HAL open science

Python pour la Science Ouverte : le Baromètre de la Science Ouverte du laboratoire AAU comme cas pratique

Laurence Bizien, Françoise Acquier

To cite this version:

Laurence Bizien, Françoise Acquier. Python pour la Science Ouverte : le Baromètre de la Science Ouverte du laboratoire AAU comme cas pratique. 2022, 10.58079/qsmd . hal-03659327

HAL Id: hal-03659327

https://hal.science/hal-03659327v1

Submitted on 4 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Python pour la Science Ouverte : le Baromètre de la Science Ouverte du laboratoire AAU comme cas pratique

---

Pour citer ce billet : Bizien, Laurence et Acquier, Françoise. « Python pour la Science Ouverte : le Baromètre de la Science Ouverte du laboratoire AAU comme cas pratique ». *Le Cresson Veille et Recherche*… mis en ligne le 28 avril 2022. Disponible sur https://lcv.hypotheses.org/17446

Démarche d'analyse de la production scientifique du laboratoire

La participation à la rédaction du rapport d'auto-évaluation demandé au laboratoire dans le cadre du bilan HCERES 2015/2020 (mai 2021) nous avait amenées à mesurer le taux d'accès des publications AAU, selon leur type d'accès ouvert et la présence de DOI.

Pour obtenir ces résultats, nous avions ajouté manuellement plusieurs indicateurs dans le groupe Zotero de l'UMR, recensant l'intégralité des productions du laboratoire. Des tableaux, réalisés à partir d'exports CSV de la base ont ainsi pu être fournis, tant sur l'accessibilité des articles que sur les éditeurs les plus représentés, mais au prix d'un fastidieux travail difficilement tenable chaque année …

Accessibilité des articles

L'accès aux articles scientifiques publiés est un critère important de la politique du laboratoire en faveur de la science ouverte. On peut noter à ce sujet (Tableau 8) que :

- Plus de la moitié des articles scientifiques de la période de référence ont été publiés dans des revues en *Open Access* (80 articles sur 140 soit 57%).
- 63% des articles scientifiques de la période (88 sur 140) ont un DOI (Digital Object Identifier).
- 57% des articles scientifiques de la période (80 sur 140) ont été déposés en texte intégral dans l'archive ouverte HAL et sont donc accessibles à travers les collections des équipes et via le site web du laboratoire : cf. https://aau.archi.fr/laboratoire-aau/publications/
- En cumulant l'accessibilité offerte par les revues en *Open Access* et les dépôts dans HAL, 80% des articles scientifiques produits par le laboratoire pendant la période de référence sont en libre accès.

	CRENAU	CRESSON	Total AAU
Publiés dans des revues en Open Access	43%	77%	57%
Avec DOI	68%	57%	63%
Déposés en texte intégral sur HAL	46%	72%	57%
En libre accès (dans revue Open Access et/ou sur HAL)	73%	90%	80%

Tableau 8. Part des articles scientifiques par indicateur d'accessibilité

Document d'autoévaluation, laboratoire AAU, 2021

Ce bilan sur 5 ans nous a également permis d'analyser notre collection HAL, grâce aux outils OcdHal et VizuHal, et ainsi de mesurer la part de dépôts en texte intégral.
Dès que Hal a proposé l'affichage du taux Open Access sous forme de widget sur les pages d'accueil des collections, nous nous sommes empressées de l'activer avec le souci de valoriser l'adhésion de l'UMR au mouvement de la Science ouverte.

Mais ce qui nous faisait vraiment envie, c'était ces baromètres dynamiques (BSO) affichés sur les sites de plusieurs universités.

Nos compétences en codage et langage python, proches du néant, ne rendaient pas possible cette valorisation et nous en étions restées là …((Pour rappel, le baromètre permet d'obtenir un ensemble d'indicateurs sous forme graphique concernant l'accès ouvert aux publications. Il a été développé en 2019 par le MESRI pour des données nationales, puis adapté par l'Université de Lorraine en 2020. Son code informatique est aménageable par les établissements et organismes de recherche.))

Baromètre Science Ouverte UVSQ [Baromètre Science Ouverte UVSQ]

L'opportunité d'un cours informatique pour adapter un code BSO AAU

A la rentrée 2021, Myriam Servières envoyait un mail aux membres de l'équipe AAU-Crenau pour solliciter des sujets de développement en Python, C et Matlab destinés aux projets d'application en binôme des étudiants de son cours d'option informatique à l'[Ecole Centrale de Nantes]. Heureusement, elle précisait que ceux qui avaient un sujet mais pas les compétences informatiques pour l'encadrer, pourraient être « client » et le sujet serait alors encadré par un membre du département info-math …

Une fiche sujet a donc été illico rédigée par Laurence Bizien, sous le titre :

« **Installer un outil de pilotage pour suivre la progression de la Science Ouverte au sein de la production scientifique du laboratoire AAU (ECN/ENSA Nantes/ENSA Grenoble/CNRS)** »

Explicitant le contexte, le travail et résultats attendus, ainsi que l'environnement et les technologies utilisées, avec une courte bibliographie. La demande consistait à réutiliser les données et le code du Baromètre français de la Science Ouverte pour écrire un nouveau code permettant d'isoler les publications du laboratoire AAU et une mise à jour annuelle du baromètre. Le modèle choisi était celui du baromètre de l'Université VSQ, car intégrant des publications sans DOI en récupérant leurs données dans HAL.

L'objectif était donc d'adapter un baromètre et son code au laboratoire afin de produire un résultat qui serait publié sur le site internet de l'UMR, le code devant également être réutilisable pour permettre une utilisation annuelle.

Calendrier

- 17/09/21 : présentation des sujets
- 20/09/21 : retour de la répartition
- *Rapports d'avancement*
  - *tous les vendredi* ***18h dernier délai – à déposer sur hippocampus en pdf***
  - Contenu :
    - Avancement
    - Programme de la semaine à venir
    - Difficultés rencontrées
    - Planning mis à jour
  - Aux encadrants **et** aux clients par mail (sauf si ils vont le chercher sur hippocampus, à décider avec eux)
- Ensemble de livrables notés (par les encadrants)
  - 01/10/21 : Cahier des charges
  - 15/10/21 : Diagrammes UML / MCD / documents de conception
  - 03/12/21 : Squelette de rapport
  - 04/01/22 : Rapport final
  - 07/01/22 : Soutenances

3

Les étudiants Alan Roussel et Alexandre Bodet ont choisi notre sujet : le 22 septembre 2021, une visio avec Myriam et le client AAU (représenté par Laurence) était organisée afin de compléter la demande et fournir davantage d'éléments sur la Science Ouverte à des étudiants pas encore « affranchis » au monde de la recherche !

Ainsi, le choix du code de l'UVSQ a été explicité par le fait que les données utilisées pour le Baromètre national (et d'autres baromètres universitaires) proviennent de sources diverses, telles que Web of Science, PubMed, HAL, les paiements d'APC et Lens.org, ne conviennent pas pour notre laboratoire.

En effet, en dépit du croisement de ces données, la liste des publications recensées n'est pas exhaustive pour AAU, car seules les publications disposant d'un DOI peuvent être prises en compte. Par ailleurs, les publications en Sciences Humaines et Sociales sont généralement sous-représentées dans ces bases de données, et **seul le code UVSQ permet d'intégrer les articles sans DOI** …

Le projet : difficultés, adaptations, discussions …

Les premières semaines ont principalement servi à la compréhension du code open source de Maxence Larrieu et à la façon de l'adapter au laboratoire, le code ayant été conçu pour s'appliquer aux données de l'UVSQ et à la taille de ses données.

Maxence Larrieu a gentiment accepté un RDV en visio fin septembre pour détailler son travail et expliquer son code à Alan et Alexandre.

Le projet récupéré prend ainsi en compte les données de nombreuses bases, dont Wos et Scopus, nécessitant un abonnement et une clé de d'identification pour l'utilisation de l'API. Or, notre laboratoire n'est pas directement abonné à ces deux bases, mais bénéficie d'un accès via BibCNRS, ce qui complique la récupération automatique ! Le choix a donc été fait de ne pas aller chercher ces données, sachant que celles-ci, peu nombreuses, sont déjà dans HAL. Néanmoins, il est toujours possible de télécharger ces données et de déposer les fichiers, récupérés manuellement, dans le dossier dédié (data/dois).

Alan et Alexandre ont d'abord automatisé la récupération des données liées aux publications pour que la liste des publications soit à jour et que l'utilisateur n'ait plus à télécharger ces fichiers. Il fallait également qu'ils paramètrent le code pour que le laboratoire puisse s'en servir simplement et enfin dynamiser le code pour le rendre pérenne. La récupération des données a été réalisée à partir de l'API HAL (HALId, titre, DOI existant), pour 4 types de publication : **Articles dans une revue, Communications dans un congrès, Chapitres d'ouvrage et Direction d'ouvrages.**

```
name,all,doi,no_doi
hal,1488,274,1214
retenu,1452,273,1179
```

API HAL https://api.archives-ouvertes.fr/docs/search

Ces données ont été enrichies par l'utilisation de l'API d'Unpaywall, pour les publications disposant d'un DOI : nom d'auteur, titre de la revue, éditeur, domaines et sous-domaines, date, statut open access, …

L'étape suivante était celle de l'ajout d'information sur les APC (article processing charge), à partir des fichiers du code de Maxence Larrieu (OpenApc, DOAJ) : il a été décidé de ne pas automatiser mais de rendre optionnelle cette requête dans les settings (_apc.py).

L'objectif était de mettre les graphiques obtenus sur une page internet aussi simplement que possible. Pour générer les graphiques, Alan et Alexandre ont utilisé la librairie pyplot, avec trois temporalités possibles : **des graphiques sur une période donnée (ex. 2015/2020), sur une année complète, sur l'ensemble des dépôts de la collection (depuis 2004 dans notre cas)**.

La modification de ce paramétrage permettra la mise à jour annuelle de graphiques en png par le laboratoire, en exécutant le fichier principal. L'accès au site internet AAU étant réservé au développeur et au webmestre, le script ne pouvant tourner sur une page, l'équipe « client » s'est alors élargie à Jérémie Bancilhon (société GO ON WEB qui héberge et maintien le site web AAU) et à Véronique Dom (chargée de communication AAU) pour discuter de l'implantation et de la mise à jour des graphiques dans le site. Laurent Charrieau (administrateur systèmes et réseau Crenau) a fourni un accès ssh à un ordinateur du laboratoire pour tester le projet et le mettre en production.

Soutenance, livrable et mise en ligne

Le rapport final nous a été envoyé le 4 janvier, avec le livrable attendu.

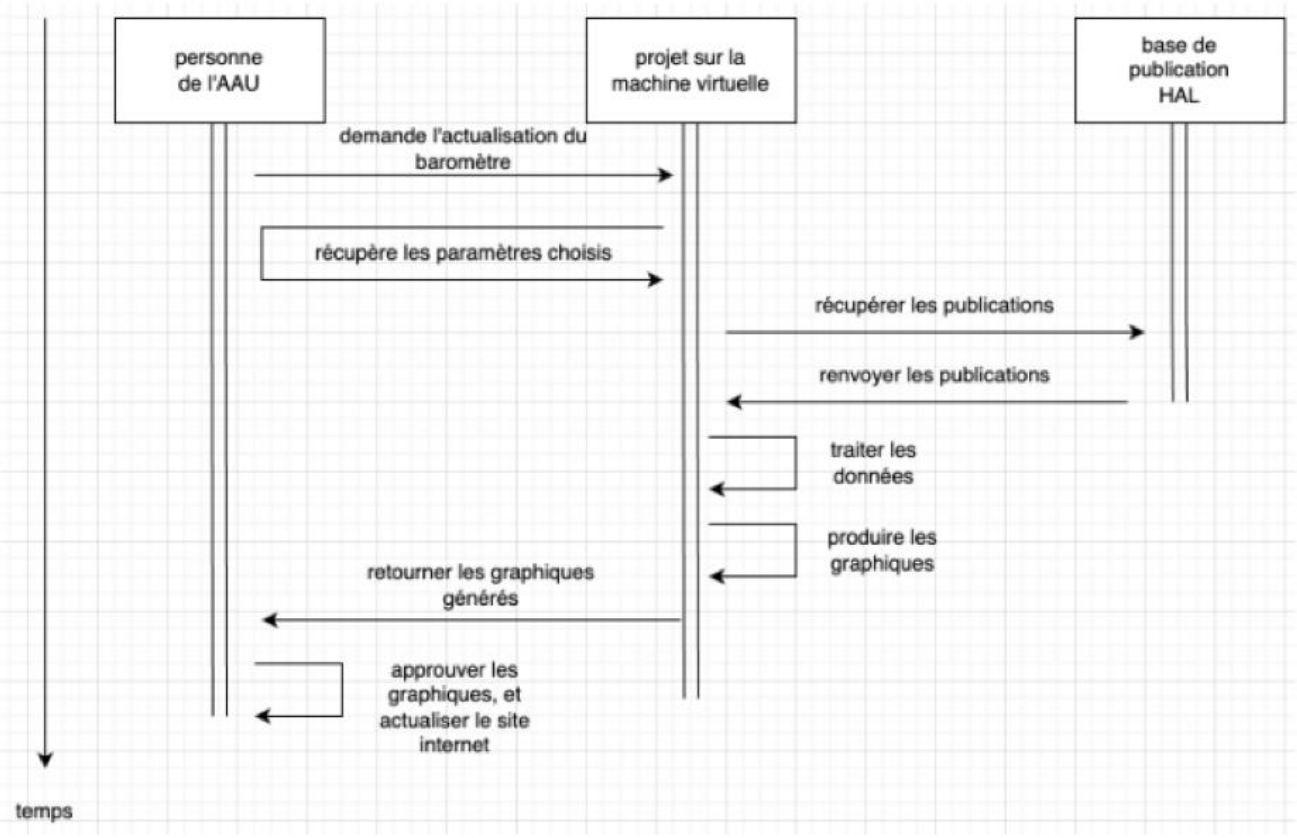


Figure 1 : Diagramme de séquence du projet

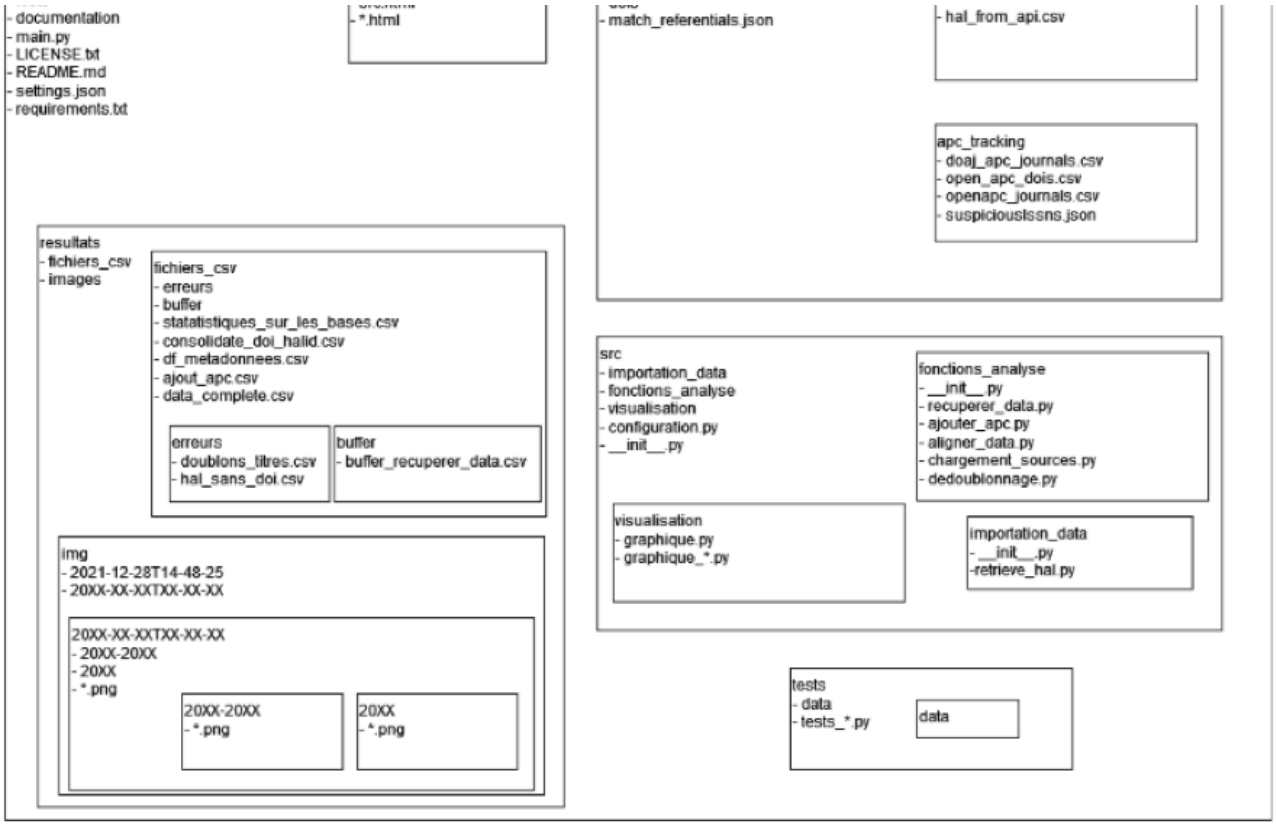


Figure 2 : Diagramme de paquetage du projet

La soutenance s'est déroulée comme prévue (mais en visio !) le 7 janvier 2022 après-midi.

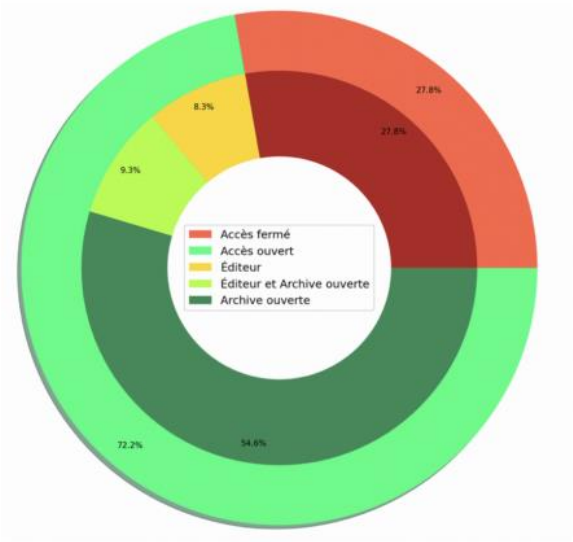
Ce beau travail, qui a réuni autour d'Alan et Alexandre une partie de l'équipe du laboratoire (certains se sont vu sollicités en cours de projet !), a permis à notre UMR d'afficher ses bons résultats, au moment où le Baromètre national 2021 était publié : https://www.ouvrirlascience.fr/le-barometre-francais-de-la-science-ouverte-fait-peau-neuve/

La version 1.0 du code AAU est ici : https://github.com/AlexandreBodet/Projet_AAU_Barometre/releases/tag/v1.0

Le github avec la documentation pour un éventuel développeur et les instructions de lancement : https://github.com/AlexandreBodet/Projet_AAU_Barometre

Après avoir fait un choix parmi les graphiques générés et rédigé un petit texte d'analyse, le BSO AAU est désormais en ligne sur le site du laboratoire.

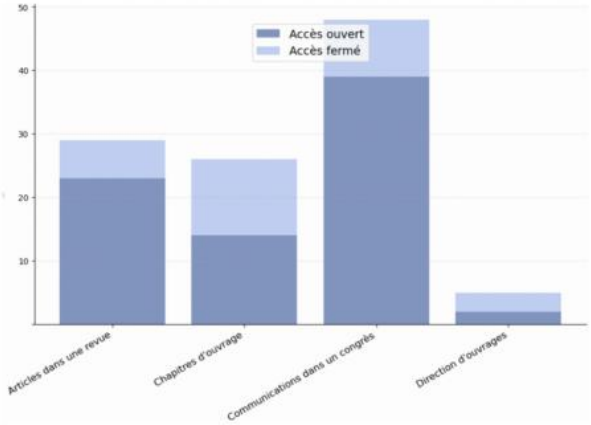
Bilan 2020 - Proportion des publications de l'UMR AAU en accès ouvert en 2020 mesurée en janvier 2022



En 2020, sur presque 120 publications, le taux d'accès ouvert est de 72,2% pour les 4 types de publications, ce qui par rapport à la moyenne des 5 dernières années offre une progression de 10%. Les publications en accès ouvert par le biais des sites éditeurs doublent par rapport à cette même moyenne, passant de 4,2% à 8,3 %.

La moyenne nationale est de 62% mais ne concerne que les publications d'articles avec DOI. « Selon l'édition 2021 du Baromètre de la Science Ouverte (BSO), 62 % des 166 000 publications scientifiques françaises publiées en 2020 sont en accès ouvert en décembre 2021 » https://www.ouvrirlascience.fr/le-barometre-de-la-science-ouverte/

Nombre de publications par genre pour l'UMR AAU en 2020 mesuré janvier 2022



Dans la continuité de la ventilation des publications par genre sur les 5 dernières années, les communications dans un congrès sont majoritaires mais il est intéressant de noter que les chapitres d'ouvrage sont de plus en plus disponibles en accès ouvert. Ce chiffre est à regarder de près les années à venir avec la mise en application des nouvelles directives du Plan S, mettant en œuvre le libre accès immédiat à toutes les publications financées par les organismes de financement dont l'ANR.