



**MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION**

*Liberté
Égalité
Fraternité*

CODES SOURCES ET LOGICIELS

12 mai 2023 – Média Normandie

Collège « Codes sources et Logiciels » du CoSO

Violaine Louvet, Grégory Miura, co-pilotes GT1

De quoi parle-t-on ?

- **Éléments issus du rapport Bothorel**

- « Un code source peut être défini comme un ensemble d'instructions exécutables par un ordinateur. »

En fait, le code source est lisible par l'humain et doit être transformé pour être exécutable par l'ordinateur

- « Un algorithme n'est pas nécessairement informatisé. »
- « De manière simplifiée, l'algorithme est une recette de cuisine, et le code sa réalisation concrète »

- **Code source vs exécutable**

- Un logiciel fonctionne grâce à du code exécutable, compréhensible uniquement par des machines (binaire).
- L'« âme » d'un logiciel est dans son code source, c'est-à-dire dans les instructions telles qu'elles sont rédigées pour être lisibles par l'humain.
- Seul le code source permet d'accéder aux informations techniques et scientifiques

En résumé

- **Algorithme** : décrit le déroulé pour la résolution d'un problème posé.
- **Code source** : mise en oeuvre et formalisation de l'algorithme dans un langage informatique (par exemple python, C++, java ...). C'est un (ou plusieurs) fichier(s) texte.
- **Exécutable** : traduction du code source (en général via un compilateur ou un interpréteur) en code binaire compréhensible par l'ordinateur.
- **Logiciel** : en général, l'ensemble environné comprenant le code source et/ou l'exécutable, et le plus souvent la documentation, des exemples d'utilisation ... et évidemment la licence associée.

Valeur « documentaire » du logiciel de recherche

- **Un enseignement sur la méthode**

- La place du logiciel entre l'individu et un contenu (document et / ou donnée) dans un contexte scientifique, **un élément d'épistémologie en voie de banalisation**

- **Un enjeu collectif de missions et de compétences**

- **Une confrontation nécessaire pour l'adaptation de nos pratiques professionnelles, l'ouverture à une autre « culture métier », la recherche de standards adaptés** visant à l'interopérabilité de nos services de description – recherche – accès - pérennisation

- **Une histoire à l'œuvre**

- Du jeu vidéo à l'intelligence artificielle générative, le traitement logiciel se révèle être **un phénomène global déjà ancien au développement rapide, doté de perspectives aux limites inconnues**

L'objet logiciel de la recherche

- **Définition proposée par le Collège Codes sources et logiciels :**

- Les logiciels de recherche sont développés pour répondre à des **besoins spécifiques de la science**. Ils sont conçus, maintenus, et utilisés par des **scientifiques (chercheurs et ingénieurs)** et institutions de recherche, éventuellement dans une dimension **internationale**.

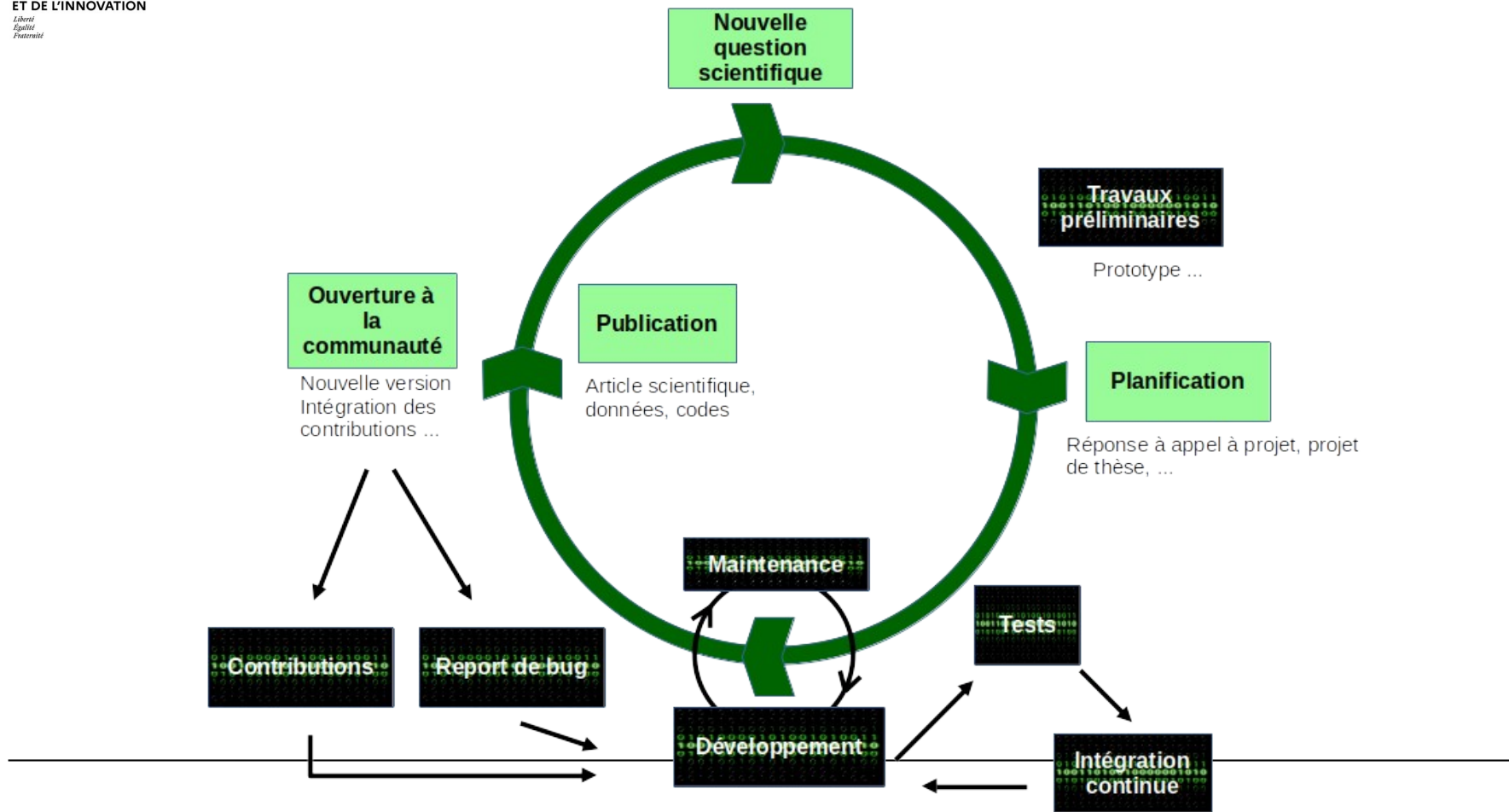
Ils peuvent découler de travaux de recherche comme ils peuvent les favoriser, notamment par des **publications avant/sur/autour/avec le logiciel**.

Ceux-ci peuvent se formaliser de différentes façons (une plateforme, un intergiciel, un workflow ou une bibliothèque, module ou greffon d'un autre logiciel) et être ainsi en **interaction dans un écosystème** ou au contraire plus **autonomes**.

Codes de recherche ≠ données de recherche

- Les données de recherche sont plutôt passives, les codes sont intrinsèquement **vivants**
 - On ne change en général pas les données, collectées dans un contexte bien défini
 - On change éventuellement la façon dont on les traite et on les analyse (grâce à des codes)
 - Les codes sont associés à une (ou des) **action(s)** : création de connaissances, transformation d'informations, visualisation, ...
 - Le code peut être réutilisé tel que, en reproduisant son environnement et toutes ces dépendances mais on a surtout envie de le **modifier pour l'adapter** à nos besoins propres ou **l'enrichir de nouvelles fonctionnalités**
- Les codes s'appuient sur des **dépendances et tout un environnement logiciel et matériel** qui évolue sans cesse
 - Cela complexifie les questions de **reproductibilité**
- Les codes représentent un travail de création, et correspondent à un **cadre juridique différent de celui des données**

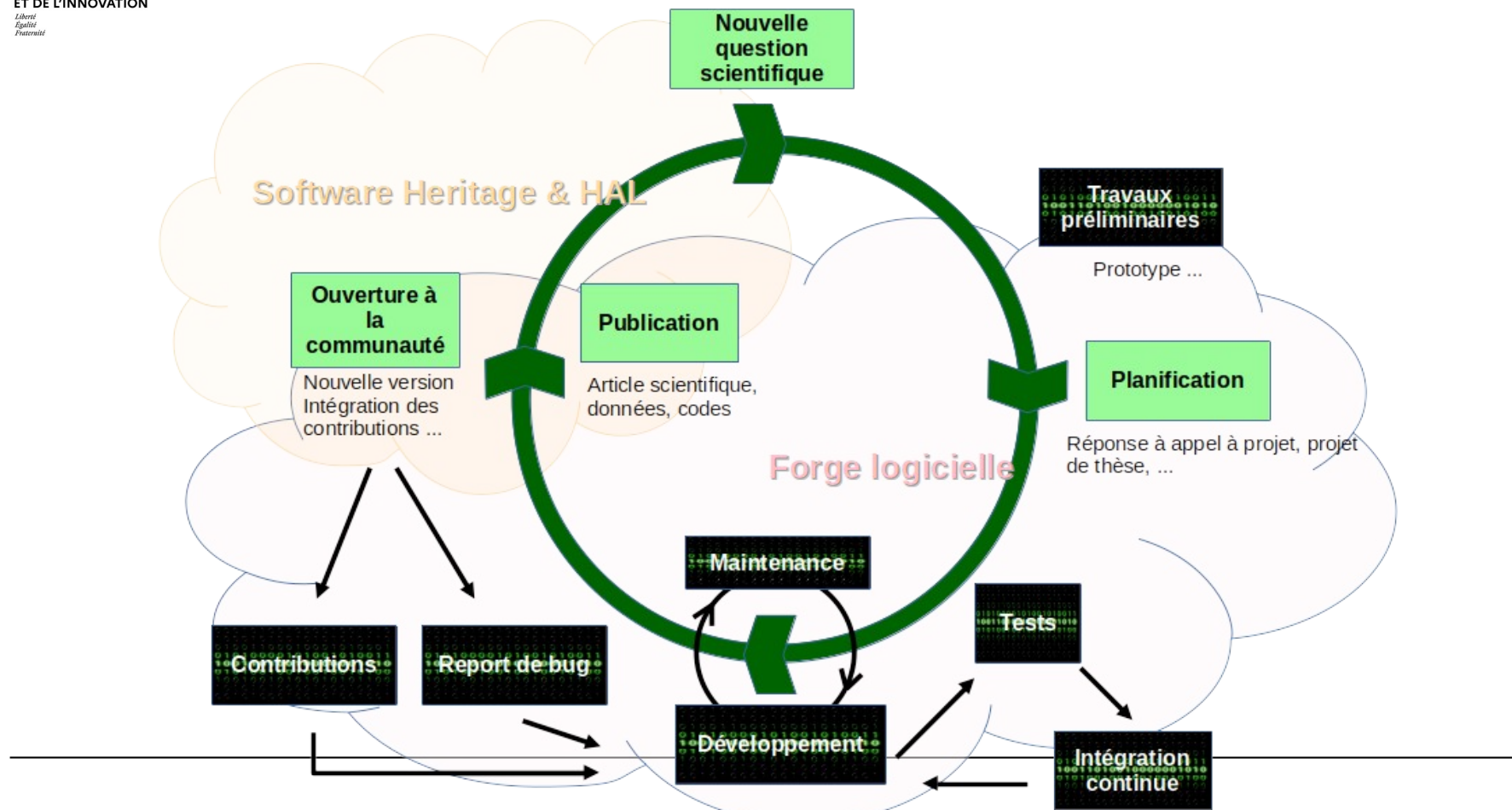
Cycle de vie d'un code de recherche



Forges logicielles

- Système de gestion de rédaction, de partage et de maintenance collaborative de fichiers texte
 - Intégrant de nombreux outils : gestion de versions (git), rédaction en ligne, gestion de tâches, gestion de documentation, suivi de bugs, forums, ...
- **Coeur névralgique de tout développement**
 - Facilite la mise en œuvre de bonnes pratiques de développement
 - Absolument indispensable quand on développe à plus de un mais particulièrement utile aussi pour les développements individuels
 - Simplifie aussi la diffusion du logiciel, son référencement, son archivage ...
 - En particulier, c'est le seul endroit où les **informations sont constamment à jour**
- Ne pas confondre les **forges logicielles et les archives de codes**
 - Software Heritage n'est pas une forge logicielle, c'est une archive de codes sources permettant la pérennisation des codes sources
 - mais pas pour leur développement collaboratif
 - Par contre SWH s'appuie sur les forges existantes
 - A l'inverse, les forges logicielles n'assurent pas la pérennisation des codes sources

Outils dans le cycle de vie d'un code de recherche



Cadre juridique du logiciel

- Le logiciel est protégé par le **droit d'auteur**
 - Droits moraux attachés à l'auteur
 - Droits patrimoniaux qui couvrent le monopole de l'exploitation, propriétés de l'administration employeur de ou des auteurs
 - Contrairement aux autres droits d'auteurs, il y a une « **dévolution automatique des droits patrimoniaux à l'employeur** »
 - Depuis le 15 décembre 2021, c'est vrai aussi pour les **stagiaires**
- Le droit d'auteur s'applique dès la **création du logiciel**, il faut pouvoir **dater** cette création
 - Par tout moyen d'horodatage, ou via un tiers par exemple dépôt APP (Agence pour la Protection de Programme)

La question des licences

- S'il n'y a pas de droit explicitement donné à travers une licence, utiliser un logiciel relève de la **contrefaçon**.
- **Pas de licence == tous droits réservés**
- Deux grands types de licence :
 - **Licences libres ou Open Source**, termes plus ou moins similaires
 - Une licence libre **ne veut pas dire libre de droit**, bien au contraire !
 - **Licences propriétaires**, donc non libre c'est-à-dire que seul l'auteur ou l'ayant droit du logiciel peut le modifier.

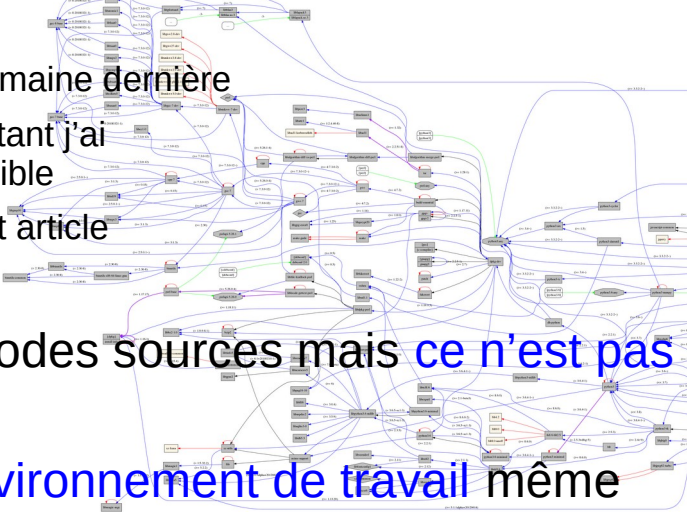
Le logiciel libre

- L'**ouverture des logiciels** développés dans le cadre de la recherche publique est un point essentiel de la **science ouverte**
- Le logiciel libre définit **4 types de libertés pour l'utilisateur** :
 - Liberté **d'exécuter** le programme, pour tous les usages.
 - Liberté **d'étudier** le fonctionnement du programme, et de l'adapter à vos besoins. Accès au code source condition requise.
 - Liberté de **redistribuer** des copies.
 - Liberté **d'améliorer** le programme et de **publier** vos améliorations, pour en faire profiter toute la communauté. Accès au code source condition requise.

Les principales licences libres

- **Sans copyleft** : la licence initiale ne s'impose pas. Permission de redistribuer et de modifier, mais aussi d'y ajouter des restrictions.
 - Exemple : BSD, Apache2, MIT
- **Copyleft faible** : la licence initiale reste, des ajouts peuvent avoir une autre licence.
 - Exemple : GNU LGPL
- **Copyleft fort** : la licence initiale s'impose sur tout. Licence dite **contaminante**.
 - Exemple : GNU GPL, EUPL

Reproductibilité

- Mes résultats ont changé ! Pourtant je n'ai touché à rien
 - Mon code ne marche plus ! Pourtant il fonctionnait la semaine dernière
 - Impossible de retrouver les résultats de cet article, pourtant j'ai réimplémenté l'algorithme car le code n'était pas disponible
 - Impossible d'installer ce code qui est référencé dans cet article
- 
- Il est évidemment **nécessaire d'ouvrir** les codes sources mais **ce n'est pas suffisant** !
 - Il faut s'assurer de retrouver le **même environnement de travail** même longtemps après (alors qu'OS, bibliothèques, compilateurs ... ont évolué)
 - **Dépendances, piles logicielles** : la complexité peut être rapidement très importante
 - Un simple « import matplotlib » cache une large chaîne de dépendances

Problématique du référencement

- Pourquoi archiver ?
 - 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
- Pourquoi signaler ?
 - Assurer la description
 - Faciliter la recherche / Rechercher facilement (par domaine scientifique par exemple) / Favoriser la réutilisation
 - Permettre la citation
 - Valoriser les logiciels
- Les logiciels sont un des piliers des processus de recherche, au côté des publications et des données et il est essentiel de les préserver
 - Inscrire cette production dans la dynamique science ouverte actuelle en ayant conscience que le logiciel de recherche est depuis toujours ancré dans l'open source

Logiciels de recherche et science ouverte



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**

Créer un **Collège des codes sources et des logiciels** au sein du Comité pour la science ouverte.

« L'ouverture des codes sources est un défi majeur pour la **reproductibilité** des résultats scientifiques. »

« La distribution des logiciels sous **licences libres** sera favorisée. »

Les axes du Plan National Science Ouverte v2

- 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** à travers un réseau d'administrateurs des données et des codes de recherche dans les différentes universités et organismes performants de recherche
 - Développer les **modèles économiques** du logiciel libre et les faire connaître au sein des services de valorisation
 - Soutenir le projet *Software Heritage* et le 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 du collège Codes sources et logiciels

- **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
 - Production de notes, recommandations, documents collaboratifs, etc.
- **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

Cinq thèmes identifiés

- Thème 1 : **Identification et mise en avant de la production logicielle de l'ESR**
 - Catalogue des logiciels
- Thème 2 : **Outils et bonnes pratiques techniques et sociales**
 - Forges logicielles
- Thème 3 : **Valorisation et durabilité**
 - Politiques de valorisation
- Thème 4 : **Liaison et animation nationale, Europe, international**
 - Politique logicielle
- Thème 5 : **Reconnaissance et carrières**
 - Prix du logiciel libre de recherche

Thème 1 : Identification et mise en avant de la production logicielle de l'ESR

- **Objectifs** : recommandations pour la construction d'un catalogue national des logiciels de la recherche
 - **Programme de travail**
 - Etude de l'existant
 - Travail sur les critères et le processus de référencement
 - Compréhension des usages et des problématiques de description (métadonnées)
 - Réflexions autour de l'accompagnement des communautés scientifiques
 - **Enquête du MESR**
-

Point d'étape du GT1 : état des lieux

- Identifier les **catalogues déjà existants** : contenu, processus de construction, pratiques des utilisateurs ...
 - Travail de recensement et d'analyse des catalogues
 - Etude des attentes des utilisateurs (scientifiques ou autres)
- Déterminer les **outils / plateformes** pertinentes pour s'appuyer sur l'existant
 - Tryptique Software Heritage / HAL / CodeMeta
 - S'appuyer sur d'autres initiatives comme le BSO
- Définir les **exigences requises** pour une appropriation par tous
 - Minimiser le travail pour le chercheur, optimiser la saisie unique pour le chercheur et la circulation des données de signalement
 - Lien avec les forges
- Comprendre et intégrer l'écosystème autour de **l'accompagnement**

La question de l'accompagnement

- Articulation au sein d'un écosystème **codes - données - publications**
- Harmoniser, fluidifier et accompagner les processus, généraliser les **bonnes pratiques de développement** et de **signalement**
- S'appuyer sur la dynamique autour des données
 - Maillage national avec les **ateliers de la donnée** au plus proche des communautés scientifiques
 - Intégrer la dimension **codes et logiciels** dans cet accompagnement
 - Mise en place d'un **GT des ateliers** sur le sujet

Enquête du MESR

- La loi n° 2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale introduit dans le même temps la nécessité que soit menée à bien [une étude d'impact de la production logicielle de l'ESR pour août 2023](#) :
 - Titre VII - Chapitre Ier - Article 163 – II
 - « Le Gouvernement remet au Parlement, dans un délai de dix-huit mois à compter de la promulgation de la présente loi, un rapport sur la production et la valorisation des logiciels libres et propriétaires issus de la recherche menée au sein des établissements publics d'enseignement supérieur, des établissements publics à caractère scientifique et technologique, des établissements publics à caractère industriel et commercial et des autres établissements »
 - Profiter du rapport attendu par le parlement pour :
 - [Quantifier l'importance des logiciels](#) dans la recherche
 - [Compléter les travaux du GT1](#) du collège qui s'est intéressé avant tout au processus de référencement
 - Les enjeux sont multiples :
 - Montrer l'importance de la [valorisation sous la forme de logiciels libres](#) dans la recherche
 - Souligner la [place essentielle du développement logiciel dans la recherche](#), dans toutes les disciplines scientifiques
 - Implication du collège et du GT1 dans le [pilotage](#) de cette enquête
-

Perspectives

- Analyser les données de l'enquête sur les usages d'un catalogue logiciels et finaliser un premier livrable autour de l'état des lieux
- S'approprier les données de l'enquête du MESR, et identifier les opportunités d'initier un premier référencement des logiciels en cohérence avec les processus sur lesquels le GT1 travaille
- Poursuivre le travail réalisé conjointement dans le cadre des ateliers de la donnée :
 - Formation, sensibilisation à l'objet logiciel
 - Travail sur des recommandations spécifiques au logiciels pour les Plans de Gestion de Données
- Poursuivre les échanges avec différents acteurs, en particulier sur le signalement, les métadonnées, les problématiques de citation ...