

Destinataires de la V2 : diffusion large

---

## Contexte

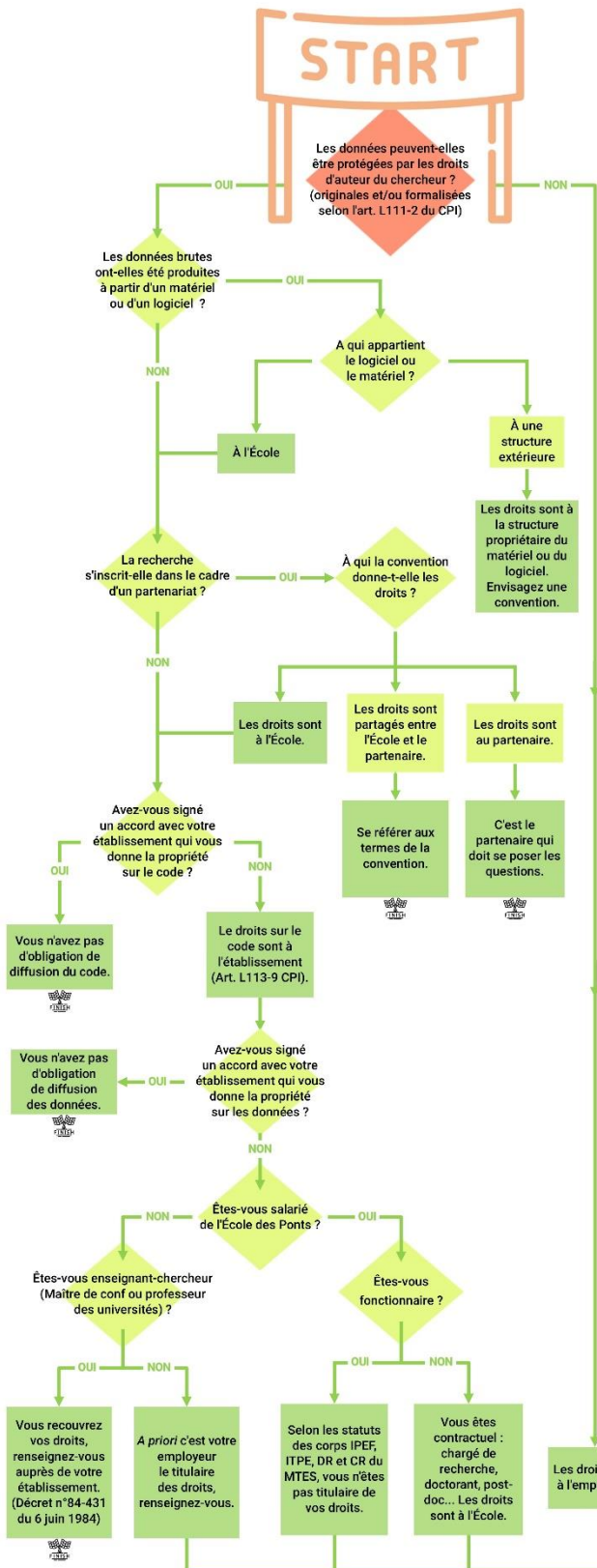
Les chercheurs de l'École des Ponts sont amenés à mettre à disposition les données qu'ils ont utilisées dans le cadre de leurs travaux de recherche. Ils peuvent vouloir le faire spontanément par souci de transparence, et ils peuvent aussi y être contraints par les exigences de l'éditeur d'une revue ou encore un financeur.

Le Pôle IST doit accompagner les chercheurs dans leur démarche de partage, à l'instar de ce qu'il a toujours fait pour ce qui relevait des publications. Afin d'être à même d'effectuer cet accompagnement, la compréhension du cadre juridique et réglementaire était nécessaire. C'est le cabinet August-Debouzy, grâce à une prestation financée par la DR, qui est intervenu pour apporter son éclairage sur le droit des données et le droit public.

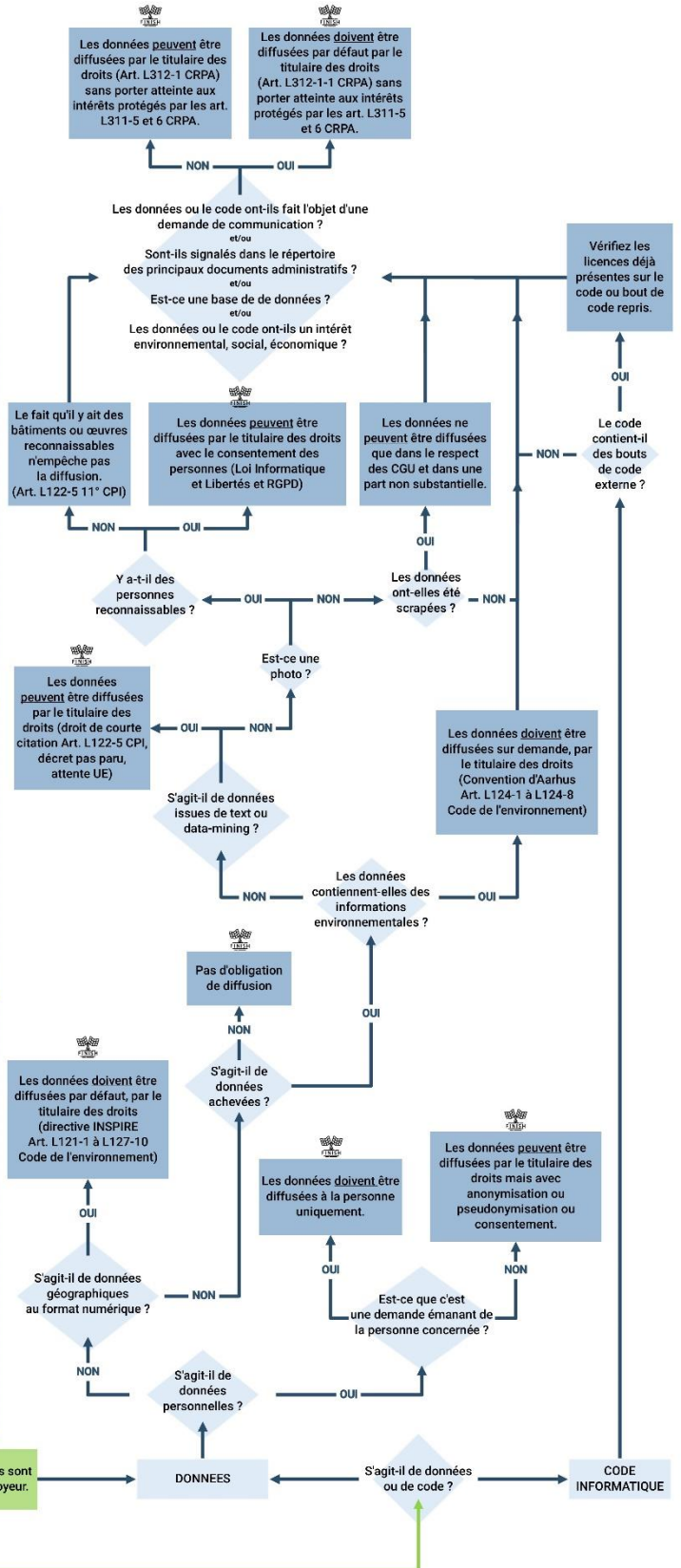
La présente synthèse a pour objectif de faire état de la situation désormais plus claire et de la présenter en se mettant à la place du chercheur pour prendre en compte la réalité de terrain de la communauté scientifique. Elle ne concerne donc que les données de la recherche et pas au sens plus large toutes les données traitées ou générées à l'École des Ponts.

Les paragraphes à venir sont la transcription du logigramme que le Pôle IST a élaboré pour aider le chercheur dans sa prise de décision. La démarche implique d'abord d'établir la titularité des droits, pour savoir si les obligations incombent au chercheur, à son employeur ou à un tiers, puis de déterminer si les données et le code informatique peuvent ou doivent être partagées par leur titulaire et selon quelles modalités.

## Titularité des droits



## Droit et obligation de diffusion



# La titularité des droits

Pour déterminer la titularité des droits sur les données examinées, il faut prendre en compte plusieurs éléments qui aboutissent, successivement, à restreindre le droit d'auteur des chercheurs et à réserver les droits d'auteur patrimoniaux à l'établissement employeur.

## Le droit d'auteur

Les œuvres de l'esprit ne sont protégées par le droit d'auteur que lorsqu'elles sont **formalisées** et porteuses d'**originalité** ([Art. L111-2 CPI](#) et jurisprudence). Les idées, les concepts, ainsi que les algorithmes ne sont donc pas protégés par le droit d'auteur : c'est leur formalisation qui le devient (texte, logiciel...).

Les données produites par des machines ou des traitements automatiques ne sont *a priori* pas porteuses d'originalité et le chercheur ne peut pas bénéficier du droit d'auteur.

Lorsque les données sont incluses dans une base de données, la structure d'une base de données peut être considérée comme une œuvre originale dont le droit d'auteur bénéficie au chercheur qui l'a élaborée (mais c'est difficile à déterminer). Dans le cas contraire, c'est le droit du producteur de base de données qui s'applique : la personne ou l'établissement qui a investi dans sa création et sa maintenance en est propriétaire ([Art. L341-1 CPI](#)).

## Le propriétaire du matériel ou du logiciel utilisé pour produire les données

Le cabinet d'avocats considère que les données brutes appartiennent au propriétaire du matériel ou du logiciel qui a servi à les produire. Les conventions entre l'École et l'établissement propriétaire ou entre l'École et les chercheurs peuvent en disposer autrement.

Lorsque le matériel ou le logiciel ont été alimentés en données par le chercheur (donc hors traitement automatique), ou lorsque les données ont été retravaillées après leur production, la propriété est de nouveau partagée entre le propriétaire du matériel ou du logiciel, et le chercheur ou l'École

Pour étayer ce droit du propriétaire du matériel ou du logiciel, le cabinet d'avocats se fonde sur un article publié par le professeur de droit Pierre-Yves Gautier : "[De la propriété des créations issues de l'intelligence artificielle](#)", dont les conclusions qui s'appliquent aux programmes d'intelligence artificielle sont extrapolées à tous les logiciels et matériels qui produisent des données. Ils qualifient ce raisonnement de solide.

## Les partenariats avec les différents organismes et établissements de recherche

Si la recherche est effectuée dans le cadre d'un partenariat, il faut vérifier que la convention ne prévoit pas que la propriété intellectuelle de tous les produits issus de ce partenariat soit

dévolue à l'entreprise. Si c'est le cas, c'est à l'entreprise de se poser les questions sur ce qu'elle peut et/ou doit faire des données. Sinon, il convient de se référer à la convention.

## Le statut des chercheurs concernés

Le cas général est que les agents de l'État n'ont pas de droits sur les œuvres produites dans l'exercice de leurs fonctions ([Art. L131-3-1 CPI](#)), sauf si de par leur statut ils ne sont soumis à aucun contrôle préalable de l'autorité hiérarchique ([Art. L111-1 CPI](#)).

Cette exception de l'article L111-1 CPI ne s'applique explicitement qu'aux enseignants-chercheurs (corps des professeurs des universités et des maîtres de conférences) :

- [Article 2 de leur décret statutaire](#) : ils jouissent d'une pleine indépendance et d'une entière liberté d'expression conformément à l'[article L952-2 du code de l'éducation](#).
- L'article L952-2 du Code de l'éducation n'est pointé spécifiquement que par le statut des enseignants-chercheurs. Il n'est pas pointé, et ne s'applique pas explicitement, pour les enseignants et les chercheurs qui relèvent d'autres corps.
- Par ailleurs cet article L952-2 du Code de l'éducation ne concerne que les fonctionnaires et pas les chercheurs contractuels, comme le précise l'[article L952-1 du Code de l'éducation](#).

Ce statut d'enseignant-chercheur n'existe pas parmi les chercheurs employés par l'École des Ponts. On distingue en effet 2 grands groupes au sein de l'établissement :

- les fonctionnaires des corps [IPEF](#) et [ITPE](#) et les fonctionnaires [Chargés et Directeurs de Recherche du MTES](#) dont les statuts ne mentionnent pas explicitement leur autonomie scientifique vis-à-vis de leur autorité hiérarchique.
- les contractuels (doctorants, post-doctorants, chargés de recherche...) qui sont soumis aux clauses de propriété intellectuelle de leurs contrats, et dont les droits sont par défaut cédés à l'École.

Cela signifie que l'École des Ponts est titulaire des droits sur toutes les données produites par les chercheurs qu'elle finance. Il lui en incombe à la fois les droits et les obligations relatives à ces données. Cependant l'École n'a pas les moyens de mettre en œuvre le partage des données de la recherche alors que les chercheurs y sont *a minima* invités pour le bien de la science et parfois même obligés par les financeurs de leurs projets (Union Européenne et bientôt ANR notamment).

Dans un souci d'efficience, pour rester en adéquation avec les pratiques scientifiques, et afin d'encourager au partage des données, l'École délègue aux chercheurs la mise en œuvre de la diffusion des données et du code informatique dans le respect de la loi.

Dès lors, le Pôle IST est en charge de la mise à jour de ses supports (à commencer par l'Espace chercheurs) et de la communication active auprès des chercheurs afin de les assister dans les dépôts.

# Le contexte juridique : obligations et modalités de diffusion

Il n'existe pas de droit unifié de la donnée. Le cadre juridique se construit en tissant des liens logiques entre des textes adoptés à des dates et dans des contextes différents.

On trouve essentiellement ce droit dans le Code des Relations entre le Public et l'Administration (CRPA), complété par la Loi pour une République Numérique du 7 octobre 2016, ainsi que dans Code de la Propriété Intellectuelle (CPI). Pour les aspects de droit privé, il faut se référer aux contrats entre les établissements et les chercheurs, aux conventions entre établissements et aux conditions d'utilisation des services et outils utilisés.

## L'achèvement des données

Il n'existe pas d'obligation de diffusion si les données ne sont pas achevées ([Art. L311-2 CRPA](#)). Deux exceptions : les données géographiques numériques qui relèvent de la directive européenne INSPIRE ([Art. L127-1 à L127-10 du code de l'Environnement](#)), et les données personnelles qui doivent être communiquées sur demande de la personne concernée.

## De la possibilité de partager à l'obligation de diffusion

Depuis la Loi pour une République Numérique, les données achevées doivent obligatoirement être publiées en ligne si au moins l'un des 4 critères suivants est rempli ([Art. L312-1-1 CRPA](#)) :

- Elles ont fait l'objet d'une demande de communication selon la procédure CADA.
- Elles sont signalées dans le répertoire des principaux documents administratifs que doit tenir l'École.
- Il s'agit d'une base de données.
- Les données ont un intérêt environnemental, social, sanitaire ou économique.

Au vu du dernier point, il paraît peu probable voire impossible que les données nécessaires à des travaux de recherche n'aient aucun intérêt ni environnemental, ni social, ni économique, ni sanitaire.

Une fois diffusées en ligne, les données et les bases de données doivent être mises à jour régulièrement.

## Les exceptions

Dans toutes les situations, la possibilité ou l'obligation de diffuser signifie qu'il a été vérifié que cela ne portait pas atteinte à l'un des intérêts énoncés par l'[article L.311-5 du CRPA](#) (secret défense) ni à la protection de la vie privée, au secret médical ou au secret des affaires d'une personne (ou de lui porter préjudice : [article L.311-6 du CRPA](#)).

## L'établissement d'une redevance

La valorisation commerciale du code et des données "brutes" ne peut se faire, au moyen d'une redevance, que pour les catégories de données fixées par décret ([Art. D324-5-1 CRPA](#)). Cela concerne uniquement les données de l'IGN, de Météo France, du SHOM, et les données numérisées des bibliothèques, musées et archives, et à condition de ne pas dépasser les coûts de production.

Il n'est donc pas possible de commercialiser des données "brutes". La seule option ouverte est de commercialiser des outils, des services ou des prestations associés aux données diffusées gratuitement. C'est un principe de libre concurrence qui s'applique, où tous les acteurs ont leurs chances à partir des mêmes données publiques.

Les outils doivent être développés par des filiales de droit privé, autrement, les données et le code tombent sous l'obligation de diffusion qui s'attache aux données publiques.

## Le choix des licences

Hors toute contrainte externe (convention, réutilisation de données ou de code sous licence), la loi impose l'utilisation de [deux types de licences](#) lors de la diffusion du code ou des données ([Art. D323-2-1 CRPA](#)) :

- Des licences permissives. Ces licences ne protègent que la paternité des auteurs et limitent leur responsabilité. Elles offrent toute liberté de réutilisation, rediffusion ou modification de licence et permettent une exploitation commerciale des données ou du code.
- Des licences avec obligation de réciprocité (ou copyleft). Ces licences obligent d'une part à conserver les conditions de la licence d'origine, et d'autre part à la propager à toute l'œuvre dérivée. Elles peuvent restreindre l'exploitation commerciale des données ou du code et ne doivent être utilisées que pour des motifs d'intérêt général et de façon proportionnée ([Art. L323-2 CRPA](#)) (même si cette justification peut s'apporter facilement).

Pour le code :

- Licences permissives : BSD, Apache, CeCILL-B et MIT Licence
- Licences avec obligation de réciprocité : Mozilla Public License, GNU GPL et CeCILL

Pour les données :

- Licence permissive : Licence ouverte Etalab
- Licences avec obligation de réciprocité : Open Database Licence

À noter que la licence permissive Etalab s'applique mieux pour les données, et la licence copyleft ODbL s'applique mieux pour les bases de données. Mais priorité doit être donnée à la licence permissive Etalab, même pour les bases de données. À l'inverse en cas de motif d'intérêt général, on peut utiliser la licence copyleft ODbL même pour des données.



# Les données

Les données sont considérées comme des **documents administratifs**. En effet, [l'article L300-2 du CRPA](#) définit les documents administratifs comme étant tous les documents "*quels que soient leur forme et leur support*". Il donne une liste qui commence par "*notamment*" : elle n'est pas exhaustive et permet donc d'englober les données. De ce fait, comme les données sont donc des documents administratifs, cela implique un droit d'accès sur demande selon la procédure CADA ([Art. L311-1 CRPA](#)), un droit de diffusion ([Art. L312-1 CRPA](#)) et une obligation de diffusion ([Art. L312-1-1 CRPA](#)).

Les données, en tant que documents administratifs, sont également considérées comme des **informations publiques** ([Art. L321-1 CRPA](#)) ce qui implique une libre réutilisation (même commerciale), sous conditions ([Art. L322-1 CRPA](#)) et gratuite ([Art. L324-4-1 CRPA](#)).

## Le cas particulier des données personnelles

La collecte et le traitement des données personnelles sont soumis au respect des conditions posées par la [Loi Informatique et Libertés](#) de 1978, et par le Règlement Général sur la Protection des Données (RGPD) transposé dans la [Loi relative à la protection des données personnelles](#) en 2018.

Pour les données hors code informatique, il convient tout d'abord de se demander s'il s'agit de données personnelles. Si c'est le cas et qu'une demande émane de la personne concernée par ces données, elles doivent être communiquées à cette personne uniquement. Dans le cas contraire, les données peuvent être diffusées mais après avoir été anonymisées ou pseudonymisées ou encore si la ou les personnes ont donné leur consentement préalable.

## Le cas particulier des données géographiques

La directive européenne INSPIRE (transposée dans les [articles L127-1 à L127-10 du Code de l'environnement](#)) vise à établir une infrastructure d'information géographique dans l'Union Européenne pour favoriser la protection de l'environnement.

Elle rend obligatoire la diffusion en ligne par les autorités publiques des données géographiques disponibles au format numérique. Elle encadre fortement l'interopérabilité des données, leur référencement, et la qualité de leurs métadonnées pour permettre une réutilisation optimale. Elle n'impose pas que les données soient achevées ou parfaites, mais les métadonnées doivent indiquer la qualité et la validité des données ([Art. L127-2 du Code de l'environnement](#)). Il existe des exceptions restreignant cette obligation de diffusion ([Art. L127-6 du Code de l'environnement](#)).

INSPIRE s'applique à 34 domaines thématiques qui concernent souvent l'École des Ponts, et se décomposent en trois groupes principaux :

- les données nécessaires au repérage sur le territoire, telles que systèmes de coordonnées, unités administratives, réseaux de transport, hydrographie, parcellaire cadastral, adresses et noms de lieux ainsi que des données relatives aux sites protégés.

- les données générales complémentaires, telles que l'altimétrie, l'occupation des terres, la géologie et l'ortho-imagerie.
- les données thématiques telles que bâtiments, vocation des sols, santé et sécurité des personnes, services d'utilité publique et services publics, données sur l'environnement (nombreuses et variées), installations industrielles, agricoles, démographie, périmètres de réglementation, données météorologiques, données maritimes, sources d'énergie et ressources minérales.

## Les données environnementales

La convention d'Aarhus (transposée dans les [articles L124-1 à L124-8 du Code de l'environnement](#)) stipule que les données environnementales doivent être diffusées si elles ont fait l'objet d'une demande. Contrairement à la directive INSPIRE, les données doivent être achevées, en revanche il existe moins d'exceptions que dans le CRPA pour opposer un refus.

## Les données issues de text ou data-mining

Que ce soit pour gagner du temps dans l'exploration de la littérature scientifique ou plus souvent pour explorer un grand volume de textes ou de données, les chercheurs recourent à des outils automatisant ce travail. Les données qui en sont issues peuvent être diffusées.

En effet, les auteurs d'œuvres ou les propriétaires de bases de données ne peuvent plus s'opposer à ce que des « *copies ou reproductions numériques* » soient réalisées « *à partir d'une source licite* » de leurs œuvres, « *en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique* » ([Art. L.122-5 10° CPI](#) créé par la Loi pour une République Numérique ; à noter qu'un décret d'application devant préciser les conditions d'application de cet article est toujours attendu.)

Les données issues de ce travail peuvent donc être diffusées, en revanche, les textes sources ne peuvent l'être que dans le respect du principe de la courte citation ([Art. L.122-5 3a° CPI](#)).

## Les photos

Les chercheurs, toutes disciplines confondues, peuvent être amenés à prendre des photos qui sont alors la base de leurs travaux. Pour savoir si elles peuvent être diffusées, il faut tout d'abord vérifier si des personnes reconnaissables y figurent. Si oui, c'est seulement avec le consentement de la personne que la diffusion pourra se faire ([Art. 9 Code Civil](#) et jurisprudence). Dans le cas contraire, et même s'il y a des bâtiments ou œuvres reconnaissables dessus, rien n'empêche la diffusion (exception de panorama : [Art. L.122-5 11° CPI](#), depuis la Loi pour une République Numérique).

## Les données scrapées

Les chercheurs pratiquent très fréquemment le scraping, autrement dit l'automatisation de l'extraction des données contenues sur un site web. Nous avons eu de très nombreux exemples : collecte des montants de loyer sur le site du Bon Coin, collecte de tweets, collecte de données relatives à de l'auto-partage etc...



Il convient tout d'abord de se référer aux conditions générales d'utilisation du site scrapé. Si elles interdisent l'extraction, il faut en conclure que le scraping d'une partie substantielle des données du site est *a priori* illicite, sur le fondement du droit *sui generis* du producteur de base de données et d'une interdiction contractuelle. En revanche, le scraping d'une partie non substantielle serait possible ([Art. L.342-3 1° CPI](#)).

Par conséquent, les données ne peuvent être diffusées que dans le respect des CGU et dans une part non substantielle. Ceci étant dit, il faut intégrer que les chercheurs retravaillent et analysent ces données scrapées avant de les publier dans le cadre de leur recherche. Le préjudice potentiel vis-à-vis du producteur des données est donc nul et le risque encouru aussi.

## Le code informatique

Le code est considéré comme un document administratif ([Art. L300-2 CRPA](#)). Les règles concernant le droit à l'accès, l'obligation et les modalités de diffusion s'appliquent donc de la même manière pour le code que pour les autres types de données.

## Spécificités de la titularité des droits pour le code

La titularité des droits sur le code dépend des éventuelles personnes extérieures qui sont intervenues dans sa production. Elle est déterminée d'une part par les licences appliquées sur les codes informatiques externes qui sont réutilisés, et d'autre part par les conventions signées en cas de collaboration entre l'École et d'autres partenaires. Il convient de se reporter à la fois aux licences déjà appliquées et aux conventions pour savoir ce qui peut être diffusé et selon quelle licence.

Au sein de l'École, à moins que le contrat signé avec le chercheur stipule que le code informatique appartient au chercheur, c'est l'École qui est titulaire des droits patrimoniaux, quel que soit le statut du chercheur ([Art. L113-9 CPI](#)).

Si le code fait l'objet d'une valorisation commerciale par l'École, les agents auteurs ne disposent que d'un droit de préférence ([Art. L131-3-1 CPI](#)). En revanche ils doivent bénéficier d'une prime d'intéressement ([Décret n° 96-858 du 2 octobre 1996](#)).

Enfin les éléments auxiliaires du logiciel (interface utilisateur, cahier des charges, documentation) sont protégés par le droit d'auteur commun, pour lequel il convient alors de vérifier la titularité des droits.

## L'obligation de diffusion

Lorsque le code appartient à l'École, il est soumis aux mêmes obligations de diffusion que les données ([Art. L312-1-1 CRPA](#)).

## Le choix de la licence

Le choix de la licence, hors éléments de code réutilisés sous licence ou dispositions des conventions de recherche, est déterminé par la loi ([Art. D323-2-1 CRPA](#)) qui impose soit la licence ouverte Etalab, permissive, soit la licence Open Database License, avec obligation de réciprocité.