



La place du logiciel dans le contexte de la science ouverte

Lucie Albaret, BAPSO
Violaine Louvet, LJK

■ Plan de la présentation

- **Enjeux de la science ouverte**
 - Quelques éléments de chronologie
 - Les textes fondateurs en France
 - Recherche reproductible
 - La place du logiciel
- **Quelques éléments historiques autour du logiciel libre**
- **Gouvernance de la science ouverte**
 - Gouvernance nationale de la science ouverte
 - Gouvernance locale de la science ouverte
 - Obligation des financeurs
 - Initiatives à l'international
- **Référencement et archivage**
 - Software Heritage
 - HAL et le catalogue logiciels
 - Liens publications, données, codes
- **Accompagnement sur le site**
 - La Cellule Data Grenoble Alpes
 - Site et contacts



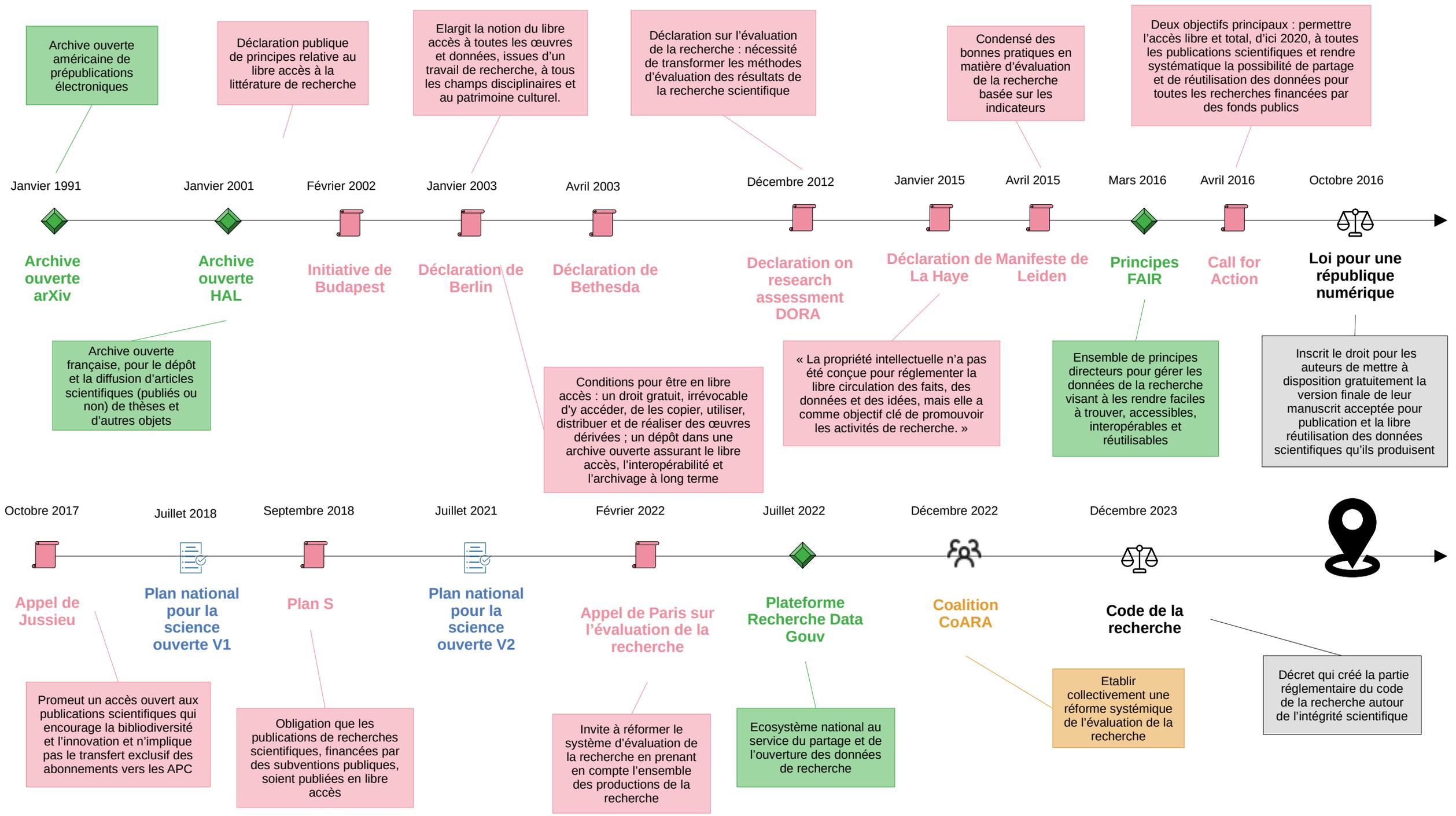
Enjeux de la science ouverte

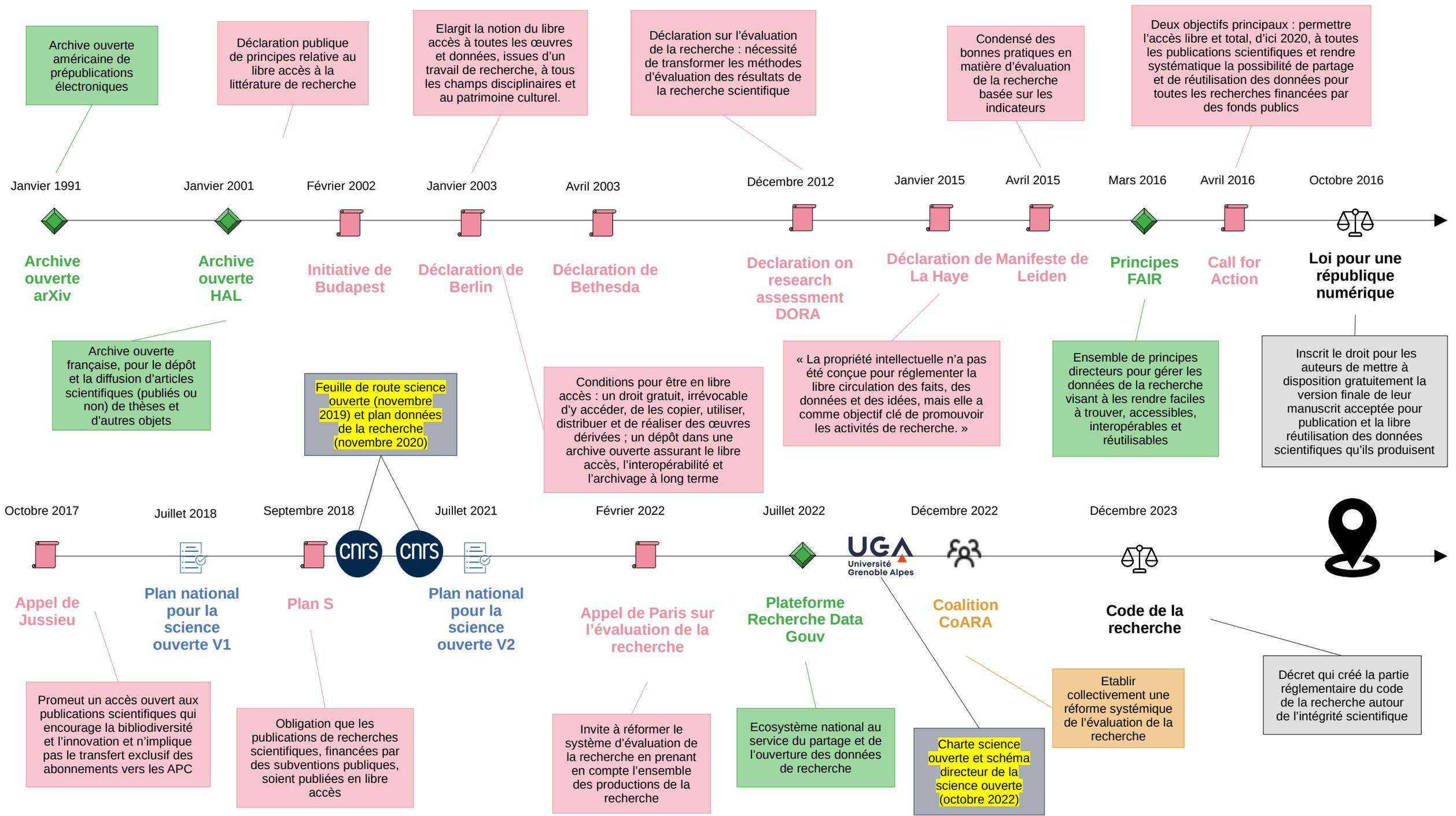
Science ouverte ?

La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique.

- Reproductibilité
- Science cumulative
- Confiance : transparence, éthique, intégrité
- Visibilité et traçabilité
- Accélération de la circulation des productions scientifiques
- Ouverture de nouveaux champs de recherche, de nouvelles collaborations
- Pérennité d'accès aux travaux scientifiques







■ Les textes fondateurs en France

Un mouvement relativement récent

- **2016** : Loi pour une république numérique. L'article 30 donne de nouveaux droits et de nouvelles libertés au chercheur pour mettre ses écrits en libre accès
- **2018 puis 2021** : Plan national pour la science ouverte
 - Généraliser l'accès ouvert aux publications
 - Structurer, partager et ouvrir les données de la recherche
 - **Ouvrir et promouvoir les codes sources de la recherche (2021)**
 - Transformer les pratiques pour faire de la science ouverte le principe par défaut
- **2023** : Décret sur l'intégrité du code de la recherche : interface entre science ouverte et intégrité scientifique. Les établissements publics :
 - Veillent à ce que les travaux de recherche qu'ils conduisent ou auxquels ils participent respectent les exigences de l'intégrité scientifique ;
 - Assurent la formation des personnels et des étudiants au respect de ces exigences ;
 - **Promeuvent la diffusion des publications en accès ouvert et la mise à disposition des méthodes, protocoles, données et codes sources associés aux résultats de la recherche ;**
 - Définissent les conditions de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en leur sein ;

Recherche reproductible

- Très liée aux questions d'éthique et à la transparence de la recherche
- **Quelques éléments de sémantique**
 - Reproductibilité : en utilisant le même code et les mêmes données, on aboutit à des conclusions/résultats/... identiques
 - Répliquabilité : en prenant un jeu de données différents et/ou une implémentation différente de la méthode, on aboutit à des conclusions/résultats/... équivalents

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Reproductibilité et science ouverte

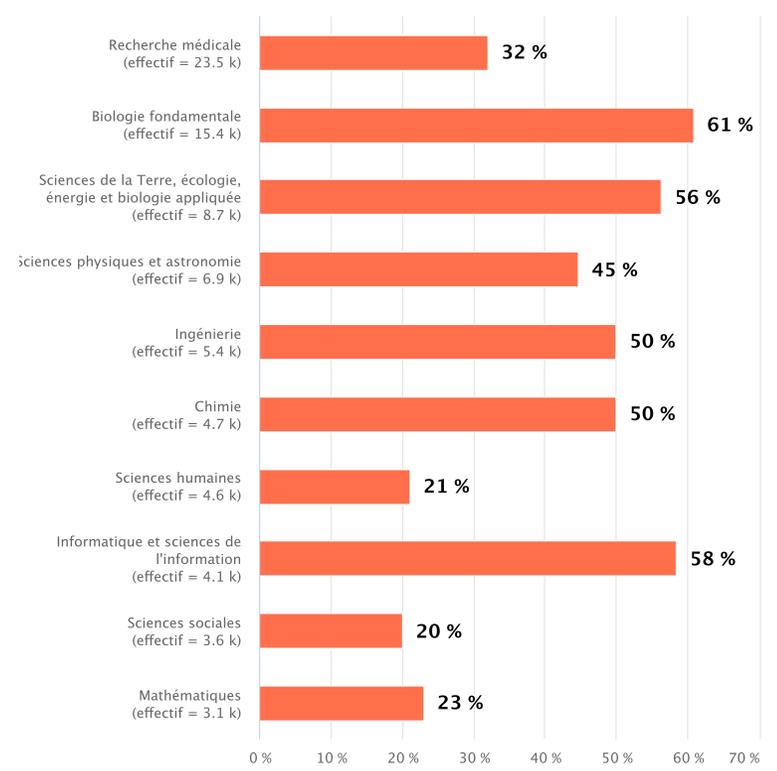
- De quoi a-t-on besoin ?
 - **article**, disponible sur une archive
 - **code**, disponible grâce à un lien web, un identifiant pérenne
 - **jeu de données**, disponible sur le web (une page ou un entrepôt)
- Les principes de la **science ouverte** offre un cadre idéal pour assurer la reproductibilité d'un travail. Mais est-ce suffisant ?
- Action de formation autour de la reproductibilité computationnelle du 1^{er} au 3 juillet :
<https://replications.sciencesconf.org/>



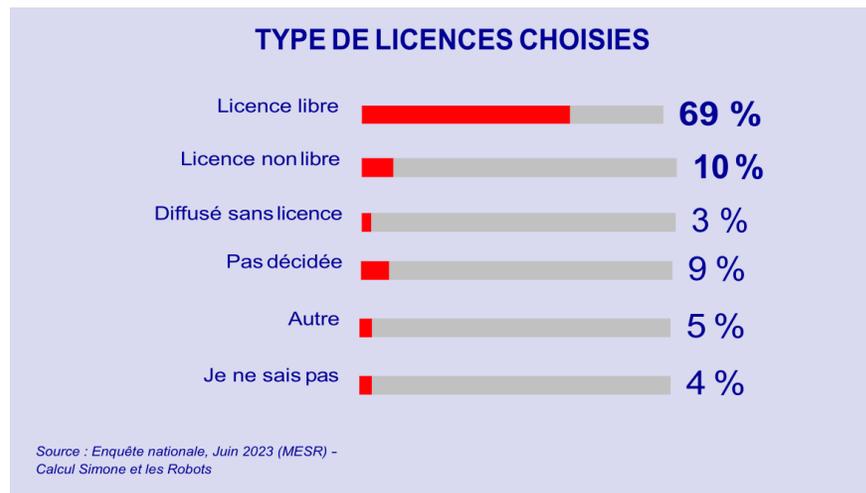
■ Importance du logiciel dans la recherche

- Le **baromètre science ouverte** intègre depuis 2 ans le sujet du logiciel et montre qu'il touche l'ensemble des disciplines
- **Rapport MESR** : état des lieux réalisé au printemps 2023 par le MESR sur la production et la valorisation des logiciels issus de la recherche.
 - 1331 réponses

Proportion de publications françaises parues en 2023 qui mentionnent l'utilisation de code ou logiciels par discipline



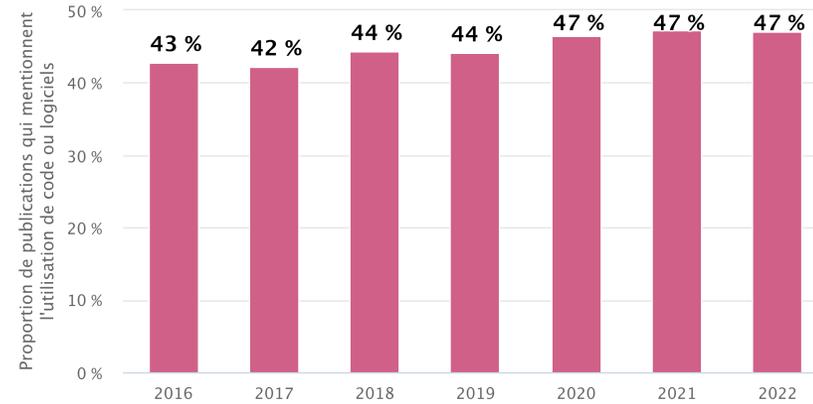
Proportion de publications qui mentionnent l'utilisation de code ou logiciel
Baromètre français de la Science Ouverte - CC-BY MESR, Sources : MESR,



Contexte à l'UGA

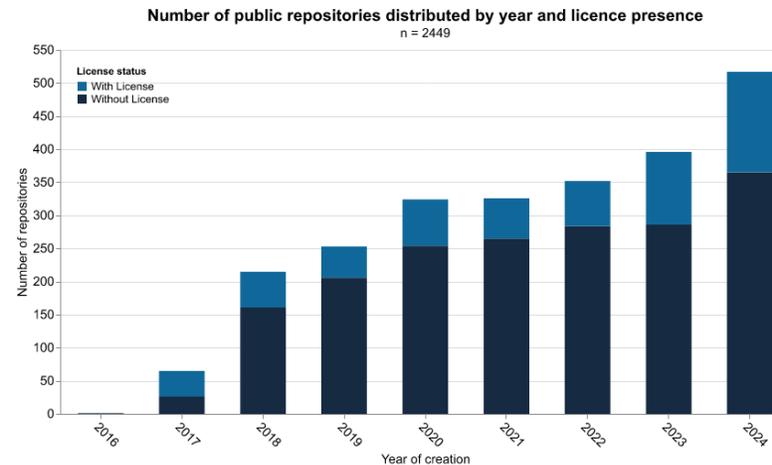
- Quelques indicateurs sur l'importance de l'activité de développement logiciels dans les laboratoires :
 - ▶ 141 notices logicielles dans le portail HAL UGA
 - ▶ 11000 utilisateurs, 20000 projets (2600 publics) sur la forge UGA
 - ▶ Le moniteur SO de l'UGA donne des informations importantes sur l'évolution sur la forge et la présence de licences
 - ▶ Près de la moitié des publications mentionnent l'utilisation de logiciels (BSO UGA)

Université Grenoble Alpes (UGA) : Proportion de publications de l'université Grenoble Alpes qui mentionnent l'utilisation de code ou logiciels par année de publication

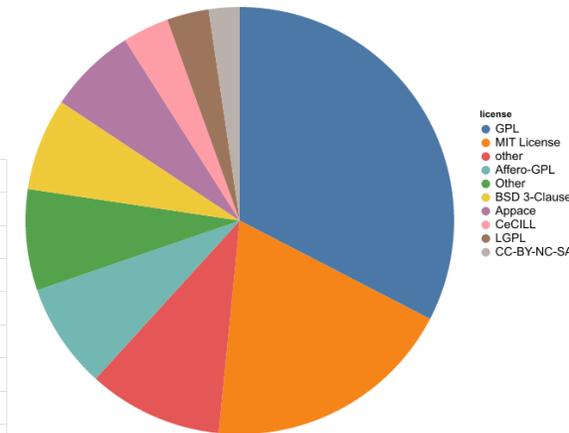


Baromètre français de la Science Ouverte - CC-BY MESR, Sources : MESR, Université Grenoble Alpes (UGA)

Most frequently used licences in public repositories
25% of repositories have a license



Moniteur SO UGA





Quelques éléments historiques autour du logiciel libre

■ Petite historique autour du logiciel libre

- Jusque dans les années 1970, seules **certaines entreprises** avaient les moyens financiers de s'acheter des ordinateurs et donc les logiciels associés.
 - Ces entreprises avaient un intérêt particulier à laisser les chercheurs et développeurs disposer de **ces logiciels** notamment pour les étudier et les modifier.
 - ♦ A cette période des débuts de l'informatique, le **partage de code source** entre chercheurs était une pratique universellement admise
 - ♦ Les programmes informatiques sont alors traités, dans les universités concernées, de la même manière que n'importe quelle **information scientifique** : mis à la disposition de tout un chacun pour étude, exploitation et amélioration
- Progressivement, avec le développement des ordinateurs personnels, **l'importance économique** des logiciels apparaît.
 - ♦ Bill Gates (création de Microsoft en 1975) publie en 1976 An Open Letter to Hobbyist dans laquelle il dénonce la faible protection des créations des développeurs.
 - ♦ En Octobre 1976, le Copyright Act est adopté aux Etats-Unis, protégeant les logiciels par le droit d'auteur américain.
 - ♦ C'est la naissance des **logiciels propriétaires**.

Philosophie libre

- L'échange libre de logiciels a donc existé depuis les **débuts de l'informatique**, sans qu'on ait défini les principes de ces partages.
- En particulier dans la communauté des **hackers** (« bidouilleurs »), désignant originellement ces jeunes étudiants en informatique du MIT, génies de la programmation
 - ➔ Un des principes de l'« éthique hacker » est que **toute information est par nature libre**

Richard Stallman

Informaticien au MIT, membre de la communauté des hackers, frustré de ne pas pouvoir modifier le logiciel qui contrôlait l'imprimante de son labo, il est alors convaincu de la nécessité pour les gens de pouvoir librement modifier le logiciel qu'ils utilisent.



■ Naissance du logiciel libre : Free Software Foundation

- Richard Stallman fonde la **Free Software Foundation** (FSF) en 1985.
- Cet organisme formalise les 4 grands principes du libre dans des **licences**, textes juridiques qui fixent les conditions pour exploiter un logiciel ou une oeuvre intellectuelle.
- La FSF soutient le développement du projet GNU depuis le début.
 - Le projet GNU concerne le développement d'un système d'exploitation complètement libre.
 - Sa version fonctionnelle s'appuie sur le noyau linux, développé indépendamment par Linus Torvalds

Les 4 libertés fondamentales du logiciel libre

- Possibilité d'utiliser l'oeuvre, pour tous les usages ;
- Possibilité d'étudier l'oeuvre ;
- Possibilité de redistribuer des copies de l'oeuvre ;
- Possibilité de modifier l'oeuvre et de publier ses modifications.

Logiciel libre ≠ logiciel gratuit

■ Logiciel libre vs open source

- Principale différence se situe dans les **valeurs que chacun porte**
 - Le mouvement du libre est un **mouvement social** qui soutient des valeurs philosophiques et politiques.
 - L'open source est plus récent et met en avant la **méthodologie de développement et de diffusion** du logiciel.
- Le logiciel libre est forcément open source. Un logiciel peut être open source sans être libre au sens de la FSF.



■ Logiciel libre et science ouverte

- Le logiciel libre est donc avant tout une **philosophie du partage des communs, de la connaissance**.
- Cette philosophie est naturellement née dans le **milieu universitaire**.
- Le mouvement de la science ouverte, en particulier autour des données, ne fait finalement que reproduire ce qui existe dans le monde du logiciel depuis le début.
- **Il n'empêche que le logiciel est régulièrement « oublié » des discours sur la science ouverte**
- Le logiciel libre est cependant **précurseur de la science ouverte** en mettant en avant dès le début de l'informatique la nécessité du partage des communs





Gouvernance de la science ouverte

Au niveau national

Comité pour la science ouverte, COSO

Assure la mise en œuvre de la [politique nationale sur la science ouverte](#)

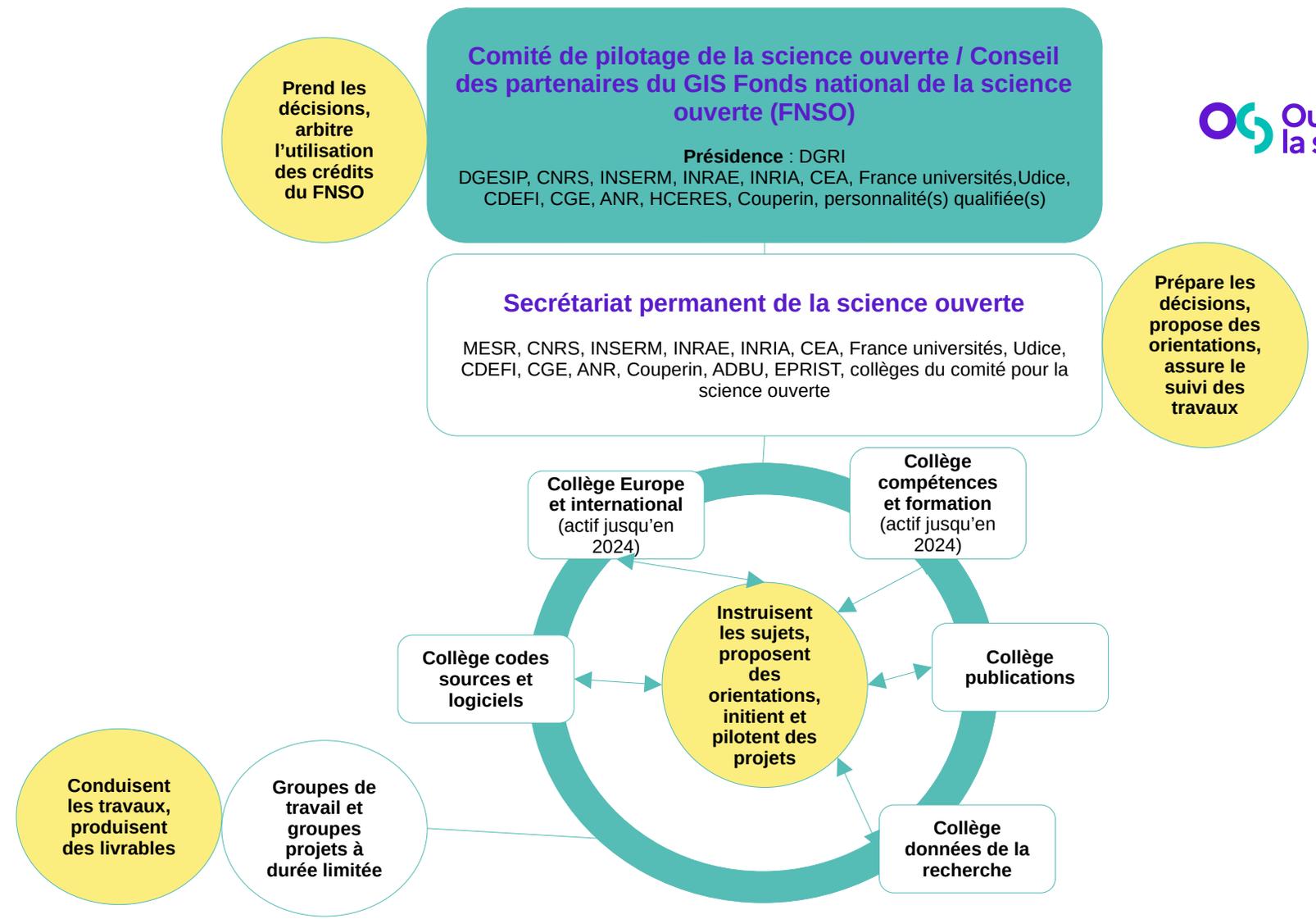
Regroupe les acteurs de l'enseignement supérieur et de la recherche pour accompagner de manière dynamique et coordonnée la mise en œuvre de la politique nationale de science ouverte.



<https://www.ouvrirlascience.fr>

- **Comité de pilotage de la science ouverte** : arbitrages stratégiques
- **Secrétariat permanent pour la science ouverte** : préparation des travaux du comité de pilotage et mise en œuvre des décisions
- **Collèges et groupes d'expertise** : instruisent les sujets et proposent des orientations.
 - Collèges : publications, données de la recherche, Europe et international, compétences et formation, codes sources et logiciels (depuis 2022)
 - Groupes d'expertise : édition scientifique ouverte, juridique.

Organigramme du COSO



Le collège codes sources et logiciels

Répondre aux ambitions du PNSO2

- Définir et promouvoir une **politique en matière de logiciels libres**
 - Produire une charte nationale pour les logiciels libres issus de l'enseignement supérieur, de la recherche et de l'innovation
 - Développer le lien entre les données et les logiciels grâce au réseau des **administrateurs des données, des algorithmes et des codes sources** dans les établissements.
- Émettre des recommandations auprès des **organismes financeurs** pour accompagner au mieux le développement logiciel.
 - Développer les modèles économiques du logiciel libre et les faire connaître au sein des services de valorisation
 - Soutenir l'infrastructure Software Heritage et la recommander pour l'archivage et le référencement du code source
- Reconnaître le code source comme une **contribution à la recherche**
 - Créer un prix du logiciel libre de la recherche
 - Accorder une plus grande reconnaissance à la production de logiciels dans la carrière des chercheurs et du personnel de soutien à la recherche
- Créer un écosystème qui relie le **code, les données et les publications**
 - Développer une coordination adéquate entre les forges logicielles, les archives de publications ouvertes, les dépôts de données et le secteur de l'édition scientifique

■ Missions et composition du collège

- Suivre, analyser et produire des avis sur les **sujets d'actualité** relatifs aux codes source et aux logiciels dans le cadre de la science ouverte
- Impulser, réaliser ou piloter des **projets opérationnels**, qui contribuent aux chantiers identifiés dans l'axe logiciel du PNSO2
- Se **coordonner** avec les autres collèges, la DINUM, et l'écosystème national et international
- **Diffuser** ses réflexions et les bonnes pratiques via la production de notes, recommandations, documents collaboratifs, etc.

Composition

- 41 personnes
- Des profils variés :
 - Recherche
 - Technique
 - Information scientifique et technique
 - Valorisation

■ Administrateurs des données, des algorithmes et des codes sources

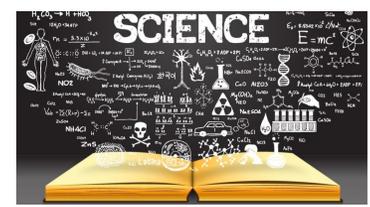
- Une **politique nationale** : un AMDAC par ministère, un ADAC par établissement
- **Rôles des ADAC**
 - Élaborer la **politique de l'établissement**
 - Accompagnement
 - Infrastructure
 - Reconnaissance
 - Incitation ...
 - Evaluer le **capital données** de l'établissement
 - Identifier
 - Valoriser
 - Travailler au **partage des données**
 - Analyse de risques
 - Opportunités
 - Coordination des services
 - **Eclairer la gouvernance** de l'établissement dans ses prises de décisions
 - Participation au **réseau national des ADAC**

Gouvernance au niveau local

- Au niveau de la **gouvernance**
 - **Portefeuille SO** : Caroline Bertonèche (VP relations internationales)
 - **Chargé de mission SO** : Lucie Albaret (BAPSO)
 - **ADAC** : Glenn Cougoulat (Directeur GRICAD)
 - **Commission SO** regroupant tous les établissements du site et les acteurs de la SO
 - Politique SO décrite dans la charte et le schéma directeur votés en 2022
- Les **textes** votés côté UGA :
 - Le **schéma directeur de la science ouverte**, voté en octobre 2022, 4 défis et 21 actions
 - Défi n°1 : Rendre plus efficace la stratégie données et codes de recherche de l'UGA grâce à une gouvernance adaptée et à une coordination des acteurs
 - Défi n°2 : Vers de meilleures pratiques : le respect des réglementations et des recommandations locales, nationales, européennes et l'engagement envers le principe des données ouvertes
 - Défi n°3 : Améliorer l'accompagnement tout au long du cycle de vie des données, codes et logiciels
 - Défi n°4 : Améliorer les éditions et publications ouvertes
 - La **charte science ouverte**, votée en octobre 2022, présente les recommandations et les engagements de l'UGA concernant la science ouverte.

Au niveau projet, les obligations des financeurs

- Obligations
- Recommandations



Publications

- Rendre disponible en accès ouvert immédiat sous une licence Creative Commons CC-BY ou équivalente, toute publication scientifique issue d'un projet.
- Dépôt du texte intégral dans HAL ou dans une autre archive ouverte.
- Obligation de diffuser en libre accès toutes les formes de publications
- Accès ouvert immédiat sans embargo par le dépôt dans une archive ouverte.
- Frais de publications dans les revues hybrides non éligibles.



Données

- Plan de gestion de données
- « Aussi ouvert que possible, aussi fermé que nécessaire »
- Plan de gestion de données.
- Gestion des données FAIR.
- Partage des données dans des entrepôts de référence selon le principe « aussi ouvert que possible, aussi fermé que nécessaire ».



Codes

- Mise à disposition sous licence libre, archivage des codes dans Software Heritage et référencement dans HAL
- Accès aux résultats nécessaire pour la validation des conclusions scientifiques.
- Accès ouvert aux logiciels

■ Initiatives au niveau international

- **EOSC, European Open Science Cloud** : initiative européenne qui vise à fédérer les services existants pour faciliter la gestion et la FAIRisation des données de recherche. Quelques (assez peu) projets intégrant la problématique du logiciel ciblés essentiellement sur la diffusion (citation, archivage, métadonnées) : ARDC pour Archive, Reference, Describe, Credit
- **RDA, Research Data Alliance** : organisation internationale sur les données de la recherche, incluant plusieurs groupes de travail sur le logiciel (FAIR4RS, Reproducibility, Policies in Research Organisations ...)
- **ReSA, Research Software Alliance** : organisation internationale sur les logiciels de recherche, pour faire avancer l'écosystème autour du logiciel à travers différentes task forces.
- Des initiatives dans plusieurs pays comme le **Software Sustainability Institute** (UK), l'**Australian Research Data Commons** (Australie), mais aussi la NASA, le CERN, ...



Référencement et archivage

■ Archivage des codes sources

Pourquoi archiver des codes sources ?

- Le code source est **fragile** :
 - Obsolescence des formats, problème matériel, dépendances à des outils (forge par exemple) qui disparaissent ...
 - La perte des codes ayant été utilisés pour de la production scientifique arrive malheureusement régulièrement
 - Les logiciels sont un des **piliers des processus de recherche**, au côté des publications et des données et il est essentiel de les préserver
-
- Archivage ≠ stockage ≠ outil de développement
 - Archivage ≡ **préservation sur du long terme**

■ Software Heritage

- Initiative dont l'objectif est de construire une **archive universelle des codes sources**
- En les collectant, les préservant et les partageant sur le **long terme**
- Lancée en 2016 par INRIA et soutenue par l'UNESCO
- Collecte de l'**intégralité des logiciels disponibles publiquement** sous forme de code source.
- Depuis des **plateformes d'hébergement de codes**, comme GitHub, GitLab.com ou Bitbucket, et des archives de paquets, comme Npm ou Pypi ...
- <https://archive.softwareheritage.org/>



Collect
Preserve
Share

■ Signalement des codes de la recherche

Pourquoi signaler des codes ?

- Assurer la description
- Faciliter la recherche (par domaine scientifique par exemple)
- Permettre la citation
- Valoriser les logiciels

- Importance des **métadonnées** pour :
 - Décrire de façon précise et contrôlée un logiciel
 - Identifier son usage
 - Spécifier le contexte
 - Créditer plus facilement les créateurs
- Importance des **identifiants pérennes** pour :
 - Garantir un lien stable à la ressource en ligne

Le triptyque forge - Software Heritage - HAL

- **Complémentarité** des deux plateformes HAL et Software Heritage
 - Grande **visibilité des logiciels** dans une démarche de science ouverte via HAL
 - **Archivage pérenne** via Software Heritage avec moissonnage automatique des forges
 - **Modération** des métadonnées pour assurer leur qualité
 - Différents formats d'export pour **faciliter la citation**



←
Moissonnage



Collect
Preserve
Share



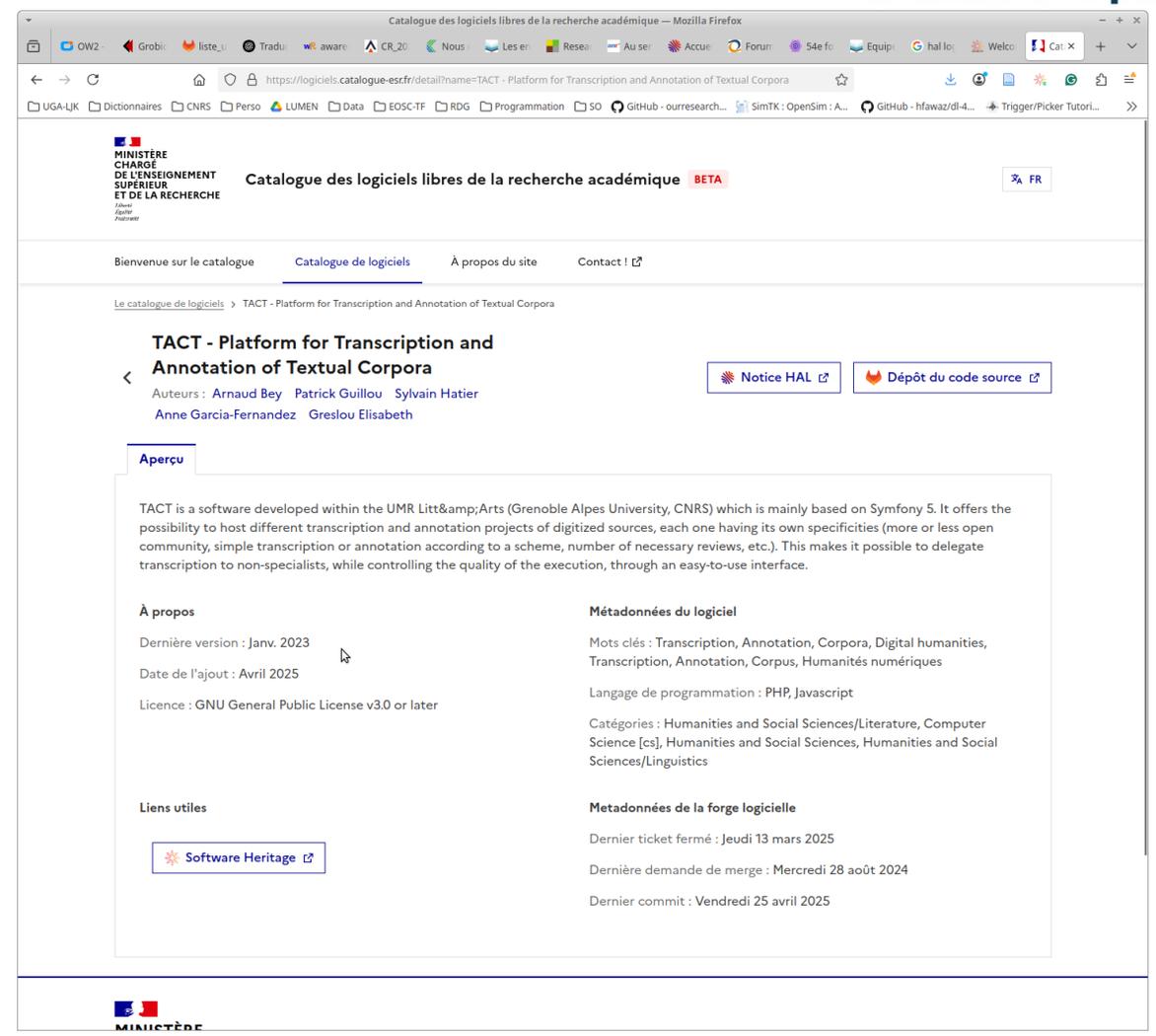
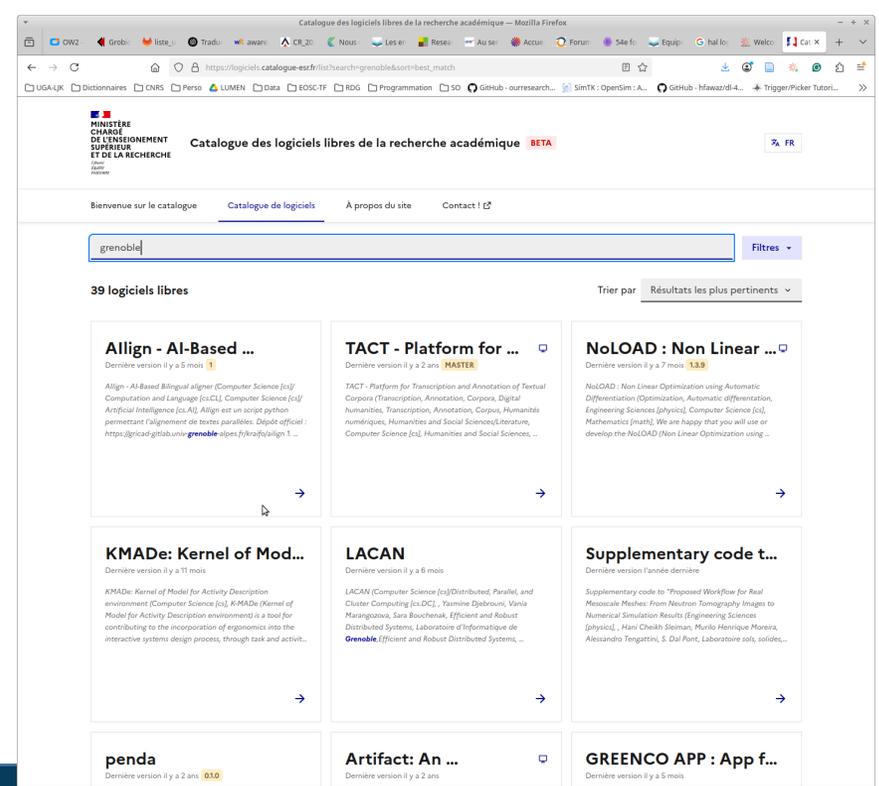
Notice logicielle
avec lien SWHID



HAL
science ouverte

Catalogue des logiciels de l'ESR

- Projet porté par le collège logiciels et codes sources du COSO
- **Extraction des notices logicielles de HAL** : modérées, alignées avec des référentiels consolidés
- **Enrichissement** des entrées avec des informations extraites des dépôts des forges reflétant le dynamisme du développement



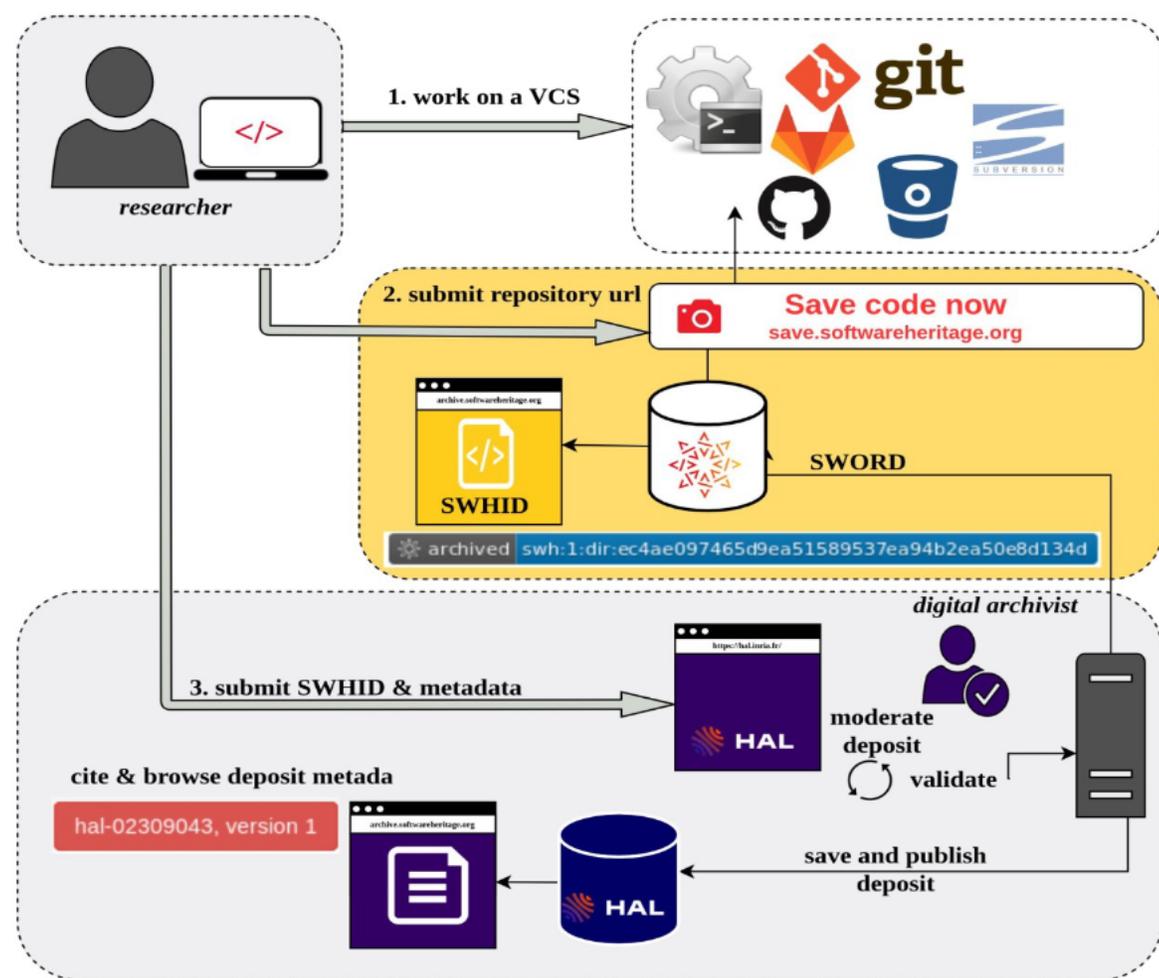
■ Le lien publications - données - codes

- L'utilisation de HAL et SWH facilite la citation du logiciel :
 - ▶ Plusieurs **formats d'exports** possibles (BibTeX, TEI, DCTERMS, codemeta.json, etc.)
 - ▶ Permet de créer les liens nécessaires entre les **publications, les données et les codes**
 - ▶ Indispensable pour la **reproductibilité**. En particulier, le SWHID désigne précisément un logiciel dans son contexte (version, commit, ...)
 - ▶ Important pour **créditer** correctement les auteurs et contributeurs

Le lien publications - données - codes

Comment faire concrètement ?

- Vérifier l'archivage sur SWH
- Dans HAL
 - Choisir le type dépôt « logiciel »
 - Indiquer le SWHID contextualisé
 - Vérifier et compléter les métadonnées
 - Auteurs/affiliations
 - Disciplines
 - Financement
 - Ressources associées
 - Publications
 - Données
 - ...



■ Le lien publications - données - codes, quelques exemples

- Simon Fernandez. WHOIS Right? - Dataset Analysis. 2023
 - Notice logicielle sur HAL
 - Entrée dans le catalogue logiciels
 - Jeu de données sur Recherche Data Gouv
- Sylvain Hatier, Arnaud Bey. LST-SHS - Plateforme d'exploration lexicale. 2023
 - Notice logicielle sur HAL
 - Entrée dans le catalogue logiciels



Accompagnement sur le site

■ La Cellule Data Grenoble Alpes

- Coordonnée par GRICAD et la BAPSO, regroupe des compétences complémentaires pour répondre à tous les questionnements et déployer les outils nécessaires pour satisfaire les besoins des communautés scientifiques du site de Grenoble dans le domaine des données et des codes de la recherche :
 - Informatique, calcul
 - Information scientifique et technique
 - Juridiques
- Labellisée atelier de la donnée de l'initiative nationale Recherche Data Gouv





OFFRE DE SERVICES CELLULE DATA GRENOBLE ALPES



Science ouverte
Grenoble Alpes



CC BY-Nicolas Lambert/GRICAD/CNRS

■ Equipe HAL UGA

- Accompagnement au dépôt dans HAL
- Expertise sur la qualité du signalement des publications
- Aide juridique
- Création de collections laboratoire ou projet
- Aide à l'export
- Administration du **portail HAL de l'UGA**



HAL
science ouverte

■ Site et contacts

- **Site science ouverte** : informations sur l'ensemble des sujets (publications, données, codes), actualités, évènements, supports des formations et des séminaires, mails de contact, fiches pratiques ...
- Contact publications :
 - équipe HAL UGA : hal.support@univ-grenoble-alpes.fr
 - Questions autour des APC, revues prédatrices ... :
sos-publications@univ-grenoble-alpes.fr
- Contact données et codes : Cellule Data Grenoble Alpes
 - support données : sos-data@univ-grenoble-alpes.fr
 - support codes : sos-codes-recherche@univ-grenoble-alpes.fr