

Les coulisses du code - Python pour les SHS

Approches cartographiques et démarche de science ouverte

Célya Gruson-Daniel, Maya Anderson-González et
Camille Moulin
9/6/22

inno CUBE

Sommaire

- **Introduction :** la recherche-action ouverte chez inno³
- **1. *Open Science Open Source (OSOS)* :** python pour de l'analyse de réseau Twitter
- **2. Préfiguration du Lab de la Science Ouverte (LabSO) :** utilisation d'Hyphe pour une cartographie de la « recherche sur la recherche »
- **Conclusion :** modèles soutenable de développement de logiciels de recherche en python

Intro

La recherche- action ouverte chez Inno³

Développer des modèles ouverts soutenables

Les recherches effectuées chez inno³ visent à **apporter une compréhension, proposer des grilles d'analyse et développer des outils** pour différentes institutions publiques et privées pour le développement de modèles ouverts soutenables.

Comprendre

Analyser

Prototyper

Développer

Accompagner

La recherche action chez inno³

- Méthodes mixtes
 - Qualitative (entretiens)
 - Quantitative (questionnaires, données numériques)
- Co-Design/UX research
 - Inclure les parties prenantes dans la conception de solutions collectives
 - S'assurer que les solutions puisse être prises en main facilement
- Science ouverte
 - Traçabilité, transparence, reproductibilité
- Réflexivité et éthique
 - Dialogue avec les acteurs de l'écosystème
 - Participation aux réflexions et événements structurants pour les thématiques de recherche du cabinet

Méthodes mixtes



- Analyse et visualisation de données quantitatives
 - questionnaires



- Analyse et visualisation de données qualitatives
 - entretiens, questions ouvertes, analyse de corpus



- Analyse et visualisation de données massives
 - Analyse de réseaux sociaux, crawling, cartographies

1

Open Science Open Source
(OSOS) : python pour de
l'analyse de réseau
Twitter

Objectifs, méthode, démarche



- **Objectif** : explorer les acteurs et les dynamiques autour de conférences à la croisée de l'*open science* et de l'*open source*.
- **Méthode** : un volet quantitatif (analyse de réseau de suivi Twitter (*Follow graph*) des participants de deux événements (ChaosCon, FOSDEM20) en soutien au travail de recherche qualitatif (observation participative + 3 entretiens avec des participants ciblés).
- **Démarche SO** : documentation du code en python, mise en conformité et partage des données et du code, bibliographie partagée...

Utilisation de python



- Collecte avec python (Twarc) en ligne de commande



- Traitement des données avec python (pandas) dans un Jupyter Notebook



Code pour structurer la liste d'adjacence du réseau de suivi OSOS

```
In [2]: from os import path as os_path
import os

In [3]: import json
import pandas as pd

In [4]: git_dir = "/home/maya/chaoscon_fosden20-research-project"
usercomplete_path = os_path.join(git_dir, 'twarc_data', 'userscomplete.jsonl')

In [5]: with open(usercomplete_path) as json_file:
    json_data = [json.loads(line) for line in json_file]

In [6]: json_data[0]

Out [6]: {'id': '953304055337668608',
'id_str': '953304055337668608',
'name': 'CHAOS5',
```

Utilisation de python



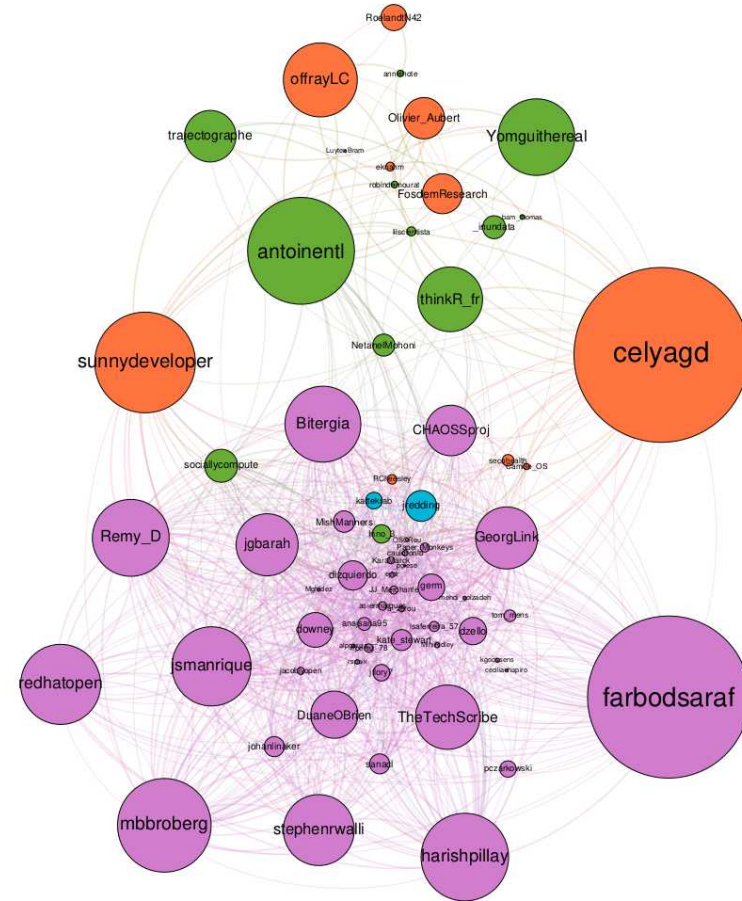
- Visualisations et l'analyse avec le logiciel open source Gephi.



- Stockage des données et des Notebooks avec le code python sur Gitlab



- Code publié sous licence GNU GPL



Retour d'expérience

- Le projet a permis de mettre le doigt sur deux principaux obstacles à surmonter pour l'utilisation de python en recherche SHS dans une démarche de science ouverte :
 - Travailler en équipe avec une complémentarité de compétences.
 - Aller au-delà d'une simple compréhension du code pour pouvoir systématiser l'utilisation d'un script dans différents contextes.

2

Préfiguration Labo de la
Science Ouverte (LabSO) :
utilisation d'Hyphe pour
une cartographie de la
« recherche sur la
recherche »

Objectifs, méthode, démarche



- **Objectif** : explorer les acteurs institutionnels autour de la « recherche sur la recherche », « metascience » et « meta-research ».



- **Méthode** : un volet quantitatif d'analyse de réseau de pages web et de liens en soutien au travail de recherche qualitatif (13 entretiens).



- **Démarche** : collecte automatisée de données massives, traçabilité du raisonnement, bibliographie partagée...

Utilisation de python



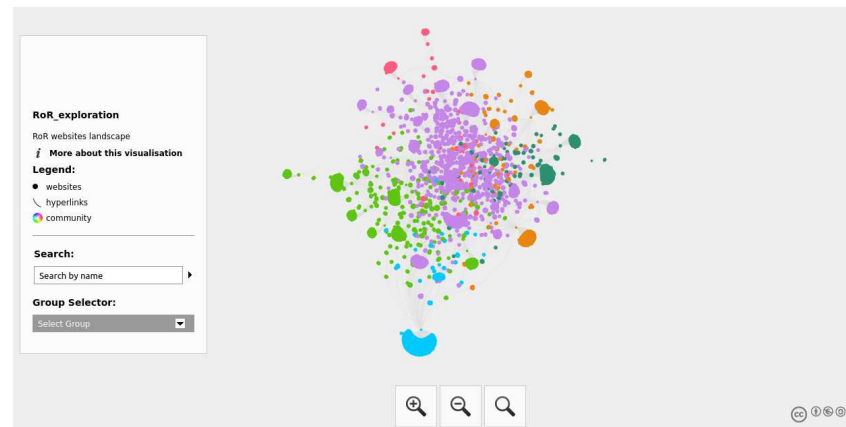
- Collecte et traitement de données avec l'outil de moissonnage open source **Hyphe**



- Visualisation et analyse avec Hyphe, puis export vers Gephi.



- Stockage des données sur un serveur interne inno³



Avantages



□ Code :

- Avoir une démarche plus transparente pour celles et ceux qui souhaitent reproduire des résultats de recherche en adaptant le code publié.



□ No code :

- Avoir une démarche plus accessible pour celles et ceux qui ne programment pas.



□ Dans les 2 cas :

- S'appuyer sur une communauté Open Source

Inconvénients



□ Code :

- Inaccessible pour celles et ceux qui ne savent pas programmer.



□ No code :

- Dépendance des évolutions ou disparition des outils.
- Boîtes noires au fonctionnement *click and play* comportant des risques non-maîtrisés de traitements de données.





Modèles soutenables de
développement de
logiciels de recherche
en python



- Une prise en considération de plus en plus grande du code source et des logiciels au sein de l'ESR
 - Plan National de la Science Ouverte (PNSO2) : reconnaissance de l'importance de l'ouverture des codes sources et logiciels pour des enjeux de reproductibilité et de valorisation de la recherche.



- Une mise en œuvre à accompagner :
 - Des mesures qui dépassent le seul milieu de l'ESR (AMDAC, Open Source Officer)
 - Un besoin de soutien des communautés de développement de logiciels open source
 - Des bonnes pratiques à développer (reproductibilité et partage du code, etc.)
 - Une sensibilisation à faire auprès des professionnels de la recherche et des services de valorisation



- Pour en savoir plus :
 - 2020 - 2021 : L'étude sur l'ouverture des codes source en recherche : [Open Source ESR](#)
 - 2021 : Plan National pour la science ouverte 2 ([PNSO2](#))
 - 2022 : Les 1ères rencontres [#FLOSS_ESR](#)
 - 2022 : The software pillar of open science ([conférence OSEC](#))

Ressources

□ OSOS

- Billet de blog sur le site d'Inno3
- Projet sur Framagit
- Open research devroom FOSDEM

□ LabSO

- Rapport sur HAL
- Billet de blog sur le site d'Inno3
- Projet sur Gitlab
- Site éditorialisé sur PubPub