



Reproductibilité des calculs scientifiques



Marie-Josée Cros, MIAT, INRA Toulouse



CONTEXTE

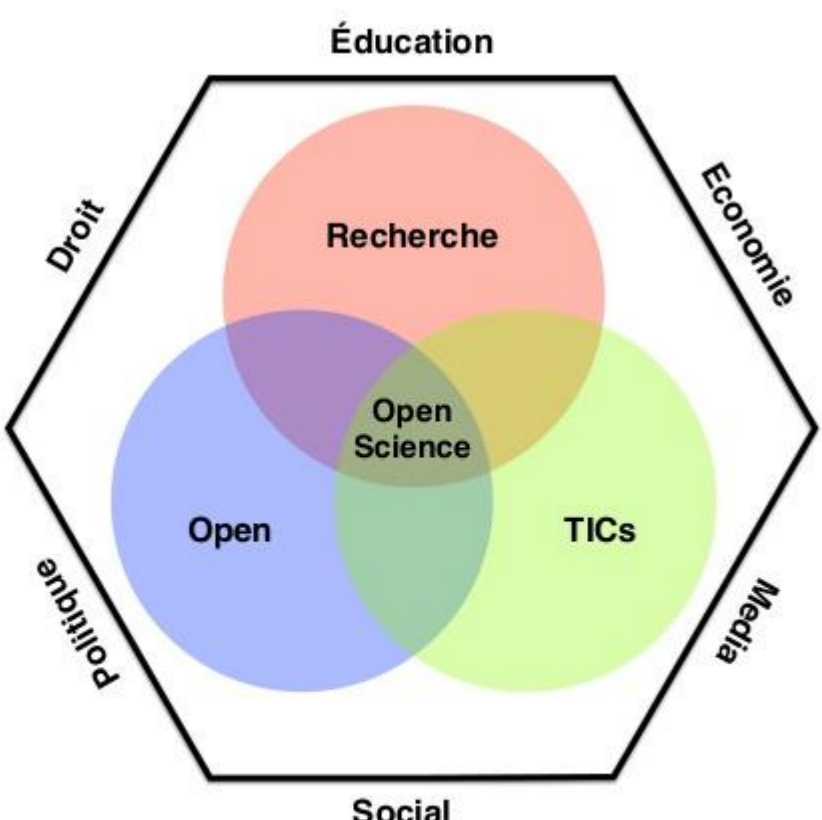


Le calcul prend une place de plus en plus importante dans la recherche. Un mouvement appelé *Recherche reproductible* se préoccupe de rendre disponible les aspects calculatoires de la recherche (code, données...) permettant de vérifier les résultats, réutiliser les codes ...

COMMENT ASSURER LA REPRODUCTIBILITE



Des mouvements pour une science reproductible, ouverte



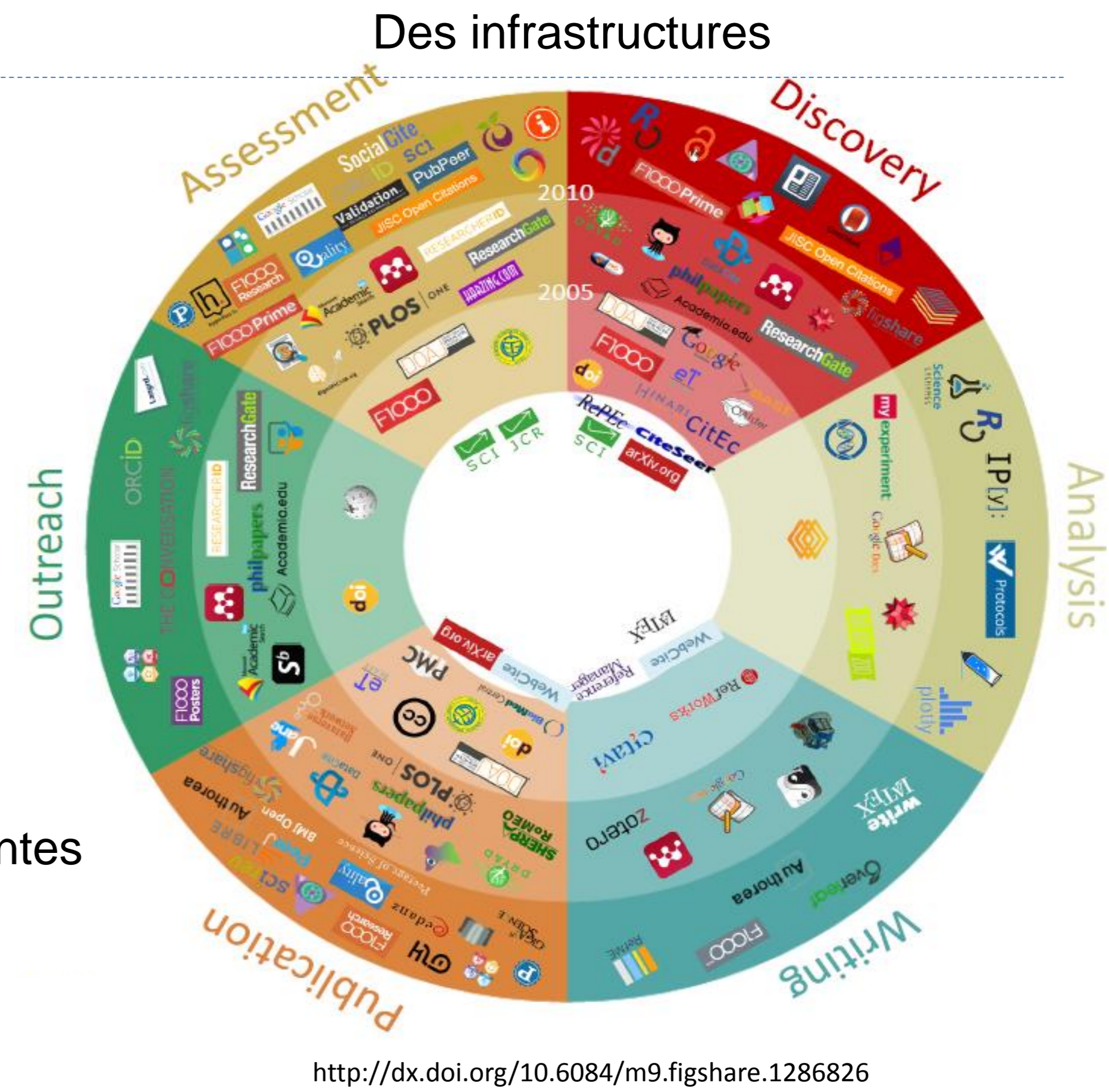
Célya Gruson-Daniel (2014) Mapping the Open Science Movement. Available on Figshare: <http://dx.doi.org/10.6084/m9.figshare.1064413>

Panton Principles
Principles for Open Data in Science
pantonprinciples.org



POLITIQUES

- Code** : disponible pour les reviewers et les lecteurs de papier
 - Copyright** : propriété du droit d'auteur et licence clairement indiqués
 - Citation** : créditer les créateurs du code utilisé ou adapté dans les publications résultantes
 - Crédit** : contributions logicielles incluses dans l'évaluation scientifique
 - Conservation** : code source disponible lors de la durée de vie utile d'une publication
- sciencecodemanifesto.org



<http://dx.doi.org/10.6084/m9.figshare.1286826>

PRATIQUES

- Ecrire les programmes pour les hommes, pas les machines
- Automatiser ou enregistrer la suite des actions réalisées
- Faire des changements incrémentaux
- Ne pas se répéter (ou répéter d'autres)
- Prévoir des erreurs
- Optimiser le code seulement lorsqu'il marche correctement
- Documenter la conception et l'objectif pas la mécanique
- Collaborer

Best practices for scientific computing [2]

Des formations



OUTILS, STANDARDS, FORMATS

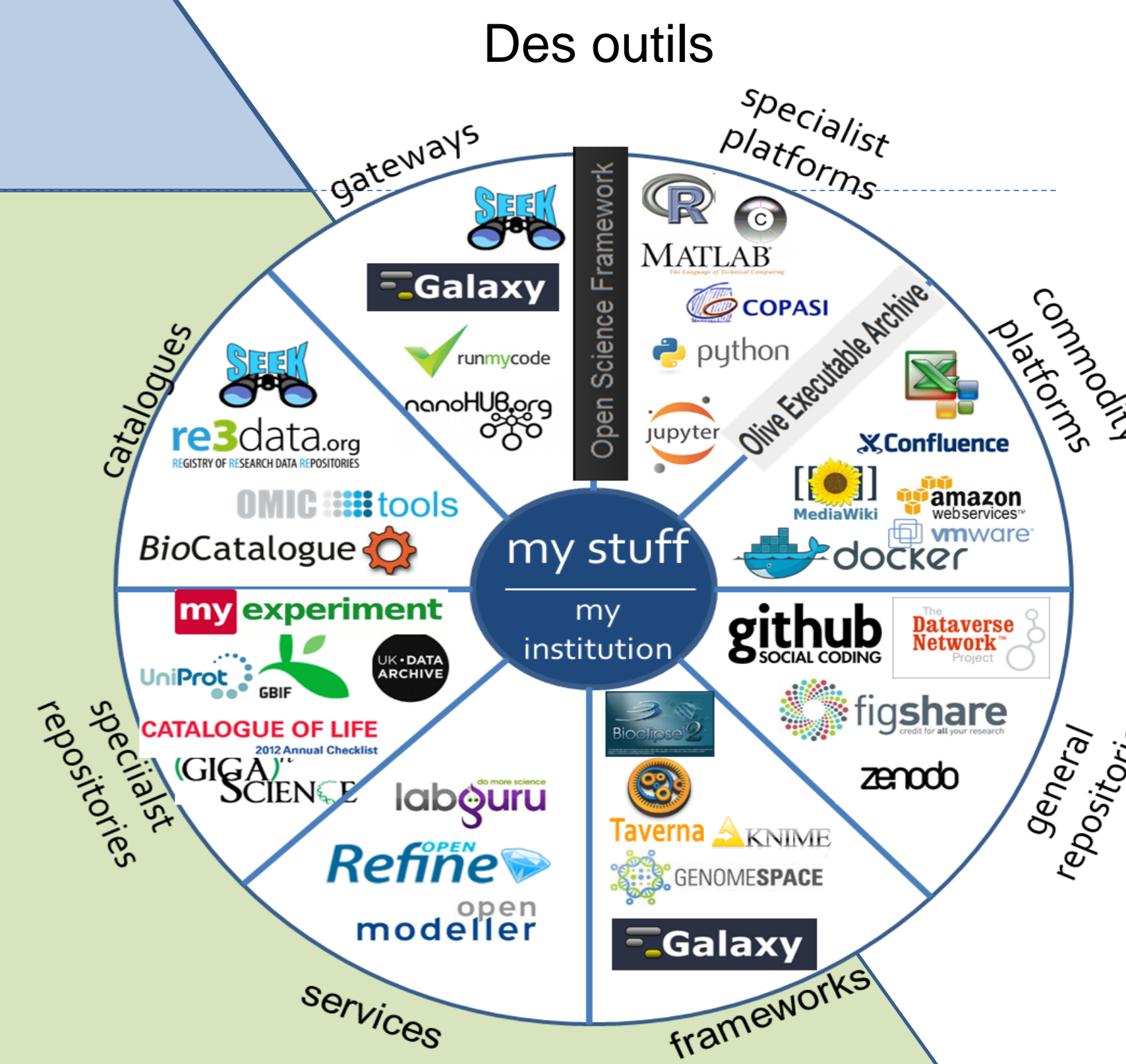
Gestionnaire de version de fichiers : Git (git-scm.com), Mercurial (mercurial.selenic.com), Bazaar (bazaar.canonical.com) et code repositories : Github.com, Bitbucket.org, Forge logiciel (sourcesup.cru.fr) ...

Programmation lettrée (*Literate Programming*) et cahier de laboratoire électronique (*Electronic lab notebook*) : Sweave (<http://leisch.userweb.mwn.de/Sweave>), knitr (<http://yihui.name/knitr>), Emacs Org mode (orgmode.org), Ipython (ipython.org), Matlab, Mathematica, Sage (sagemath.org)...

Workflow Tracking environment : VisTrails.org, Taverna.org.uk, Galaxy (galaxyproject.org), Sumatra (neuralensemble.org/sumatra/), Kepler (kepler-project.org), Pegasus (pegasus.isi.edu)...

Capture d'environnement : virtual machine, Linux package (CDE pgbovine.net/cde.html), Docker.com ...

Sites de dépôt : FigShare.com, Zenodo.org, ResearchCompendia.org, dataverse.org, Dryad (datadryad.org), RunMyCode.org, MyExperiment.org, recomputation.org, Open Science Framework (osf.io) ...



UNE DEMARCHE TANT INDIVIDUELLE QUE COLLECTIVE



Apprendre, pratiquer, s'améliorer !

- Individu** : passer à la programmation lettrée, tenir un cahier de labo électronique, gérer les versions, tester, rendre accessible données/code/processus
- Laboratoire** : encourager description intelligible / transparence / échanges
- Réseau** : partager pratiques, échanger, former, définir des standards, des formats
- Institut / financeur de recherche** : pondérer l'importance des publications, proposer des infrastructures
- Publieur de recherche** : afficher l'ouverture d'une publication, évaluer aussi la réalisation (pas seulement le résultat)

Bibliographie

[1] Reproducibility and Scientific Research. Carole Goble, Open Data Manchester, 2015. (<http://fr.slideshare.net/carolegoble/open-sciencemcrgoble2015>)
 [2] Best Practices for Scientific Computing. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, Haddock SH, Huff KD, Mitchell IM, Plumbley MD, Waugh B, White EP, Wilson P. PLoS Biology 12(1), 2014.
 [3] Ten simple rules for Reproducible Computational Research. Sandve GK, Nekrutenko A, Taylor J, Hovig E. PLoS Computational Biology 9(10), 2013.