

L'évaluation des logiciels de la recherche : CDUR - nouveaux protocoles d'évaluation

Teresa Gomez-Diaz [*], Tomas Recio [**]

[*] CNRS - Laboratoire d'informatique Gaspard-Monge

[**] Universidad de Cantabria - MATESCO

Cette œuvre est mise à disposition selon les termes de la licence
Creative Commons Attribution 4.0 International

<http://creativecommons.org/licenses/by/4.0/>

Séminaire LIGM, 5 novembre 2019



Plan

1 Contexte

2 Objectif de cette présentation

- Protocole CDUR

3 Les concepts

- Logiciel de la recherche (RS)
- Auteur d'un logiciel de la recherche
- Publication d'un logiciel de la recherche (RS papers)
- Référence et citation

4 Sur l'évaluation

- Contextes d'évaluation
- Deux méthodes d'évaluation

5 Protocole(s) CDUR

6 Conclusions

Contexte (1/2) : Logiciels LIGM

- Évaluation HCERES (2013-18) au LIGM (février 2019)
- Mission Logiciels LIGM (juin 2006) :
 - étudier la mise en place de services (suivi de versions, publication, etc.) associés au développement logiciel au sein du laboratoire,*
 - avec pour **objectif de favoriser leur visibilité***
- Production importante de logiciels : **66** listés pour 2013-18
- Nombre important de logiciels diffusés sous licence libre : **50**/66

Question : comment sont pris en compte les logiciels de la recherche dans l'évaluation : d'un laboratoire, d'un chercheur, d'un candidat... ?

Contexte (2/2) : L'évaluation de la science et la Science ouverte

Question : Pourquoi les évolutions sont si lentes ?

- Déclaration de Budapest (BOAI, 2002), définition d'*accès libre*
<http://www.budapestopenaccessinitiative.org/>
- 2 Conférences en France avec présentations J.C. Guédon :
 - ▶ Journées Open Access et évaluation de la recherche (Toulouse, 10/2016) <https://openeval2016.sciencesconf.org/>
 - ▶ Carrefour de l'IST (Nancy, 03/2017) <https://carist.sciencesconf.org/>
- 2 Rapports de groupes d'experts pour l'EC :
 - ▶ [9] (2017) Cabello Valdes C, Rentier B, et al. : Evaluation of research careers fully acknowledging Open Science practices. Rewards, incentives and/or recognition for researchers practicing Open Science.
 - ▶ [11] (2019) Guédon JC, Jubb M, et al. : Future of Scholarly Publishing and Scholarly Communication. Report of the Expert Group to the EC.
 - ▶ [9, 11] Software is now included in Open Science best practices.

The conclusion is actually simple : the evaluation of research is the keystone [11].

Objectifs de la présentation (1/2)

Après l'étude de la situation, on constate que la production de logiciels n'est pas correctement prise en compte dans l'évaluation de la recherche, pourtant cette production peut prendre une place importante.

Il est nécessaire de **faire évoluer les pratiques d'évaluation**, et cela dans le cadre des évolutions de la Science ouverte.

Gomez-Diaz T. and Recio T.,

On the evaluation of research software : the CDUR procedure

[version 1 ; peer review : 1 approved, 1 approved with reservations].

F1000Research 2019, 8 :1353, <https://doi.org/10.12688/f1000research.19994.1>

Revue F1000Research, avec *open peer review* + section *Science Policy Research*.

Version 1 publiée le 5 août 2019, V2 en cours.

+1 an d'écriture, double expertise, 75 références (80 en V2), 73 n. bas de page (85).

[49, V2] (2019) Druskat S. Software and their Dependencies in in Research Citation Graphs

Research policy researchers

Note1 : RS veut dire *research software* o logiciel de la recherche

Note2 : [nb] indique la référence dans l'article (en V1) avec ce *nb*.

Objectifs de la présentation (2/2)

L'article propose le protocole **CDUR** pour prendre en compte la production des logiciels de la recherche dans l'évaluation.
Il est flexible pour s'adapter aux différents contextes d'évaluation, tout au long de la vie scientifique.

Il y en a quatre étapes :

- (C) Citation** mesure la bonne identification du RS en tant que production de la recherche, implique nom, auteurs, version, dates... informations bien établies
- (D) Dissemination** toute diffusion a ses objectifs et un public cible, ce point regarde les bonnes pratiques suivies
- (U) Use** point dédié aux aspects purement "logiciel" **du RS** : produit des résultats corrects, facilite la réutilisation
- (R) Research** point dédié aux aspects purement "recherche" : qualité du travail scientifique, publications associées au RS, impact

Plan

1 Contexte

2 Objectif de cette présentation

3 Les concepts

- Logiciel de la recherche (RS)
- Auteur d'un logiciel de la recherche
- Publication d'un logiciel de la recherche (RS papers)
- Référence et citation

4 Sur l'évaluation

5 Protocole(s) CDUR

6 Conclusions

Concept : logiciel de la recherche (RS) (1/2)

- [2] (1994) Partha D, David PA : Toward a new economics of science

*there may be **important positive spillovers** across projects in the form of “learning effects” [...] including the development of computer software for performing data processing, storage, retrieval and network transmission.*

- [12] (2011) Kelly D : An Analysis of Process Characteristics for Dev. Scientific Soft.

Scientific software is defined by (1) it is developed to answer a scientific question ; (2) it relies on the close involvement of an scientific expert ; and (3) it provides data to be examined by the person who will answer that question ...

- [16] (2011) TGD : Article vs. Logiciel : questions juridiques et de politique scientifique...
(2009) TGD : Guide laboratoire pour recenser ses développements logiciels (PLUME)

logiciel du laboratoire tout programme utile pour faire avancer la recherche, qui a été produit avec la participation d'un membre du laboratoire. Il arrive souvent que des publications de recherche soient associées.

- [14] (2012) Sletholt MT, Hannay JE, et al. : What Do We Know about Scientific Software Development's Agile Practices ?

software developed by scientists for scientists

- [15] (2016) Hettrick S : Research Software Sustainability

Research software is developed within academia and used for the purposes of research : to generate, process and analyse results. This includes a broad range of software, programs written by researchers for their own use.

- [10] (2018) NASA Committee : Open Source Software Policy Options for NASA Earth and Space Sciences

Research software – that is, the software that researchers develop to aid their science...

Concept : logiciel de la recherche (RS) (2/2)

- [12] (2011) Kelly D : An Analysis of Process Characteristics for Dev. Scientific Soft.
 - ▶ **exclut** ce qui peut être inclus dans d'autres définitions :
[...] control software whose main functioning involves the interaction with other software and hardware ; user interface software [...]; and any generalized tool that scientists may use in support of developing and executing their software, but does not of itself answer a scientific question.
 - ▶ l'importance de l'**exactitude** :
If the software gives the wrong answer, all other qualities become irrelevant
- [16] (2011) TGD : Article vs. Logiciel : questions juridiques et de politique scientifique...
les définitions ne tiennent pas compte de **l'état du RS** :
"en projet", "fini", diffusé, qualité, portée, taille, documenté, maintenu, utilisé uniquement par une équipe pour réaliser une publication, ou dans plusieurs laboratoires

Conclusion de la définition de RS :

- ce qui est fait : code (ensemble de fichiers, bien identifié),
- qui le fait : auteur(s), mais aussi contributeur ou expert scientifique,
- pourquoi faire : recherche, science, ie. article(s) associés,
- **important** : qualité des résultats scientifiques produits.

Concept : auteur d'un logiciel de la recherche

Que veut dire auteur d'un RS ?

[16] (2011) TGD : Article vs. Logiciel : questions juridiques et de politique scientifique...

- concept légal : écriture du code

l'auteur écrit le code

- concept scientifique : apport d'expertise

sans l'expert scientifique le code n'existerait pas

- il peut y avoir d'autres contributions

documentation, correction de bugs, test, maintenance...

Définition d'auteur d'un RS :

- selection de 3 rôles dans l'article (limites peuvent être floues) :
 - ▶ (i) responsable du RS,
 - ▶ (ii) contributeur principal ou important (écriture du code),
 - ▶ (iii) contributeur mineur (écriture du code ou autre participation).

Les personnes qui n'écrivent pas du code peuvent avoir un % de participation décidé par l'équipe.

Concept : publication d'un RS (RS papers)

L'article étudie certaines publications dédiées aux RS, présentent les logiciels, avec des procédures de *software peer review*.

- Journal of Open Research Software (JORS)
- The Journal of Open Source Software (JOSS)
- Research Ideas and Outcomes (RIO)
- Software Impacts
- SoftwareX
- et aussi : (2010) Image Processing On Line Journal (IPOL)

Voir la liste de N. Chue Hong, Software Sustainability Institute (SSI)

<https://www.software.ac.uk/resources/guides/which-journals-should-i-publish-my-software>

En France, Projet PLUME (2006-2013) <https://projet-plume.org/> :

- publication de fiches RELIER (description courte de RS), avec lien vers les publications associées, 358 RS (fr), 116 RS (en)
- publication de fiches de *logiciel validé par la communauté de l'ESR au sens de PLUME* et mis en production sur au moins 3 sites, stats : 96 RS sur 406
- pas de *software peer review*
- classification thématique, mots clés, interface de recherche

Concepts : référence et citation

[34] (2013) Pontille D, Torny D : La manufacture de l'évaluation scientifique ...

[...] la différence entre référence et citation : l'acte de référence relève d'un auteur donné alors que la citation est une nouvelle propriété, éventuellement calculable, du texte source. Selon P. Wouters (1999), ce renversement a radicalement modifié les pratiques de référencement et littéralement créé une nouvelle "culture de la citation"

Une référence **fixe** nom, auteur, date, et identifie le RS tant qu'object scientifique.

L'article retient 3 types de référence :

- RS paper, avec *software peer review*,
- un article de recherche "classique" qui décrit le logiciel,
- une "référence" : auteurs(s), nom du RS, description courte, version, date, url.

À noter que :

- Il peut avoir plusieurs références associées à un RS.
- Identifications plus complètes : metadonnées, CITATION files...
- Software Citation Group, Software Citation Implementation Working Group.

Plan

- 1 Contexte
- 2 Objectif de cette présentation
- 3 Les concepts
- 4 Sur l'évaluation**
 - Contextes d'évaluation
 - Deux méthodes d'évaluation
- 5 Protocole(s) CDUR
- 6 Conclusions

Contextes d'évaluation

Les différents contextes d'évaluation :

- thèse, recrutement, évolution de carrière
- articles, publications, *peer review*
- participation à conférences, workshops, (sélection)
- financement de projets : demande, étapes, fin de projet
- établir des réseaux de collaboration, contexte international

En général, la première évaluation qui entre en jeu est l'**autoévaluation**.

Toute diffusion à des objectifs, un public cible :

- ce résultat sera dans un preprint ou dans une publication ?
- ce projet sera-t-il financé ?
- ces chercheurs voudront participer à ce projet, à cet article ?
- les décisions évoluent dans le temps... et peuvent évoluer aussi selon la prochaine évaluation en vue (par ex. j'ai besoin de + d'articles...)

Deux méthodes d'évaluation

En gros, il y en a deux méthodes d'évaluation :

[8] (2016) Mårtensson P, Fors U, et al. : Evaluating research : A multidisciplinary approach to assessing research practice and quality (62 réfs.)

- étudier la qualité : quels critères ?
- indicateurs et métriques : lesquels ? [11, 34, 45, 46, 47, 48]
 - ▶ l'utilisation du facteur d'impact des revues
 - ▶ l'utilisation de métriques en général
- facteur social

[49] (1999) Martin U : Computers, Reasoning and Mathematical Practice
[...] the community's "social knowledge" :

the methods of checking the proof are social rather than formal.

Voir les "EC Expert reports" [9, 11] :

- ils donnent plein de recommandations
- il faut tenir compte de la production de RS
- [9] Open Science Career Assessment Matrix (OS-CAM)
- [11] comment établir les comités d'évaluation (par ex.)

Plan

- 1 Contexte
- 2 Objectif de cette présentation
- 3 Les concepts
- 4 Sur l'évaluation
- 5 Protocole(s) CDUR**
- 6 Conclusions

Protocole(s) CDUR

Conçu pour aider les chercheurs évalués, les comités d'évaluation, les décideurs.

- (C) Citation** mesure si RS bien identifié en tant que produit de la recherche, référence, ou bien metadataset, aussi bonnes pratiques pour citer les autres RS ou travaux ;
point légal : auteurs, affiliations, participation
- (D) Dissemination** les bonnes pratiques de diffusion suivies en fonction de la politique scientifique du contexte d'évaluation ;
[29] (2014) TGD : Free software, Open source software, licenses...
point Science ouverte, légal : licences
- (U) Use** point dédié aux aspects purement "logiciel" **du RS** : produit des résultats corrects, facilite la réutilisation ; peut regarder aussi bonnes pratiques logiciel : exemples, docs., test, facilite install., voir le code, lancer le RS, normes de qualité, dépôt APP, valorisation, startup...
point reproductibilité, validation des résultats obtenus
- (R) Research** point dédié aux aspects purement "recherche" : qualité du travail scientifique, algorithmes proposés et programmés, publications, utilisateurs, collaborations et projets associées au RS...
point impact de la recherche

Flexibilité d'application : chaque ensemble décideurs/comité d'évaluation établi **sa propre procédure**, adaptée aux objectifs fixés et au contexte d'évaluation.

Conclusions

Ouvrir le débat, est-ce que ce type de procédure sera adoptée ?

Si on veut réussir à faire évoluer les pratiques en Science ouverte, RS...

il est nécessaire de faire évoluer les méthodes d'évaluation pour en tenir compte.

Conséquences attendues :

en accord avec [11] (2019) Guédon JC, Jubb M, et al. : *world brain vision*

- maximize [RS] accessibility and usability,
- support and expand range of contributions (equity, diversity, inclusivity criteria),
- support community building, and
- promote high-quality research with heightened integrity

Mais aussi :

Pour un labo : liste de RS fait partie de la production scientifique après évaluation.

Augmenter la transparence dans les méthodes d'évaluation de RS.

[4] (2016) Howison J, Bullard J : Software in the scientific literature (p.15) :

Clearly, a policy is only as good as its enforcement.