

OUVREZ-LA !

**Données et publications
ouvertes pour une science
partagée**

**Signaler et archiver son logiciel
dans HAL et Software Heritage**

Soraya Demay

Administratrice des données, algorithmes et codes sources de recherche de l'Université de Toulouse
Responsable de la cellule données de la Recherche
Département Appui à la Recherche et Science ouverte – SCD

Bastien Plazolles

Ingénieur de recherche CNRS – Laboratoire Géosciences Environnement Toulouse

Les logiciels de recherche sont ...

... des objets de recherche

... des outils pour la recherche

... des résultats de recherche

... garants de la reproductibilité de la science

... utilisés dans toutes les disciplines

- Données

- Publications

- Logiciels



➤ 3 piliers de la recherche contemporaine

➤ Le signalement et l'archivage des logiciels sont un enjeu majeur pour la science ouverte

Cadre institutionnel

2^{ème} Plan National sur la Science Ouverte – 2021

Troisième axe

Ouvrir et promouvoir les codes sources
produits par la recherche

Mesures

- 7** Valoriser et soutenir la diffusion sous licence libre
des codes sources issus de recherches financées sur fonds publics
- 8** Mettre en valeur la production des codes sources
de l'enseignement supérieur, de la recherche et de l'innovation
- 9** Définir et promouvoir une politique en matière
de logiciels libres

Collège codes sources et logiciels du Comité pour la Science Ouverte - 2022

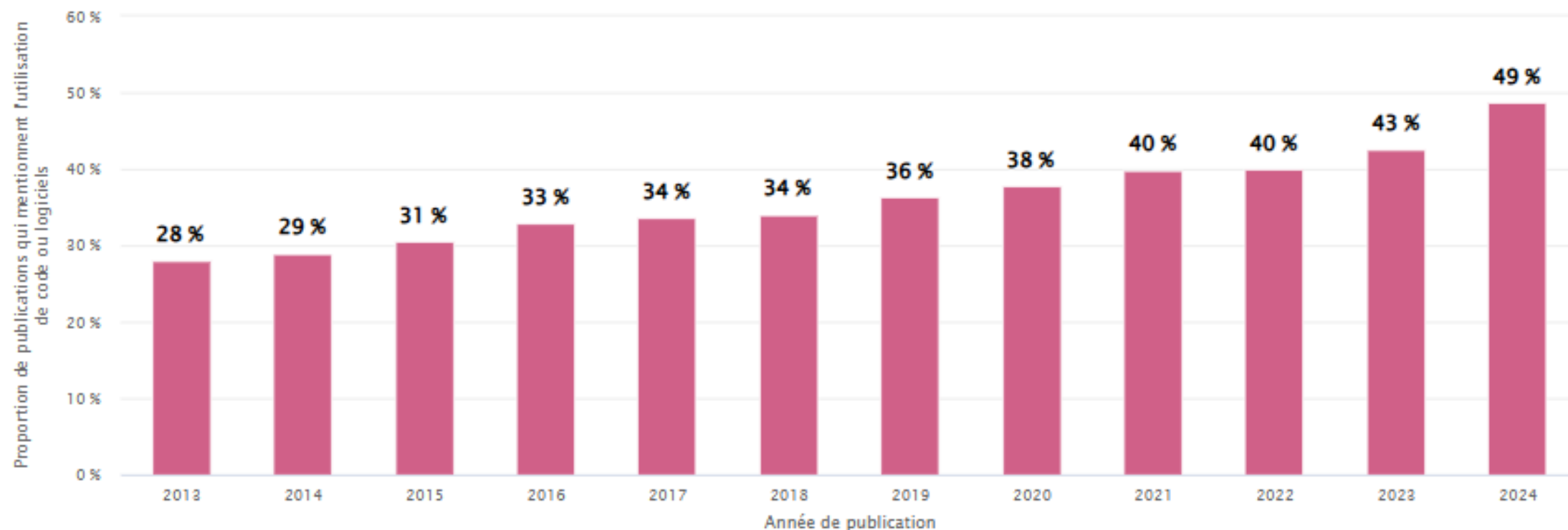
- Accompagnement le développement de logiciels libres et ouverts
- Production de métriques
- Veille sur les initiatives
- Production de ressource



Visibilité des logiciels de recherche aujourd'hui

Version [bêta]

Proportion de publications françaises qui mentionnent l'utilisation de code ou logiciels par année de publication



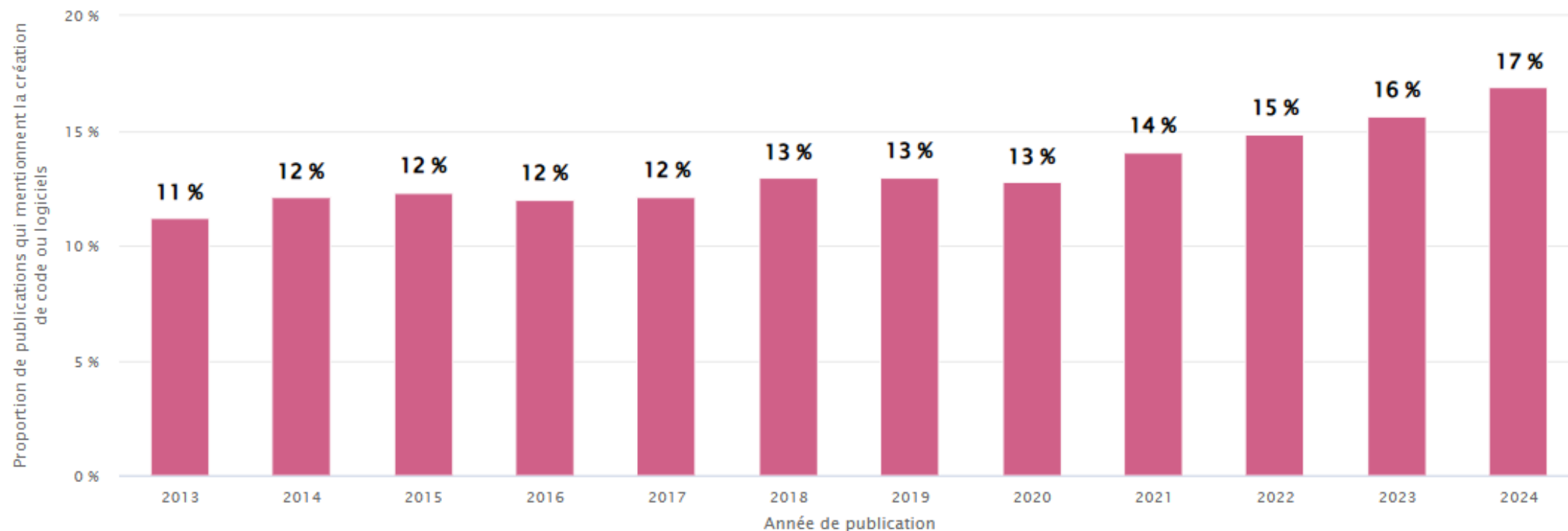
Baromètre français de la Science Ouverte – CC-BY MESRE

<https://barometredelascienceouverte.esr.gouv.fr/>

Visibilité des logiciels de recherche aujourd'hui

Version [bêta]

Proportion de publications françaises qui mentionnent la création de leurs code ou logiciels par année de publication



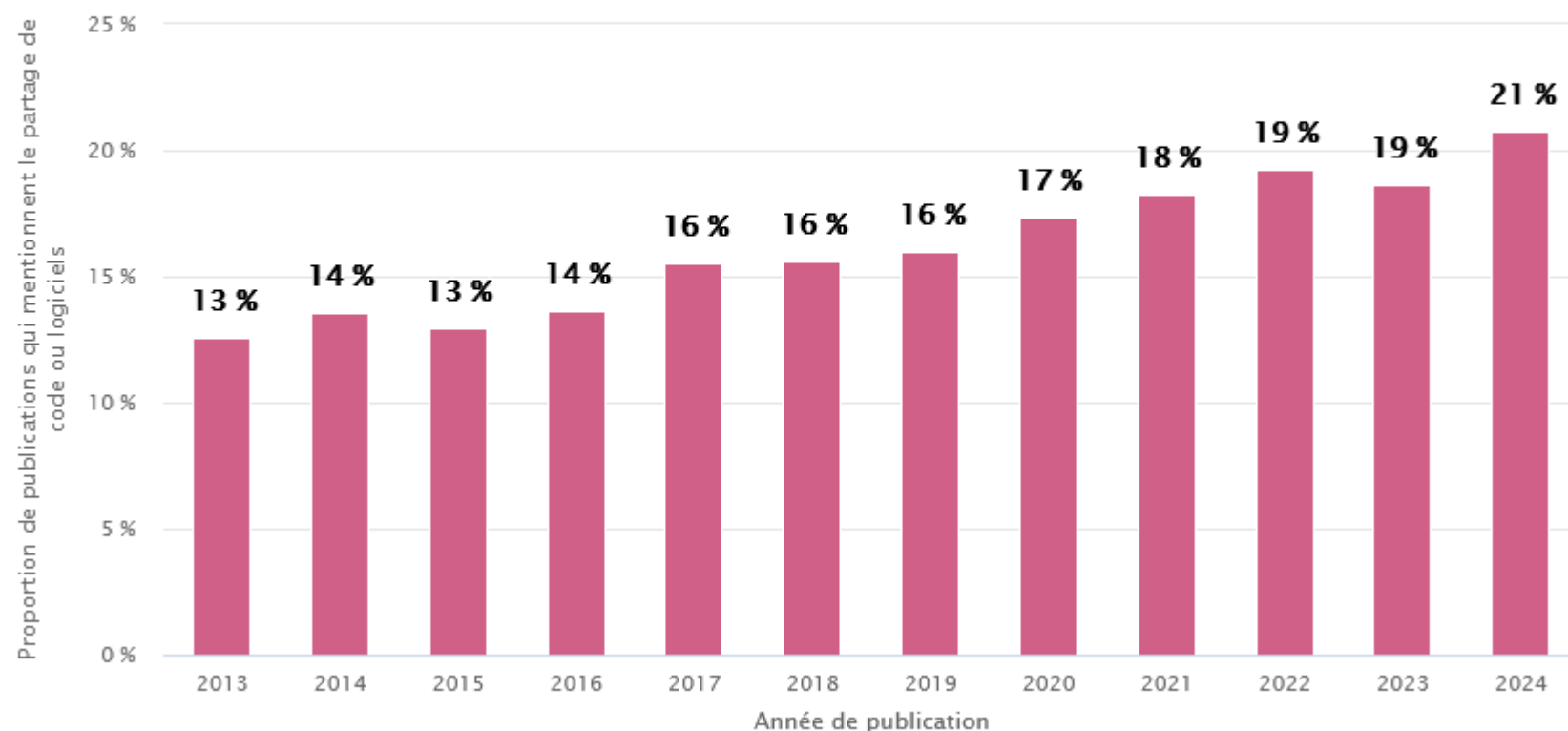
Baromètre français de la Science Ouverte - CC-BY MESRE

<https://barometredelascienceouverte.esr.gouv.fr/>

Visibilité des logiciels de recherche aujourd'hui

Version [bêta]

Proportion de publications françaises qui mentionnent le partage de leurs code ou logiciels par année de publication



Baromètre français de la Science Ouverte – CC-BY MESRE

➤ La sauvegarde et le partage des codes sources restent un défi

<https://barometredelascienceouverte.esr.gouv.fr/>

Les forges logicielles

Forges privées

- GitHub (Microsoft)
- BitBucket
- Gitea

Forges institutionnelles

- INRIA
- INRAE

Forges communautaires libres

- Free Software Foundation

+ **Système de contrôle de version**

+ **Développement collaboratif**

- **Fragilité des sauvegardes**

➤ Les forges ne sont pas des solutions pour l'archivage pérenne des logiciels de recherche

Pourquoi archiver les codes sources de la recherche ?

Objectifs de l'archivage pérenne

- Conservation
 - Accessibilité
 - Intelligibilité
-
- Sur plus de 30 ans
 - L'archivage va au-delà de la sauvegarde

Avantages

Eviter la perte

- Obsolescence
- Disparition de forge

Valorisation

- Visibilité
- Citabilité
- Datation et protection

S'inscrire dans une démarche de science ouverte

Qu'est-ce qu'un code source ?

Logiciel

Programmes informatiques

Code source

Instructions :

- compréhensibles par les humains
- exécutables par les machines

Source code matters!



"The source code for a work means the preferred form of the work for making modifications to it." — GPL Licence

Hello World

Program (excerpt of binary)

```
4004e6: 55
4004e7: 48 89 e5
4004ea: bf 84 05 40 00
4004ef: b8 00 00 00 00
4004f4: e8 c7 fe ff ff
4004f9: 90
4004fa: 5d
4004fb: c3
```

Program (source code)

```
/* Hello World program */
#include<stdio.h>

void main()
{
    printf("Hello World");
}
```

Roberto Di Cosmo - CC-BY-SA 4.0

Software Heritage: an essential facility for the digital age 24/10/2017 3 / 27

https://www.informatics-europe.org/images/ECSS/ECSS2017/Slides/ECSS2017_Cosmo.pdf

Où archiver les codes sources ?



Collecter, préserver et rendre accessible les logiciels open source produits dans le monde entier

- Archive créée en 2016 dédiée aux codes sources libres – INRIA / UNESCO
 - Répond à la crise de la reproductibilité / Forges non pérennes
 - Pour tous les rédacteurs/utilisateurs de code source quel que soit leur secteur
- 70% de la production scientifique française dans Software Heritage vient de GitHub



Deux modes d'archivage :

- Archivage automatisé : collecte (moissonnage) systématique et actualisée des codes sources disponibles sur certaines plateformes (Github, Gitlab, etc.)
- Archivage manuel : dépôt ou signalement volontaire par les utilisateurs

- ✓ Archivage pérenne
- ✓ Délivre un identifiant unique (SWHID)
- ✓ Conserve l'historique technique



- Archive ouverte nationale pluridisciplinaire, gérée par le CCSD
 - Objectif : permettre un accès libre et gratuit aux résultats de recherche
 - Dépôt par la communauté scientifique française
- ✓ Grande visibilité des dépôts
 - ✓ Qualité des métadonnées (modération)
 - ✓ Archivage pérenne

Dépôt des logiciels intégré - 2018
(collaboration Software Heritage, Inria, CCSD)

Lien entre publication et logiciel intégré - 2021

Quoi ? Où ? Comment archiver un logiciel ?

Logiciel

Programmes informatiques

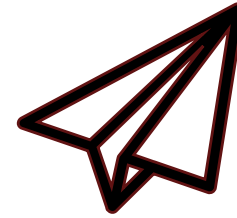
Code source

Instructions :

- compréhensible par les humains
- exécutables par les machines

modifiable, réutilisable, diffusable

➤ **Archivé**



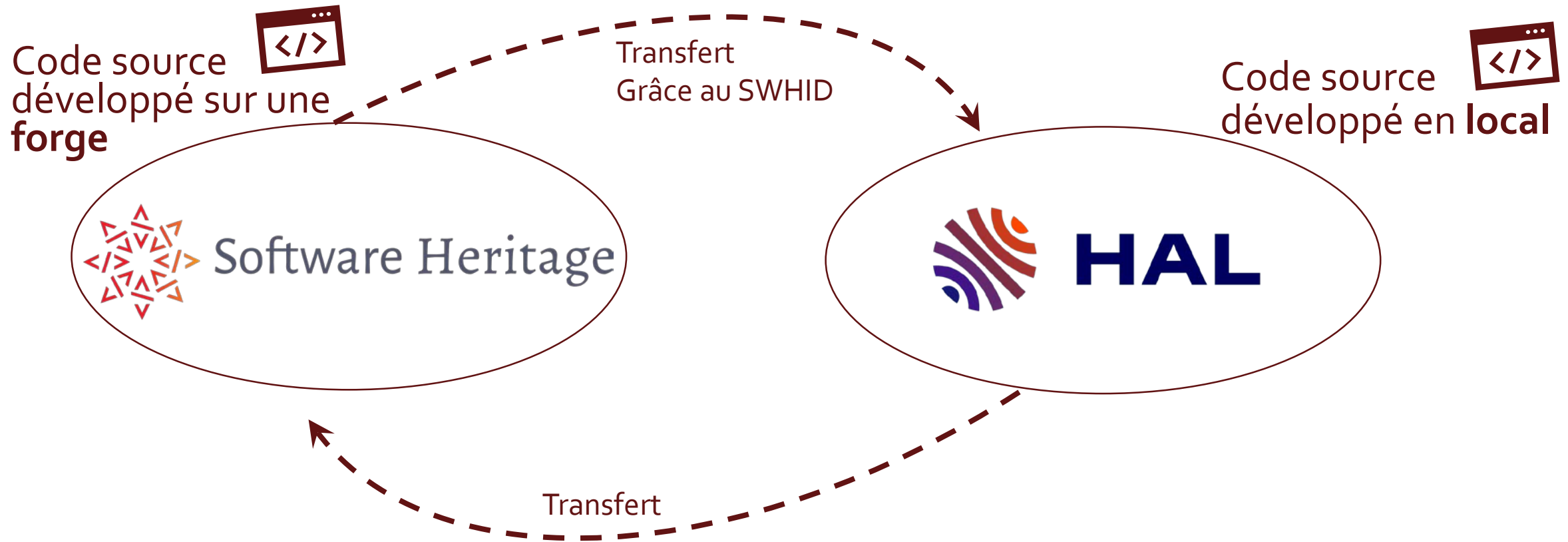
Software Heritage



HAL

- Garantissent la pérennité de l'archivage
- Services complémentaires

Complémentaires et connectées



Dépôt du code source dans Software Heritage et HAL



Meilleure **visibilité** du logiciel

Facilite **identification** et **citabilité**

Bonne description du logiciel, large éventail de métadonnées

Modération



Archivage pérenne du code source avec génération d'un identifiant unique (**SWHID**)

Accès à l'**historique de développement**

Point d'entrée unique vers des logiciels de plateformes différentes

Archivage automatisé à partir de forges publiques

Mise à jour régulière des collections

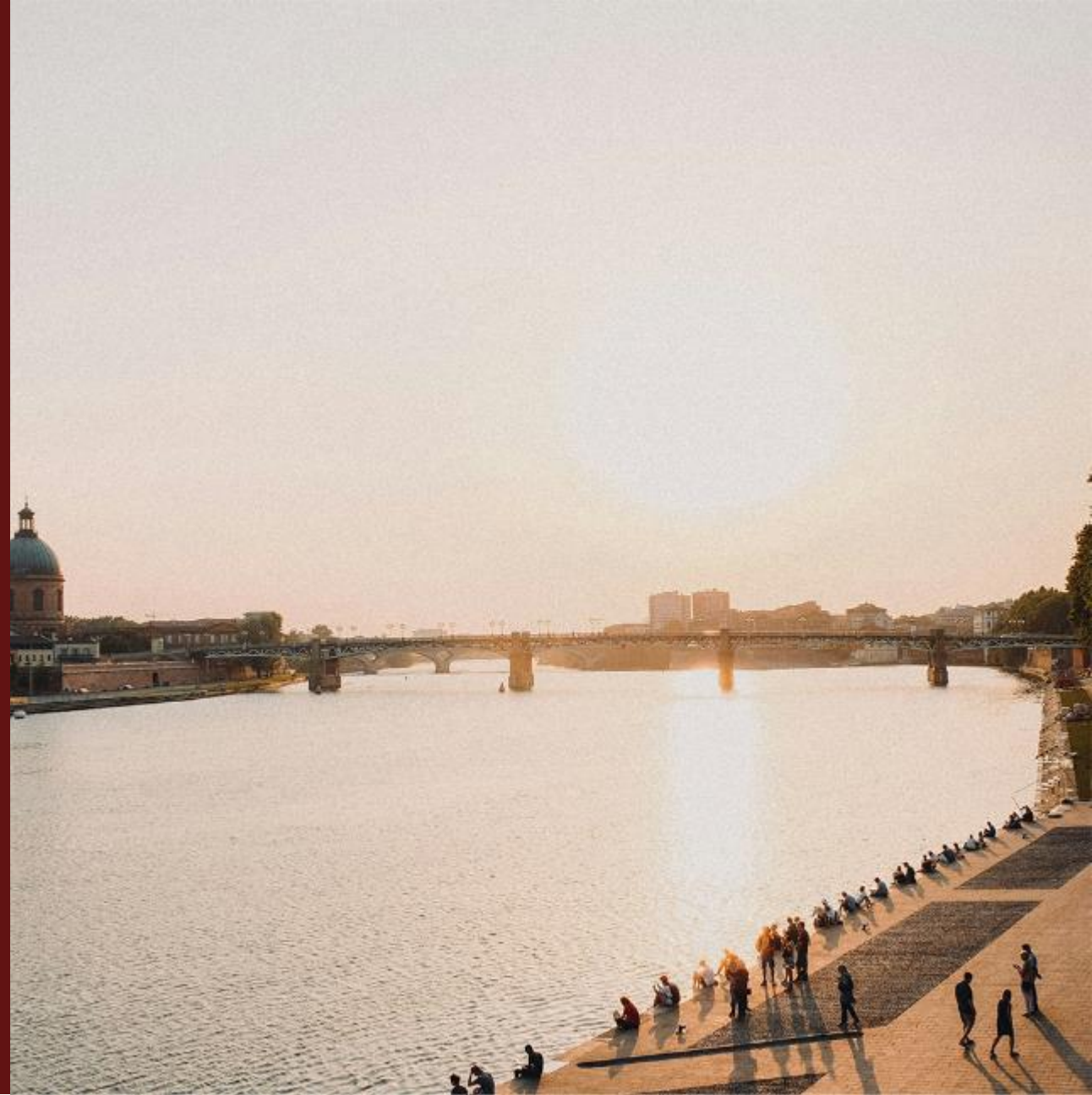
Les deux services sont complémentaires / Transfert d'un dépôt de l'un à l'autre

Bastien Plazolles

Ingénieur de recherche au GET

Equipe ER3 : géophysique et géodésie spatiale

Retour d'expérience sur l'archivage de codes sources dans HAL et Software Heritage



Des questions ?

**Merci pour votre
attention !**