



HAL
open science

L'ouverture des codes source et logiciels: état des lieux et enjeux

Camille Maumet

► **To cite this version:**

Camille Maumet. L'ouverture des codes source et logiciels: état des lieux et enjeux. Journée du réseau de l'atelier rennais de la donnée (ardoise), Jun 2024, Rennes, France. hal-04632167

HAL Id: hal-04632167

<https://inria.hal.science/hal-04632167>

Submitted on 3 Jul 2024

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.

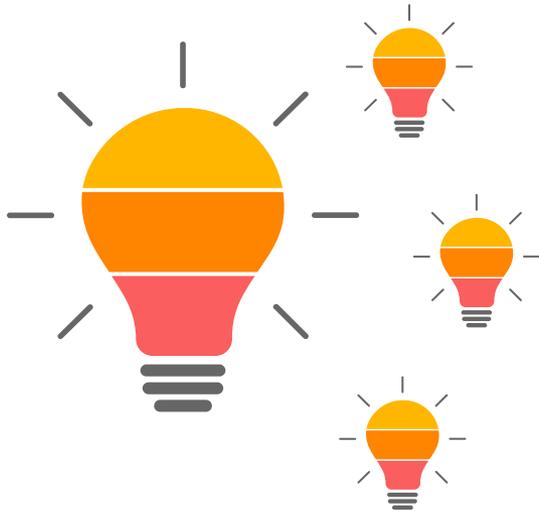


Distributed under a Creative Commons Attribution 4.0 International License

Journée du réseau ARDoISE
25 juin 2024



ARDoISE
Atelier rennais
de la donnée



Ouverture des codes source et logiciels état des lieux et enjeux

Camille Maumet

Inria, Univ Rennes, Inserm, CNRS, IRISA



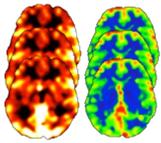
Une étude en **sciences expérimentales**



Acquisition



Pré-traitements



Mesures dérivées



Statistiques



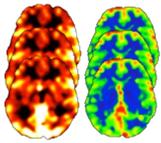
Une étude en **sciences expérimentales**



Acquisition



Pré-traitements



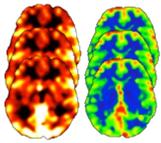
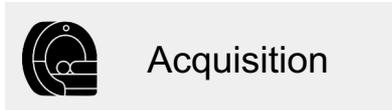
Mesures dérivées



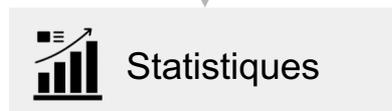
Statistiques



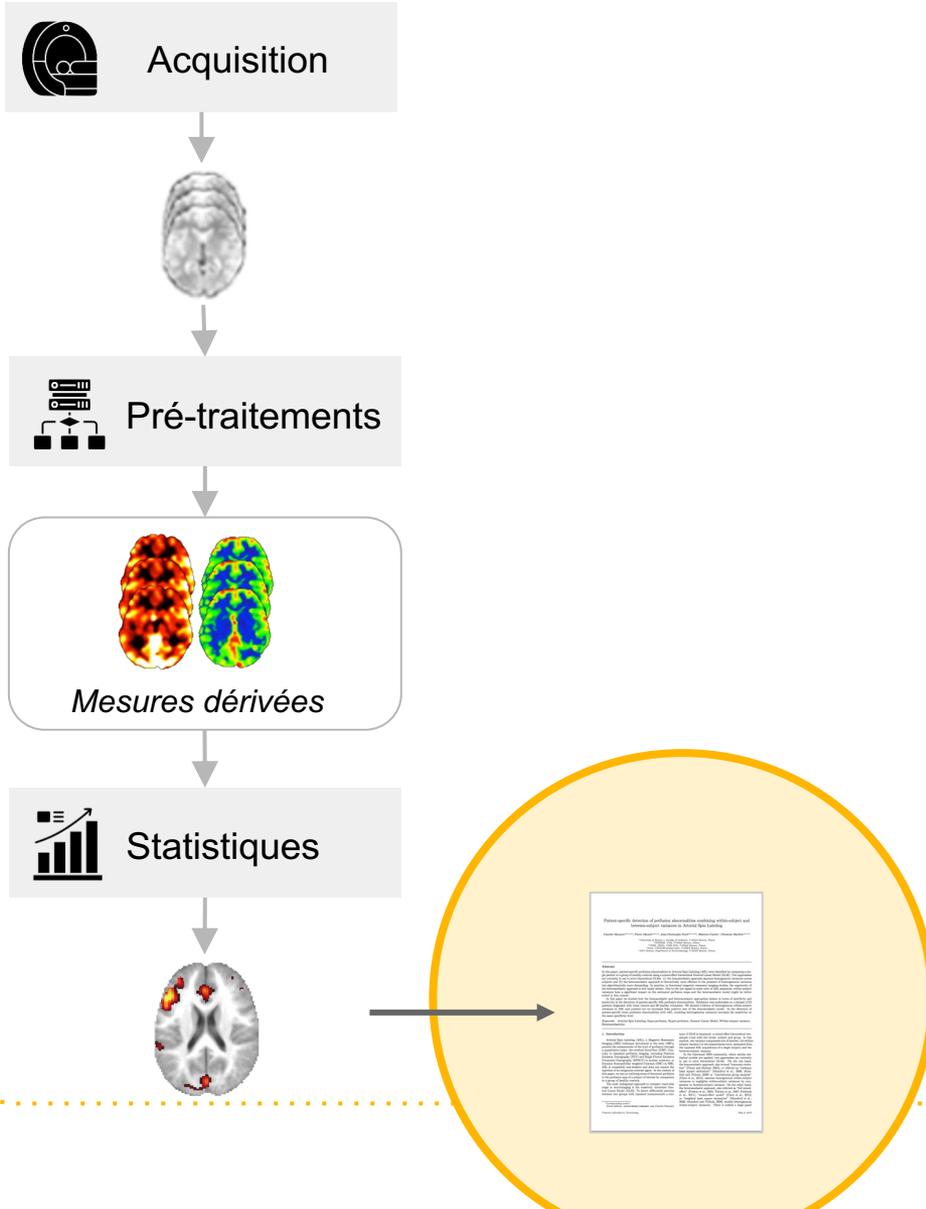
Une étude en **sciences expérimentales**



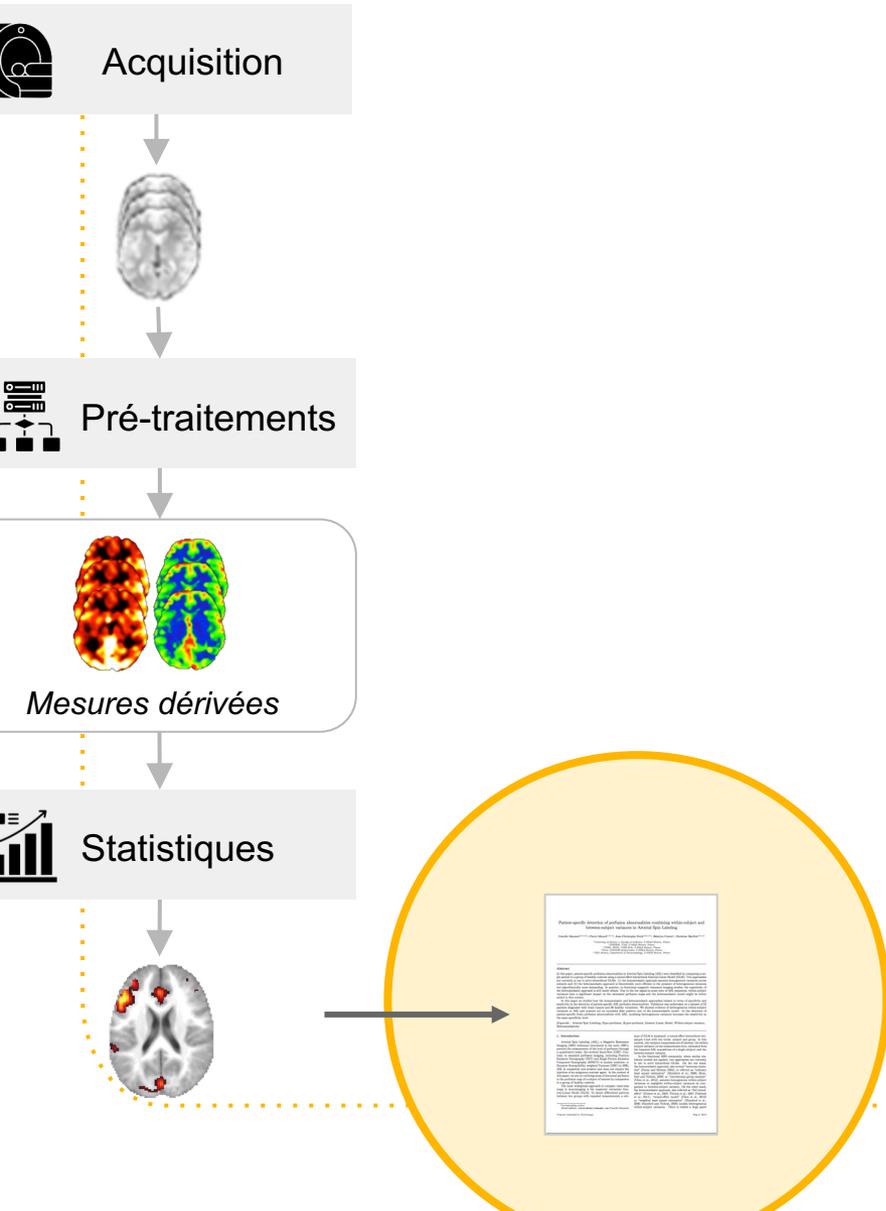
Mesures dérivées



Une étude en **sciences expérimentales**



Une étude en **sciences expérimentales**



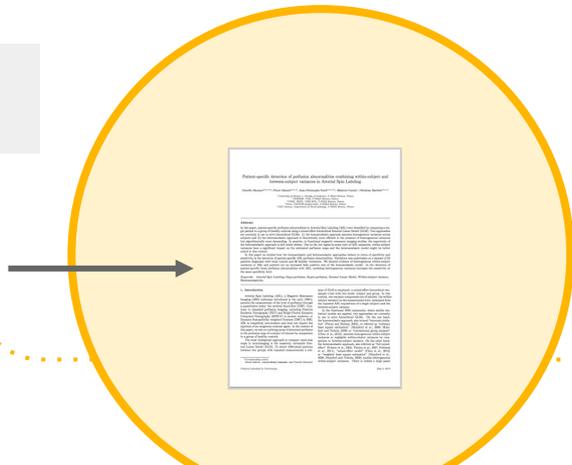
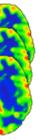
Une étude en **sciences expérimentales**

sition

tements

privées

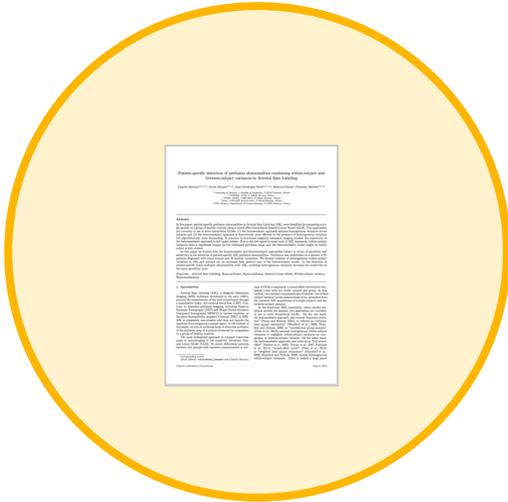
ques



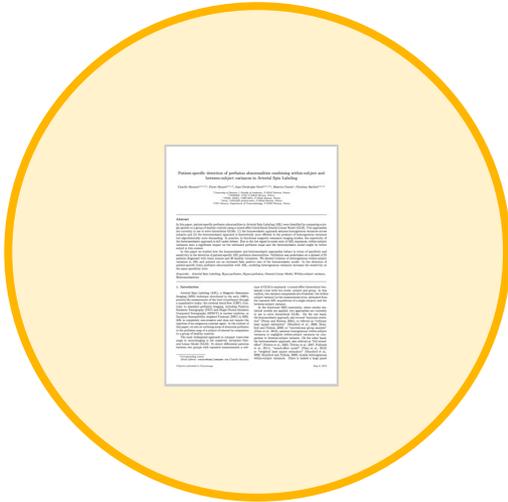
Une étude en **sciences expérimentales**



Une étude en **sciences expérimentales**



Une étude en **sciences expérimentales**



Quelle version du logiciel SPM ?

Quel algorithme de recalage ?

Comment le masque du cerveau a-t-il été calculé ?

Une étude en **sciences expérimentales**



Quelle version du logiciel SPM ?

Quel algorithme de recalage ?

Comment le masque du cerveau a-t-il été calculé ?

Hum... **Je n'arrive pas à reproduire** les résultats originaux...

Est-ce que j'ai bien compris la méthode utilisée

... est-ce parce que j'ai utilisé un jeu de données différents

... ou une plateforme de calcul différente

... un autre aspect de l'analyse qui diffère

Une étude en **sciences expérimentales**



Quelle version du logiciel SPM ?

Quel algorithme de recalage ?

Comment le masque du cerveau a-t-il été calculé ?

... est-ce parce que j'ai utilisé un jeu de données différents

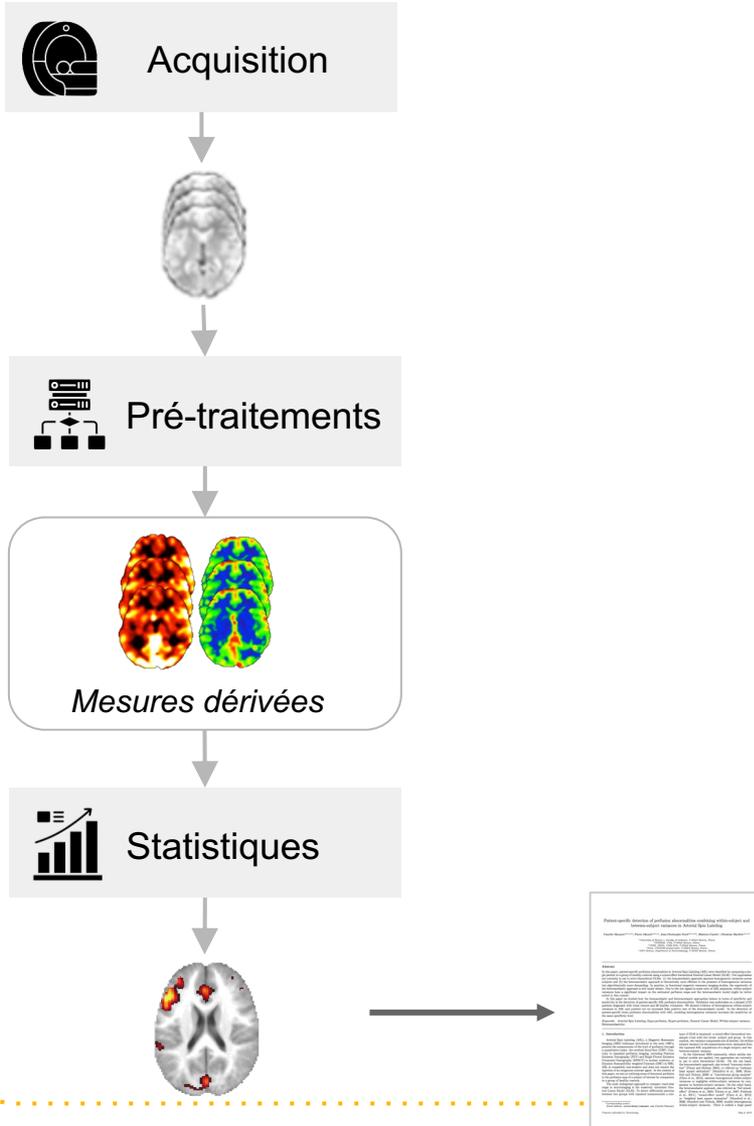
... j'ai utilisé une plateforme de calcul différente

... un autre aspect de l'analyse qui diffère

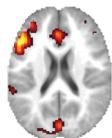
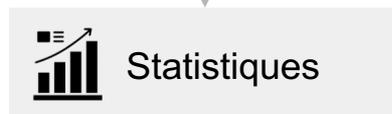
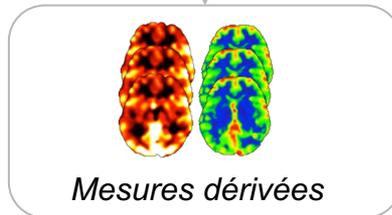
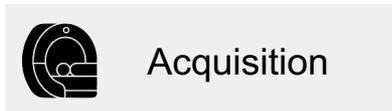
Hum... **Je n'arrive pas à reproduire** les résultats originaux...



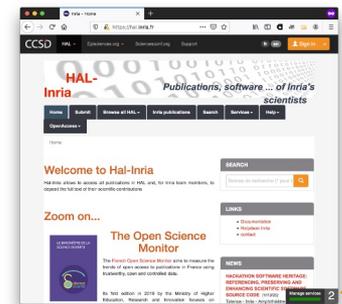
Vers une **science reproductible**



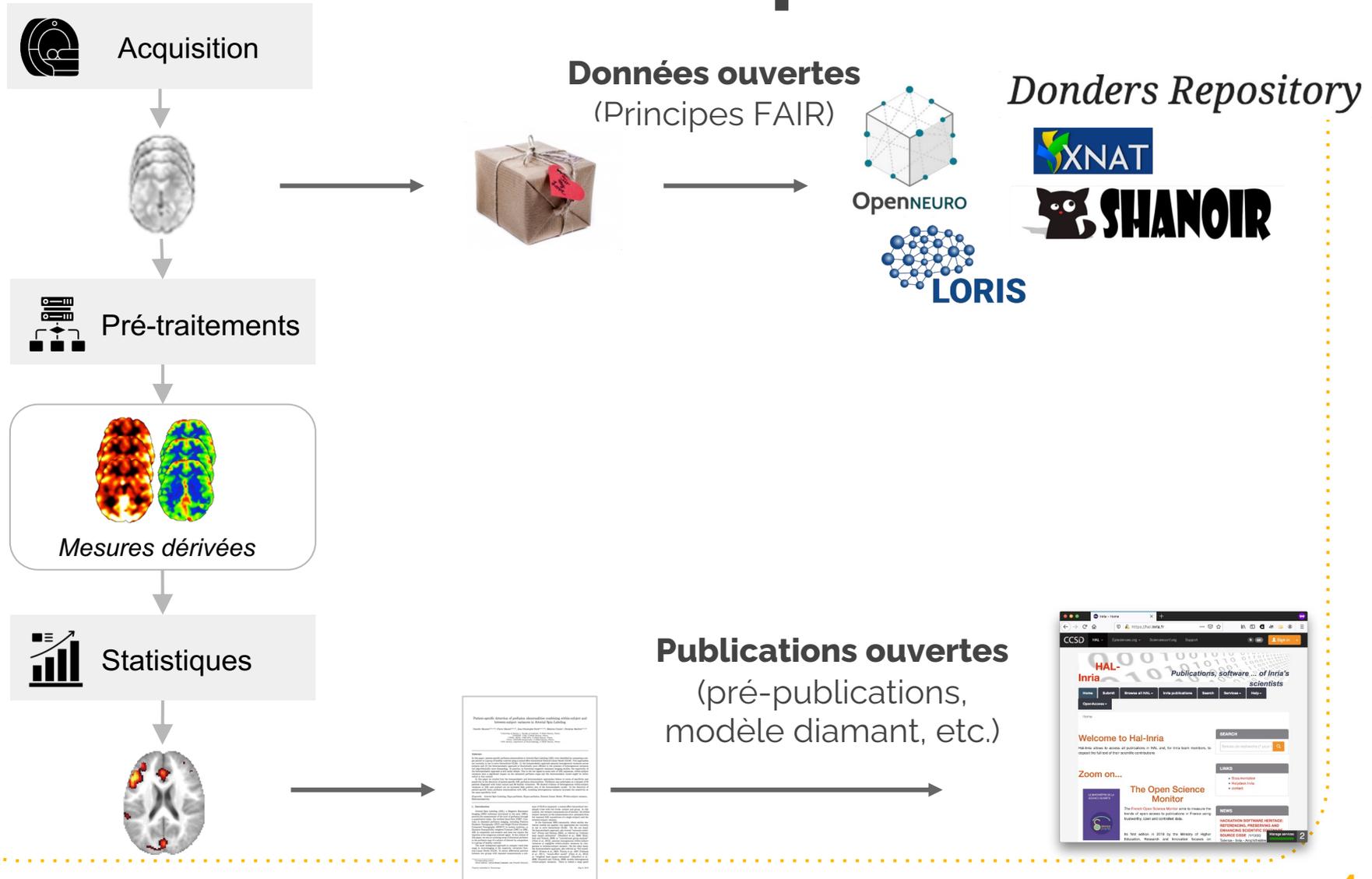
Vers une science reproductible



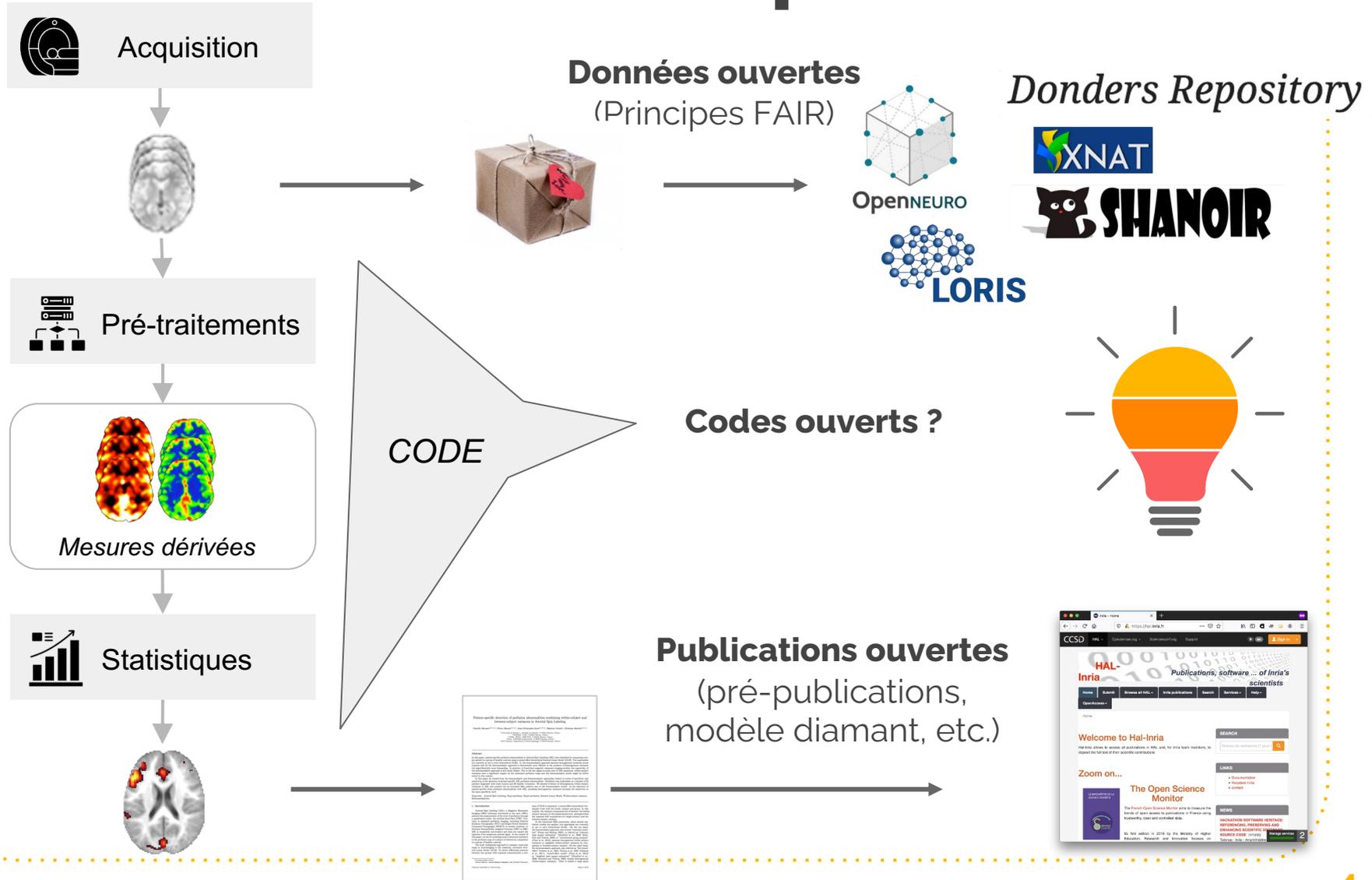
Publications ouvertes
(pré-publications,
modèle diamant, etc.)



Vers une science reproductible



Vers une science reproductible



Vers une **science reproductible**

Reproductibilité computationnelle

Possibilité de reproduire des **résultats strictement identiques** en utilisant les **même données** et le **même code**.

Replicabilité

Possibilité de reproduire les **mêmes conclusions scientifiques** en utilisant des **données similaires** et une **méthode similaire**.

Vers une **science reproductible**

Reproductibilité computationnelle

Possibilité de reproduire des **résultats strictement identiques** en utilisant les **même données** et le **même code**.

Replicabilité

Possibilité de reproduire les **mêmes conclusions scientifiques** en utilisant des **données similaires** et une **méthode similaire**.

Codes sources ouverts



Open source
Transparence

- Avec une **licence open source**, **accessible publiquement** et **citable**.

OS Ouvrir
la science !



comment ouvrir ?

Si vous souhaitez partager votre logiciel et le rendre utilisable par d'autres, vous devez en organiser l'accès sur les plans juridique et technique. Il vous faut également réfléchir au devenir du code source du logiciel sur le long terme : maintien d'une activité de recherche à son sujet ou mise en place d'une démarche active de valorisation.

Choisir une ou plusieurs licences

Le choix de la licence est une étape importante de l'ouverture, qui peut avoir des conséquences importantes sur les évolutions du logiciel. La première étape consiste à identifier qui détient les droits patrimoniaux sur le logiciel; lorsque les auteurs ne sont pas les ayants droit, le choix de la licence ne leur revient pas.

La personne, groupe ou entité à qui revient la responsabilité de choisir la licence peut s'engager dans une démarche d'ouverture complète, en utilisant une licence de logiciel libre. Le département **ETALAB** de la **DINUM** fournit une **liste de licences utilisables pour les productions de l'administration publique française**. Dans les autres cas, on peut se référer aux listes maintenues par l'**Open Source Initiative** ou la **Free Software Foundation**.

Dans certaines situations, il est possible d'utiliser des licences plus restrictives, qui n'autorisent que la consultation du code ou restreignent l'utilisation par des tiers. Par exemple, on peut limiter l'usage à un cadre académique pour la seule reproductibilité scientifique.

Si le logiciel utilise des briques préexistantes, il faut s'assurer :

- de la compatibilité juridique de la licence choisie avec celles des différentes briques préexistantes;
- de la compatibilité des licences des différentes briques entre elles.



HAL Documentation

Accueil Guide utilisateur FAQ Tutoriels Questions juridiques Usual HAL API TripleStore Support HAL

Guide utilisateur document

Déposer le code source d'un logiciel

Le dépôt de logiciel dans HAL permet sa citabilité, l'archivage étant pris en charge par l'Infrastructure Software Heritage. HAL et Software Heritage proposent des services complémentaires.

Le recours à des logiciels est constant dans la plupart des disciplines académiques, mais il s'avère que les logiciels peuvent être difficiles à identifier pour des utilisateurs comme pour des développeurs qui voudraient procéder à des adaptations. Par ailleurs, de nombreux articles mentionnant des logiciels contiennent des liens brisés, ce qui entrave la pleine compréhension des travaux scientifiques.

Un logiciel déposé dans HAL est plus facile à identifier et préservé sur le long terme. Par ailleurs, HAL développe une palette de services permettant à un auteur de **lier des publications, des logiciels et des jeux de données**. Inversement, si le logiciel est déjà archivé dans Software Heritage, son dépôt dans HAL est facilité. De plus, tout l'historique de développement du logiciel sera aussi accessible.

Quel logiciel déposer dans HAL/Software Heritage ?

Qu'appelle-t-on un logiciel de recherche ?

Quels types de logiciels sont éligibles ?

Quels fichiers déposer ?

<https://www.ouvrirlascience.fr/science-ouverte-codes-et-logiciels/>

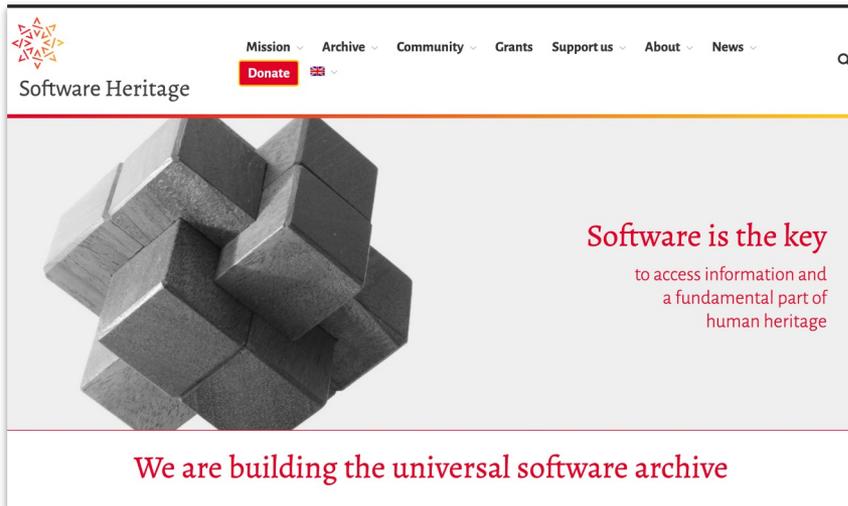
<http://doc.hal.science/deposer/deposer-le-code-source/>

Codes sources ouverts



Archivé
Perennité

- **Archivé** pour la **préservation à long terme** (e.g. Software Heritage, Zenodo)



zenodo

<https://zenodo.org/>

<https://www.softwareheritage.org/>

Codes sources ouverts



Bonnes pratiques logicielles

Reutilisabilité

- En suivant **les bonnes pratiques logicielles**
 - **Organisation:** README, hiérarchie de fichiers pour organiser le code
 - **Versionnement:** forge logicielle pour versionner le code.

```
.
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|  -- birds_count_table.csv
|-- doc
|  -- notebook.md
|  -- manuscript.md
|  -- changelog.txt
|-- results
|  -- summarized_results.csv
|-- src
|  -- sightings_analysis.py
|  -- runall.py
```

[Wilson et. al,
PlosCompBio 2017]



 **Ouvrir
la science !**

Par le [Comité Science Ouverte](#) (COSO)



Codes sources ouverts

Open source
Transparence

- Avec une **licence open source** et **mis à disposition** (dépôt HAL).

Archivé
Perennité

- **Archivé** pour la **préservation à long terme** (Software Heritage, Zenodo)

**Bonnes pratiques
logicielles**
Reutilisabilité

- **Suit les bonnes pratiques logicielles**
 - **Organisation:** README, hiérarchie de fichiers pour organiser le code
 - **Versionnement:** forge logicielle pour versionner le code.

Problèmes ouverts

Au-delà du code source, les **environnements** logiciels ont aussi un impact

→ *docker, Guix, Nix*

La reproductibilité computationnelle à long terme (**10 ans**) est encore une question ouverte.

Perkel, Jeffrey M. "Challenge to scientists: does your ten-year-old code still run?." *Nature* 584.7822 (2020): 656-659.

Vers une **science reproductible**

Reproductibilité computationnelle

Possibilité de reproduire des **résultats strictement identiques** en utilisant les **même données** et le **même code**.

Replicabilité

Possibilité de reproduire les **mêmes conclusions scientifiques** en utilisant des **données similaires** et une **méthode similaire**.

Vers une **science reproductible**

Reproductibilité computationnelle

Possibilité de reproduire des **résultats strictement identiques** en utilisant les **même données** et le **même code**.

Replicabilité

Possibilité de reproduire les **mêmes conclusions scientifiques** en utilisant des **données similaires** et une **méthode similaire**.

Ré-utilisabilité

Possibilité de réutiliser certains **artefacts produit par une étude de recherche** à de nouvelles fins.



Vers **des logiciels** ouverts



Open source

- Choisir une licence **open source**
- Rendre le code **disponible au téléchargement**
- Utiliser une **archive** pour la préservation à long terme.

Vers **des logiciels** ouverts



Qui peut
**accéder et
utiliser** le
logiciel?



Open source

- Choisir une licence **open source**
- Rendre le code **disponible au téléchargement**
- Utiliser une **archive** pour la préservation à long terme.



Vers des logiciels ouverts

Qui peut
**accéder et
utiliser** le
logiciel?



Open source

- Choisir une licence **open source**
- Rendre le code **disponible au téléchargement**
- Utiliser une **archive** pour la préservation à long terme.

Qui peut
contribuer ?



Développement ouvert

- Accepter des **contributions extérieures** (e.g. Github PR)
- Être transparent au sujet des choix de **développement** (e.g. issues, project board)

Vers des logiciels ouverts



Qui peut
**accéder et
utiliser** le
logiciel?



Open source

- Choisir une licence **open source**
- Rendre le code **disponible au téléchargement**
- Utiliser une **archive** pour la préservation à long terme.

Qui peut
contribuer ?



Développement ouvert

- Accepter des **contributions extérieures** (e.g. Github PR)
- Être transparent au sujet des choix de **développement** (e.g. issues, project board)

Qui peut
décider des
directions
futures ?



Gouvernance ouverte

- Transparence sur le **processus décisionnel**.
- Possibilité de **rejoindre la direction du projet**.

Vers une science plus **collaborative**

Au-delà des solutions techniques, la science ouverte apporte un changement de paradigme

- Mettre des ressources à la **disposition de la communauté scientifique**
- **Réutiliser des ressources existantes** (codes, jeux de données, etc.) créées par d'autres.

vers une recherche plus collaborative



Vers une science plus **collaborative**

Au-delà des solutions techniques, la science ouverte apporte un changement de paradigme

- Mettre des ressources à la **disposition de la communauté scientifique**
- **Réutiliser des ressources existantes** (codes, jeux de données, etc.) créées par d'autres.

vers une recherche plus collaborative



Prix science ouverte du logiciel libre de la recherche

<https://www.ouvrirlascience.fr/les-candidatures-pour-le-prix-science-ouverte-du-logiciel-libre-de-la-recherche-2024-sont-ouvertes/>

Image credits: Community by Vkw.studiogood, [Commons Wikimedia](#)



Journée du réseau ARDoISE
25 juin 2024

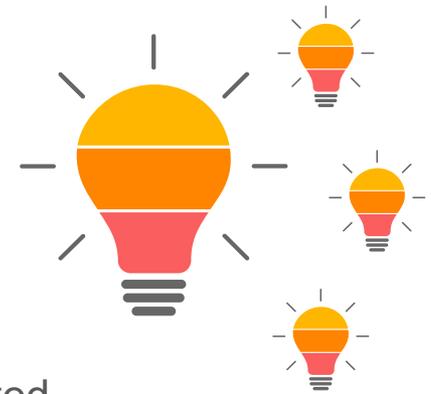


ARDoISE
Atelier rennais
de la donnée

L'ouverture des codes source et logiciels : état des lieux et enjeux

Merci

Camille Maumet



Credit: Presentation template and icons by SlidesCarnival, adapted



@cmaumet.bsky.social

@cmaumet@fediscience.org



@cmaumet