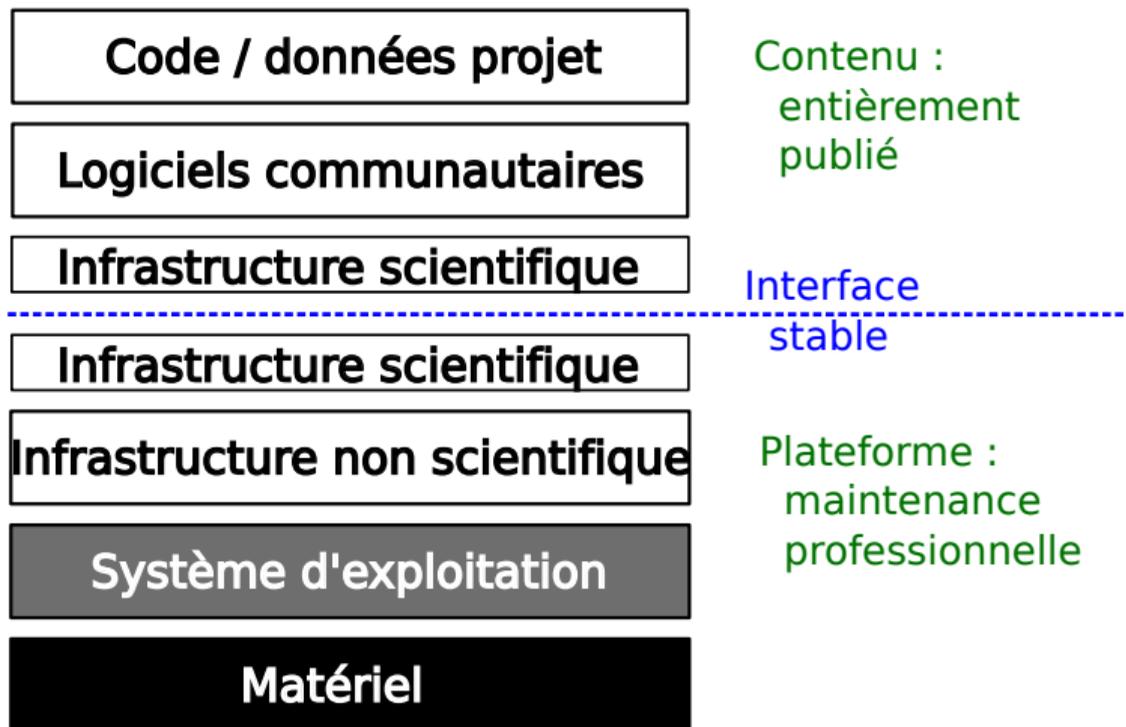
<http://www.activepapers.org/>

Deux publications:

- A data and code model for reproducible research and executable papers
ICCS 2011: [Procedia Computer Science 2011, 4:579](#)
- ActivePapers: a platform for publishing and archiving computer-aided research
[F1000Research 2015, 3:289](#)

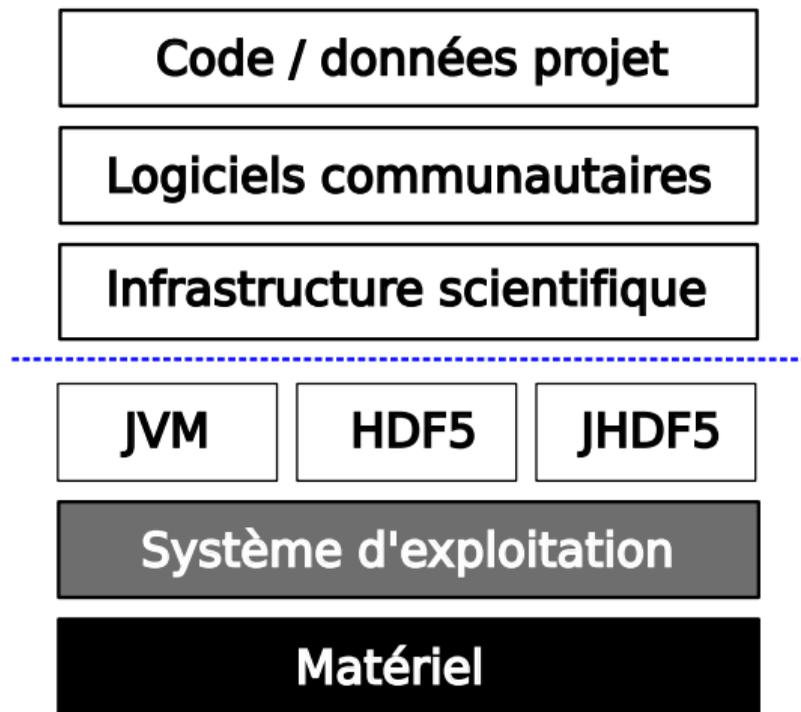
La pile logicielle en calcul scientifique

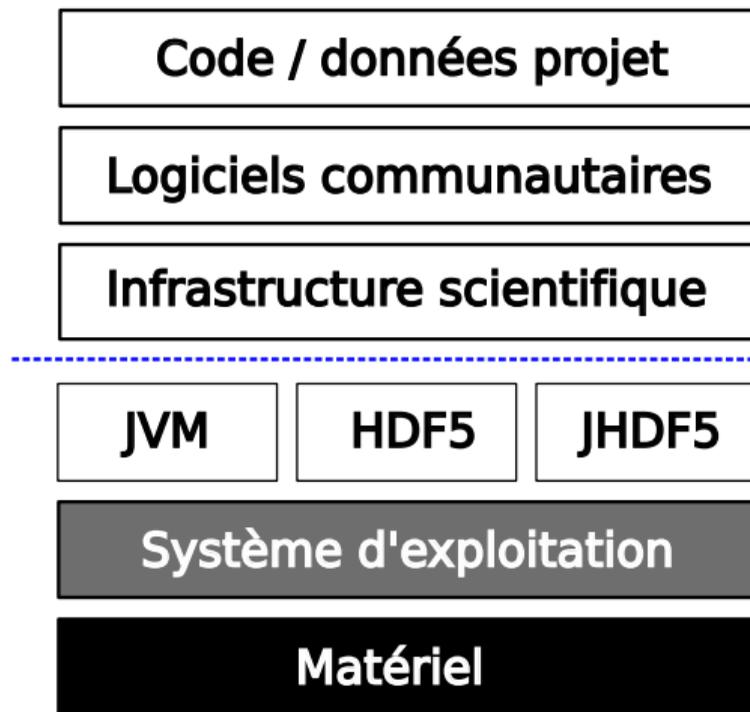




Média numériques pour la science

	Plateforme	Interface	Contenu
Article	Lecteur PDF \LaTeX Word ...	Format PDF	Fichiers PDF
Vidéo	Lecteur MP4 Caméra Montage ...	Format MP4	Fichiers MP4
Calcul	Inspecteur de données Editeur Exécution ...	Format ActivePapers	Fichiers ActivePapers





Peu de code
scientifique pour JVM

Code / données projet

Logiciels communautaires

Infrastructure scientifique

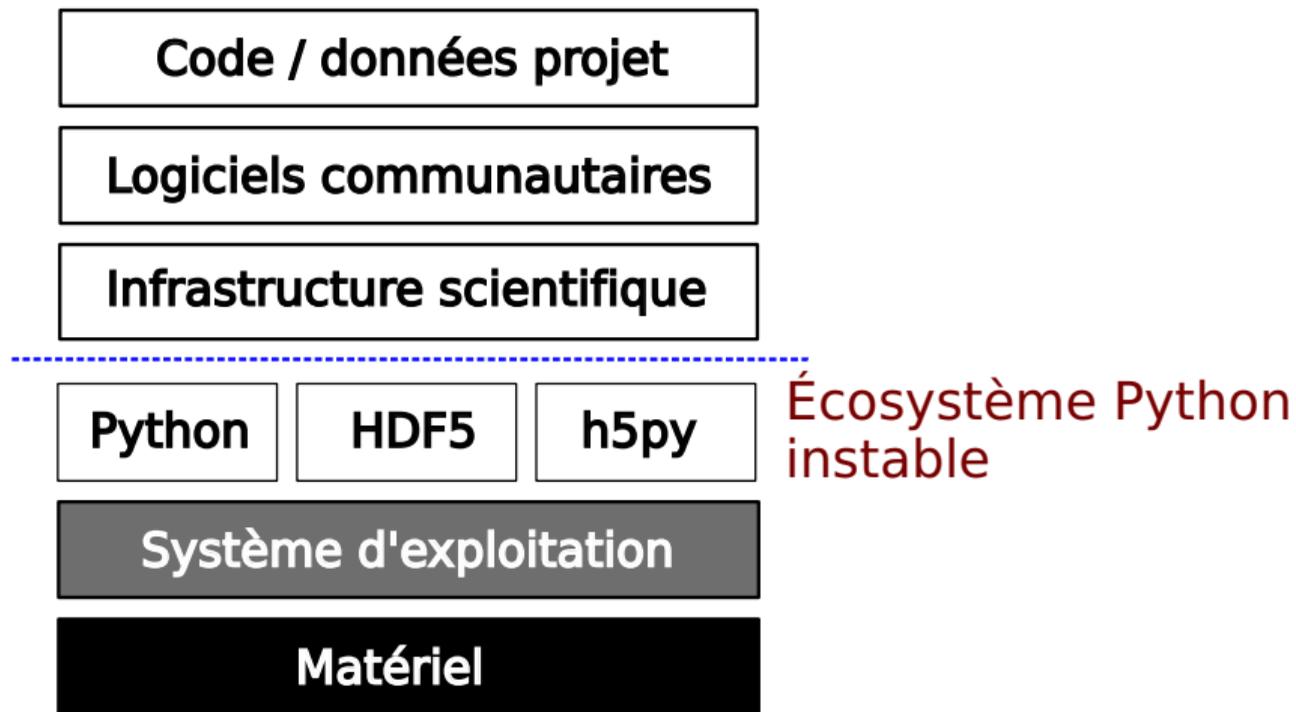
Python

HDF5

h5py

Système d'exploitation

Matériel



Code / données projet

Logiciels communautaires

Infrastructure scientifique

Guix

Linux + POSIX

Matériel

Code / données projet

À intégrer

Logiciels communautaires

Infrastructure scientifique

Guix

Linux + POSIX

Matériel

Rien n'empêche de faire du calcul parallèle déterministe ...

Le calcul parallèle

Rien n'empêche de faire du calcul parallèle déterministe ...

... sauf l'infrastructure dominante en HPC (MPI etc.)

Le calcul parallèle

Rien n'empêche de faire du calcul parallèle déterministe ...

... sauf l'infrastructure dominante en HPC (MPI etc.)

Pour l'arithmétique à virgule flottante, il restera une dépendance de l'architecture de la machine.



- 1 La reproductibilité du calcul déterministe peut être assurée automatiquement.
- 2 Une chaîne d'outils fiables réduirait largement la nécessité de refaire des calculs coûteux.
- 3 Travaux en cours, on va y arriver!
- 4 Le calcul parallèle peut être rendu déterministe, mais pour l'instant c'est difficile.