





Vive  
les communs  
numériques !

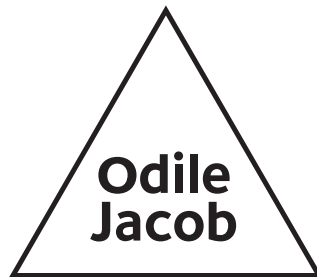


Serge Abiteboul  
François Bancilhon

# Vive les communs numériques !

Logiciels libres, Wikipédia,  
le Web, la science ouverte, etc.

Préface de Gérard Berry



Collection « Informatique et monde numérique »  
dirigée par Gérard Berry

© ODILE JACOB, mars 2024  
3, rue Auguste-Comte, 75006 Paris

[www.odilejacob.fr](http://www.odilejacob.fr)

ISBN : 978-2-4150-0798-0

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3°a, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

# Préface

*par Gérard Berry,  
professeur émérite au Collège de France*

Depuis bien longtemps, on appelle « communs » les biens mis à la disposition des habitants d'un lieu donné mais avec des règles à respecter, le cas emblématique étant celui des espaces communs cultivables, découpés en parcelles. Le mot vient du latin *communis*, qui a le sens de « bien commun », ce qui signifie en droit « commun à tous, donc qui ne peut pas être approprié ». Par exemple, la commune est un espace partagé par ses habitants et, en Angleterre, la Chambre des communes est censée les représenter explicitement. Mais, avec les développements de l'informatique et surtout d'Internet, le terme et la notion ont été repris et généralisés, du matériel à l'immatériel, leur donnant une extension bien plus considérable. C'est tout cela qui est conté en détail dans ce livre, écrit par deux auteurs parfaitement au fait de ces mécanismes depuis toujours, et eux-mêmes créateurs de bases de données et de logiciels libres.

Le livre analyse d'abord la très grande variété des communs numériques. L'un des plus importants est sans doute Internet, né à la toute fin du xx<sup>e</sup> siècle, soit quarante ans après la naissance des premiers ordinateurs. C'est un réseau de réseaux permettant aux gens

de partager rapidement et simplement toutes sortes d'informations. Tout peut s'y propager à grande allure et à coût quasi nul. Ce ne sont pas les très nombreux réseaux physiques qu'il utilise qui sont concernés, car ils sont en général propriétaires : Internet n'est en fait qu'une couche protocolaire au-dessus des techniques individuelles des réseaux physiques, qui les rend interopérables au point d'en ignorer complètement l'existence. Qui sait par combien et par quel type de réseaux physiques passe un message de Paris à Sydney, et qui a besoin de le savoir, si tant est que cela soit même possible ?

Internet a permis de diffuser, à très grande échelle et à coût très faible, données brutes, informations tirées de ces données, logiciels, etc. – en bref, beaucoup de ce qui est numérisé, et dans quasiment tous les secteurs. Parmi les données communes, citons les données scientifiques sur le climat, l'astronomie, la cartographie, etc., de plus en plus rendues publiques, celles sur ses membres de la communauté scientifique et leurs publications, mais aussi celles sur les habitants d'un pays, leur santé, sur les entreprises, etc., souvent elles aussi rendues publiques par les États. Pour les connaissances extraites des données brutes ou déjà connues par ailleurs, citons bien sûr Wikipédia, une encyclopédie de grande classe, comme le soulignent les auteurs : « Elle est alimentée chaque jour par plus de 100 000 contributeurs à travers le monde. Elle est visitée chaque mois par près de 500 millions de visiteurs et propose plus de 60 millions d'articles dans plus de 300 langues. » Qui peut lutter ? Mais citons aussi les recettes de cuisine, ou toutes les vidéos mises en ligne par des personnes comme vous et moi sur les sujets les plus variés. Pour les logiciels libres, eux-mêmes d'une grande diversité, citons des logiciels de base comme Linux, le plus sûr des systèmes d'exploitation résultant d'un effort très structuré de milliers de contributeurs sur toute la planète mais avec une organisation très bien définie, beaucoup de langages de programmation, des logiciels de retouche photo ou de traitement du son créés par de petites équipes et mis à disposition de tous, et bien sûr le Web. Tout cela n'est pas commercial mais est rendu



facilement accessible, car on a le monde sous la main à partir de n'importe quel ordinateur. Et la motivation de très nombreux acteurs désirant partager leurs connaissances ne cesse de grandir : un aspect très nouveau des communs numériques est qu'ils résultent d'actions de nombreux individus qui ne se rencontreront ni se connaîtront jamais.

Mais quelles sont les motivations des communautés humaines qui fabriquent ces communs numériques, et comment influent-elles sur leur gouvernance ? Qu'a-t-on le droit de faire ou non avec ces communs ? Quelle est la différence avec les outils commerciaux, et quels sont les rapports avec eux ? Comment les communs sont-ils préservés dans un monde technique qui évolue si vite ? Qui les freine et qui les pousse ? Ces aspects sont étudiés en détail dans ce livre. Souvent, ils sont livrés avec une licence d'utilisation qui définit les droits et devoirs d'utilisation selon plusieurs types de contraintes reflétant les idéologies variées du monde bouillonnant de leurs créateurs : totalement libres, libres avec devoir de citation, logiciels libres mais ne pouvant être incorporées que dans d'autres logiciels également libres, etc. Le monde commercial n'est d'ailleurs pas exclu du jeu. Certaines grandes sociétés comme Microsoft participent activement au développement de communs dont la qualité est souvent meilleure que celle des produits qu'elles vendent, mais qui ne sont pas vraiment dans leur domaine de référence quant à l'argent qu'elles gagnent. Dans un autre domaine, des sociétés de télécommunication s'entendent pour faire à toutes ce qu'elles ne pourraient pas réaliser seule et le considérer comme un commun, par exemple au niveau des logiciels de plus en plus complexes pour accéder aux antennes : elles y gagnent toutes.

Il s'agit donc d'un grand sujet du monde moderne, au développement très rapide favorisé par l'extension partout de l'informatique et des réseaux. Pourtant, ce sujet est rarement analysé en détail. C'est exactement ce trou que vient combler ce livre parfaitement renseigné. Alors, bonne lecture !



## Mise en bouche

En janvier 1973, Liv cherche à mieux comprendre ce qui s'est passé pendant le putsch de la Brasserie mené par Adolf Hitler cinquante ans plus tôt. Elle consulte, chez ses grands-parents, l'*Encyclopédie Larousse* en 21 volumes. En janvier 2023, cinquante ans plus tard, Liv retombe sur la même question. Elle tape « putsch de la Brasserie » dans la barre de son moteur de recherche. La première réponse la conduit à un article de Wikipédia sur le sujet. Elle surfe un bon moment sur la Toile, approfondissant sa compréhension du sujet. Des masses de connaissances sur le sujet sont à portée de clic.

Qu'est-ce qui, entre ces deux époques, a profondément transformé la façon dont Liv accède à la connaissance ? Deux phénomènes distincts et complémentaires. Le premier est la conquête du monde par le numérique : les informations sont maintenant numériques et accessibles avec des outils numériques, notamment sur la Toile. Le second est le remplacement des encyclopédies d'antan conçues et réalisées par des experts, par une encyclopédie construite par une multitude de bénévoles. On est passé de produits privés réservés à ceux qui ont les moyens de payer à des communs numériques ouverts à tous gratuitement.

Nous nous intéressons ici à ce deuxième phénomène, celui des communs numériques qui, dans la tradition de l'économie sociale et solidaire, s'appuient sur le partage et la solidarité.

## *Pourquoi ce livre ?*

À travers ce livre, nous voulons d'abord partager de belles histoires sur les communs numériques qui nous apprennent tant sur les humains, leurs rêves et sur la puissance du partage. Nous voulons raconter ce formidable champ d'expérimentation d'idées, d'expériences et d'innovation. Et puis nous voulons convaincre que l'idée de commun numérique peut aider à répondre à un grand défi de nos sociétés : comment diffuser à tous informations, culture et connaissances ? Enfin, nous espérons que ce livre participera à développer une culture commune entre des domaines d'applications diversifiés, parfois cloisonnés, à mutualiser les efforts, à faire que tous bénéficient des expériences de chacun.

Nous tenons à signaler que nous nous revendiquons de la grande famille des militants des communs numériques. Nous assumons nos biais.

Pour nous approcher du sujet des « communs numériques », commençons par nous intéresser aux « communs ».

## *Les communs*

Nous nous sommes habitués depuis longtemps à vivre dans une société accordant une place importante à la propriété privée, qui nous protège contre l'arbitraire, par exemple, du prince qui un temps pouvait s'approprier notre maison sans barguigner. Pourtant, depuis toujours, les humains ont appris à gérer collectivement des ressources, à les partager. C'est le cas des poissons de l'étang communal de Champs-Romain, une commune de Dordogne. Cet étang communal, géré par une association, est ouvert à la pêche pour tous. Si prévalait une politique de « laisser-faire » total, quelqu'un pourrait pêcher tous les poissons de l'étang et en priver les autres.

Alors, on fait de l'étang un commun en établissant des règles pour pouvoir collectivement utiliser la ressource tout en la préservant, en la pérennisant.

La mise en commun est le point de départ des travaux d'Elinor Ostrom à la fin du xx<sup>e</sup> siècle. Selon Wikipédia (2022) : « Les *communs* sont des ressources partagées, gérées et maintenues collectivement par une communauté ; celle-ci établit des règles dans le but de préserver et de pérenniser ces ressources, tout en fournissant aux membres de cette communauté la possibilité et le droit de les utiliser, voire, si la communauté le décide, en octroyant ce droit à tous. »

Longtemps, ces communs ont été fondés sur des ressources physiques comme des pâturages ou des rivières. Ils sont au cœur des conflits des *enclosures*, qui ont vu en Angleterre à partir du xvi<sup>e</sup> siècle de grands propriétaires clôturer des champs communaux pour se les approprier aux dépens de petits paysans qui y faisaient, par exemple, paître des animaux.

### *Les communs numériques*

Les communs trouvent dans le numérique une puissance incroyable. Illustrons-les par trois superbes réussites : 1) la description du contenu de plus de 1 million de produits alimentaires disponibles en consultation pour tous, grâce au travail de la communauté Open Food Facts ; 2) les protocoles d'Internet et du Web ; 3) les plus de 11 000 revues scientifiques en accès libre et gratuit représentant plus de 2 millions d'articles.

Les logiciels ont été assez tôt mis en commun et ouverts, avec des exemples emblématiques comme la suite de logiciels de Berkeley (Berkeley Software Distribution) et le système d'exploitation Linux. Cette mise en commun a véritablement révolutionné l'industrie numérique. Deux domaines ont rapidement vu se développer la mise en commun de textes, images ou vidéos parfois sous des licences

Creative Commons : la science et l'éducation ouvertes. Les données ouvertes se sont aussi déployées dans la sphère publique, poursuivant le triple rôle de rendre plus transparent le fonctionnement de l'État et des collectivités territoriales, d'améliorer la gestion de la cité, et de créer des opportunités d'innovation pour les entreprises.

Mais, au-delà de ces champs emblématiques, les mises en commun numériques éclosent un peu partout. Dans la société, avec des initiatives citoyennes comme OpenStreetMap, une alternative à Google Maps, et également dans des champs professionnels entiers comme l'agriculture, la santé ou la justice. Et puis, quand on s'interroge sur le territoire des communs numériques, le sujet arrive vite sur Internet et le Web, qui sont bien des ressources mises en commun.

Les mouvements d'ouverture et de mise en place de communs numériques ont clairement le vent en poupe. Leur emprise s'étend dans le monde entier. Introduits au départ dans des pays démocratiques développés économiquement, ils se développent de façon similaire dans la plupart des pays, notamment, la Chine, la Russie, ou l'Inde.

Mais qu'est-ce que le numérique a de si spécial pour être particulièrement bien adapté au concept de commun ? D'abord, le fait qu'un bien numérique n'est *a priori* pas rival. Si je dispose d'une copie d'un logiciel (et si sa licence m'y autorise), je peux en faire mille copies, un million de copies, le logiciel reste disponible pour mille, un million de personnes supplémentaires. Il s'agit d'un bien non rival.

La non-rivalité des biens numériques est pour beaucoup dans le succès considérable des communs numériques. Mais cela n'en est pas la seule cause. Les succès des communs numériques viennent aussi de ce que le numérique facilite les échanges entre les personnes, le travail collaboratif, et ouvre de nouveaux espaces à des partages massifs de données, d'information, de connaissances, encore inconcevables dans le monde d'avant.

## *Comment met-on en commun ?*

Le point de départ est une *ressource* qu'une *communauté*, un certain nombre d'individus veulent mettre en commun pour la gérer collectivement. Ces communautés qui animent des communs numériques présentent une grande variété depuis des militants (par exemple pour OpenStreetMap), à des organismes publics (pour l'ouverture des données publiques ou la science ouverte), à des entreprises à la recherche de plus d'efficacité (pour l'*open design*). La communauté décide d'une *gouvernance* et fait appliquer des règles de fonctionnement qui vont animer le commun. La communauté est souvent stratifiée entre des contributeurs, des utilisateurs, et des membres avec des rôles plus spécifiques comme contrôler ou valider les contributions.

Une question centrale est celle du modèle économique. Comment faire vivre un système, quand la ressource est mise à disposition de tous, le plus souvent gratuitement ? Il faut le faire en évitant toute confusion avec le faux gratuit du modèle privé qui suit plutôt la règle : « Si c'est gratuit, c'est que vous êtes le produit ! » Pour les modèles économiques des communs comme pour les communautés, la variété est de mise depuis la valorisation de services autour des communs dans le cas de certains logiciels *open source* à l'appel aux dons pour Wikipédia.

## *Pourquoi participe-t-on à un commun ?*

Par altruisme parfois, mais pas seulement. Nous verrons que les réponses sont multiformes et dépendent de la nature du commun, du champ d'application. Pourtant les contributeurs doivent y trouver leur compte ; sinon ils arrêteraient de contribuer. Si pour les communs matériels, l'accueil essentiel est celui de la surconsommation

qui conduit à l'épuisement de la ressource, pour les communs numériques non rivaux c'est celui de la motivation de contribuer.

Enfin, les membres des communautés de communs numériques mettent souvent en avant que travailler dans un commun modifie les méthodes de travail et la façon de travailler, et que cela a également des impacts sur le produit de leur travail.

### *Quel futur pour les communs numériques ?*

Si les nombreuses réussites de communs numériques montrent que ça marche, si les chiffres mettent en évidence la croissance de l'ensemble de ces communs, leurs modèles économiques sont parfois fragiles, leurs gouvernances délicates. La coexistence avec le monde du privé ne va pas sans frictions. Des entreprises privées peuvent apporter des contributions importantes aux communs. Mais des entreprises, parfois les mêmes, les concurrencent aussi, les utilisent sans vergogne. Les communs doivent se défendre contre des tentatives d'OPA de la propriété privée, contre des formes modernes d'*enclosures*. C'est l'une des conditions de leur survie et de leur développement parallèle et en concurrence avec l'économie privée. Nous sommes convaincus que les engagements de leurs communautés leur permettront de continuer à se développer, à nous surprendre, et nous émerveiller.



PARTIE A

# Les communs numériques



S'ils ont été étudiés depuis le xx<sup>e</sup> siècle, les communs sont bien plus anciens. Que sont-ils ? Quelles nouvelles dimensions prennent-ils avec le numérique ? Et quels sont ces nouveaux espaces à conquérir aux frontières des communs numériques ?



## CHAPITRE 1

# Les communs matériels

Nous traitons dans cet essai de « communs numériques ». Mais considérons brièvement ces communs matériels qui leur pré-existaient et qui vont nous permettre de dégager les grands traits philosophiques, économiques et juridiques des communs en général.

Pour illustrer la notion de commun physique, prenons l'exemple de la *foggara* en Algérie ; on retrouve des structures semblables ailleurs dans le monde comme la *khettara* marocaine ou le *qanat* iranien plurimillénaire. La *foggara* est une galerie souterraine qui achemine l'eau pour des besoins d'irrigation. Dans une région désertique qui manque d'eau, vivre est pour les humains un défi qui ne se surmonte que collectivement. Les habitants du désert vont creuser des canaux souterrains pour amener l'eau jusqu'aux points de cultures et aux oasis sans qu'elle s'évapore. Comme ces canaux ont une importance vitale pour la collectivité, il faut qu'ils soient entretenus, protégés, mis au service de la collectivité. En particulier, l'eau est partagée équitablement avec une espèce de peigne pour la répartir en plusieurs rigoles. Un comité de sages, appelé Djemaa, est chargé de la gouvernance du système.

Autre exemple : la gestion des forêts au Népal confiée à des communautés locales qui ont un intérêt direct à ne pas les surexploiter. Ces organismes locaux contrôlent l'écosystème, de l'entretien de la

forêt, à la récolte du bois et à sa vente. Les profits sont aussi mis au service de la communauté. On a pu observer les effets positifs d'une telle politique où la gouvernance est décentralisée, confiée à des autorités locales.

Dans ces deux cas, les ressources sont mises en commun. Ce sont des biens communs en opposition avec les biens privés et les biens publics. Les biens privés sont typiquement des biens que l'on associe facilement à une personne (ses habits, sa maison, son jardin, ses outils pour un artisan) ou à une entreprise (ses locaux, ses machines, ses stocks). Les biens publics sont gérés par l'État ; ce sont, par exemple, les équipements militaires ou les routes. La frontière entre les deux mondes n'est pas rigide : des hôpitaux ou des écoles peuvent être publics ou privés.

Pour les communs, on pourrait parler de propriété collective, mais on se tromperait de cible. L'accent pour les communs n'est pas tant sur la propriété que sur les usages et la gestion : que peut-on faire avec les ressources ? Comment sont-elles gérées ? La *foggara* n'appartient à personne, ce qui compte, ce sont les usages qu'on peut en faire. Que la forêt népalaise soit un bien public ou pas, ce n'est pas le sujet, ce qui compte, ce sont ses usages et la façon dont on la gère.

Les communs sont donc des ressources gérées par une communauté. Les ressources sont « non exclusives », c'est-à-dire qu'on ne laisse personne se les approprier, on les laisse disponibles à tous les membres de la communauté, voire à un cercle plus large encore. La communauté optimise leur utilisation tout en les protégeant, c'est-à-dire en évitant leur épuisement, édicte des règles pour ce faire et les fait respecter. Notre essai traite donc de « biens communs gérés collectivement », ou de communs collaboratifs selon la terminologie de Paul Chemla traduisant un essai de Jeremy Rifkin<sup>1</sup>. Selon la

---

1. Jeremy Rifkin, *La Nouvelle Société du coût marginal zéro. L'Internet des objets, l'émergence des communaux collaboratifs et l'éclipse du capitalisme*, Les Liens qui libèrent, 2014.

terminologie d'Elinor Ostrom, ce sont des *common-pool resources*, que nous pouvons traduire par « réservoirs communs de ressources » avec Benjamin Coriat et Hervé Le Crosnier. Nous parlerons ici simplement de « communs », dans l'esprit du terme anglais *commons*.

Précisons les contours d'un commun :

- un commun est une ressource adaptée à une gestion collective ;
- il s'appuie sur une communauté, un groupe de personnes, qui vont gérer les usages de cette ressource ;
- cette communauté définit une gouvernance et des règles qu'elle va faire vivre. La vie et l'utilisation harmonieuse de la ressource dépendent du respect de ces règles par la communauté, ainsi que de sanctions et de mécanismes de résolution de conflits.

Cette définition couvre une grande diversité de situations, selon la nature de la ressource, la composition et la structure de la communauté, la forme de la gouvernance et des règles, les accroches sur le droit comme la nature des licences, les menaces d'appropriation, etc.

Mais examinons le terme « communs ». En 2011, dans un article d'un ouvrage intitulé *Les biens communs de la connaissance*<sup>1</sup>, Alain Rey parle de l'étymologie des « communs ». Il considère l'origine latine : « Le mot clé, en latin, pour exprimer le partage, les relations entre un grand nombre d'humains, s'agissant de biens, de besoins, d'obligations, c'est *communis*. » Dans le droit romain, un bien commun (*res communis*) est une catégorie juridique qui s'applique à des entités comme l'eau, l'air, le rivage de la mer, qui sont utilisées par tous, que personne n'est en droit de s'approprier, ni un individu ni même l'État. On retrouve bien dans l'origine du mot tous les ingrédients des communs dans le sens qui nous intéresse ici. Il est vrai que la même racine a aussi conduit à « communications » ou « communisme ».

La France est divisée en près de 36 000 communes. Dans ces communes les biens communaux sont un héritage des communs :

---

1. Alain Rey, *La Guerre des communs. Libres savoirs, les biens communs de la connaissance*, C&F Éditions, 2011.

le foncier appartient à la municipalité, les droits d'usage et d'exploitation aux habitants (exemple : les affouages). Ainsi, selon Nadine Vivier, en 1998, ces biens communaux « occupent encore aujourd'hui 10 % du territoire<sup>1</sup> ».

On trouve bien des origines à la notion de communs dans les champs agricoles partagés, les pâturages et les forêts qui ont été pendant des siècles ouverts à tous. Ce sont par exemple les « biens communaux » en France, les *commonlands* anglais ou les *grazing rights* (« droit de pâture ») des westerns états-uniens. L'histoire des communs est parfois celle de combats. On découvre tragiquement l'importance d'un commun quand d'autres essaient de se l'approprier. C'était l'origine du mouvement contre les « enclosures » en Angleterre.

Comme mentionné plus haut, le mot « enclosure » (un anglicisme) désigne l'action d'enclorre un champ et, par glissement, la parcelle de terrain clôturée de haies ou de murs. Dans l'Angleterre du *xvii*<sup>e</sup> siècle, et surtout de la révolution industrielle, les enclosures désignent le mouvement par lequel de grands propriétaires s'approprient les terres communales et en chassent les paysans malgré une mobilisation de défense des communs très forte. Ils clôturent ces champs, privant les paysans de la possibilité d'y faire paître leur bétail et ou de s'y ravitailler en bois. En favorisant la propriété privée face au partage de biens communs, en transformant les paysans en employés agricoles, en marins, et surtout en ouvriers au moment de la révolution industrielle, l'enrichissement d'une classe de grands propriétaires favorise le passage à une agriculture intensive.

Si on parle surtout du mouvement des enclosures anglais, un mouvement similaire a également eu lieu en France. Comme en Angleterre, un argument est que les terres des communs sont mal utilisées par les paysans. C'est un argument avancé en France à la

---

1. Nadine Vivier, « Les biens communaux en France de 1750 à 1914. État, notables et paysans face à la modernisation de l'agriculture », *Ruralia*, 1998, 2.



fin du XVIII<sup>e</sup> siècle par les physiocrates (dont l'un des fondateurs est Mirabeau), qui propagent une pensée économique et politique parfois considérée comme annonçant le libéralisme économique, et qui voient l'agriculture comme le moteur essentiel de l'économie.

Une thèse se dégage. Si ce mouvement des enclosures a eu lieu, peut-être était-il nécessaire pour mieux utiliser les champs, pour optimiser la ressource ? Peut-être cette idée de « communs » est-elle juste inefficace, peut-être ne fonctionne-t-elle pas ? Pour en revenir à l'étang mentionné dans l'introduction, l'être humain ne va-t-il pas prendre autant de poissons qu'il pourra, de manière égoïste, et ne va-t-il pas épuiser les ressources halieutiques de l'étang ? On parle de la « tragédie des communs ». L'*Homo œconomicus* optimiserait son intérêt personnel et ignorerait systématiquement l'intérêt collectif dans le cas des communs. C'est la thèse d'un article très repris de Garrett Hardin : « The tragedy of the commons ». Cette théorie économique affirme que chaque personne dans une situation de compétition pour l'accès à une ressource limitée va privilégier son intérêt individuel, ce qui va conduire collectivement à une surexploitation de la ressource. Tout le monde finit par y perdre.

Alors, les communs, une mauvaise bonne idée ? Non, des réalisations de communs comme les *foggaras* ou les forêts népalaises ont montré que cela pouvait fonctionner. L'argument de Hardin a été depuis détricoté : son erreur fondamentale a été de supposer que chaque individu était libre d'utiliser la ressource comme il le souhaitait. La gouvernance impose des contraintes décidées par la collectivité, acceptées par tous, et des règles et des punitions limitent les comportements de chacun. Et puis l'argument de Hardin s'appliquait à des biens rivaux et tombe donc pour les communs numériques dont nous allons parler ici, qui sont non rivaux.

Elinor Ostrom étudiait des communs matériels locaux qui intéressent une communauté géographiquement localisée. Les questions environnementales ont maintenant mis l'accent sur les « communs environnementaux ». Nous sommes entrés dans l'anthropocène,

où l'avenir même de l'humanité est mis en péril par les effets des activités humaines sur la biosphère. On attache donc une attention particulière à la mise en commun de l'eau, des forêts, de l'air. Il s'agit alors d'établir de nouvelles communautés fondées sur la responsabilité de leurs membres-citoyens, et d'inventer de nouvelles formes de gouvernance. On se place ici aussi volontairement en dehors de la dualité biens publics/biens privés, déjà mise à mal auparavant, par exemple, par l'économie sociale et solidaire. On s'éloigne de la propriété pour mettre l'accent sur les usages.

Dans cet essai, nous ne nous intéressons ni aux communs matériels, ni à de tels communs environnementaux, mais seulement aux communs numériques.

**Elinor Ostrom** (1933-2012) est une États-Unienne, spécialiste d'économie politique. Elle a obtenu en 2009 le prix Nobel d'économie pour son « analyse de la gouvernance économique, en particulier des communs ». Elle et Esther Duflo (2019) sont les seules femmes prix Nobel d'économie. Professeure à l'Université d'Indiana, elle s'est intéressée à la gestion des ressources et a étudié les interactions entre les écosystèmes et les individus. En particulier, elle a montré comment des communautés pouvaient gérer des ressources épuisables sans qu'il soit nécessaire de faire appel au gouvernement pour protéger ces ressources.

Au départ, elle s'intéressait uniquement aux communs fonciers. Pourtant, avant même d'obtenir le prix Nobel, en 2007, elle élargit ses études à d'autres formes de communs, et en particulier aux connaissances. Elle coédite avec Charlotte Hess un livre qui dessine le domaine : *Understanding Knowledge as a Commons*. On y trouve notamment des contributions d'acteurs du logiciel libre, des données et des revues ouvertes.

## CHAPITRE 2

# Les communs numériques

Le monde numérique et en particulier la diffusion de la connaissance en format numérique prennent deux formes : l'une s'appuyant sur la propriété, les brevets et la génération de revenus, l'autre collaborative et basée sur le partage et les communs. L'imprimerie avait déjà permis une forme de partage dans les bibliothèques. Le numérique a provoqué une véritable lame de fond des communs avec notamment la science ouverte, les ressources éducatives libres, les réseaux numériques, etc.

Une rivière ou une forêt est un bien tout ce qu'il y a de plus physique, tangible, matériel. Par contraste, dans le cas du numérique, certains parlent de « commun immatériel ». Pourtant, des objets numériques comme des logiciels, des livres, des photos, des vidéos ont besoin de supports physiques : ils sont stockés sur des disques, transformés par les processeurs de *data centers*, transportés dans des réseaux de communications et visualisés sur des écrans. Les disques, les *data centers*, les réseaux, les écrans sont bien concrets, et ont une empreinte environnementale. C'est pourquoi nous préférons éviter le terme de « commun immatériel », et lui préférons celui de « commun numérique ».

Mais revenons sur cette notion essentielle de « rivalité » que nous avons déjà mentionnée. Un bien physique est « rival » :

sa consommation par un agent réduit la quantité de bien disponible pour les autres agents. Les poissons d'un étang sont des biens rivaux. *A contrario*, pour les objets numériques, on parle de biens non rivaux. La consommation d'un bien numérique par un agent ne change en rien les possibilités de consommation par d'autres agents. Ses coûts de production et de maintenance sont marginaux. C'est bien en cela que les communs numériques diffèrent de communs plus traditionnels.

### *Des biens non rivaux*

La non-rivalité des communs numériques change la focale. Pour des biens matériels, la propriété joue un rôle central. Pour des biens numériques, les droits d'usage prennent le pas. Par exemple, les licences d'usage prennent une importance particulière tant pour les logiciels (licences libres), que pour les contenus multimédias (Creative Commons) ou les données (licences de données ouvertes). Surtout, puisque ces communs sont non rivaux, on peut en multiplier les usages, aller jusqu'à en « ouvrir » l'accès : logiciels *open source*, données ouvertes, science ouverte, éducation ouverte, etc.

Comme une œuvre numérique peut être facilement reproduite, il est difficile d'en interdire le partage, d'exclure certains de son usage. Pourtant, certains s'y essaient pour en tirer plus de profits. Pour cela, ils s'appuient sur des mécanismes légaux défendant la propriété intellectuelle, par exemple le copyright. La technique même est appelée en renfort avec les DRM (*digital rights management*) censés empêcher une reproduction pourtant *a priori* inhérente au numérique. Pour les œuvres d'art numériques, on peut confier la preuve de propriété à des NFT (*non fungible token*), des jetons cryptographiques qui identifient la propriété d'objets numériques. L'objet peut être dupliqué, mais le NFT détermine à qui il appartient. En faisant cela, on cherche

artificiellement à apporter à des biens numériques non rivaux, le caractère rival et propriétaire de biens matériels.

Que devient la notion de commun quand on la transpose de biens matériels à des biens qui ne le sont pas ?

Pour un bien rival, le risque est que la demande soit trop forte. C'est le cas s'il y a trop de pêcheurs à l'étang communal de Champs-Romain : il n'y a pas de poissons pour tout le monde et on épuise la ressource. Cela ne se produit pas pour un logiciel libre ; mille, cent mille, un million d'utilisateurs peuvent le télécharger sans qu'il y ait pénurie. Pour un bien non rival, le risque est donc différent ; c'est celui d'une offre trop faible. Quelle est l'incitation à produire cette ressource s'il n'y a pas de retour financier ou autre de l'utilisateur au producteur ? Quelle est l'incitation à écrire des articles dans Wikipédia, à développer du logiciel libre ? Nous verrons que, selon les domaines, la question de l'offre trop faible trouve toute une gamme de réponses qui conduiront à des modèles économiques originaux.

Les communs numériques tirent leur succès de la non-rivalité. Parce que l'encyclopédie Wikipédia est non rivale, on peut la mettre en accès libre et en faire bénéficier des milliards de personnes. Bien sûr, il faut des serveurs pour répondre à tous ces utilisateurs, mais le coût par utilisateur est faible et on peut le couvrir avec des souscriptions d'une petite partie de ces utilisateurs. Mais l'apport du numérique aux communs ne s'arrête pas là.

En permettant des communications électroniques entre les membres d'une communauté, le partage de données entre eux, leur collaboration à grande échelle, le numérique ouvre un monde nouveau à la gestion de communs, agrandit considérablement le champ des possibles. Quand on passe d'une communauté dans un canton où les membres peuvent se retrouver facilement devant un verre ou un bon repas à une communauté de milliers de personnes dispersées dans le monde entier, on change la nature même du système.

Wikipédia ne pourrait pas exister sans le numérique. Cette encyclopédie est indissociable du numérique, tout comme Pl@ntnet ou Open Food Facts, des communs dont nous reparlerons plus loin, qui sont intrinsèquement numériques.

À côté des communs numériques, il faudrait aussi parler de communs dont la ressource n'est pas numérique mais que le numérique facilite rend possible. Dans l'enseignement, c'est Faire école ensemble<sup>1</sup> et dans l'agriculture Open Source Seeds<sup>2</sup>.

### *Des communs de la connaissance*

Une expression voisine, préférée par certains comme Hervé Le Crosnier, est « commun de la connaissance », qui indique clairement l'horizon : mettre en commun les connaissances humaines. Si on prend le mot « connaissance » dans un sens large incluant tous les savoirs comme ceux des paysans ou des artisans, les connaissances scientifiques, les matériaux éducatifs, les créations culturelles, toutes les informations numérisées, les bases de données, les logiciels, etc., on est très près des communs numériques, parce que ces connaissances sont de nos jours de plus en plus systématiquement numériques.

Nous avons préféré l'expression « communs numériques » d'abord parce que les communs de la connaissance n'ont pas à se limiter au numérique. Un théorème mathématique est un commun de la connaissance sans rien avoir de numérique. Surtout, nous l'avons préféré parce que le numérique ne s'est pas limité à améliorer des communs de la connaissance qui existaient avant l'informatique et l'ère numérique. Il a aussi donné naissance à de nouveaux types de ressources de connaissance comme les logiciels. Il a inventé de

---

1. <https://faire-ecole.org/>. **URL NON VALIDE**

2. <https://www.opensourceseeds.org/en/home>.

nouvelles formes de communs, les réseaux numériques, plus spécifiquement Internet et le Web. Il a participé à la transformation totale des communs de la connaissance, voire à l'invention d'autres qui ne se conçoivent que numériques. Il est véritablement au cœur du sujet. Tout cela explique notre choix de l'expression « commun numérique ». Cela dit, les communs dont nous allons parler sont massivement des communs de la connaissance (au sens large).

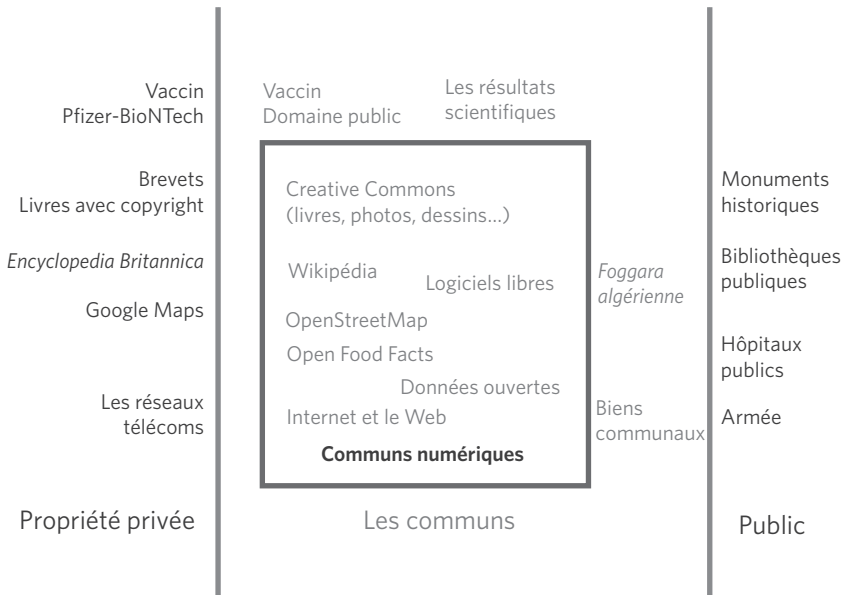
### *Une terre de conflits*

Si le numérique conduit à de nouvelles déclinaisons des communs et à de nouveaux modèles, cela passe souvent, comme pour les communs traditionnels, par des combats. S'il rend possible des utopies d'échange et de partage universels, on voit sa route ralentie, voire bloquée, par la propriété intellectuelle et les pressions commerciales. Les communs numériques se trouvent eux aussi confrontés à des tentatives d'enclosure.

Dans les premiers temps de l'informatique, les logiciels étaient gratuits, seuls les matériels étaient vus comme des biens privés, payants. Quand on achetait un ordinateur, on récupérait un paquet de logiciels qui allaient avec. Même si ce n'était pas articulé ainsi, on pouvait voir les codes numériques comme des biens communs au même titre que les théorèmes mathématiques ou les algorithmes. Le prix des machines était astronomique, et on considérait que les logiciels étaient inclus parce qu'ils étaient indispensables pour les faire fonctionner. La situation a changé quand les prix ont baissé. Surtout, les logiciels se sont éloignés du langage machine, et on a pu installer le même logiciel sur des machines différentes. Certains ont réalisé alors qu'on pouvait faire des profits en rendant les logiciels payants. Cela a même conduit à des brevets sur le logiciel aux États-Unis.

On a donc assisté pour les logiciels à un mouvement dans le plus pur style des enclosures. En réaction, des développeurs se sont organisés pour défendre l'accès à tous les logiciels et ont introduit les logiciels libres. Cela a résulté en des tensions, notamment autour de normes et de logiciels clés comme les systèmes d'exploitation ou les suites bureautiques.

Un scénario semblable s'est réalisé pour Internet et pour le Web. Basé au départ sur une idée de partage universel, dans le plus pur esprit de communs avec leur propre gouvernance, on a vu de très grandes entreprises y prospérer, y édifier des silos et s'approprier les ressources mises en commun par d'autres. Cela a résulté en l'extension du domaine du privé au détriment des communs.



**Figure 1.** Les communs entre propriété privée et biens publics.



## CHAPITRE 3

# Aux frontières des communs numériques

Si le concept de commun numérique a un potentiel considérable, il existe en revanche un risque à voir des communs partout, par exemple dans toute activité économique collective. Dans un kibboutz, on retrouve bien la notion de ressource partagée, la terre. Mais ce sont surtout l'activité économique et le travail qui sont partagés. De même, si avec l'éducation populaire on est amené à mettre en commun des ressources éducatives, le cœur du sujet est ailleurs. Dans de tels cas, on sort des communs.

L'économie sociale et solidaire est également proche de l'esprit des communs. Mais on parle d'entreprises, d'organisations, impliquées dans des activités économiques avec une forte fibre sociale. Les liens sont suffisamment forts pour questionner la frontière avec les communs. Si le sujet est passionnant, nous le laisserons à d'autres plus au fait de cette économie particulière.

Si, comme nous l'avons dit, un commun numérique se construit autour du triptyque ressources-communauté-gouvernance, il inclut bien sûr d'autres dimensions. Leurs communautés d'utilisateurs sont en général ouvertes, ne fonctionnant pas comme des clubs privés. La transparence de la gouvernance semble aussi faire l'unanimité.

Mais tout cela suffit-il à déterminer ce que sont les communs ? Nous concluons cette première partie avec trois sujets qui soulèvent des controverses.

### *Ouverts et gratuits ?*

Pour beaucoup, ces deux qualités sont indispensables pour être un commun. Pourtant, nous pensons que, même si les deux notions « être ouvert » et « être un commun » sont voisines, elles ne coïncident pas. Des données médicales peuvent être partagées dans un commun, mais elles ne sont pas ouvertes pour protéger leur confidentialité. Le fichier Sirene, Système national d'identification et du répertoire des entreprises et de leurs établissements, est ouvert mais, comme nous le verrons, on peut questionner le fait que ce soit un commun même si ce sont des données ouvertes, du fait de l'absence d'une véritable communauté gérant la ressource.

Nous avons également vu que même si les communs numériques sont souvent gratuits, « être gratuit » et « être un commun », cela ne veut pas forcément dire non plus la même chose. Des manuels éducatifs développés en commun, véritablement des communs, sont payants. Google Search ou le réseau social de Facebook sont gratuits. Ils n'en sont pas, pour autant, des communs.

Reste que les communs numériques sont le plus souvent ouverts et gratuits. D'autres exigences semblent aussi leur coller à la peau comme celle que la ressource ne soit pas figée, et qu'elle soit entretenue par la collectivité. Des contre-exemples existent comme un roman en Creative Commons : l'auteur préfère souvent qu'on ne puisse pas modifier son œuvre. On peut choisir d'inscrire de telles injonctions en dur dans la définition des communs, ou pas.

### *Un commun doit-il être éthique ?*

La question a été soulevée, notamment lors de l'utilisation de logiciels libres par Palantir pour fournir des services à l'Immigration and Customs Enforcement Agency (ICE) des États-Unis. Ces services ont été utilisés pour des expulsions d'enfants à la frontière mexicaine en violation de la Déclaration universelle des droits de l'homme des Nations unies.

Les communs numériques sont souvent des projets humanistes basés sur le partage et la collaboration. Mais faut-il inclure cet humanisme dans le cahier des charges ? L'encyclopédie Wikipédia propage des valeurs humanistes. C'est un commun. Mais on s'interroge sur les valeurs d'une encyclopédie comme Metapedia, une encyclopédie d'extrême droite qui défend la supériorité de la race aryenne et le suprémacisme blanc, entre autres. Est-ce qu'un commun peut sentir le soufre ?

Les licences de logiciels libres ont choisi d'ignorer tout jugement moral. Une licence GPL peut interdire certains usages d'un code, comme de distribuer une version modifiée du code sans fournir son code. Mais elle n'interdit pas une utilisation non éthique de ce code, comme son utilisation dans le logiciel d'un drone tueur. On peut voir dans ce choix une conséquence d'une forme de philosophie libertarienne, primauté à la liberté de l'utilisateur de faire ce qu'il veut avec le logiciel. S'il doit y avoir un jugement moral, il doit se situer ailleurs.

La reconnaissance faciale pour surveiller les musulmans ouïghours ou l'utilisation de drones pour exécuter des terroristes sont deux activités discutables, parmi d'autres, qui s'appuient sur des logiciels libres. On comprend bien qu'un développeur soit mal à l'aise avec l'idée que son code puisse servir à cela. Mais la régulation de l'usage des technologies, même si elle peut être considérée comme

souhaitable, ne relève pas du libre ou de l'*open source*, et une clause les interdisant contredirait l'esprit des licences correspondantes.

Il est clairement difficile de spécifier les applications considérées comme éthiques et celles qui ne le seraient pas. Des licences sont pourtant proposées pour interdire de telles utilisations, comme par exemple, la licence Hippocratic de l'Organization for Ethical Source. Cette licence introduite par Coraline Ada Ehmke propose d'ajouter à la licence *open source* du MIT la clause suivante :

Le logiciel ne peut pas être utilisé par des individus, des entreprises, des gouvernements ou d'autres groupes pour des systèmes ou des activités qui mettent en danger, nuisent ou menacent activement et sciemment le bien-être physique, mental, économique ou général d'individus ou de groupes défavorisés.

Même si cette licence s'appuie sur une licence *open source*, elle n'est pas elle-même *open source* selon sa créatrice ou selon l'Open Source Initiative, pas plus qu'elle n'est libre pour la Free Software Foundation.

L'idée de « faire le bien » avec la technologie, qui se développe avec des mouvements comme Tech for Good ou Data for Good, ne peut que s'enrichir au contact des communs numériques, et ne peut qu'enrichir l'idéal des communs. Mais les deux sujets peuvent être considérés comme indépendants.

### *Le moteur de recherche de Google est-il un commun ?*

En tant que système économique, les communs numériques coexistent avec le modèle de l'entreprise libérale. Comment ne pas s'émerveiller de voir les Wikipédia, OpenStreetMap, Linux, Open Food Facts, les ressources éducatives libres, et autres communs

numériques, s'épanouir à côté des Google, Amazon, Facebook, Apple, Microsoft, et autres fleurons du capitalisme flamboyant ? Quand les communs numériques prônent le partage et l'ouverture, ces entreprises gigantesques avec leur puissance financière enferment leurs utilisateurs dans des silos.

Un moteur de recherche comme celui de Google se nourrit essentiellement de ressources qu'il ne produit pas : les données publiques du Web, par exemple, et celles de Wikipédia qui arrive souvent parmi les premières réponses. Comme le coucou s'approprie le nid des autres, Google utilise des ressources de communs et bénéficie de l'écosystème des communs, en y contribuant modestement vu les profits qu'il en tire. Google aurait pourtant largement les moyens de contribuer de manière plus significative à Wikipédia, mais il ne le fait pas. Il est légitime de penser que des entreprises qui utilisent massivement un commun devraient être légalement tenues d'y contribuer. On trouvait d'ailleurs une telle proposition dans le projet de loi pour une République numérique qui a été soumis à consultation en 2016 : l'introduction du concept d'un domaine public informationnel visait à protéger une appropriation abusive des communs. Cette proposition a été écartée avant que la loi ne soit votée. Pourtant, il faudrait bien que des lois protègent les communs comme d'autres protègent la propriété privée, au risque sinon que la multiplication des acteurs qui en profitent sans y contribuer ne mette en péril les communs.

Notons que nous ne voulons surtout pas stigmatiser ici une entreprise particulière. Nous aurions pu prendre comme exemple le réseau social Facebook, le commerce en ligne d'Amazon, l'Apple Store, etc. À chaque fois, les ressources d'une communauté, celle des utilisateurs, des vendeurs de produits, des développeurs d'applications, sont utilisées par un acteur très puissant.

Maintenant, oublions quelques instants le modèle économique de Google pour nous concentrer sur l'usage de son moteur de recherche. On le consulte gratuitement, pas si différemment de la manière dont

on consulte Wikipédia. Ce service privé est devenu incontournable, quasi monopolistique, au point qu'on en arrive presque à le voir comme un service public. Pour certains, aller sur le Web consiste à aller sur le moteur de recherche de Google. Ne pourrait-on pas aussi considérer ce moteur de recherche comme un commun ?

Attention ! Là, il nous faut revenir sur son modèle économique. D'abord, sa gratuité est sujette à caution car nous payons avec notre attention en étant exposés à des pubs. Et nous payons cher si on considère les énormes profits que cette entreprise réalise. Et puis l'absence totale de contrôle par la communauté de ses utilisateurs pose problème. Il est surtout clair que ce service sert d'abord les intérêts de ses clients et de ses actionnaires, avant ceux de ses utilisateurs. En cela, il est fondamentalement différent d'un commun numérique, même s'il peut en sembler assez proche d'un point de vue des usages.

Mais peut-être pourrions-nous rêver que le moteur de recherche de Google se mette un jour au service de ses utilisateurs en proposant un ordonnancement neutre des résultats, en permettant à ses utilisateurs de participer, etc., c'est-à-dire qu'il se transforme véritablement en commun. Il est peu probable que le moteur de recherche de Google évolue naturellement, de son plein gré, dans ce sens. Mais, les règlements européens comme le *Digital Market Act* ou le *Digital Service Act* nous font faire de petits pas dans cette direction. Ce serait une manière pour ce moteur de recherche de rendre un peu de la valeur qu'il tire des ressources communes qu'il utilise sans contribuer en proportion aux communs qui les lui fournissent. Il se rapprocherait un peu de l'idéal des communs numériques. Nous avons le droit de rêver qu'un jour, du fait de son importance pour la société, le moteur de recherche de Google soit contraint de se transformer en commun.

Et tant qu'à rêver, nous pouvons aussi imaginer que d'autres services hyperdominants, comme des réseaux sociaux, des places de marché, etc., soient amenés par la loi et la régulation à se rapprocher de communs, au service de leurs communautés. Peut-être, mais la distance avec les communs est encore importante.

## *Pour conclure sur la définition des communs*

Le débat sur la définition exacte des communs est intéressant pour savoir de quoi on parle exactement. Dans le texte, nous basons notre définition sur le triptyque ressources-communauté-gouvernance.

Pour certains, d'autres aspects sont indispensables pour pouvoir se revendiquer comme un commun, notamment l'ouverture de la communauté, mais parfois aussi la transparence et la gratuité. Pour d'autres, un commun doit avoir une utilité sociétale, se doit d'être éthique. D'autres enfin incluent dans la notion de commun une idée de mouvement, de dynamique sociale, de combat peut-être. L'accent porte moins sur le fait qu'une communauté partage une ressource que sur la « mise en commun » (*commoning*) même de la ressource. On est un bien commun, on devient un commun. Bien sûr, tout cela conduit à des communs particulièrement intéressants.

Nous avons choisi de considérer aussi des systèmes qui ne satisfont pas à tous ces critères parce qu'ils présentent, selon nous, également leur propre intérêt. Nous considérerons même des cas où le triptyque n'est pas totalement présent, parce qu'ils peuvent constituer des étapes importantes vers la mise en commun. Nous aboutissons ainsi à une large palette de situations qui agrège de nombreuses expériences riches et intéressantes depuis les marges des communs jusqu'au cœur du dispositif que nous voulons raconter.





PARTIE B

# Les objets des communs numériques



Nous distinguons ici cinq grandes catégories qui couvrent peu ou prou l'ensemble des communs numériques : les données, l'information, les logiciels, les connaissances et les réseaux. Cette catégorisation propose une première classification assez naturelle de ces communs, selon la nature des ressources qu'ils concernent. Nous en considérons une autre en partie D, basée sur leurs usages.

Commençons par distinguer données, information et connaissances en reprenant une formulation précédente :

Une donnée est une description élémentaire, typiquement numérique pour nous, d'une réalité. C'est par exemple une observation ou une mesure.

À partir de données collectées, de l'information est obtenue en organisant ces données, en les structurant pour en dégager du sens. En comprenant le sens de l'information, nous aboutissons à des connaissances, c'est-à-dire à des « faits » considérés comme vrais dans l'univers d'un locuteur, et à des « lois » (des règles logiques) de cet univers<sup>1</sup>.

Et pour reprendre un exemple de Serge :

Des mesures de température relevées chaque jour dans une station météo, ce sont des données. Une courbe donnant l'évolution dans

---

1. Serge Abiteboul, *Sciences des données : de la logique du premier ordre à la Toile*, Collège de France/Fayard, 2012.

le temps de la température moyenne dans un lieu, c'est une information. Le fait que la température sur Terre augmente en fonction de l'activité humaine, c'est une connaissance<sup>1</sup>.

À partir de données, d'informations, ou de connaissances, on peut construire des communs. Pour les données, cela peut être des données de géolocalisation partagées sur OpenStreetMap ou les composants d'un produit alimentaire partagées sur Open Food Facts<sup>2</sup>. Pour les informations, cela peut être les articles en accès ouvert de *The Conversation* ou de Wikipédia. Pour les connaissances, l'exemple le plus emblématique est la mise en commun d'ontologies, c'est-à-dire, d'ensembles structurés de concepts et de leurs relations, décrivant un domaine spécifique. Un autre est la conception ouverte, qui met numériquement la description de la conception d'un objet à la disposition de tous.

Évidemment, la distinction entre ces trois domaines est des plus floues. Les communs les plus intéressants sont souvent aux frontières, voire dans le passage d'un domaine à l'autre :

- Le statut des logiciels, qui tiennent une place considérable dans les communs, n'est pas si clair. Un logiciel comme Linux peut être vu comme de l'information ; on peut « lire » des bouts de son code. Mais son contenu peut aussi être vu comme une collection de connaissances accumulées pour contrôler les ressources d'un ordinateur.
- Un article scientifique tient bien sûr de l'information, car ses auteurs informent de leurs résultats. Mais on espère qu'il contribue aux connaissances.
- Wikipédia est une encyclopédie, donc une compilation de connaissances. Mais ces connaissances sont des textes destinés à des humains. Qu'on butine Wikipédia ou qu'on atteigne ses pages à

---

1. *Ibid.*

2. Tous les exemples cités dans ce paragraphe sont décrits en détail plus loin.

partir de moteur de recherche, on est essentiellement confronté à une grande collection de textes, d'informations. Mais quand une base de connaissances formelles est extraite de ces textes, on passe du monde de l'information à un monde de pures connaissances.

L'organisation de notre première promenade dans les communs numériques résulte donc d'une classification et de choix relativement arbitraires. Nous l'entamons avec les données. Puis nous considérons l'information, avant de passer aux logiciels qui ont souvent servi de modèles pour les communs numériques. Nous abordons ensuite les connaissances, les ontologies et la conception ouverte. Nous concluons cette partie avec les réseaux de télécommunications, Internet et le Web, qui sont à la base du partage de données, d'information et de connaissance.



## CHAPITRE 4

# Les données

Depuis son invention en Mésopotamie à la fin du IV<sup>e</sup> millénaire avant J.-C., nous utilisons l'écriture pour représenter des données. Avec le numérique, nous avons fait un saut considérable dans la production automatique et le stockage de données. Les données numériques traitent de tous les aspects de nos vies, de notre monde, de notre société, du commerce, de la santé, de la politique, de la météo, de la nature, des religions, etc. Elles sont de plus en plus produites massivement par des capteurs et des logiciels. Des logiciels les analysent et les présentent visuellement, entre autres, pour expliquer une réalité comme le déclin des petites villes, ou pour proposer des choix comme celui d'un itinéraire qui évite les bouchons à une automobiliste.

Certaines données sont réservées à un usage personnel. D'autres sont utilisées par un État ou une entreprise. D'autres enfin, ouvertes à tous, au service de la collectivité, tiennent une place centrale dans de nombreux domaines comme les sciences ou le gouvernement.

**LES FORMATS ET LE STOCKAGE.** Les données ouvertes sont stockées dans des fichiers avec des formats dits *ouverts*, c'est-à-dire conformes à des spécifications visibles et utilisables par tous. S'il n'y a pas consensus sur un format unique de données, certains formats émergent comme UTF-8 (pour les formats de texte brut),

CSV (.csv), JSON (.json), XML (.xml). Des domaines spécifiques proposent leurs propres formats comme KML (.kml) et SHP (.shp) pour les données géographiques.

### *L'ouverture des données*

Les corpus de données ouvertes sont accessibles sur le Web à partir de portails d'organismes, de territoires, d'États. Par exemple, les données publiques ouvertes sont recensées en France dans le portail data.gouv.fr. Les données de ces corpus sont parfois immuables, ce sont des archives. Souvent, elles sont « vivantes », en ce qu'elles sont modifiées et/ou historisées pour tenir compte des évolutions naturelles des phénomènes qu'elles mesurent ou pour corriger des erreurs. Quand on dispose de données ouvertes, on peut aussi vouloir offrir des services à partir de ces données ; cela peut consister notamment à proposer des requêtes qui permettent de les utiliser sans avoir à charger des jeux entiers de données.

Dans le cours de ce livre, nous allons rencontrer des ouvertures massives de données, en particulier par :

- l'État et les collectivités : c'est un domaine emblématique de l'utilisation de données ouvertes ;
- les sciences : les scientifiques partagent les fruits de leurs travaux les uns avec les autres. Le site FAIRsharing.org recense un très grand nombre de bases de données scientifiques ouvertes de la chimie, à la biologie, à la médecine, à l'archéologie ;
- les entreprises privées : certaines ouvrent également leurs données.

Nous détaillerons ces usages dans des chapitres spécifiques, ainsi que leurs motivations.



## *La production par la multitude de données ouvertes*

Pour obtenir des données, on peut aussi s'adresser à une foule d'utilisateurs et leur demander de contribuer. On parle de données produites par « les foules », par la « multitude », en anglais : *crowd-sourcing*.

Le numérique a montré que la multitude savait produire des données en dehors de tout circuit officiel, indépendamment des gouvernements, de la presse ou des partis politiques. Quand on fait appel à la multitude pour recueillir et maintenir des données, celle-ci répond si le sujet l'intéresse. C'est ce qui s'est passé avec Open Food Facts, où chaque contributeur scanne des listes d'ingrédients de produits alimentaires, ou avec OpenStreetMap, où chacun participe à l'enrichissement de cartes géographiques. Dans les deux cas, les données sont produites et mises à jour en continu ; elles sont validées par la communauté qui assure la gouvernance du projet. Ces données ne sont pas au service de quelques-uns ou d'une entreprise, mais au service de la société. L'objectif est de contribuer aux communs de la connaissance.

Pour les données ouvertes produites par l'État, on part le plus souvent de données existantes que l'on met en commun en s'appuyant sur tout un processus réglementé d'ouverture. Dans le cas du *crowd-sourcing*, les données sont produites directement avec l'intention d'en faire un commun. C'est la multitude des contributeurs qui va construire le corpus de données ; la valeur du tout est le produit du travail de fourni de chacun. Parfois, l'histoire se termine par l'acquisition du commun par une entreprise, comme celle de la base de données de films IMDb (Internet Movie Database), rachetée par Amazon en 1998, ou celle du site de la *Digital Photography Review* (une référence pour les appareils photos et les photos numériques) aussi rachetée par Amazon en 2007 et fermé depuis.

Dans le monde d'avant, les « experts » pouvaient affirmer que le nuage de Tchernobyl ne menaçait pas la France (pour caricaturer une communication officielle optimiste qui s'est révélée en total décalage avec la réalité). À l'époque déjà, une association (la CRIIRAD) s'était créée pour faire connaître la vérité sur la contamination radioactive du territoire français. Aujourd'hui, des réseaux d'organisations citoyennes comme Safecast, née après la catastrophe nucléaire de Fukushima Daiichi en 2011, publient des données ouvertes pour la surveillance de l'environnement, et en particulier sur les rayonnements ionisants et la qualité de l'air. Les données sont récoltées collectivement avec une qualité professionnelle, ce qui permet leur utilisation par des scientifiques et des gouvernements. En 2020, 120 millions d'observations ont été relevées dans plusieurs centaines de villes dans le monde. Par ces mouvements, des citoyens produisent collectivement des données qui sont mises au service de la collectivité.

De telles formes de *crowdsourcing* citoyen nous situent clairement dans le cadre des communs. Les individus fournissent aussi des informations à des plateformes privées, parfois sans le savoir, simplement par les choix qu'ils font. Cela peut être l'appréciation d'un service d'hôtellerie (Booking), de restauration (Tripadvisor), de vidéo (Netflix), de musique (Deezer), etc. Ces données servent aux systèmes de recommandation de ces plateformes. Même s'il s'agit aussi de partage de connaissances, nous ne les considérons pas dans le périmètre de notre essai, parce que l'objectif n'est là pas de servir la collectivité, et que cette collectivité n'est pas du tout associée à la gouvernance de la ressource.

**OpenStreetMap.** Créé en 2004, OpenStreetMap produit et maintient une base de données géographiques libres du monde en s'appuyant sur le système GPS. C'est un peu le Wikipédia des cartes. (Wikipédia est souvent la référence dans le monde des données produites par la multitude.) Des citoyennes et des citoyens, des collectivités, des associations et des entreprises utilisent les données d'OpenStreetMap, pour concevoir par exemple des cartes thématiques en accès libre. Selon OpenStreetMap, en 2022, la communauté compte plus d'un million et demi de contributrices et de contributeurs dans le monde. Comme celui de Wikipédia, son modèle économique est basé sur le travail des bénévoles et des donations.

Au départ du projet, on trouve un informaticien britannique, Steve Coast. Avant OpenStreetMap, les bases de données géographiques étaient sous le contrôle des États, et il fallait payer pour y accéder. On pouvait leur reprocher que le citoyen qui avait déjà payé pour la construction de la carte avec ses impôts devait payer une seconde fois pour les utiliser. Aujourd'hui, quelques entreprises se partagent le domaine de la cartographie numérique comme Google, Apple et Bing. OpenStreetMap est le seul commun à contester leur domination.

## CHAPITRE 5

# L'information

Des personnes produisent de l'information sous forme de textes, d'images, de musiques, de vidéos.

**LES FORMATS ET LE STOCKAGE.** Les formats d'information sont très variés et dépendent évidemment du média concerné. Pour le texte, cela peut être OpenDocument Text (.odt), Hypertext Markup Language (.html), Document PDF (.pdf), Livre numérique ePub (.epub), etc. Pour les photos et images, cela dépend principalement de la technique de compression : JPG, PNG, GIF, etc. Pour les vidéos, les formats sont souvent propriétaires, c'est-à-dire que ce sont des spécifications définies et contrôlées par des intérêts privés. Des formats propriétaires comme ceux de Word (.docx) ou Excel (.xls) sont aussi largement utilisés pour le partage d'information du fait de la popularité du logiciel correspondant. Un format de Word, Office Open XML, est devenu ouvert, en sous-format de XML, sous la pression politique.

Le plus souvent, les auteurs tiennent à ce que leur propriété intellectuelle soit reconnue sur leurs œuvres, et que celle-ci soit placée sous la protection du droit d'auteur (droit français) ou du copyright (plus international). Si Albert Cohen restera éternellement l'auteur de *Belle du seigneur* (le droit moral), les droits patrimoniaux de son roman qui en déterminent les droits d'exploitation prendront fin

soixante-dix ans après la mort de l'auteur, en 2051. Un auteur peut aussi choisir de faire de son œuvre un commun.

## *Le droit d'auteur*

Jusqu'au Moyen Âge, l'œuvre littéraire ou la création artistique sont souvent anonymes, artisanales. Le droit d'auteur date vraiment en Europe de la Renaissance et surtout de l'invention de l'imprimerie. Les temps modernes l'ont vu s'exacerber. Des auteurs vont aujourd'hui jusqu'à interdire, au nom du droit d'auteur, la production par d'autres de fictions basées sur leurs personnages (la *fan fiction*). Pourtant, quand le droit d'auteur était moins affirmé, des auteurs, et non les moindre comme Shakespeare, empruntaient volontiers leurs personnages, voire leurs histoires, à des auteurs antérieurs.

Le droit d'auteur est une notion de droit français. Sa cohabitation avec le droit international du copyright est complexe. Des législations différentes d'un pays à l'autre rendent difficile l'interprétation de ce droit en dehors de la France.

De manière paradoxale, avec ce renforcement du droit d'auteur, le numérique rebat les cartes en favorisant la réalisation d'œuvres plus collectives, en questionnant la propriété intellectuelle d'objets quand ces objets deviennent numériques.

La propriété d'un livre n'est pas de même nature que celle d'une pomme ou d'un poisson. Quand on a mangé une pomme, elle n'est plus. Quand on a lu un livre, il existe heureusement encore. La mise en commun est donc simplifiée. On peut en faire des copies, et la même copie peut être lue séquentiellement plusieurs fois, ce qui a conduit aux bibliothèques. Le numérique accentue le caractère non rival du livre. Il est possible aujourd'hui de fournir la même œuvre numérique à des millions de personnes à coût quasiment nul. Pourtant, les éditeurs insistent pour faire payer chacun pour le livre comme s'il bénéficiait d'un produit unique en s'appuyant sur

des mécanismes artificiels comme les DRM (*digital right management*) qui rendent plus difficile le partage et la réplique d'œuvres numériques. L'édition est confrontée à une évolution rapide de son métier du fait du numérique ; les modèles économiques traditionnels ne sont plus forcément bien adaptés.

Le numérique permet des souscriptions à des bibliothèques pour la musique (comme Spotify ou Deezer) et les films (comme Netflix ou Prime Video). Amazon propose aussi un abonnement qui permet une lecture illimitée de livres parmi plus d'un million de titres. Surtout, s'appuyant sur l'énorme masse de textes déjà dans le domaine public, des sites proposent gratuitement des livres : le précurseur, le Projet Gutenberg, mais d'autres également comme Ebooks Gratuits ou Gallica de la BNF. Nous sommes clairement dans une époque de transition qui cherche une voie pour pleinement profiter de la diffusion des œuvres que permet le numérique tout en rétribuant correctement les créateurs.

### *L'œuvre libre*

L'œuvre libre, ou le contenu libre, est une œuvre dont aucune contrainte n'est imposée à la diffusion. Chacun peut distribuer autant de copies qu'il le souhaite d'une telle œuvre, image, texte, musique, etc.

Une œuvre libre est souvent une ressource figée. Une nouvelle édition d'une telle œuvre ne modifie ou ne supprime pas la précédente. La liberté s'entend comme une liberté de la diffuser. On n'a pas le droit en général de la modifier. Dans certains cas, cette liberté peut être poussée plus loin avec le droit de modifier l'œuvre en question, de s'en servir pour créer d'autres œuvres, voire même de la traduire sans même avoir à citer l'auteur initial.

Le mouvement des Creative Commons s'est précisément développé en rupture avec l'accent mis sur la propriété et sa défense par

le copyright. Son idée est d'apporter plus de flexibilité à la distribution des œuvres en évitant le choix binaire entre propriétaire ou totalement ouvert. Créé par des personnes comme Lawrence Lessig, ce mouvement a conduit à proposer toute une gamme de licences de libre diffusion, parfois appelées « licences ouvertes ». En choisissant sa licence, on peut décider de ne protéger que certains des droits associés classiquement à la propriété d'une œuvre. Par exemple, la licence Creative Commons BY-NC-SA demande de citer l'auteur (BY), de ne pas tirer un profit commercial de l'œuvre sans autorisation de l'auteur (NC), et d'utiliser la même licence si on rediffuse l'œuvre (SA). Nous reviendrons plus en détail sur la question des licences en partie C.

Ce mouvement est loin d'être anecdotique. Il s'est, par exemple, imposé dans les encyclopédies. Les contenus de Wikipédia, un succès planétaire, sont en Creative Commons. Nous verrons que la mise en commun d'œuvres (articles, livres, etc.) est devenue prépondérante dans des domaines comme la science ouverte ou l'éducation ouverte. Les contenus d'un blog du *Monde* comme Binaire ou d'un magazine comme *The Conversation* le sont également.

Nous verrons que le Web est une source considérable d'œuvres en Creative Commons (1,6 milliard d'œuvres en 2020).

Pour ce qui est de la préservation patrimoniale d'œuvres numériques, que ce soient des communs ou pas, une partie du travail est réalisée par les États, par exemple par la Bibliothèque nationale de France et l'Institut national de l'audiovisuel en France, et par la Library of Congress aux États-Unis. Une association se distingue dans ce domaine, Internet Archive. Selon Wikipédia, elle détiendrait en 2022 dans sa Wayback Machine : 34 millions de livres et de textes, 7 millions de films, de vidéos et d'émissions de télévision, 14 millions de fichiers audio, 4 millions d'images, 1 million de fichiers multimédias, 2 millions de clips télévisés et plus 681 milliards de pages Web dans la Wayback Machine.

## *Militantisme : l'art et la culture libre*

On peut regrouper sous la dénomination d'« art libre » tout une gamme de mouvements artistiques se dégageant du droit d'auteur pour atteindre un large public, et mettant leurs œuvres à disposition gratuitement. Ce mouvement se décline dans tous les arts : musique, cinéma, littérature, peinture, etc. On peut aussi lui rattacher des formes d'art comme l'art urbain (*street art*) ou le théâtre de rue. Cela ne se limite donc clairement pas au numérique.

On parle parfois de « libérer » la culture. L'idée est de pouvoir sans contrainte distribuer, accéder, modifier des œuvres d'art. Dans le cadre du numérique, la thèse est la suivante : si j'ai acheté un livre, une musique, un film, en format numérique, je peux en faire ce que je veux. Cela relève de mes libertés fondamentales. Je dois aussi pouvoir m'inspirer des œuvres auxquelles j'ai accès pour en créer de nouvelles. En grossissant le trait, les droits d'auteur et le droit de propriété s'effacent devant le droit d'accès à la culture, celui de créer en toute liberté, la liberté d'expression. Sans surprise, ce mouvement lutte par exemple contre les DRM (gestion des droits numériques).

On appelle libristes les militants qui luttent pour limiter la propriété intellectuelle dans le champ de l'art, de la culture, des connaissances. Leur militantisme peut les conduire à violer des lois. On pourrait mettre dans cette catégorie des personnalités géniales selon les uns, sulfureuses pour d'autres comme Alexandra Elbakyan ou Aaron Swartz ; nous les retrouverons dans le chapitre Sciences ouvertes, un domaine où le libriste est particulièrement populaire. Leur combat est proche de celui d'autres militants qui veulent donner un accès libre à des données ou des informations politiques traditionnellement tenues secrètes, comme Edward Snowden qui a mis en accès libre en 2013 des informations classées top secret de la NSA (National Security Agency).



## CHAPITRE 6

# Le logiciel

Des algorithmes permettent de spécifier comment résoudre des problèmes. Un programme informatique, un logiciel, c'est la transcription d'un algorithme dans un langage formel exécutable par un ordinateur. C'est l'ordinateur qui va exécuter le logiciel, donc l'algorithme. Les logiciels tiennent une place particulière parmi les communs numériques, d'abord du fait de leur succès en tant que commun, mais également, parce que les logiciels libres et *open source* ont souvent servi de modèles pour la mise en commun d'autres objets numériques de toutes natures et enfin parce qu'ils sont souvent présents comme composants d'autres communs.

Un logiciel a deux facettes :

- un code source qui a été écrit par des développeurs ;
- un code objet, le résultat de la compilation du code source, qui s'exécute sur une machine.

Vous pouvez prêter un livre à une amie, qui peut le lire ou en photocopier des passages, et vous le rendre (ou pas). Même chose pour un article, une photo, un journal, ou un code source (si on dispose de quelques notions d'informatique). Mais un code objet, c'est un animal différent. Il est à peu près incompréhensible pour un être humain.

Un développeur peut inclure dans son code des mécanismes pour contrôler son exécution, par exemple, n'autoriser cette exécution que sur certaines machines. Le propriétaire tire ainsi plus de profits de la vente de son code. C'est ce qui s'est passé jusqu'à la fin des années 1990, l'industrie du logiciel s'étant développée en s'appuyant sur le secret des codes sources. Pendant toutes ces années, le développement de l'informatique a été principalement tiré par les milieux académiques. Les améliorations de codes venues de l'industrie n'avaient souvent que des impacts incrémentaux parce qu'elles restaient enfermées dans chaque logiciel. Et puis Richard Stallman est arrivé, qui a apporté l'idée révolutionnaire que les codes sources des programmes informatiques, comme les théorèmes de mathématiques, pouvaient être la propriété de tous. Chacun pouvait alors en faire l'usage qu'il souhaitait : l'exécuter, le lire, le comprendre, le modifier, le réutiliser, et le diffuser. Son idée révolutionnaire consiste à voir des logiciels comme des communs numériques.

Cette idée simple a permis à la technologie du logiciel de réaliser au XXI<sup>e</sup> siècle des progrès considérables. Avec le logiciel libre, on peut construire des bibliothèques mondiales de logiciels réutilisables fondées sur des communautés qui les maintiennent et accompagnent les nouveaux utilisateurs. Le logiciel libre et l'ouverture du code source ont véritablement révolutionné l'industrie du logiciel en multipliant sa capacité à produire.

Tout développeur à accès à des bibliothèques gigantesques de logiciels ouverts, la plus grande étant GitHub. On remarquera que GitHub a été acheté en 2018 par Microsoft pour 7,5 milliards de dollars, ce qui montre bien que le développement en commun de logiciels n'est plus du seul fait d'idéalistes comme Stallman : l'industrie s'est emparée du sujet.

**Richard Stallman** (né en 1953) étudie à Harvard et devient *research assistant* au laboratoire d'intelligence artificielle du MIT en 1975. C'est un militant, un des plus influents défenseurs des logiciels libres et critiques des logiciels propriétaires et des DRM (qui limitent l'utilisation de ressources numériques). Il est à l'origine de la suite de logiciels libres GNU et de la licence GPL de logiciel libre. Il a également beaucoup contribué au développement de logiciels très populaires comme l'éditeur GNU Emacs ou le compilateur C de GNU (GCC).

**GitHub** est une entreprise qui propose de l'hébergement de logiciels. Ses services sont gratuits pour les logiciels libres. GitHub est devenu le plus important dépôt de logiciels libres ou privés au monde. En 2020, GitHub avait 40 millions d'utilisateurs et plus de 190 millions de référentiels de logiciels.

GitHub utilise le logiciel de gestion décentralisé de versions Git créé en 2005 par Linus Torvalds, auteur du noyau Linux. Le service GitHub aide les développeurs dans la gestion de leurs logiciels. Il assure le contrôle d'accès, le suivi des bugs, des demandes de nouvelles fonctionnalités, des tâches, des versions temporaires, etc. Il offre aussi des fonctionnalités de réseau social pour tenir compte du fait que le développement collaboratif de logiciel est une activité extrêmement sociale.

GitHub a été lancé en 2008. Sa base de données a grossi très vite jusqu'à son acquisition par Microsoft en 2018. Elle continue de grossir.

Pour ce qui est de la préservation patrimoniale des logiciels, la fondation Software Heritage s'y consacre depuis 2016 et archive déjà en 2022 des milliards d'artefacts logiciels distincts (fichiers, répertoires, révisions, versions).

Il nous faut maintenant faire une distinction subtile entre logiciels libres, en anglais, *free software*, et logiciels *open source*, littéralement, logiciels dont le code source est ouvert. Si les deux expressions dénotent des concepts voisins, elles tiennent de deux philosophies différentes. L'histoire va nous permettre de préciser leurs différences.

À l'origine de l'informatique (dans les années 1950 et 1960), l'objet commercial est le matériel (*hardware*), c'est cela qu'on paie. Le logiciel (*software*, un jeu de mots sur *hardware*) n'est là que pour faire fonctionner le matériel ; il est fourni gratuitement avec le matériel en question. Puis on s'aperçoit qu'il est possible de faire fonctionner le même logiciel sur des matériels différents. Les utilisateurs prennent alors l'habitude de s'échanger des logiciels, de les modifier pour les adapter à de nouveaux matériels. Comme ils ont l'impression de ne payer que le matériel, ils n'ont même pas conscience du fait que les logiciels sont libres ou *open source*.

Puis, dans les années 1970 et 1980, la place du logiciel devient de plus en plus importante. Les entreprises en font commerce, les « éditeurs de logiciels » naissent en tant que nouvelle catégorie d'entreprise informatique et la notion de logiciel propriétaire, accompagné d'une licence d'utilisation et d'un prix, s'impose.

### *Logiciel libre et logiciel open source*

C'est en réaction à cette tendance qu'en 1983 Richard Stallman lance le projet GNU (Gnu is Not Unix) pour produire un système d'exploitation complet et libre. C'est ce système qui est à l'origine des systèmes d'exploitation modernes, dont Linux. Il crée la Free Software Foundation, donne une définition du logiciel libre en 1986 et propose une licence pour ces logiciels, la GPL (General Public Licence) en 1989. À partir de là, le mouvement s'amplifie et le concept de logiciel libre se répand, principalement dans le

monde académique. Pourtant, dans les années 1990, cela reste un mouvement minoritaire de militants. Les choses changent à la fin du siècle dernier quand Netscape décide d'ouvrir l'accès au code de son navigateur Internet et de créer Mozilla pour coordonner le développement de futurs logiciels libres.

En février 1998, Eric Raymond pense que ce basculement est le signe du passage d'un mouvement de militants à un mouvement industriel. Il publie un article intitulé « Goodbye, "free software" ; hello, "open source" », dans lequel il propose une stratégie pour faire accepter le concept de logiciel libre par l'industrie. Pour lui, ce qui empêche l'industrie de s'appropriier l'idée de mise en commun de logiciel, c'est le terme *free software* qui est ambigu et qui fait peur. Est-ce qu'on parle de *free* comme dans *free beer* (une bière gratuite) ? Ou de *free* comme dans *free speech* (la liberté d'expression) ? L'industrie, qui doit rentabiliser ses investissements, est mal à l'aise avec la gratuité. Pour Raymond, le terme *open source* pourrait emporter le soutien de la grande masse des entreprises. Ce terme obtient d'ailleurs le soutien de grands noms du domaine comme Jon « Maddog » Hall (un militant connu du libre), Bruce Perens (le porteur du projet Debian), et Linus Torvalds (le créateur du projet Linux) lui-même. Ce changement de terminologie va permettre de passer du petit club des militants initiaux du libre à une adoption massive de l'*open source* par l'industrie du logiciel.

Dans le passage du libre à l'*open source*, l'esprit a fondamentalement changé, les adjectifs libre et *open source* s'appuyant sur des valeurs essentiellement différentes. Richard Stallman ne s'y trompe pas, qui continue à promouvoir le logiciel libre, en commençant, par exemple, toutes ses conférences sur le sujet en France par : « Je peux expliquer le logiciel libre en trois mots : *liberté, égalité, fraternité.* » Ce sont de grands principes qu'il défend avec le logiciel libre, des principes absents des logiciels *open source*.

À partir de ce moment, deux communautés coexistent (et collaborent), celle du logiciel libre et celle de l'*open source*, chacune

avec ses spécificités, ses motivations, sa gouvernance, ses licences, ses modèles économiques, ses propres narratifs. Ces deux communautés vont promouvoir toutes les deux des logiciels ouverts, même si leurs points de vue philosophiques restent essentiellement différents. Dans l'ouverture du code, les tenants de l'*open source* voient une méthode efficace de développement de logiciel, les tenants du libre voient des impératifs éthiques, le moyen de défendre les libertés de l'utilisateur. On utilise parfois l'acronyme FOSS (pour *free and open source software*) pour ce que nous avons nommé simplement *open source*, et FLOSS (pour *freelibre and open source software*) pour ce que nous avons appelé libre. Si des conflits ont pu surgir entre ces deux communautés, elles coexistent finalement souvent relativement paisiblement, au point que les termes *open source* et *libre* sont parfois même utilisés indifféremment.

Concrètement, les logiciels libres ou *open source* ont des propriétés très proches : leur code est ouvert, ils sont gratuits, ils sont construits par des communautés, même si ces communautés peuvent avoir des cultures et des motivations différentes. Un tel logiciel, *open source* ou libre, est régi par une licence, définie par une entreprise ou une communauté. Ce qu'on peut en faire dépend de sa licence, et les licences peuvent être plus ou moins contraignantes dans les deux cas.

Nous rencontrerons souvent de tels logiciels, *libres* ou *open source*, dans cet ouvrage ; ils nous serviront souvent d'ailleurs de fil rouge pour présenter différentes facettes des communs numériques. Nous espérons apporter des débuts de réponses à différentes questions : Qu'est-ce qui fait des logiciels des objets tout particulièrement aptes à devenir des communs ? Est-ce parce qu'ils se retrouvent à la confluence des ambitions les plus nobles de partage et de liberté (des logiciels libres) et des recherches d'efficacité, de performance (des logiciels *open source*) ? Leur plus grand succès n'est-il pas de mettre l'efficacité au service de la société ?

**Framasoft.** Depuis 2004, Framasoft est un réseau consacré principalement au logiciel libre. Il a démarré autour de l'idée d'encourager l'utilisation de logiciels libres dans l'enseignement. Son nom est dérivé de Fr@m@net (pour FRANçais et MATHématiques en intraNET), un projet éducatif interdisciplinaire. C'est un acteur incontournable du logiciel libre en France. Framasoft soutient la création de logiciels libres, et propose des applications très populaires comme Framapad, un carnet de notes collaboratif, ou Framadate, une alternative à Doodle.

Plus récemment, l'association s'est positionnée sur le déploiement de services libres installés sur ses serveurs ou des serveurs de partenaires (en mode SaaS, *software as a service*). Le but est de rendre disponibles ces services au plus grand nombre.

## CHAPITRE 7

# Les connaissances

Les humains découvrent et transmettent des connaissances. C'est ce qui permet à l'humanité de progresser. Pendant longtemps, ces connaissances ont été transmises oralement, puis elles ont été diffusées dans les langues parlées et écrites (chinois, arabe, latin, français, anglais, etc.) Par exemple, aux alentours du VIII<sup>e</sup> siècle avant J.-C. en Inde, le Baudhayana capture en Sanscrit l'état des connaissances de l'époque en géométrie.

### *Les connaissances sous forme de texte*

Au XVIII<sup>e</sup> siècle, Diderot et d'Alembert conçoivent et dirigent l'ambitieux projet de réunir l'ensemble des connaissances humaines, donnant naissance à leur encyclopédie qui sera rédigée par une équipe d'experts. Cette approche perdure dans l'histoire, et de nombreuses encyclopédies commerciales voient le jour (l'*Encyclopædia Britannica*, *Encyclopédie Universalis*, etc.). Au début du XXI<sup>e</sup> siècle, Wikipédia réinvente la notion d'encyclopédie en remplaçant les experts en nombre par nature limité, par une multitude de passionnés. Nous raconterons plus loin le combat entre l'*Encyclopædia Britannica* et Wikipédia.



### **Wikipédia et Jimmy Wales**

Créée en 2001, **Wikipédia** a inventé une nouvelle façon de produire une encyclopédie, en s'appuyant sur une multitude d'éditeurs qui ne sont pas forcément des spécialistes des sujets, mais des passionnés qui vérifient leurs sources et les mises à jour des autres. Dans une absence apparente de hiérarchie, la communauté est en fait assez bien organisée. L'accès à l'encyclopédie est gratuit, Wikipédia vivant du travail gratuit de ses bénévoles et des donations de ses soutiens. L'accès à une encyclopédie était limité dans l'ancien monde à ceux qui pouvaient s'offrir un ou deux volumes de synthèse, ou l'édition complète. La troisième édition de l'*Encyclopædia Britannica* de 1797 et son supplément de 1801 consistaient en vingt volumes. Le domaine de Wikipédia est beaucoup plus large, et des mises à jour sont réalisées en permanence. Selon Wikipédia:Statistiques : « Elle est alimentée chaque jour par plus de 100 000 contributeurs à travers le monde. Elle est visitée chaque mois par près de 500 millions de visiteurs et propose plus de 60 millions d'articles dans plus de 300 langues. Plus de 25 000 articles sont créés par jour sur les différentes versions linguistiques de Wikipédia et on compte plus de 10 millions de modifications par mois. » On parlera plus loin de son modèle économique fondé sur les dons de ses utilisateurs.

**Jimmy Wales** (né en 1966), un homme d'affaires américain, est le cofondateur avec Larry Sanger de Wikipédia. Il a eu un rôle essentiel dans le développement de la galaxie Wikimedia, et du projet Wiki, une application Web pour l'édition collaborative de pages Web. Plus récemment, il a lancé WT:social, une plateforme de microblogging qui vise à concurrencer Twitter et où la modération est réalisée collectivement

## *Les connaissances formalisées*

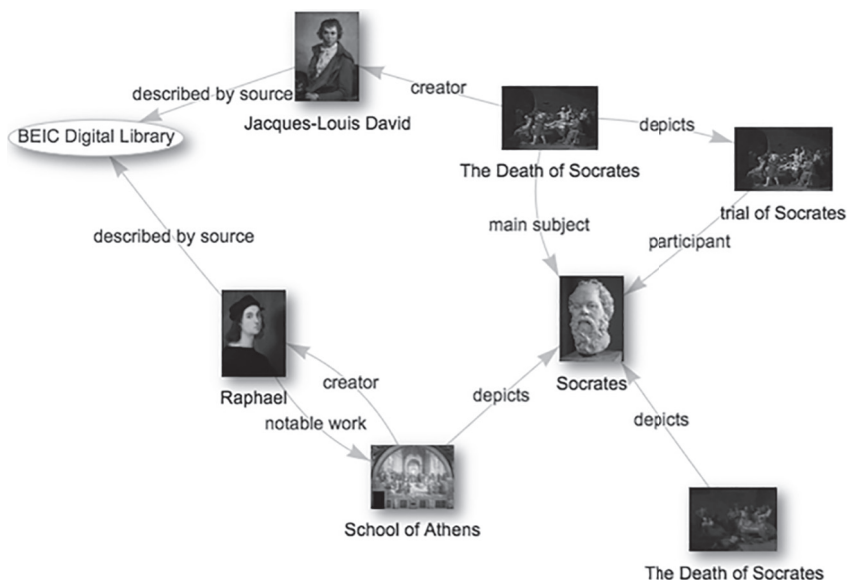
On peut transcrire des connaissances sous forme numérique. Cela peut trivialement consister à numériser un texte, par exemple, le Baudhayana. Mais, dans certains cas, on peut aller plus loin en

spécifiant des connaissances dans des langages formels et en se servant de ces spécifications pour guider des logiciels. Pour illustrer, considérons le cas des formules mathématiques. On peut utiliser le langage LaTeX pour préciser comment une formule mathématique doit être présentée à des humains. Mais on peut aussi utiliser MathML (le Mathematical Markup Language) du W3C (le World Wide Web Consortium, qui gère le Web et dont nous parlerons plus loin). MathML permet de spécifier, séparément, la présentation d'une formule mathématique (*presentation MathML*) et son contenu, son « sens » (*content MathML*). En s'appuyant sur le contenu, un programme peut ensuite évaluer la formule ou la composer avec d'autres formules. Le tableau 1, probablement parfaitement incompréhensible pour le commun des mortels, illustre LaTeX et MathML.

**Tableau 1.** LaTeX et MathML.

Formule	LaTeX	PresentationMathML (début de l'expression)
$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$	<code>x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}</code>	<code>&lt;math mode= "display" xmlns="http://www. w3.org/1998/Math/ MathML"&gt; &lt;semantics&gt; &lt;mrow&gt; &lt;mi&gt;x&lt;/mi&gt; &lt;mo&gt;=&lt;/mo&gt; &lt;mfrac&gt; &lt;mrow&gt; &lt;mo&gt;±&lt;/mo&gt; &lt;math form="prefix"&gt;&amp;minus; &lt;/math&gt; &lt;mi&gt;b&lt;/mi&gt; ... </code>

Un autre exemple de représentation de connaissance est proposé sous forme graphique (et en anglais) en figure 2.



(Crédit : Wikimedia Commons/Fuzbeado CC BY-SA 4.0 DEED.)

**Figure 2.** Schéma de connaissance de Wikidata illustrant Socrate [Wikipédia].

Pour formaliser des connaissances, il est d'abord indispensable de fixer les terminologies au moyen d'une « ontologie ». En informatique, une ontologie définit un ensemble de termes (comme « Socrates » dans l'exemple précédent), de concepts (comme artiste et œuvre), de relations entre ces concepts (comme *notable work*) et de leurs propriétés (comme *described by source*). En définissant une ontologie, les experts d'un domaine spécifique se mettent d'accord sur une terminologie pour parler de leur domaine. À partir de là, ils sauront exprimer plus précisément de quoi ils parlent. On pourra aussi écrire des logiciels pour les aider, par exemple, à annoter des textes ou à faire des recherches dans des corpus de textes. Il s'agit bel

et bien d'un passage obligé pour faire réaliser des tâches complexes de traitement de connaissances à des machines.

Si de telles ontologies sont de plus en plus présentes dans la recherche scientifique, si elles prennent leur place dans l'industrie, arrivent-elles jusqu'au grand public ? Oui, même si le grand public n'en a pas toujours conscience. Supposons que l'on pose au moteur de recherche Google la question : « Quel âge avait Coluche quand il est mort ? » En plus de la suite habituelle de propositions de pages de réponses, le moteur retourne la réponse « 42 ans ». Cette réponse a été obtenue en interrogeant une base de connaissances que Google a construite, le Google Knowledge Graph. C'est cette même base de connaissances qui lui permet d'accompagner ses pages de résultats, d'« infoboxes », ces petits encadrés qui présentent un condensé d'informations sur un sujet. Ce sont également ces bases de connaissances qui permettent le fonctionnement des assistants vocaux.

Comment de telles bases sont-elles construites ?

- Le système YAGO<sup>1</sup> (Yet Another Great Ontology) est automatiquement extrait de sources d'information ouvertes, en particulier de Wikipédia. Il a été développé à l'Institut d'informatique Max-Planck de Sarrebruck et à Télécom Paris. Il est en accès libre.
- Wikidata est une base de connaissances multilingue éditée par la multitude de personnes hébergée par la Wikimedia Foundation. Elle est également en accès libre.

On notera que, pour construire sa propre base de connaissances, Google a utilisé Wikipédia. Mais si Google offre bien une API<sup>2</sup> pour interroger un graphe de connaissances dans l'esprit de celui de la figure 2, le graphe lui-même n'est pas en accès ouvert et les

---

1. <https://yago-knowledge.org/>.

2. Une API, pour *application programming interface* (« interface de programmation d'application » en français), est une interface logicielle qui permet à un logiciel d'interagir avec un autre logiciel, en échangeant des données et des commandes.

sources de ces connaissances ne sont pas citées. Le fait que Google Knowledge Graph utilise des sources ouvertes mais qu'il ne soit pas lui-même ouvert questionne. On peut voir cela comme un pillage de communs numériques. En 2022, le moteur de recherche Qwant utilise aussi Wikipédia pour ses infoboxes, mais explique qu'il tire ses informations de Wikipédia et propose de soutenir l'encyclopédie, comme on le voit dans cet extrait d'infobox de Qwant.



**Alan Turing**  
Mathématicien et cryptologue britannique  
(1912-1954)

Alan Mathison Turing, né le 23 juin 1912 à Londres et mort le 7 juin 1954 à Wilmslow, est un mathématicien et cryptologue britannique, auteur de travaux qui fondent scientifiquement...

[Read more on Wikipedia](#)

**Date et lieu de naissance** : 23 juin 1912 à Maida Vale, Royaume-Uni

**Date et lieu de décès** : 7 juin 1954 à Wilmslow, Royaume-Uni

**Professions** : informaticien ou informaticienne, mathématicien, professeur d'université, cryptographe, logicien, statisticien ou statisticienne, marathonnier, chercheur en intelligence artificielle

**Parents** : Julius Mathison Turing, Ethel Sara Stoney

**Site Web** : [www.turingarchive.org](http://www.turingarchive.org)

Wikipedia.org • How to contribute?

**Figure 3.** Extrait de l'infobox d'Alan Turing sur Qwant avec référence explicite à Wikipédia.

La retranscription de connaissances sous forme numérique peut suivre des chemins très différents. Dans les sciences, cela peut passer par l'utilisation de cahiers d'expériences numériques pour suivre pas à pas la science en train de se faire. Cela peut être également la description et le suivi détaillé des processus mis en œuvre dans la recherche. L'idée principale est de diminuer voire d'éliminer le flou pour que les résultats de l'expérience et les différentes étapes d'un cheminement qui mène à une connaissance ne laissent pas de place à l'imprécision, pour qu'elles puissent facilement être reproduites mécaniquement.

**LES FORMATS ET LE STOCKAGE.** Pour les ontologies, on peut utiliser RDF, le langage de description des contenus pour le Web sémantique, ou le langage OWL pour Web Ontology Language, qui permet de décrire la sémantique d'un domaine sous forme de règles logiques.

**Logipedia.** À l'initiative de Gilles Dowek, une communauté se construit depuis 2019 pour créer une encyclopédie de preuves mathématiques. Le but à long terme est de construire une encyclopédie qui contient toutes les démonstrations mathématiques disponibles. Un informaticien combine des bouts de programmes informatiques pour résoudre son problème. De même, on imagine qu'un mathématicien puisse combiner des bouts de preuves pour prouver le résultat sur lequel il travaille. Un tel outil constitue un aspect important de l'évolution des mathématiques, car il permet d'envisager la construction de démonstrations beaucoup plus complexes.

De la même façon qu'il existe de nombreux langages de programmation, il existe plusieurs systèmes de preuves. Le but de Logipedia est aussi de leur permettre d'interopérer. En s'appuyant sur un standard commun de preuves, on permettra d'utiliser dans la même preuve des éléments de preuves développés avec des systèmes différents.

Les buts sont véritablement de mieux mettre en commun les preuves mathématiques disponibles, et d'aider les mathématiciens à construire de grandes preuves.

On a vu comment on pouvait mettre en commun des données, des contenus multimédias, des logiciels, des connaissances. Voyons maintenant comment mettre en commun des « conceptions ».

## *La conception ouverte*

*« Give ordinary people the right tools, and they will design and build the most extraordinary things. »*

Neil GERSHENFELD.

La conception dont il est question ici est celle d'objets physiques. Le point de départ est le savoir-faire technique, l'artisanat, le bricolage, la bidouille, mais de plus en plus leur rencontre avec les communs numériques, et en particulier, le logiciel libre.

Prenons un objet en trois dimensions. On peut modéliser mathématiquement sa forme et la capturer par une représentation numérique. On peut ensuite demander à une imprimante 3D de la reproduire. Quand la ou les personnes qui ont conçu un objet mettent volontairement la spécification numérique de cet objet en accès ouvert, on parle de « conception ouverte », *open design* en anglais. Comme pour le logiciel libre ou la science ouverte, le partage de connaissances est à l'œuvre : on utilise les conceptions des autres, leurs idées, on les enrichit de ses propres idées, on modifie, on adapte, on améliore et on offre à la communauté le résultat de ses efforts. En ouvrant une conception, on la libère, en laissant à chacun la possibilité de fabriquer l'objet.

Le problème de la conception des objets est bien entendu ancien. Au Moyen Âge, il était central pour le compagnonnage. L'histoire de ce mouvement est complexe et multiforme. On en trouve la trace aussi loin qu'au xv<sup>e</sup> siècle, mais le terme lui-même n'apparaît dans la langue française qu'à la fin du xviii<sup>e</sup>. On retrouve un certain nombre de traits des communs dans les diverses formes de compagnonnage (elles sont nombreuses) : la communauté bien entendu, l'intérêt commun qui est celui d'apprendre, les règles de fonctionnement et très clairement un mode de partage de la connaissance suivant des

modes très particuliers (la plupart des compagnons ne sachant ni lire ni écrire).

Au XVIII<sup>e</sup> ou au XIX<sup>e</sup> siècle, il n'était pas rare de partager des informations sur les processus de fabrication d'objets, de machines ou de systèmes. Puis on a vu fleurir les brevets pour protéger la propriété intellectuelle. Pour augmenter ses profits et se protéger des contrefaçons, l'industrie s'est mise à cacher la conception, en ne laissant voir que les produits finis. Comme pour le logiciel ouvert, le résultat de la conception ouverte est souvent aujourd'hui un code numérique, même si le but est ici la production d'un objet physique.

Plus récemment, le mouvement Maker a émergé dans les années 2000. Dans le sillage de la conception ouverte, on voit surgir une multitude de lieux : *fablabs*, *makerspaces* et tiers-lieux divers. On recense aujourd'hui de l'ordre de 1 700 fablabs dans le monde, dont 400 en France (chiffre 2019). Ils reçoivent des financements publics ou de fondations.

Un fablab (pour « laboratoire de fabrication ») est un atelier qui propose à ses membres de fabriquer des objets à partir de spécifications. Pour cela, il est équipé de logiciels de conception (architecture, photo, dessin) et d'outils contrôlés par ordinateur comme des imprimantes 3D, des découpeuses laser, des fraiseuses numériques, etc. Tout cela, bien que très technique, est bien loin de la production de masse, plus dans l'esprit du laboratoire que dans celui de l'atelier industriel. Avec l'utilisation d'outils numériques, la frontière entre les experts et les amateurs se brouille. La conception ouverte démocratise la création et le prototypage d'objets techniques en la mettant à la portée de l'amateur qui peut concevoir et fabriquer ses propres objets, en se réappropriant la technique. Il n'est plus nécessaire de disposer de compétences pointues et d'outils sophistiqués. La conception devient collective, pilotée ou parfois même réalisée par des utilisateurs.

Tout cela est fort sympathique, mais sommes-nous encore dans les communs ? La référence aux logiciels libres souvent présente



semble l'indiquer. Les communautés sont bien identifiées (fablabs, makerspaces, hackerspaces, etc.) avec leurs règles. Et puis on partage des ressources : d'abord les outils, mais aussi les savoir-faire, les spécifications de conception. La communauté est typiquement ouverte et accueillante. Mais des règles sont nécessaires pour entretenir des machines souvent fragiles, difficiles à mettre au point, comme des imprimantes 3D, et aussi pour protéger les utilisateurs. Le fablab a donc une charte de gouvernance, la Fab Charter. La communauté dépasse même l'espace du lieu unique dans des partages à plus grande échelle. Pour cela, les fablabs sont organisés mondialement en réseau sous l'égide de la Fab Foundation. Ses membres essaient souvent d'utiliser les mêmes machines, de partager les expériences. Nous sommes donc bien dans les communs numériques.

**Neil Gershenfeld**, né en 1959, est un physicien et informaticien américain, professeur au Massachusetts Institute of Technology (MIT) et directeur du Center for Bits and Atoms également au MIT. Il est à l'origine de la création des fablabs, avec Bakhtiar Mikhak, à la fin des années 1990. Le but était de familiariser les enfants et le public avec la technologie, de les rendre capables de concevoir et fabriquer leurs propres objets. Le premier fablab a été fondé en 2001 par le Media Lab du MIT, en collaboration, entre autres, avec le Center for Bits and Atoms. On trouve maintenant des fablabs un peu partout dans le monde.

Les outils numériques démocratisant la conception et la fabrication d'objets, ce mouvement ne pouvait être ignoré de l'industrie. L'*open design* a donc commencé à essaimer dans les entreprises, en permettant des interactions surprenantes avec des communautés fondées sur le partage. Il ne s'agit pas de concurrencer la production industrielle classique ; on se situe le plus souvent dans le cadre de production de prototypes ou d'un petit nombre d'objets. Il s'agit

plutôt de proposer une alternative moins coûteuse qu'une production traditionnelle d'objets manufacturés, de servir parfois des causes sociales ou écologiques, ou encore d'aider des pays émergents.

Un exemple intéressant dans ce contexte est celui d'Arduino. Arduino est une entreprise d'origine italienne qui conçoit et fabrique des microcontrôleurs et leurs logiciels sous licences libres<sup>1</sup>. Leurs produits sont populaires auprès d'amateurs ou de professionnels ainsi que dans le milieu éducatif pour la construction d'objets numériques. Arduino permet la construction facile et peu coûteuse d'objets qui interagissent avec leur environnement à l'aide de capteurs et d'actionneurs. Le partage de spécifications est au cœur du fonctionnement de sa communauté.

Un tel partage est aussi à la base de communautés comme Thingiverse et son site Web. Les membres y publient des fichiers de conception sous licences libres qui permettent à d'autres de piloter la construction d'objets avec des outils standards, notamment les imprimantes 3D de MakerBot, l'entreprise qui gère le site Web.

Pour un dernier exemple de conception ouverte, l'entreprise Wikispeed propose des voitures *open source*. Les spécifications de construction de voitures et leurs fichiers CAO (conception assistée par ordinateur) sont accessibles librement. On notera que la conception de ces voitures suit les idées modernes de conception de logiciel comme le *framework scrum* et fait parfois même appel à la communauté pour que cette dernière participe à la conception.

Pour ce qui est des formats de stockage, de nombreux langages sont utilisés pour le *design*. Ils dépendent souvent du logiciel utilisé, par exemple les fichiers AutoCAD pour le logiciel du même nom. Des formats comme DFX ou DWG sont parfois utilisés pour échanger des données entre deux systèmes.

---

1. Un microcontrôleur est un circuit intégré qui inclut les éléments de bases d'un ordinateur, son processeur, des mémoires et des interfaces d'entrées-sorties.

## CHAPITRE 8

# Les réseaux comme bien commun

« Les réseaux comme bien commun » est une devise de l'Arcep, l'autorité française indépendante chargée (entre autres) des télécoms. Ce sont des réseaux numériques, Internet et le Web qui rendent aujourd'hui possible l'accès mondial aux connaissances, le développement de l'économie numérique, d'un monde numérique. Dans ce chapitre, nous considérerons des liens entre ces réseaux et les communs.

Dans un premier temps, les acteurs historiques, les entreprises de télécommunications comme AT&T ou France Télécom ont proposé des communications analogiques sur les infrastructures physiques qu'ils développaient. C'était l'époque du « 22 à Asnières ». Ces réseaux ont peu à peu cédé la place à des réseaux numériques. En se déployant sur toute une gamme de support, fibre optique, câble, téléphones cellulaires, liaisons satellitaires, les réseaux numériques sont devenus une clé de voûte du monde numérique. De grandes entreprises du numérique, notamment les GAFAM (Google, Apple, Facebook, Amazon et Microsoft), ont rejoint les acteurs traditionnels des télécoms pour participer au développement des infrastructures de communications numériques.

Ces réseaux numériques ont vu converger trois classes de services qui se sont développés à l'origine isolément : les télécommunications,

l'audiovisuel, et l'informatique. On tend aujourd'hui à distinguer dans les réseaux numériques plusieurs niveaux :

1. le niveau des services applicatifs (les applications Web ou autres) ;
2. les services d'infrastructure (sécurité, contrôle, intelligence artificielle, etc.) ;
3. l'infrastructure numérique (les services qui permettent par exemple de stocker et de transmettre des données) ;
4. l'infrastructure physique (antenne, câble, centre de données, etc.).

Les réseaux numériques, qui à leurs débuts (au temps d'Arpanet<sup>1</sup>) étaient surtout publics, sont devenus principalement privés. Certains appartiennent à des entreprises privées comme Orange, SFR, Bouygues et Free en France qui proposent au grand public des accès payants à Internet (les réseaux télécoms). D'autres appartiennent à de grandes entreprises numériques, souvent les GAFAM. Ils servent alors de support aux activités propres de ces entreprises ou sont monayés pour des applications professionnelles, par exemple, Amazon Web Services ou Microsoft Azure. Sur les réseaux numériques, s'est déployé Internet.

**INTERNET** est un réseau de réseaux informatiques. En informatique, le terme « réseau » désigne un ensemble d'ordinateurs reliés entre eux et à des terminaux (des ordinateurs, des smartphones, etc.). Un réseau permet d'échanger des données. Certains des réseaux qui composent Internet appartiennent à des entreprises privées, d'autres à des organismes publics. Certains sont des réseaux locaux (LAN, pour *local area network*), d'autres des réseaux distants (WAN, pour *wide area network*). Internet permet le déploiement d'applications comme le courrier électronique, les échanges de fichiers, ou les vidéoconférences. L'une de ces applications est particulièrement populaire, le Web.

**LE WEB** est un réseau de contenus et de services. Il est implanté sur Internet.

---

1. Arpanet est le premier réseau informatique à transfert de paquets de données conçu dans les années 1960 aux États-Unis. Cette technique de communication sera à la base d'Internet quelques années plus tard.

Que peut-on dire des contenus et des services qu'on trouve sur Internet et le Web ? Ce sont parfois des communs, mais ils sont souvent propriétaires (possédés et contrôlés par des entreprises privées). En fait, ce n'est pas simple de s'y retrouver. Par exemple, regardons ce qui se passe quand on cherche des illustrations pour un site Web ou une plaquette promotionnelle. On découvre assez vite que beaucoup de photos accessibles sur le Web sont propriétaires. Mais des sites<sup>1</sup> proposent aussi des photos en Creative Commons. La licence Creative Commons (quand elle est bien spécifiée) précise pour une ressource si on peut la réutiliser gratuitement (droit patrimonial) et, quand c'est le cas, qui créditer (droit moral).

Les réseaux numériques, Internet, et le Web sont des lieux de rencontre entre des entreprises privées et le domaine public. Ils sont devenus incontournables, essentiels, pour tous les internautes, toutes les entreprises. En cela, on peut choisir de les voir comme des biens communs à toute l'humanité même si les entreprises privées y occupent une place considérable.

### *Ces réseaux sont-ils des communs ?*

Considérons le triptyque ressources-communauté-gouvernance.

Les ressources partagées sont bien identifiées : des réseaux qui nous permettent à tous de vivre dans un monde numérique.

La communauté des utilisateurs est bien identifiée pour ces réseaux : toute l'humanité connectée. Les communautés de contributeurs diffèrent selon les réseaux. Pour les réseaux numériques, ce sont des entreprises privées comme les GAFAM ou des opérateurs télécoms. Pour Internet, ce sont tous ceux qui font fonctionner les réseaux : les opérateurs télécoms qui fournissent l'accès à Internet, les transitaires (qui relient les différents réseaux internationaux) comme Liberty

---

1. Par exemple, <https://wordpress.org/openverse/>.

Global, les centres d'interconnexion Internet (IXP) comme France IX, les réseaux de diffusion comme Akamai, etc. Ce sont typiquement de grosses entreprises qui réalisent des investissements considérables, comme de faire passer des câbles de fibres optiques (contrôlés en grande partie par des GAFAM) à travers les océans. Pour le Web, ce sont les fournisseurs de contenus et de services, depuis la prof de biologie qui propose un site Web à ses élèves, jusqu'à Netflix ou Amazon.

Pour les télécoms, la gouvernance est définie depuis longtemps ; il s'agit de l'Union internationale des télécoms et d'instances de régulation nationales. La gouvernance définit des choix d'affectation de fréquences, des standards de communication. À chaque État ensuite de piloter les télécoms sur son territoire. Pour l'Union Européenne, les choix sont collectifs.

Pour les réseaux numériques hors télécoms, la gouvernance est celle des grandes entreprises qui suivent les lois du marché.

Pour Internet et le Web, la gouvernance est souple, imprécise, diffuse. On pourrait les qualifier de communs à gouvernance floue. On remarquera que malgré (grâce à) l'absence d'une gouvernance forte, ils fonctionnent plutôt bien et même efficacement, ce qui est déjà en soi extraordinaire vu la complexité de la tâche ; ils ont même superbement résisté à la pandémie du Covid-19 comme d'ailleurs les réseaux numériques en général, et les réseaux télécoms en particulier.

Tout le monde est conscient de l'importance sociétale considérable prise par ces réseaux. Nous citerons trois questions sociétales fondamentales qu'ils soulèvent :

- L'inclusion numérique : dans un monde où tout passe par ces réseaux, il est impossible d'accepter que certains en soient exclus pour des questions de fracture sociale (comme l'impossibilité de payer un abonnement), géographique (zones non connectées) ou cognitive (illettrisme et illettronisme).
- La question environnementale : ces réseaux, avec leurs centres de données, les communications, et surtout avec les terminaux des

utilisateurs et autres objets connectés, ont des impacts environnementaux importants et croissants.

- La sécurité : il s'agit de protéger les internautes, les applications, les serveurs, contre les intrusions, fraudes, cyberattaques, usurpation d'identité, etc.

Du fait de l'importance de ces questions, on comprend bien que les États et les communs ont des places essentielles à tenir dans ces réseaux, mais chacun à sa place.

### *Les réseaux télécoms mobiles*

Les ondes radios sont utilisées avec une large gamme de technologies et de fréquences à l'origine d'une diversité d'usages : téléphone mobile, TNT, téléphonie et télévision satellitaire, bluetooth, wi-fi, citizen-band, faisceau hertzien, GPS, etc. Ces ondes, qui transcendent les frontières, sont des biens rivaux. Les différentes utilisations sont en compétition, ce qui peut conduire à des brouillages. Même pour un usage spécifique, plusieurs utilisateurs peuvent se trouver en compétition. Pour s'en convaincre, souvenez-vous des moments où, dans un grand rassemblement de personnes, vous n'avez pas réussi à appeler ou envoyer des photos, vidéos ou autres. Ça arrive de moins en moins souvent du fait de l'accroissement des capacités des réseaux mobiles mais cela vous donne une idée du phénomène à l'œuvre : les ondes peuvent saturer, ce sont bien des ressources rivales.

Pour éviter des problèmes de brouillage, notamment aux frontières, les États définissent les grandes lignes de l'utilisation des fréquences dans un organisme international, l'Union internationale des télécoms. Nationalement, l'État s'approprie les fréquences en les plaçant dans le domaine public. Il peut choisir de les gérer directement, par exemple, pour l'armée. Parfois, il confie la gestion de certaines fréquences, moyennant finance ou certains engagements, à des

entreprises privées ; c'est ce qui se passe en France de la 2G à la 5G. Mais les communs peuvent également être convoqués. Examinons deux facettes des télécoms où c'est le cas, le wi-fi et l'Open RAN.

**LE WI-FI : LES FRÉQUENCES EN COMMUN.** Nous sommes en 1985 et la FCC (le régulateur des télécoms états-unien) décide que les usages sont « libres » sur ce qui est alors considéré comme des bandes inexploitable, des *junk bands*, autour de 2,4 et de 5,5 gigahertz. Cela conduit à l'émergence du wi-fi sur ces mêmes bandes. Pour ces fréquences qui intéressent alors peu de monde (notamment du fait de brouillage possible par les fours à micro-ondes), il est alors décidé de les laisser ouvertes dans le monde entier, à tous, sans affectation préalable. À l'époque, leur utilisation s'est développée du fait d'avancées technologiques : on a appris à faire coexister un grand nombre d'utilisateurs sur une même fréquence. Comment ? En s'appuyant sur des standards de communications proposés par l'IEEE<sup>1</sup> et poussés par un consortium d'entreprises, la Wi-Fi Alliance. Des terminaux arrivent à communiquer sur ces bandes de fréquences sans qu'il y ait de brouillage, grâce aux protocoles et aux technologies utilisées. Cela transforme la vision selon laquelle une bande doit être attribuée à un usage donné, à un utilisateur précis.

Le wi-fi est une technique très simple de partage parce qu'elle n'exige pas de « synchronisation » entre les machines qui communiquent entre elles (de l'adjectif « synchrone » qui indique une réalisation dans le même temps ou à des intervalles de temps égaux). Le wi-fi est en compétition avec les techniques synchrones de type GSM, beaucoup plus complexes mais de meilleure qualité. C'est parce qu'elles sont complexes, qu'on ne peut les laisser libres et qu'elles sont confiées à des opérateurs de télécommunications.

Comme sur Internet, le message transmis par wi-fi est coupé en petits paquets. Les terminaux utilisent des fréquences disponibles

---

1. L'Institute of Electrical and Electronics Engineers (IEEE), ou, en français, Institut des ingénieurs électriciens et électroniciens, est une association professionnelle.



parmi celles allouées au wi-fi. Typiquement, un particulier ou une entreprise va installer son réseau local wi-fi et différents objets pourront échanger entre eux, voire utiliser à partir de ce réseau une passerelle vers Internet. Et cela n'empêchera pas d'autres personnes ou entreprises de le faire. La réglementation des bandes libres impose une faible portée et donc de plus faibles puissances des points d'accès. Du coup, le spectre est bien utilisé.

Le succès est international et place le wi-fi en concurrence directe pour les fréquences avec la téléphonie cellulaire sur laquelle des opérateurs télécoms sont chargés de gérer des fréquences qui leur sont individuellement attribuées (en les monnayant par exemple avec des forfaits). En 2022, par exemple, un débat porte sur l'affectation des fréquences hautes de la bande 6 gigahertz entre biens privés (la téléphonie cellulaire pour la 5G ou 6G) ou biens communs (le wi-fi). La question est complexe parce qu'il peut s'agir d'usages différents, de la connexion en mobilité pour le premier, à la connexion à un réseau fixe pour le second. Il peut aussi s'agir dans les deux cas de la même fonctionnalité : on peut, depuis son salon, visionner la même vidéo en 4G ou en wi-fi en passant par sa box.

**OPEN RAN : MATÉRIEL ET LOGICIEL *OPEN SOURCE*.** Un *radio access network* est un environnement matériel et logiciel pour accéder avec des ondes radios à un réseau télécom, par exemple en 2, 3, 4 ou 5G. Traditionnellement, des industriels comme Nokia, Ericsson, et Huawei, construisent de tels réseaux. Typiquement leurs matériels et leurs logiciels n'interopèrent pas entre eux, donc un opérateur télécom doit choisir l'un de ses fabricants pour déployer une station radio (donc son réseau) dans un lieu spécifique.

L'Open RAN marque un changement de philosophie dans les architectures de réseaux mobiles : il permet de faire collaborer des composants de différents industriels. Différentes fonctions classiquement regroupées dans le même matériel sont séparées ; des fonctions peuvent être virtualisées, c'est-à-dire qu'au lieu d'être réalisées par des composants matériels, elles sont confiées à des logiciels. La véritable

révolution est que les interfaces entre ces nouveaux composants sont ouvertes et interopérables. L'opérateur télécom n'a plus à choisir un fabricant unique ; il peut préférer intégrer des composants de plusieurs d'entre eux.

Un groupe d'industriels, au sein de l'alliance O-RAN, fondée en 2018, définit les spécifications ouvertes de tous les composants et les interfaces entre eux.

L'intérêt principal de l'Open RAN est de pouvoir sortir de l'enfermement résultant d'un unique fabricant, et en particulier, de réduire les coûts. L'Open RAN offre les avantages que nous avons vus des logiciels *open source* : on peut partager les compétences et les expériences. En particulier, on peut plus facilement embarquer dans le réseau local des techniques nouvelles comme de l'apprentissage automatique. Évidemment, de tels systèmes sont complexes et les défenseurs des systèmes propriétaires argumentent qu'on ne pourra pas optimiser aussi bien les performances dans un contexte ouvert, ni atteindre les mêmes niveaux de sûreté. Les déploiements de réseaux ouverts sont encore limités. Mais ils se multiplient. Par exemple, le groupe Rakuten (connu pour sa filiale d'e-commerce) déploie un nouveau réseau de communications en Open RAN au Japon.

## *Internet et le Web*

Si les communs sont présents dans les réseaux numériques des GAFAM à travers les logiciels *open source* que ces réseaux utilisent, s'ils sont encore loin de dominer dans les télécoms, ils sont bien installés dans Internet et le Web depuis leurs débuts.

Internet est né en 1969, quand le premier lien entre l'Université de Californie à Los Angeles (UCLA) et le Stanford Research Institute (SRI, centre de recherche sur le campus de l'Université de Stanford) a été établi. Le Web est né en 1990 comme un prototype au CERN et est devenu public en 1993. À leur naissance,

les deux réseaux, Internet et le Web, n'intéressaient que quelques poignées d'idéalistes, de chercheurs et d'ingénieurs. On imaginait mal leur croissance future. Leur gouvernance était alors totalement informelle, principalement technique, et tout se passait si bien qu'on a assisté à leur incroyable succès. De fait, la gouvernance consistait essentiellement à se mettre d'accord sur des protocoles libres : TCP/IP pour Internet et HTML/HTTP pour le Web. Si personne ne parlait alors de communs, c'étaient bien à notre avis des communs qui se mettaient en place. Il est important de regarder plus en détail comment Internet et le Web fonctionnent, et plus particulièrement leur gouvernance.

**INTERNET.** Internet est un réseau de réseaux. Chacun des réseaux qui le compose définit sa propre façon de fonctionner. Le fonctionnement général est défini à partir d'échanges électroniques, sur des listes de courriers et surtout par le partage de documents, nommé partage de RFC (*request for comments*). À ses débuts, la gouvernance dépendait de l'ARPA<sup>1</sup>. Petit à petit, la communauté s'est emparée des commandes. Elle est passée sous la responsabilité de l'IETF (Internet Engineering Task Force) qui rassemble les groupes de travail chargés des standards et de l'IAB (Internet Architecture Board) qui propose les choix d'architecture. La véritable gouvernance d'Internet, notamment les spécifications techniques des protocoles comme TCP/IP est donc confiée à l'IETF. L'IETF choisit de façon transparente et collégiale comment ces spécifications évoluent. C'est un groupe informel sous les auspices de l'Internet Society<sup>2</sup>, une association internationale fondée en 1992. L'Internet Society, sert en quelque sorte de garant moral et technique, fournit un socle juridique et participe à la promotion de l'Internet à travers le monde.

---

1. L'ARPA est le prédécesseur de la DARPA (Defense Advanced Research Projects Agency), agence du département de la Défense des États-Unis chargée de la recherche et développement des nouvelles technologies destinées à un usage militaire.

2. <https://www.internetsociety.org/>.

Un autre aspect de la gouvernance est l'adressage des serveurs : les noms de domaines (par exemple inria.fr) et les adresses IP (par exemple 128.93.162.83). C'est un organisme d'origine américain, l'ICANN<sup>1</sup>, qui est responsable des bases de données qui gèrent ces adresses et associent adresses IP et noms de domaines.

*We reject kings, presidents and voting ; We believe in rough consensus and running code*<sup>2</sup> (Dave Clark, 1992).

Nous rencontrons ici une forme de gouvernance particulière. Le cœur du réacteur consiste en des groupes de travail qui sont chargés de points techniques à résoudre et doivent faire émerger un « consensus approximatif ». Une fois ce consensus approximatif atteint, les travaux de développement de code sont lancés, et des expérimentations. Les dissensions sont enregistrées si elles existent, et on peut revenir sur le consensus si on rencontre des difficultés avec les choix qui ont été faits. Pour qu'un standard devienne officiel, il faut que plusieurs développements indépendants aient démontré qu'il fonctionne parce que les codes peuvent interopérer.

En 1996, l'Internet Architecture Board publie *Les Principes architecturaux de l'Internet*. Dans l'introduction, on notera : « Le but de la communauté est la connectivité, l'outil est le protocole IP et l'intelligence est aux extrémités du réseau plutôt qu'en son sein. » Il s'agit d'abord de permettre des échanges entre les participants de cette communauté (la connectivité). L'importance du protocole est mise en avant ; nous retrouverons cet aspect avec le Web. Le dernier point – l'intelligence est aux extrémités du réseau plutôt qu'en son sein – est vraiment le choix technique qui va faire le succès d'Internet et du Web. Le réseau est mis au service des applications déployées sur les ordinateurs situés aux extrémités.

1. L'ICANN, Internet Corporation for Assigned Names and Numbers, est une société à but non lucratif : <https://www.icann.org/>.

2. « Nous rejetons les rois, les présidents et le vote ; nous croyons dans le consensus approximatif et le code qui marche. »

Les protocoles d'Internet ne cessent d'évoluer. Une évolution récente est, par exemple, le passage du protocole IPv4 à IPv6, rendu nécessaire en partie parce qu'on commençait à manquer d'adresse IP. Cette mise à jour qui pourrait paraître relativement bénigne (augmenter la taille des adresses, donc le nombre d'adresses IP disponibles) est compliquée du fait qu'il faut migrer un nombre considérable de réseaux et de machines vers le nouveau protocole, que chaque élément du réseau est indépendant, et que tout doit se passer sans interrompre Internet.

Jean-François Abramatic, ancien président du W3C (World Wide Web Consortium), explique le succès d'Internet par cinq raisons principales<sup>1</sup> :

1. Un soutien continu des pouvoirs publics américains (pendant plus de vingt ans), comparé au soutien limité et épisodique d'initiatives similaires en Europe.
2. L'équilibre et les coopérations entre la communauté de l'informatique et celle des télécommunications.
3. L'émergence rapide d'une communauté scientifique et technique bâtie sur la culture du consensus, sur l'expérimentation concrète, sur une vision universelle.
4. La circulation des idées favorisée par la très grande mobilité des hommes et les fréquents échanges entre le monde académique, les organisations publiques et les entreprises privées.
5. L'inclusion des entreprises privées très tôt (dès la fin des années 1980) dans le développement d'Internet.

On voit bien dans cette liste l'importance que joue la communauté dans son développement et sa diversité.

**LE WEB.** Le World Wide Web, « la Toile » en bon français, est au départ un système hypertexte mondial sur Internet, permettant d'accéder à des applications distantes. Le protocole du Web est HTTPS pour *hypertext transfer protocol secure* (HTTP, dans sa

---

1. Communication privée.

version sans authentification ni chiffrement). Ce protocole a été défini à ses débuts par Tim Berners-Lee à partir du langage d'hypertexte SGML. Le Web tient sa puissance de son universalité : tous les utilisateurs peuvent s'y connecter et toutes les entreprises peuvent y participer. Il y a un intérêt véritable à définir des standards communs, c'est le travail de W3C. W3C est cogéré par le MIT aux États-Unis, l'Ercim en Europe (à l'initiative de l'Inria en France), l'Université Keiō au Japon, et l'Université Beihang en Chine. Au début du Web, un débat a eu lieu sur la gratuité de l'accès au Web et un certain nombre d'acteurs commerciaux ont quitté la table quand cette gratuité s'est imposée comme la règle.

À noter que la gouvernance assez floue et distribuée d'Internet et du Web freine l'avancement sur des sujets comme la réduction de l'empreinte environnementale, surtout quand ils deviennent de plus en plus des lieux de tensions entre les entreprises et entre les États.

**Tim Berners-Lee** (né en 1955) est un informaticien britannique, principal inventeur du World Wide Web. Quand il était employé par le CERN dans les années 1990, il cosigne avec Robert Cailliau, un informaticien belge, le premier texte de présentation du Web : « WorldWideWeb : proposition pour un projet hypertexte ». En 1989, il met en œuvre la première communication réussie entre un client et un serveur du protocole de transfert hypertexte (HTTP) *via* Internet. Vers 1992, Tim Berners-Lee choisit de léguer son invention (spécifications et codes sources) au domaine public. L'accord du CERN est obtenu en 1993. Tim Berners-Lee fonde ensuite et préside le World Wide Web Consortium qui définit les standards du Web.

Il a reçu le prix Turing 2016 « pour avoir inventé le World Wide Web, le premier navigateur Web, et les protocoles et algorithmes fondamentaux permettant au Web de s'étendre ».

## *Les enclosures par les entreprises*

Pour les entreprises, la tentation est grande de conquérir des parts de marché sur Internet et le Web. Une manière de faire consiste à étendre les fonctionnalités d'un langage normalisé du Web, par exemple HTML, ou de proposer son propre idiome dans un domaine particulier. Si on arrive à imposer sa technologie, on en tirera des profits, on arrivera peut-être à exclure des concurrents. C'est ce qui s'est passé pour les tableurs (*spreadsheet* en anglais). Le logiciel Excel de la suite bureautique Microsoft Office s'est imposé, de par son succès, comme un standard de fait, y compris pour des échanges d'information.

Une guerre économique se joue aussi à travers les applications des smartphones (les *apps*). On peut voir dans ces apps la construction de nouveaux territoires. Elles entrent en compétition avec le Web : plutôt que de vous laisser naviguer sur le Web dans un environnement ouvert, les apps vous enferment dans un silo. De fait, les fonctionnalités qu'elles proposent sont en grande partie disponibles dans HTML5, une version récente du langage HTML. Si les apps permettent d'améliorer l'expérience utilisateur, notamment en simplifiant les parcours, elles servent beaucoup à enfermer l'utilisateur, à échapper à l'ouverture du Web. La multiplication des Apps avec leur stratégie de silo participe certainement à un affaiblissement du Web et d'Internet. L'utilisateur se rend sur Facebook ou sur YouTube et perd le sens d'être dans un monde de partage comme Internet et le Web.

Des tensions se cristallisent autour d'une règle fondamentale d'Internet : la neutralité du réseau. Cette règle interdit toute discrimination *positive* ou *négative* à l'égard de la *source*, de la *destination* ou du *contenu* de l'information transmise sur le réseau. Cette règle impose l'égalité entre tous les flux de données. En particulier, un contenu quel qu'il soit, film primé à Cannes ou photo de chaton, que son

auteur soit célèbre ou inconnu, public ou privé, sera acheminé de la même manière. Et le contenu sera acheminé à la même vitesse qu'il soit émis ou reçu par une entreprise puissante comme Microsoft ou Google, ou par une PME ardéchoise ou un individu lambda dans la Creuse. On peut voir cette règle comme l'application de principes fondateurs d'Internet : la liberté d'expression et de communication, la liberté d'accès au savoir et de partage. Cette neutralité est notamment menacée par une convergence d'intérêts entre des entreprises privées qui voudraient faire payer pour l'utilisation du réseau (des fournisseurs d'accès Internet comme Orange ou Free) et d'autres qui accepteraient de payer pour bénéficier d'un accès privilégié au réseau (de gros fournisseurs de contenus comme Netflix ou Spotify).

La neutralité protège les utilisateurs et les entreprises moins puissantes en évitant qu'Internet ne se transforme en un réseau à plusieurs vitesses, en une jungle économique où les plus faibles se feraient écraser par les plus forts et les utilisateurs perdraient toute liberté de choix. En cela, c'est un moteur de l'innovation comme on a pu le voir pendant l'histoire encore courte du réseau.

Il est intéressant d'observer comment les États s'emparent de la question. En Europe, c'est un règlement de 2015 qui protège la neutralité des réseaux. La France a été relativement motrice dans l'adoption de ce principe avec notamment la loi pour une république numérique en 2016. La neutralité du Net est d'ailleurs maintenant gravée dans un règlement européen. Pourtant, si la neutralité du réseau progresse dans le monde, elle est encore loin d'être acceptée universellement.

### *Les enclosures par les États*

Depuis que les gouvernements ont découvert les potentiels économiques et politiques d'Internet, ils supportent mal de n'avoir que peu de contrôle sur cet espace et cherchent à y prendre du pouvoir.



On a assisté à cela dans l'ICANN avec des conflits autour des noms de domaines et leurs utilisations. Pour la France, le domaine « .fr » est administré par l'Afnic (Association française pour le nommage Internet en coopération) qui a longtemps eu une politique assez restrictive pour ce qui est de l'attribution de noms de domaine.

Des pays totalitaires cherchent à toujours plus contrôler « leur » Internet et « leur » Web. Cela passe par l'obligation pour accéder à l'Internet mondial de transiter par des serveurs qui filtrent certains contenus et bloquent certaines applications. Vu de certains pays comme la Russie, Internet au-delà des frontières se réduit comme peau de chagrin, en particulier en période de crise, comme pendant la guerre en Ukraine. Tout est également fait pour dissuader l'utilisation de services étrangers comme Google comme moteur de recherche et Facebook comme réseau social, au profit de services nationaux.

Le Forum sur la gouvernance de l'Internet (FGI), un groupe de gouvernance multipartite sur les questions de gouvernance d'Internet, a été créé par le secrétaire général des Nations unies en juillet 2006. Il essaie avec difficulté d'empêcher les enclosures d'Internet par des États ou des groupes d'États. Il n'est pas arrivé à empêcher, de fait, une fragmentation d'Internet en différents sous-réseaux relativement isolés comme celui de la Chine ou de la Russie.

### *Des réseaux au service des humains*

Internet et le Web sont des ressources partagées par des communautés importantes. Mais on voit bien la complexité de la question de leur gouvernance. Si on laisse les grandes entreprises du Web faire ce qu'elles veulent, on bascule les réseaux dans le domaine privé au détriment de la liberté de choix des internautes et on les transforme en jungle commerciale. Si les gouvernements poussent le curseur trop loin, on place les réseaux sous la coupe des gouvernements.

Entre contrôle privé et contrôle public, peut-on conserver un espace pour ces communs ?

Des instances internationales comme l'ONU penchent vers plus de contrôle des deux réseaux, afin d'éviter qu'ils ne deviennent des zones de non droit. C'est dans cet esprit qu'a été mis en place le Forum sur la gouvernance de l'Internet (IGF), un forum mondial multipartite pour discuter de la gouvernance d'Internet. Comme il n'a pas de pouvoir décisionnaire, il est peu probable qu'il puisse résoudre les problèmes.

Plus concrètement, l'Europe avance avec une série de règlements et directives : sur la protection des données personnelles (Règlement général sur la protection des données, RGPD, adopté en 2016), la protection des internautes (*Digital Services Act*, DSA, adopté en 2022), la régulation de la concurrence (*Digital Markets Act*, DMA, en négociation début 2023), entre autres. Le Royaume-Uni adopte des lois dans un esprit assez semblable. On commence à observer aussi des mouvements dans ce sens aux États-Unis. Peut-être une conjonction des planètes pourrait-elle permettre la mise sur pied d'une régulation des réseaux qui ne soit pas leur mise au pas, mais au contraire les protège des tentatives d'enclosure par les États ou des entreprises ?

Internet et le Web sont en perpétuelle mutation. Des propositions pour des refontes plus radicales sont également proposées. Un problème existentiel du Web actuel est la centralisation et la domination par quelques plateformes. C'est à cela que s'attaque Tim Berners-Lee dans son projet Solid<sup>1</sup> basé sur un contrôle plus poussé des données personnelles par les utilisateurs et une meilleure protection de la vie privée. D'autres projets proposent de s'appuyer sur la technologie des *blockchains*<sup>2</sup> pour qu'on puisse garder plus de contrôle sur ses propres données.

---

1. <https://solid.mit.edu/>.

2. La *blockchain* est une technologie de stockage et de transmission d'informations, transparente, sécurisée et fonctionnant sans organe central de contrôle (CNIL).

Internet et le Web sont des réalisations fantastiques au service des humains. On aimerait que ce soit plus clairement des communs placés au service de tous.

### *Et demain les métavers...*

La prochaine étape des réseaux sociaux comme Facebook, YouTube, ou TikTok pourrait être les métavers. Ce sont des espaces virtuels partagés sur Internet où les utilisateurs peuvent se rendre pour se rencontrer, communiquer, en utilisant une ou plusieurs identités (avatars). Les métavers redéfinissent les communications entre individus. Les réseaux sociaux essaient de nous enfermer dans des silos. Avec les métavers, nous pourrions nous retrouver enfermés encore plus dans tous ces enclos. Si aucun métavers ne s'impose comme enclos mondial, ce qui est probable, il faudra imaginer des ponts pour passer de l'un à l'autre, de l'interopérabilité entre les métavers.

Quand ces réseaux sociaux sont utilisés par des centaines de millions, voire des milliards de personnes, la question se pose de faire des réseaux sociaux, en particulier dans leur version métavers, des communs. Autre question essentielle : pourquoi un utilisateur de Facebook ou de TikTok ne peut-il pas échanger de contenus avec un utilisateur d'un autre réseau comme on se parle entre Orange et Deutsche Telekom ? Des législateurs européens partagent ce point de vue et ont commencé à écrire dans le droit avec le DMA des débuts d'obligation d'interopérabilité. Dans les passages d'un métavers à l'autre, il faudrait qu'un utilisateur puisse conserver ses avatars, ses « propriétés ». Des initiatives comme le Metaverse Standards Forum, Open Metaverse Interoperability ou Letacircle œuvrent dans cette direction. Elles s'appuient souvent sur des technologies comme la *blockchain* pour éviter toute dépendance envers un acteur particulier.



PARTIE C

# Le fonctionnement



Comment se déploient les communs ? On s'aperçoit rapidement de l'absence de modèle unique. En plus de différences liées à la nature des ressources mises en commun, on observe une étonnante diversité suivant, notamment, les spécificités des communautés, les formes de gouvernance, et les modèles économiques choisis. Le déploiement de communs s'accompagne souvent de transformations des métiers, parfois de la création de véritables écosystèmes.





## CHAPITRE 9

# La communauté

Un commun numérique est défini d'abord par une communauté, qui produit et partage des ressources numériques, animée par un intérêt partagé pour ces ressources. La communauté est au cœur du commun. C'est son fonctionnement que nous étudions dans ce chapitre.

### *L'esprit de la communauté*

Il est intéressant de comparer l'esprit qui guide le fonctionnement de communs et de le comparer à celui qui guide les entreprises privées.

Dans les deux cas, on peut mettre en avant un principe : « Le client est roi. » C'est vrai dans l'entreprise privée car de la satisfaction du client dépend la réussite de l'entreprise. C'est vrai aussi pour le commun, le but affiché étant d'apporter des ressources de qualité à un maximum d'utilisateurs (même si on ne les appelle pas des « clients »). Pourtant, ce principe est à relativiser dans les deux cas :

- Pour les communs, il faut distinguer la communauté des utilisateurs (« clients ») de la communauté des contributeurs, même si ces deux communautés ne sont pas nécessairement disjointes.

Des contributeurs déçus se désintéressent du projet, ce qui fait baisser la production. Une pierre angulaire est donc la satisfaction de la communauté des contributeurs pour qu'ils continuent à contribuer.

- Pour l'entreprise privée, la satisfaction des employés est importante mais pas essentielle : on peut licencier des employés, en embaucher d'autres. Par contre, on ne peut pas se passer des profits. La responsabilité première du chef d'entreprise est de faire croître la valorisation de l'entreprise, donc il faut faire croître les profits. La pierre angulaire est de ce fait la recherche de productivité pour dégager des profits.

En conséquence, on assiste dans le privé à une uniformisation des cultures. Les employés passent facilement d'une entreprise à une autre, et on adopte rapidement des procédés qui améliorent la productivité. Dans les communs, les communautés des contributeurs sont typiquement aux commandes et imposent leurs choix, leurs méthodes. Ces communautés tiennent souvent beaucoup à leurs particularités, leurs façons de travailler, et sont même prêtes à sacrifier l'efficacité pour travailler comme elles aiment le faire.

### *Quelle communauté ?*

Elinor Ostrom met la communauté au centre du commun. Les communautés varient en taille, en démographie, en organisation, en motivations, etc. La communauté est souvent au centre d'un cercle vertueux : quand elle grandit, cela conduit à plus de contributions, ce qui mène à plus de possibilités pour les membres, d'où une plus grande attractivité qui la fait grandir.

On peut souvent distinguer plusieurs communautés, typiquement, de contributeurs et d'utilisateurs. Bien entendu, le nombre d'utilisateurs est dans la plupart des cas beaucoup plus grand que celui des contributeurs : toute personne qui s'est connectée sur Internet est

d'une certaine façon un utilisateur de Linux, mais chaque version de Linux n'est produite « que » par un millier de développeurs. Wikipédia est consulté chaque année par plusieurs milliards de personnes, mais moins de 130 000 personnes y ont contribué dans les trente derniers jours (pour la version anglaise, chiffres début 2022).

Regardons la démographie de quelques communautés de contributeurs sur lesquelles nous disposons d'informations relativement précises.

**OPENSTREETMAP** a 1,5 million de contributeurs dans le monde. Si on se focalise sur la France, les contributeurs sont répartis de façon plus ou moins uniforme, mais sont à majorité des urbains. L'âge varie de 15 à 71 ans, la médiane étant à 35 et la moyenne à 38 ans. Moins de 10 % sont des femmes, et 60 % sont à bac+5. Ce biais de genre de la communauté OpenStreetMap est malheureusement partagé par de nombreux communs numériques. L'une des raisons en est peut-être le mot « numérique » porteur aujourd'hui d'un fort biais de genre : depuis l'école, l'informatique (à la fois la science et la technique, qui sous-tend le numérique) attire beaucoup plus les garçons que les filles, qui pensent (bien sûr à tort) ne pas être douées pour cette discipline.

**WIKIPÉDIA.** La dernière étude démographique sur la communauté des wikipédiens date de 2010. Elle est mondiale et porte sur plus de 176 000 personnes dont 54 000 sont des contributeurs. La moyenne d'âge est 25 ans et la médiane 22 ans (les contributeurs étant légèrement plus âgés). Parmi les contributeurs, 87 % sont des hommes et 13 % des femmes (chiffres arrondis). Enfin, 55 % des contributeurs avaient un diplôme universitaire ou équivalent.

**LE LOGICIEL LIBRE** a une communauté importante par sa taille : les 14 millions de développeurs actifs sur GitHub nous en donnent une estimation. Un sondage sur 5 500 contributeurs de GitHub en 2017 montre que 95 % déclarent être des hommes et 3 % des femmes, un ratio encore plus mauvais que celui de l'industrie du logiciel en général. Une raison peut en être la relative dureté et violence de la culture du logiciel libre. Les critiques du code ayant contribué aux

projets ont tendance à tourner rapidement aux attaques personnelles. Dans ce même sondage, 50 % des personnes interrogées disent avoir été témoins d'interactions toxiques et 18 % disent avoir souffert de ces interactions négatives. Les réponses dédaigneuses, les conflits et les termes désagréables sont cités comme les troisième, quatrième et sixième problèmes les plus importants dans le monde du logiciel libre.

À côté de ces traits généraux, on peut considérer des traits plus spécifiques développés dans des projets distincts avec leurs propres règles de fonctionnement. Regardons-en deux :

- La *Fondation Linux* impressionne par ses chiffres : plus de 700 000 contributeurs, plus de 18 000 entreprises contributrices. Il existe beaucoup de fondations *open source* ou logiciel libre, et la Fondation Linux est clairement la plus importante. Créée à l'origine pour gérer le projet Linux, elle a progressivement étendu ses responsabilités à plus de 400 projets et gère plus de 3 400 répertoires de code. Parmi les projets les plus visibles après Linux : Kubernetes, Node.js, Let's Encrypt, et Call for Code. Les contributeurs sont souvent des ingénieurs.
- *Scikit-Learn* est une communauté qui vient du monde de la recherche et qui a son origine à l'Inria. Elle inclut des personnels de l'Inria, des universitaires, des chercheurs et des développeurs, qui sont financés par les sponsors du projet et des externes, mais les contributeurs se comptent en dizaines, plutôt qu'en milliers. Les contributeurs sont souvent des chercheurs.

**Linux** est une famille de systèmes d'exploitation *open source*. Les systèmes d'exploitation sont indispensables pour gérer les ressources matérielles et logicielles d'un ordinateur. Linux s'inspire du système d'exploitation Unix conçu en 1969 aux Bell Labs d'AT&T, aux États-Unis, par Ken Thompson, Dennis Ritchie, Douglas McIlroy et Joe Ossanna. Le noyau du système Linux a été proposé en 1991 par Linus Torvalds.

Pour qu'un ordinateur soit effectivement utilisable, il faut ajouter au système d'exploitation un grand nombre d'applications, par exemple des outils de gestion de courrier, des logiciels bureautiques, etc. Une « distribution de Linux » consiste en Linux et des centaines de logiciels libres, formant un ensemble cohérent de logiciels.

À partir de tous les logiciels *open source* disponibles, on peut librement construire sa propre distribution, même s'il faut, pour y arriver, disposer de solides compétences techniques. De nombreuses distributions Linux ont vu le jour. Debian est développée et maintenue par une communauté de volontaires. À l'inverse, Ubuntu est construite par une entreprise privée, Canonical. La distribution Mandrakesoft/Mandriva a disparu, mais son code subsiste notamment avec un de ses *forks*, Mageia. Mentionnons enfin une autre distribution commerciale, Red Hat.

**Red Hat** est une société américaine de logiciels fondée en 1993 spécialisée dans les logiciels *open source* aux entreprises. Elle a été rendue célèbre par son système d'exploitation Red Hat Enterprise Linux. Red Hat fournit entre autres des produits de stockage, du *middleware*, des applications, des produits de gestion. C'est une des entreprises qui contribuent massivement au noyau Linux. Elle s'est développée plus récemment dans le *cloud* avec des produits comme l'OpenShift Container Platform. Red Hat édite en outre une distribution *open source* de Linux, Fedora.

En 2019, Red Hat a été rachetée par IBM pour 34 milliards de dollars. C'est une filiale indépendante qui a conservé tous ses liens avec son écosystème.

## *Pourquoi contribuer ?*

Les motivations pour contribuer à des communs varient énormément suivant les cas, depuis des considérations purement altruistes jusqu'à des intérêts strictement pécuniaires. Un programmeur bénévole de Framasoft développant un logiciel pour l'éducation populaire n'a pas les mêmes motivations qu'un ingénieur de Microsoft payé pour développer un logiciel éducatif diffusé en *open source*. Même dans des cas comme Wikipédia, les motivations peuvent être variées : du prof de maths qui peaufine une entrée sur la règle de trois, à la responsable de la com qui améliore la page de son P-DG. Pour ce qui est des agents de l'État, ils peuvent avoir comme mission de contribuer à des communs. Le plus souvent, des personnes contribuent à des communs par choix. Même, par exemple, pour des agents de l'État à qui on a donné la tâche de publier des données ouvertes, on s'aperçoit qu'ils ont parfois choisi le service dans lequel ils travaillent parce qu'il leur permet de contribuer à un commun.

Le point de départ de la motivation de la communauté est souvent un intérêt commun pour la ressource. Mais, quand on parle d'un intérêt commun pour la ressource, parle-t-on d'un intérêt collectif de la communauté ou d'un intérêt individuel partagé par tous les individus qui la composent ? Dans le livre *Gödel, Escher, Bach : An Eternal Golden Braid*, Richard Hofstadter prend l'exemple de la fourmi, de la fourmilière et du fourmilier<sup>1</sup>. Le fourmilier est un animal qui se nourrit de fourmis, et, ce faisant, contribue à maintenir la fourmilière en bon état. Et Hofstadter de conclure : la fourmi n'aime pas le fourmilier (on la comprend), mais la fourmilière aime le fourmilier. Sans aller jusqu'à des situations aussi dramatiques,

---

1. Richard Hofstadter, *Gödel, Escher, Bach : An Eternal Golden Braid*, Basic Books, 1979 ; traduit en français par Jacqueline Henry et Robert French, *Gödel, Escher, Bach. Les brins d'une guirlande éternelle*, Dunod, 1985.

un décalage similaire peut se produire entre les intérêts de la communauté d'un commun numérique, et ceux des individus qui la composent et qui ont leur motivation personnelle pour participer au commun.

Pour ces communs qui reposent sur le travail d'un grand nombre de bénévoles, par exemple le recueil de données par la multitude, le sujet de la motivation est essentiel. Pourquoi passer du temps à relever des données pour une carte, à récolter des données pour une base de données des aliments, ou à améliorer un article d'une encyclopédie ?

Les motivations des contributeurs peuvent aussi varier dans le temps : à la création du commun, il faut avoir une vocation de pionnier puis, quand le commun est bien établi, il faut savoir pérenniser la ressource, organiser la communauté, ce sont deux exercices d'une nature différente.

Les motivations des membres de la communauté ont souvent plusieurs facettes. Certaines tiennent de l'altruisme, la réalisation d'un objectif noble : les bénévoles considèrent d'abord qu'ils accomplissent quelque chose d'utile pour la société. D'autres motivations sont plus liées au « bénéfice » personnel du participant, par exemple, à l'intérêt intellectuel ou social de sa participation. Regardons au travers de quelques exemples ces motivations collectives et individuelles.

Pour **OPENSTREETMAP**, la mission est de créer le meilleur ensemble de données cartographiques du monde grâce à sa communauté et de le mettre gratuitement à la disposition de tous sous une licence libre. Mais considérons plus en détail les motivations individuelles des contributeurs, elles vont de la passion des cartes, au plaisir de la course à la précision, à l'exhaustivité et à l'actualisation. Vu de l'extérieur, cette obsession pour certains de cartographier pourrait même être assimilée à un TOC (trouble obsessionnel compulsif). Mais elles sont aussi liées à des thématiques précises qui intéressent le contributeur : par exemple, les premiers contributeurs à l'origine d'OpenStreetMap en Grande-Bretagne étaient

des cyclistes qui cherchaient à construire une carte adaptée à leurs besoins, carte qui manquait alors. D'autres contributeurs peuvent être des passionnés du patrimoine architectural ou du street art, et vouloir répertorier les plus belles contributions à ces domaines. De nombreux développeurs préfèrent OpenStreetMap à Google Maps parce qu'ils peuvent plus facilement combiner dans leur application les points d'intérêt particuliers à leur métier.

**LES DONNÉES OUVERTES PUBLIQUES** (*public open data*) tiennent d'un mouvement descendant. Il part de l'État et des acteurs publics, plutôt que de la base, au contraire des données ouvertes *crowd-sourcées* qui proviennent de la base. Clairement, sans le soutien et l'action de l'État, il n'y a pas de données publiques ouvertes. L'intérêt général des données ouvertes publiques est connu : optimisation du fonctionnement de l'État, amélioration de la qualité des données utilisées par la puissance publique, réduction de la distance entre l'État et ses administrés et enfin nouvelles opportunités entrepreneuriales à partir des données libérées. Nous reviendrons plus en détail sur ce sujet.

Pour les serviteurs de l'État impliqués dans les données ouvertes, la participation au commun tient à leur travail professionnel, même si on retrouve chez certains des formes de militantisme. Mais à cette activité de l'État s'associe une base militante de données ouvertes : des associations comme Libertic, la Fing, ou l'OKFN, des regroupements de territoires ou d'organismes public comme Open Data France, et bien sûr des entreprises, comme OpenDataSoft en France, ou Socrata aux États-Unis (acquise par Tyler Technologies). De claires motivations militantes contribuent à l'ouverture et l'utilisation des données ouvertes.

La mission de **WIKIPÉDIA** est de « créer et de distribuer une encyclopédie gratuite de la meilleure qualité possible à chaque individu du monde dans sa propre langue ». Mais qu'est-ce qui motive les plus de 18 000 wikipédiens ayant fait au moins une modification sur Wikipédia en français en mars 2022 ? De nombreuses études ont été



faites sur les motivations de ces contributeurs. On a découvert que les plus fortes relevaient du « concept de soi » comme, par exemple : « J'aime partager la connaissance, ce qui me donne le sens d'un accomplissement personnel. » Une autre étude, qui voulait classer les motivations à partir de concepts définis, a trouvé que les plus forts étaient « amusement », « idéologie » et « valeurs » et les plus faibles « carrière », « social » et « protection ». Enfin, beaucoup mentionnent le caractère addictif de l'exercice. On comprend qu'en plus de participer à une belle œuvre, les wikipédiens trouvent concrètement du plaisir à contribuer à l'encyclopédie.

Considérons maintenant les logiciels. Il faut évidemment distinguer entre les tenants des logiciels *open source* et ceux des libres. Pour l'*open source*, les énoncés de missions (*missions statements*) que déclarent la Fondation Linux, la Fondation Mozilla ou la fondation Apache ressemblent, dans l'esprit, à ceux des sociétés privées conçus par des spécialistes du marketing : « démocratiser le code », « créer un meilleur Internet » ou « fournir du logiciel pour le bien du public », etc. Pour ce qui est de l'intérêt commun qui soude les membres de la Fondation Linux, faut-il l'imaginer comme la combinaison des intérêts des 15 membres « platine » et des 18 membres « gold », qui sont parmi les sociétés privées les plus puissantes de la planète ? Pour ces sociétés, l'intérêt commun principal est de développer conjointement, de façon efficace (donc avec une bonne productivité) des technologies informatiques de qualité (performantes et fiables) que chacune d'entre elles pourra utiliser pour son intérêt commercial propre. Donc le système capitaliste a su identifier dans l'idée originale des innovateurs du logiciel libre un outil qui lui permettait d'être plus efficace et plus productif. On est loin du rêve de liberté de Richard Stallman. Mais, après tout, dans la notion d'intérêt commun de la communauté, rien ne précise que cet intérêt doit être altruiste ou louable.

Qu'en est-il des développeurs individuels qui participent aux projets de cette fondation ? Dans un sondage conduit en 2020

par la fondation Linux et un laboratoire de Harvard, auquel ont répondu environ 1 200 développeurs *open source* (600 ayant répondu à l'ensemble du questionnaire), avec une distribution géographique mondiale (27 % des États Unis, 12 % d'Allemagne et 7 % de France), la partie sondage sur la motivation individuelle de ces contributeurs nous apprend que les trois premières motivations des contributeurs ne sont pas financières : ajouter une nouvelle fonctionnalité ou réparer un bug, apprendre de nouvelles choses et faire un travail créatif et agréable ont été classées dans le top 3. Être payé pour travailler est classé toujours dans les 3 plus faibles motivations. En fait, 75 % des contributeurs ont un travail rémunéré et 52 % sont précisément payés pour contribuer.

Du côté du **LOGICIEL LIBRE**, nous trouvons des développeurs individuels, qui prennent du plaisir à développer du code utile à la société et de qualité, et qui considèrent que donner à chacun la liberté d'utiliser leur logiciel est une contribution au bien de l'humanité, indépendamment du retour financier qu'ils peuvent en retirer, ou pas. Un bon exemple d'une telle communauté est Mageia. Mageia développe et diffuse une distribution Linux, née d'un *fork* de la distribution Mandriva en 2010 (voir l'encadré pour une explication du concept de fork). Les participants de Mageia sont tous des bénévoles, le projet reçoit quelques donations et produit chaque année (ou presque) une version nouvelle de la distribution. Ici, pas d'intérêt commercial, mais simplement l'objectif de « fabriquer de superbes outils pour tout le monde » et de « devenir et faire vivre une communauté crédible et reconnue dans le monde du logiciel libre ». On est bien dans l'esprit du logiciel libre.

**Fork.** Les conflits ne sont pas rares dans les communautés de communs numériques. Ils se résolvent le plus souvent par la discussion, le débat, voire des votes. Quand les deux visions sont irréconciliables, cela peut conduire à un *fork*.

Le *fork* est l'analogie pour les logiciels à ce qui s'est passé pour la SFIO au congrès de Tours en 1920. Suite à un désaccord de fond, le parti s'est scindé : une partie de ses membres l'a quitté pour fonder le Parti communiste, Léon Blum restant « garder la vieille maison » avec les autres. Des participants à un logiciel libre, qui ne sont pas satisfaits de la direction prise par la communauté, ou de son mode de fonctionnement, peuvent proposer un *fork* (en français « fourche » ou « embranchement »). On assiste alors à un schisme de la communauté. Chaque branche prend une copie du logiciel ; les deux copies vont ensuite évoluer indépendamment et s'éloigner progressivement l'une de l'autre. Les deux communautés sont alors plus faibles numériquement, tant pour les contributeurs que pour les utilisateurs.

Nous allons rencontrer plusieurs exemples de *forks* dans la suite de notre essai.

Dans le numérique, les employeurs ont du mal à recruter. Les travailleurs passent souvent d'un emploi à un autre, parce qu'il est facile de trouver un emploi. Ils le font souvent pour donner un sens à leur travail, pour que leur emploi corresponde mieux à leurs valeurs. C'est pourquoi ils sont souvent attirés par les communs. Un autre argument est celui de valoriser son travail sur le long terme. Le développeur d'un logiciel propriétaire pourra plus difficilement monnayer les compétences qu'il a acquises dans son métier qu'un développeur de logiciel libre qui pourra montrer ce qu'il a réalisé, voire continuer à travailler sur son propre code dans un autre emploi.

Quand la tension pour recruter est forte, cet attrait des employés pour les communs peut encourager les employeurs à se diriger vers des solutions fondées sur des communs.

## Les licences

Les produits des communs numériques, que ce soient les articles de Wikipédia, les données d'OpenStreetMap ou d'Open Food Facts, les ressources éducatives libres ou les logiciels Linux et Mozilla, sont régis par des licences qui protègent leurs auteurs en spécifiant les droits et devoirs de ceux qui les utilisent. Faisons un rapide tour d'horizon de ces licences.

Pour comprendre leur fonctionnement, il faut revenir sur le droit d'auteur.

Le droit d'auteur (dans le droit français) instaure le monopole d'exploitation de l'auteur d'une œuvre sur celle-ci. Il inclut le droit moral et le droit patrimonial. Le droit moral impose le respect du nom, de l'œuvre et de la qualité du créateur. Le droit patrimonial concerne l'exploitation commerciale de l'œuvre et est limité dans le temps : soixante-dix ans après le décès du titulaire du droit d'auteur, la création tombe dans le domaine public. Les livres, les photographies, les sculptures, les partitions de musique, les sons MIDI<sup>1</sup>, ou encore les logiciels et les sites Internet entrent dans le champ d'application du droit d'auteur.

---

1. Le MIDI, pour *musical instrument digital interface*, est un format de fichier (correspondant à un protocole) pour la musique.

Le copyright est une forme de propriété intellectuelle qui donne à son propriétaire le droit exclusif de distribuer une œuvre, en général pour un temps limité. L'œuvre peut être littéraire, artistique, éducative ou musicale. Le copyright est là pour protéger l'expression originale de l'œuvre, mais pas l'idée elle-même. Le copyright protège uniquement les œuvres fixées sur un support matériel (fichier informatique, dessins...), et ce que l'œuvre soit publiée ou non. De son côté, le droit d'auteur protège toutes les créations de l'esprit fixées sur un support matériel ou non.

Le domaine public réunit toutes les créations auxquelles ne s'applique aucune propriété intellectuelle. Ça peut être le cas parce que cette propriété a expiré, mais également parce qu'elle a été expressément abandonnée par leurs auteurs. En France, on l'a dit, un livre tombe dans le domaine public soixante-dix ans après la mort de son auteur. Donc les œuvres de Jules César ou d'Honoré de Balzac sont dans le domaine public. Un album de Gilberto Gil diffusé en 2004 avec l'aide du magazine américain *Wired* a été déplacé hors du champ de la propriété strictement privée avec une licence Creative Commons. On peut diffuser les titres de cet album à condition de ne pas en faire pas d'utilisation commerciale. Ces limites le situent hors du domaine public.

**LE LOGICIEL.** Les logiciels libres et *open source* sont réglementés par des licences, qui précisent les droits et devoirs des utilisateurs. Comme le domaine est divisé entre les tenants du logiciel libre et ceux du logiciel *open source*, les licences se sont alignées suivant cette fracture, en s'organisant autour de deux organismes : la Free Software Foundation (FSF) et l'Open Source Initiative (OSI).

Pour le logiciel libre, la FSF, une association à but non lucratif, fondée en 1985 par Richard Stallman, définit les licences libres. Face au copyright, les militants du libre ont inventé avec humour le mot *copyleft*, son pendant ouvert qui insiste sur les droits plutôt que sur les devoirs. Le copyleft consiste à donner le droit de distribuer ou de modifier un objet numérique avec la contrainte que

le même droit doit être appliqué aux produits dérivés créés à partir de cet objet : le copyleft est viral ! La GPL (GNU Public Licence) a été définie par Richard Stallman en 1989 pour s'appliquer aux logiciels développés dans le cadre de projet GNU. C'est bien Richard Stallman avec cette licence qui a initié le processus d'utilisation libre des ressources logicielles et, également, la liberté d'apporter des améliorations au logiciel.

À l'origine, une licence avait été développée pour chaque logiciel du projet GNU : une pour Emacs, une pour le Debugger et une pour le compilateur C. Stallman a unifié les trois licences et produit une licence générique qui pouvait s'appliquer à tout logiciel, ce qui facilitait le partage de code par plusieurs logiciels. Une V2 de la GPL a paru en 1991 et une V3 en 2007, avec l'aide d'un juriste, Eben Moglen du Software Freedom Law Center. Une version affaiblie de la GPL, la LGPL (*lesser GPL*) permet de pas appliquer le principe de la viralité aux autres composants logiciels liés à l'œuvre en GPL. Avec la LGPL, on peut construire un système informatique qui contient des composants *open source* et des composants propriétaires sans que la GPL ne s'applique aux composants propriétaires, ni au système résultant.

En France, la licence CeCILL, créée par le CEA, le CNRS et l'Inria, est une adaptation de la GPL au droit français.

Pour le logiciel *open source*, l'OSI (Open Source Initiative), une autre association à but non lucratif, fondée en 1998 par Eric Raymond, Bruce Perens et Michael Tiemann, appuie sa définition de l'*open source* utilisée dans ses licences sur les caractéristiques suivantes :

- redistribution libre et gratuite ;
- code source disponible ;
- produits dérivés autorisés ;
- aucune discrimination à l'égard de personnes ou de groupe ;
- aucune discrimination du contexte d'usage (e.g. pas de discrimination pour la commercialisation).

Donc l'*open source* n'impose pas la viralité, ce qui la rend plus attractive pour les industriels. Ces licences sont qualifiées de « permissives », parce qu'elles imposent moins de contraintes.

Il existe de nombreuses licences *open source* largement utilisées. Parmi les plus importantes on trouve :

- La licence Apache, produite par l'Apache Software Foundation et qui s'applique à tous les logiciels produits par la Fondation qui date de 2004.
- La licence MIT, une dénomination un peu abusive parce que le MIT utilise de nombreuses licences et qui recouvre notamment la licence X11 (du logiciel de gestion de fenêtrage X) et la licence Expat (utilisée notamment par XML) ; ces licences datent de la fin des années 1980.
- La licence BSD, qui désigne en fait une famille de licences, venant à l'origine de Berkeley Software Distribution, un système d'exploitation de type Unix. La version d'origine (1988) contenait une clause de publicité, qui a été critiquée et supprimée dans les versions suivantes.

Quelle est la « part de marché » de chacune d'entre elles ? Il est difficile d'avoir des chiffres précis : ils dépendent trop de la façon dont on mesure (nombre de logiciels ou usage des logiciels) et de la source de la mesure (GitHub ou une mesure précise des licences de chaque composant de chaque logiciel). Statista classe les premières de la façon suivante : Apache, MIT, GPL, BSD.

Si l'on regarde la dynamique, il semble que l'adoption massive et récente de l'*open source* par l'industrie du logiciel s'est traduite par une perte de part de marché de la GPL et de ses versions, qui étaient majoritaires à l'origine, au profit des licences plus permissives, mieux adaptées au goût des industriels.

**CREATIVE COMMONS.** La licence Creative Commons qui s'applique à toutes les œuvres a été conçue en 2001 par Lawrence Lessig et Eric Eldred. Elle a été créée dans le contexte états-unien, et des adaptations à d'autres systèmes légaux ont été développées,

par exemple en Chine. La licence Creative Commons se définit par quatre paramètres :

- *attribution* (BY) : si on rediffuse l'œuvre, on doit créditer l'auteur de sa contribution (ce paramètre est obligatoire) ;
- *share-alike* (SA) : si on rediffuse l'œuvre, on doit le faire sous une licence au moins aussi libre que celle de l'œuvre originale (c'est la viralité) ;
- *non commercial* (NC) : pour utiliser commercialement l'œuvre, on doit obtenir l'accord de l'auteur ;
- *no derivative* (ND) : on ne peut pas utiliser un morceau de l'œuvre dans une autre œuvre (par exemple, l'échantillonnage, le *sampling*, est interdit).

On définit une version de la licence Creative Commons en choisissant un ou plusieurs des paramètres. Par exemple, la licence CC BY-SA spécifie qu'on doit créditer l'auteur ou le licencieur et qu'on doit redistribuer les œuvres dérivées sous une licence au moins égale à CC BY-SA.

Enfin, la licence CC0 place l'œuvre carrément dans le domaine public. L'œuvre peut être copiée, modifiée, distribuée et représentée, même à des fins commerciales, sans avoir besoin de demander d'autorisation.

Pour le cas des données ouvertes, une licence spécifique a été définie, l'Open Database Licence (ODbL), qui est une licence CC BY-SA. Cette licence est utilisée par OpenStreetMap et Open Food Facts. En France, Etalab a défini sa propre licence, la Licence Ouverte qui est aussi une CC BY.

**Lawrence Lessig** (né en 1961) est un juriste américain passionné par le numérique. Il a été professeur de droit à la Harvard Law School, où il a fondé le Center for Internet and Society. Il a défendu la liberté sur Internet, s'opposant à des interprétations strictes du copyright. Il est également fondateur et président du conseil d'administration de l'organisation Creative Commons.



## Gouvernance des communs

Chaque commun numérique met en place sa propre organisation et sa gouvernance spécifique. Cette gouvernance détermine le fonctionnement du commun, depuis le rôle des individus dans la communauté, jusqu'à l'optimisation de l'utilisation de la ressource. Cette gouvernance dépend bien entendu de la nature de la ressource gérée par le commun, mais aussi de la culture de la communauté.

La dépendance à la nature de la ressource n'est pas surprenante. On ne gère pas de la même façon la production d'un logiciel, d'un jeu de données, d'une encyclopédie, d'un support pédagogique, etc. La nature de la ressource soulève des problèmes spécifiques, et conduit donc à des processus différents.

Mais la gouvernance dépend aussi de la culture de la communauté, parce que c'est cette dernière qui la choisit. Nous avons vu les différences de culture entre les communautés de logiciel libre et celles de l'*open source*, même si l'objet produit par le commun est à chaque fois un logiciel ouvert. De même, il n'y a pas « une » communauté Wikipédia, mais « des » communautés : la gouvernance de la version francophone est différente de celle de la version anglophone par exemple. Pour la communauté, la manière de produire la ressource commune est aussi importante que la ressource elle-même.

L'accent mis sur les choix de la communauté est compréhensible. Si la communauté est mécontente, frustrée, ses membres vont mettre moins de cœur à la tâche, quitter la communauté, voire former une communauté concurrente. La primauté de la communauté sur le produit est même un principe affiché dans certaines communautés, comme la Fondation Apache.

La gouvernance est plus ou moins formalisée et évolue dans le temps. Ainsi, à la formation du commun, elle est relativement légère et informelle, et avec le temps, elle se précise et se complexifie souvent. En revanche, les communs adoptent en général et conservent un principe de transparence : les règles sont explicitées et publiques.

Pour illustrer ces différents aspects, nous allons plus précisément considérer des communautés de *crowdsourcing* et des communautés de logiciels.

## *Le crowdsourcing*

**LA COMMUNAUTÉ D'OPENSTREETMAP** est organisée horizontalement. Il n'y a que deux niveaux de contributeurs : la grande masse des contributeurs de base et un petit un groupe de 10 personnes au niveau mondial qui arbitre les conflits. L'association a un conseil d'administration qui se réunit régulièrement avec la communauté pour établir les priorités et les projets. La communauté pratique l'ouverture radicale, c'est-à-dire que les discussions et les décisions sont prises en public. La communauté utilise Loomio, un logiciel *open source* qui permet à une communauté de se gérer de façon participative et transparente.

Les règles de base affirmées par OpenStreetMap sont (*verbatim*) :

- Une communauté ouverte, à la gouvernance horizontale, où les règles sont claires et s'appliquent à tous. La seule barrière à l'entrée est celle du respect de ces règles.

- *Community first* : même en ayant un objectif final de production de données, l'approche adoptée doit se focaliser sur *la communauté avant tout*. Animer ses membres et produire des outils pour faciliter son implication doivent être des sujets centraux.
- La création de véritables services utiles basés sur le commun. C'est peut-être le point le plus important et le véritable indicateur de succès. Par exemple, le succès de la Base Adresse Nationale est ainsi mesurable en mesurant le nombre d'appels à son API de géocodage.

**LA COMMUNAUTÉ WIKIPÉDIA** insiste également sur la transparence. Ceux qui s'intéressent à cette encyclopédie et veulent l'étudier ou comprendre son fonctionnement disposent d'un corpus de données passionnant : l'historique de tous les articles et les débats rattachés, ainsi que l'historique des règles et les débats qui ont mené à leur choix. Pour le français, toutes ces données sont consultables sur le site de l'encyclopédie, <http://fr.wikipedia.org/>. C'est un actif fabuleux pour les chercheurs qui étudient ce processus.

Dans Wikipédia, la communauté s'organise autour de règles décidées collectivement : tout volontaire peut contribuer à l'encyclopédie ou aux règles qui régissent son édition. Un contributeur se retrouve facilement en désaccord avec d'autres wikipédistes et cela peut amener à des dialogues complexes, voire conduire à l'intervention d'« autorités » de la communauté pour résoudre le conflit. Les administrateurs, qui ont surtout au départ plutôt un rôle technique, coordonnent les éditeurs et peuvent être appelés à intervenir, aller jusqu'à verrouiller un article devenu une zone de conflit.

Chacun peut contribuer à la production, critiquer, remettre en question les modifications des autres, le tout sous le contrôle d'un outil d'édition collaborative, le wiki. Chaque participant, inscrit ou anonyme, peut participer à l'édition sans même d'étape de validation. Les modifications sont immédiatement en ligne, dans des articles accompagnés par des listes de suivi et des pages de discussions. Les

articles ont eu en moyenne plus de cinquante éditions successives, par près d'une vingtaine de contributeurs.

Un certain nombre de règles sont énoncées, par exemple, dans la version française de Wikipédia, on trouve :

- ne rien inventer (ce n'est pas le lieu pour un travail inédit) ;
- indiquer ses sources ;
- rester neutre<sup>1</sup> ;
- être précis.

Pour ne pas décourager les débutants, des « antirègles » (par exemple « ignorer toutes les règles ») sont aussi énoncées.

Il y a séparation des pouvoirs entre une autosurveillance participative et un mécanisme de punition et de justice. La première fonctionne par démocratie directe et la deuxième par démocratie représentative. En parallèle, un système de résolution des conflits permet de gérer les guerres d'éditions : quand plusieurs personnes ou groupes se disputent sans fin sur un sujet. L'encyclopédie dispose aussi d'un système judiciaire avec des magistrats élus qui tranchent si nécessaire. Mais c'est une situation rare : quelques dizaines de cas par an pour la version francophone par exemple. Finalement, un mécanisme de sanctions existe pour les perturbateurs éventuels.

Pour la gouvernance, le mode de décision est celui du « consensus approximatif » (inspiré de l'IETF<sup>2</sup>), parfois implémenté par sondage. L'encyclopédie en langue française a développé une gouvernance autonome dès 2002, même si les règles de la langue anglaise contiennent d'avoir une influence forte.

**OPEN FOOD FACTS** est un exemple de communauté relativement jeune, dont la gouvernance est en cours de formalisation. Elle utilise un *slack*<sup>3</sup> avec 300 canaux. Pour son fonctionnement, la commu-

---

1. Des points de vue opposés sont présentés, en essayant de rester proportionnels à leurs audiences scientifiques respectives.

2. L'Internet Engineering Task Force est un organisme de standardisation d'Internet.

3. Slack est un outil de collaboration sur le *cloud* qui permet, entre autres, des dialogues entre ses utilisateurs (organisés par sujet, les canaux) et le partage de documents.

nauté s'inspire de OpenStreetMap et de Wikipédia, mais est encore moins organisée que ses modèles.

## *Les logiciels*

Avant de mieux expliquer la différence entre logiciel libre et *open source*, il est important de comprendre les rôles essentiels des participants à de grands projets dans les deux cas :

- Un *contributeur* est une personne qui apporte sa pierre à l'édifice : cela peut être en signalant un bug, en signalant une erreur dans la documentation, en traduisant la documentation ou une partie du logiciel, etc.
- Un *mainteneur* est une personne qui produit du code et propose des modifications ou des ajouts au logiciel.
- Un *committer* (de *to commit*, « valider ») est une personne qui a le pouvoir de valider des ajouts ou des modifications, de les répercuter dans une version utilisable par tous du code.

Il y a un grand nombre de communautés de logiciel libre et chacune a défini sa propre gouvernance. Leur organisation est étroitement liée aux objets que ces communautés construisent et maintiennent. Le développement d'un grand logiciel tel que Linux, Apache ou PostgreSQL (des millions de lignes de code) est un processus difficile et complexe. Le logiciel, qui est au cœur du fonctionnement de *systèmes critiques*, doit pouvoir fonctionner sans défaut et toute mise à jour doit être sûre et préserver le fonctionnement du système. Un système critique est un système où les pannes ou les dysfonctionnements posent de graves problèmes et dont le fonctionnement doit donc être fiable et sûr. Internet en est un exemple. Sans minimiser l'effort que représente le développement d'une encyclopédie, d'un jeu de données *crowdsourcé* ou d'un support de cours, la complexité d'un grand logiciel exige des talents particuliers d'organisation et une discipline rigoureuse.

Pour résoudre les conflits potentiels, certaines communautés ont introduit le rôle de « dictateur bienveillant à vie » (*benevolent dictator for life*), rôle que d'autres communautés honnissent. C'est le titre donné (avec un certain humour) à une personne unique qui a le dernier mot en cas de désaccord sur une décision à prendre pour l'évolution du logiciel. L'idée est que le diktat est nécessaire pour trancher, mais qu'il doit être « bienveillant » pour être accepté par la communauté. Il est normalement à vie (sauf démission) parce qu'aucun processus n'est prévu pour remettre cette position en question. En cas de conflit insurmontable, on en arrive à un *fork*, dont nous avons déjà parlé.

Pour comprendre la gouvernance de ces communautés, nous décrivons le cas du noyau Linux, de la Fondation Apache (qui regroupe un grand nombre de projets) et d'un projet particulier issu de la recherche publique, Scikit Learn.

**LE NOYAU LINUX.** Sa gouvernance s'est progressivement établie dans le temps, structurée en quatre niveaux :

- Linus Torvalds, le « dictateur bienveillant à vie » ;
- les lieutenants de confiance ;
- le *pool* de développeurs ;
- la communauté Open Source, qui bien que n'apparaissant pas toujours dans le schéma de gouvernance en fait réellement partie et dispose d'une véritable influence.

Le fonctionnement est par nature en réseau : tout le monde peut contacter tout le monde. Le flux d'information s'étend à l'ensemble de la communauté, en incluant les entreprises, les chercheurs en informatique, ou les simples utilisateurs. La communication entre toutes ces personnes est publique et visible de tous.

Les développeurs qui veulent soumettre une modification l'adressent normalement à un lieutenant, mais ils peuvent aussi la soumettre directement à Linus. Les lieutenants sont spécialisés par domaine technique. Linus et les lieutenants ont ces rôles particuliers parce que leur compétence technique est reconnue par l'ensemble de

la communauté. La structure est guidée par la recherche systématique de la qualité et de la flexibilité.

Le fonctionnement de cette communauté n'est clairement ni égalitaire ni démocratique. La force réside dans la puissance de la multitude de développeurs compétents. Mais l'autorité centrale avec son pouvoir ultime de décision est indispensable pour que le système fonctionne correctement. On peut dire que Linus Torvalds a su mettre en place une « cathédrale » de développeurs de talent pour gérer le « bazar » du développement du produit, pour faire référence à un article célèbre d'Eric Raymond qui opposait le fonctionnement hiérarchisé du développement du logiciel propriétaire (la cathédrale), à l'organisation décentralisée du logiciel open source (le bazar). Certains critiques voient en Linus, un dictateur. Après tout, il en a le titre !

**LA FONDATION APACHE** gère 351 projets ; une poignée d'entre eux concentrent une grande partie de l'activité des développeurs. Ce sont plus de 9 000 développeurs et presque 2 500 *committers*. Chacun de ces 351 projets a sa propre gouvernance, mais ils partagent tous des principes communs, ceux de la méthode Apache (*the Apache way*) :

- Méritocratie : tout le monde peut participer, mais l'influence de chacun est basée sur son mérite publiquement reconnu, ses contributions passées à la communauté. Le mérite est associé à l'individu, indépendamment de son statut (par exemple, employé ou employeur d'une entreprise participante).
- Une communauté de pairs : la communauté est formée de personnes et non d'organisations, ayant toutes le même poids quel que soit leur statut ou leur entreprise. En particulier, il n'y a pas de dictateur bienveillant à vie.
- Communication ouverte : toutes les communications sur le code et les prises de décision sont publiques.
- Consensus : les décisions sont prises par accord et consentement du plus grand nombre.

- La communauté prime le code : il est plus important d'avoir une communauté en bonne santé qu'un code de qualité, parce qu'une communauté forte pourra toujours régler des problèmes, alors qu'une communauté malade aura du mal à maintenir le code.
- Flexibilité : il existe de nombreuses façons de mettre en œuvre la Méthode Apache et la fondation n'en impose aucune.

**Scikit Learn.** La communauté ouverte inclut des universitaires, des chercheurs et des développeurs, et fonde ses règles sur le mérite et le consensus. Sa gouvernance est transparente. En la simplifiant, en voici les règles :

- Les contributeurs sont des membres de la communauté qui contribuent au projet. Cela inclut les développeurs de code, mais aussi l'équipe de communication chargée de faire connaître le produit, son usage et sa marque.
- Les contributeurs expérimentés sont ceux qui ont obtenu le droit d'étiqueter et de clore des problèmes sur GitHub. Tout contributeur peut devenir membre de l'équipe après avoir démontré ses compétences et une assiduité de participation. Tout développeur ou membre de l'équipe des contributeurs expérimentés peut proposer un nouveau membre, auquel cas l'ensemble des développeurs est sollicité. Normalement, l'unanimité est attendue, mais une majorité des deux tiers est en fait suffisante.
- Les *core developers* sont des membres de la communauté qui ont fait montre d'un fort engagement dans la communauté. Ils font des revues de code, ils intègrent les travaux d'autres contributeurs dans le logiciel.
- Le comité technique est composé de *core developers* qui ont des responsabilités supplémentaires pour le bon fonctionnement du projet et un rôle stratégique. À ce jour il est composé de huit membres.
- Les processus pour promouvoir les membres d'un statut à un autre sont démocratiques.



## L'organisation autonome décentralisée

Le numérique favorise le fonctionnement de communs. En s'appuyant sur une organisation autonome décentralisée ou DAO (pour *decentralized autonomous organization*), on peut aller encore plus loin dans l'organisation numérique de la communauté. La DAO est une organisation autour d'une *blockchain* et de *smart contracts*. La *blockchain* (chaîne de blocs en français) propose un registre géré de façon distribuée sur Internet sans recours à aucune autorité centrale. Les *smart contracts* (contrats intelligents en français) consistent en du code qui s'exécute automatiquement quand certains critères sont satisfaits. Ils permettent de définir et faire appliquer les règles de fonctionnement et la gouvernance de la communauté. La transparence du fonctionnement de la communauté est garantie par la *blockchain* et par l'accès au code des *smart contracts* ; chacun peut suivre ce qui se passe dans la communauté.

Les DAO sont relativement récentes. L'encre n'est pas sèche. Dans l'esprit, la DAO permet d'organiser le partage d'une ressource utilisée par une communauté, et le *smart contract* définit sa gouvernance. Nous sommes bien dans le cadre des communs numériques. C'est pourquoi nous avons choisi d'en parler, même brièvement,

en précisant que notre compréhension de cette technologie et de ses usages est encore vraiment parcellaire.

Le projet « The DAO », premier exemple d'une DAO, avait pour but d'organiser une très importante campagne de levée des fonds de capital-risque en 2016. Il permettait à des investisseurs de s'organiser pour lever des fonds en partageant les risques et les profits, sans avoir besoin de faire confiance à l'un d'entre eux pour gérer le tout. Le système fonctionnait sans nécessiter une autorité particulière au-delà de celle du code que tous partageaient. Le système était basé sur la *blockchain* Ethereum. Peu de temps après son lancement, un tiers des fonds (en cryptomonnaie) a été détourné par des hackers mal intentionnés, ce qui a conduit à une partition de la *blockchain* Ethereum (un *fork*). Quelques mois plus tard, la valeur du fonds était retirée des places d'échanges de cryptomonnaies, signant de fait la fin de l'aventure.

Des DAO ont également été utilisées pour des motifs différents comme le financement collectif de logiciels ou la gestion d'organisations caritatives.

Decentraland est un autre exemple plus récent de DAO. C'est un monde numérique virtuel en 3D où les utilisateurs peuvent créer des avatars, acheter des biens immobiliers virtuels ou des objets et créer des entreprises (voir la discussion sur les métavers plus loin). Les utilisateurs possédant un jeton de cryptomonnaie MeANA peuvent participer à la gestion de la plateforme.

On peut aujourd'hui trouver des systèmes pour déployer facilement des DAO comme la plateforme Aragon.

La gouvernance des DAO soulève des difficultés. Leur fonctionnement est gravé dans leur code. Tout doit être prévu avant le lancement, y compris les moyens de faire évoluer le *smart contract*. Les membres peuvent être amenés à voter pour faire évoluer les règles de la communauté, de l'admission de nouveaux membres, à la gestion des conflits, au partage des profits comme des pertes. Des aspects de la gouvernance peuvent également être délégués à un

groupe d'experts. Les évolutions du *smart contract* sont complexes et lentes à mettre en place. Un imprévu peut donc avoir des conséquences dramatiques car la communauté n'a pas le moyen de réagir rapidement.

Pour des logiciels libres ou des données ouvertes, par exemple, il existe des licences, donc un cadre juridique bien défini. De même, une association ou une fondation déposent des statuts officiels. De par sa nouveauté, une DAO n'a encore aujourd'hui aucune forme juridique. Si le modèle doit perdurer, il lui faudra développer son propre cadre juridique.

La *blockchain* joue un rôle central dans les DAO. Certains voient d'ailleurs cette technologie jouer un rôle essentiel dans les évolutions de communs que nous avons déjà mentionnés, Internet et le Web :

- Les métavers sont présentés par certains comme le futur d'Internet. En s'appuyant sur les technologies de réalité virtuelle et augmentée, les métavers proposent des réseaux de mondes virtuels. On peut s'inquiéter de leurs impacts sur la psychologie des utilisateurs, et sur l'environnement, mais ce n'est pas le sujet ici. Une difficulté qu'ils soulèvent est de permettre à un individu de garder, entre plusieurs de ces mondes, son identité, ses objets numériques (par exemple, des avatars) et ses droits. Les métavers pourraient utiliser des *blockchains* pour ce faire.
- Le Web3 est aussi proposé comme la prochaine étape du Web. La *blockchain* est utilisée à la base d'une infrastructure pair à pair, sécurisée pour que les utilisateurs du réseau puissent publier sur le Web de l'information, tout en gardant un contrôle direct sur leurs données. Ils n'ont plus à les confier à des serveurs, comme c'est le cas dans le Web actuel.

Bulles spéculatives, éléments de marketing, véritables avancées du numérique ? Il n'est pas simple de se faire un avis sur les métavers et le Web3.

Les DAO proposent des pistes intéressantes pour organiser des communs numériques. La technologie est relativement récente et son avenir n'est pas clair. Ce qui manque peut-être encore, c'est une *success story* véritablement socialement positive pour être convaincante.

## CHAPITRE 13

# Le modèle économique

Toute organisation, même la plus axée sur le bénévolat, a besoin de se financer. Pour le cas de communs numériques, une certaine infrastructure informatique peut-être nécessaire avec des serveurs, des réseaux télécom, etc. En outre, ces communs exigent souvent des développements informatiques parfois conséquents, cela dans une économie mondiale où les développeurs sont une ressource chère et en forte demande. Ils doivent donc trouver des solutions pérennes pour financer les coûts de la production de leurs logiciels et leurs coûts opératoires. Comme pour la gouvernance, nous allons voir qu'il n'y a pas de solution unique ni de solution idéale, mais une grande diversité de situations. Nous allons aussi rencontrer une grande disparité, en termes de volumes financiers, entre des communs riches et des communs plus frugaux, voire spartiates.

Avant de nous lancer dans le sujet, considérons la question de la gratuité. Pour beaucoup, un commun numérique doit être gratuit. L'idée sous-jacente est que tout un chacun doit pouvoir y accéder gratuitement ; le prix de l'accès ne peut être un obstacle à l'utilisation de la ressource. Pourtant, ces communs ont des coûts, même si le coût marginal d'un utilisateur est en général très faible. Tout l'art des modèles économiques va être d'aboutir à une gratuité ou quasi

gratuité. C'est, par exemple, le cas de l'archive d'articles scientifiques arXiv. L'accès à un article est gratuit. Mais on paie pour obtenir toute la collection.

### *Le financement public*

Commençons par deux communs fortement financés par les États : la science ouverte et les données ouvertes publiques.

La science ouverte est un mouvement à la fois descendant et ascendant. Il correspond souvent, au départ, aux aspirations de scientifiques qui aimeraient que leurs résultats soient le plus largement accessibles. Les scientifiques construisent des communautés qui défendent les accès ouverts, mettent en évidence la supériorité de ce modèle. Ce mouvement est ensuite amplifié par des organismes de recherche, des ministères ou des organismes supranationaux comme l'Unesco ou la Commission européenne. Chacun de ces organismes dispose donc de ses propres financements et pilote le déploiement progressif de la science ouverte. La victoire des communs est confirmée par des obligations légales.

Dans le cadre de l'accès ouvert aux revues scientifiques, plusieurs modèles économiques sont en compétition avec parfois des bras de fer entre les éditeurs scientifiques qui aimeraient conserver leurs marges importantes et les États qui voient dans les communs des moyens pour un meilleur fonctionnement de la recherche et pour empêcher les coûts d'exploser.

Différentes voies sont proposées. On utilise des couleurs pour les distinguer. En simplifiant :

- Dorée : le modèle doré transfère la charge financière du lecteur vers l'auteur en donnant un accès libre aux lecteurs et en faisant payer l'auteur. C'est la solution que les éditeurs scientifiques ont trouvée pour conserver leur marge tout en ouvrant leur lectorat. Un des

inconvenients de ce modèle est le conflit potentiel entre le maintien de la qualité scientifique et la recherche du profit par l'éditeur. Ce modèle a notamment conduit à l'émergence d'éditeurs prédateurs, prêts à publier tout et n'importe quoi, sans véritable évaluation par les pairs, pour faire de l'argent.

- Diamant : le modèle est gratuit tant pour les lecteurs que pour les auteurs. Les auteurs conservent leur droit d'auteur avec des licences de type Creative Commons. Le système peut être financé par des subventions publiques ou par des contributions de grandes bibliothèques ou de grands centres de recherches, qui *souscrivent pour que cela soit ouvert*. Cela conduit à un problème récurrent des communs, celui des « coucous ». Comment empêcher que certains ne profitent de la publication ouverte en refusant de contribuer alors qu'ils le pourraient ?
- Platine : le texte brut est disponible en accès ouvert (*freemium*) mais le lecteur peut vouloir payer l'éditeur pour des services supplémentaires comme un PDF interactif. C'est ce qu'on trouve par exemple dans des revues d'OpenEdition.
- Verte : les scientifiques publient leurs résultats dans des archives ouvertes. Cela peut être en parallèle avec une publication sur le site d'un éditeur après parfois une durée d'embargo pour laisser la primeur de l'information pour quelque temps à l'éditeur.

Observons que, dans les modèles doré et platine, les articles sont sous le contrôle des éditeurs et pas des scientifiques.

Les scientifiques publient aussi parfois des « prépublications » : ce sont des versions préliminaires de leurs articles, comme la version envoyée à l'éditeur de la revue. Et pour ce qui est du droit d'auteur, on voit se développer des réticences à l'abandonner aux éditeurs de revues. Par exemple, en France, les chercheurs du CNRS n'ont plus le droit de l'abandonner.

L'*open data* public est principalement dans les mains de l'État, qui supporte le plus gros de la charge de l'ouverture des données,

en finançant, par exemple en France le service Etalab. Pour le recueil de données, les collectivités territoriales et les entreprises qui bénéficient de délégation de services publics sont essentielles. Enfin, l'État a pris conscience qu'un vrai défi était le passage des données ouvertes aux communs. Il encourage donc la création de communautés autour de données publiques pour les faire vivre et évoluer.

## *Le crowdsourcing*

Considérons maintenant les communs basés sur une forte communauté de volontaires bénévoles, comme Wikipédia, OpenStreetMap et Open Food Facts. Dans ces trois communs, le gros du travail (rédaction des articles, relevé des cartes et collection de données sur les aliments) est effectué par des foules de bénévoles. Wikipédia et OpenStreetMap ont mis en place des fondations, et Open Food Facts s'est structurée en association.

Ces trois exemples permettent d'illustrer des différences entre ces communs :

- les revenus de la fondation Wikimedia étaient de 163 millions de dollars en 2021, et la fondation employait 550 personnes ;
- les revenus d'OpenStreetMap en 2021 étaient de 226 000 livres et la fondation, entièrement gérée par des bénévoles, n'avait pas d'employé ;
- nous ne connaissons pas le montant des revenus d'Open Food Facts, mais l'association avait en 2021 quatre employés à plein temps.

**WIKIPÉDIA** vit essentiellement de dons individuels, qui couvrent 87 % du budget annuel de l'ordre de 175 millions de dollars (exercice 2022-2023). Le reste vient d'institutions philanthropiques, d'entreprises ou de fondations. Wikipédia conduit annuellement des campagnes d'appel aux dons. Ces campagnes ont des résultats qui augmentent chaque année à peu près linéairement. La fondation



Wikimedia dépense moins que ce qu'elle gagne chaque année avec les dons, et a accumulé un trésor de guerre de plusieurs centaines de millions de dollars à ce jour, ce qui garantit son indépendance. La contribution moyenne de ceux qui soutiennent la fondation est de 15 dollars. Donc, le modèle est clairement pérenne et les utilisateurs soutiennent leur encyclopédie. Pour mettre ces chiffres en perspective, le site affiche 5,6 milliards de visiteurs uniques pour le mois de juin 2021, donc la marge de progression est potentiellement énorme.

**OPENSTREETMAP** a adopté un modèle plus proche du bénévolat pur et un budget bien plus frugal. Ses ressources viennent de donations individuelles et d'organismes comme la Croix-Rouge américaine.

**OPEN FOOD FACTS** fonctionne sur un modèle assez proche de celui d'OpenStreetMap et reçoit des subventions de divers organismes publics et de fondations caritatives, notamment Santé Publique France, Google, Ford foundation, Mozilla, Union Européenne, Perl Foundation, Fondation Free, OVH cloud, Afnic et Xwiki. La Fondation fait aussi appel aux dons du public pour soutenir son action.

### *Logiciel libre et open source*

Examinons ici les projets de Linux et Mozilla, sachant que nous rencontrerons dans cet essai de nombreux autres logiciels libres comme Scikit Learn, OpenOffice, Libre Office, ou Elasticsearch.

**LINUX.** Le développement de Linux est piloté par la Fondation Linux, laquelle gère effectivement le projet Linux, mais aussi de nombreux autres projets *open source* (jusqu'à 750 projets), lesquels sont exploités, commercialement ou pas par d'autres entités. Donc la Fondation Linux ne tire pas de profit particulier du système d'exploitation Linux, mais c'est un véhicule qui le développe pour

l'ensemble des parties prenantes, dont un grand nombre sont des sociétés commerciales.

La fondation Linux est financée par des contributions d'entreprises. Ses revenus annuels pour 2021 étaient de 177 millions de dollars. Ses ressources viennent pour 80 % de la participation des membres de la fondation et de dons directs aux projets, et pour 20 % environ de services (formation et organisation d'événements). Parmi les quinze membres « platine » (ceux qui contribuent le plus), on trouve 8 entreprises états-uniennes, 3 japonaises, 2 chinoises, 1 européenne et 1 sud-coréenne. Ces entreprises financent la fondation, mais contribuent en outre au développement de ses projets *via* leurs développeurs.

Le projet Linux lui-même mobilise plus de 1 000 développeurs, ce qui, vu le coût actuel moyen d'un développeur de noyau, dépasse déjà largement le budget de la fondation. Il n'y a pas de mécanisme de commercialisation de Linux à l'intérieur de la fondation, par opposition à ce que l'on trouve par exemple chez Mozilla. Donc il faut voir la Fondation Linux comme une mécanique mise en place par les acteurs de l'industrie pour développer ensemble une technologie d'infrastructure commune. L'esprit est bien celui de l'*open source*, et pas celui du logiciel libre.

Il faut surtout retenir que la grande majorité des développeurs concernés est payée, soit par la fondation, soit par un employeur qui contribue ainsi au développement de Linux ou du projet de la fondation Linux. Le modèle économique est donc stable et pérenne.

**MOZILLA.** Mozilla, surtout connue pour son navigateur Firefox, est structurée en une fondation (Mozilla Foundation), laquelle a une filiale commerciale (Mozilla Corporation). Mozilla a par ailleurs d'autres projets actifs, comme Bugzilla (un système de suivi de bugs), Thunderbird (un client de mail) et Gecko (un moteur de rendu pour pages Web). Elle s'est même essayée à produire un système d'exploitation pour objets connectés (notamment *smartphones*), Firefox OS, un projet arrêté en 2016.

La communauté Mozilla a été créée en 1998 par des anciens de Netscape et a initialement fourni la technologie de navigateur à Netscape et à AOL. En 2003, la Fondation Mozilla a été mise en place pour piloter le projet. En 2005, Mozilla Corporation a été créée pour commercialiser les produits.

De tous les communs que nous avons considérés, Mozilla est clairement celui qui bénéficie du plus important financement. En 2012, ses revenus étaient de 163 millions de dollars. Une énorme partie, 85 %, provenait d'un accord avec Google pour faire de Google le moteur de recherche par défaut de Mozilla. Cet accord a finalement été remplacé par un accord portant sur trois ans pour un total de presque 1 milliard de dollars à la suite d'enchères face à Bing et Yahoo! Depuis la situation s'est dégradée et 250 personnes ont été licenciées en 2020, année où le salaire de la P-DG Mitchell Baker était de 3 millions. On n'est plus dans le bénévolat.

## CHAPITRE 14

# communs *vs* entreprises

Une entreprise privée est un système économique qui organise la production et la distribution de biens à des consommateurs. Les communs proposent une autre forme de système économique qui sert un but similaire. Comme nous allons le voir maintenant, le choix de l'un ou de l'autre a un impact important sur le mode de production des objets, mais également sur le mode de distribution, et sur le produit final.

À un certain niveau de description, les entreprises et les communs sont assez semblables : ils produisent des biens ou des ressources et les distribuent. Existe-t-il des différences entre les processus qu'ils utilisent ? Existe-t-il des différences de nature entre leurs produits ? Ce sont des questions que nous nous posons dans ce chapitre.

### *Un processus de production différent*

À partir de la fin de 2021, en préparation notamment de cet ouvrage, nous avons interviewé des spécialistes des communs numériques (logiciel libre, *open data*, science ouverte, ressources éducatives libres, etc.). Lors de ces premières interviews, un sujet nous a intrigués : plusieurs intervenants ont mentionné que l'adoption de

communs conduisait à de nouvelles façons de produire les ressources. Nous nous sommes alors mis à questionner systématiquement nos interlocuteurs à ce sujet. La transformation importante de la manière de produire les ressources mises en commun a été régulièrement mentionnée.

**LE TRAVAIL COLLECTIF.** La manière de produire devient plus collective, fondée souvent sur la collaboration entre des individus, parfois en très grand nombre. Il est évident que cela a des impacts sur la façon de produire. On s'appuie sur le travail des autres. On doit mieux expliquer ce qu'on fait, car son travail sera utilisé par d'autres. Chacun observe et vérifie le travail des autres.

« **LA CATHÉDRALE ET LE BAZAR** ». C'est le titre d'un essai d'Eric Raymond dans lequel il compare le développement de logiciels de manière conventionnelle, suivant une organisation hiérarchique, une structure de cathédrale, avec le développement de logiciels *open source* par des développeurs organisés sans hiérarchie, en bazar. Même si cela n'est pas une règle absolue, les membres de communs sont rarement fans de structures rigides. Ils privilégient souvent une organisation en bazar, laissant plus de flexibilité, plus d'autonomie à chacun. C'est particulièrement le cas pour les communs par essence altruistes développés par des idéalistes qui ne sont pas venus là pour marcher au pas.

L'organisation du travail est affectée. Le travail lui-même est transformé, comme nous allons le voir à travers quelques domaines clés.

**LOGICIEL LIBRE.** Quand un grand nombre de développeurs participent à la réalisation d'un logiciel comme Linux, MySQL ou LibreOffice, ils le font différemment parce qu'ils font partie d'une collectivité, et parce qu'ils construisent une nouvelle version en améliorant la précédente. La mise en commun, l'audit permanent du code par la communauté, améliore en permanence la qualité. Cette mise en commun améliore aussi la productivité : les développeurs logiciels qui assemblent des composants libres avec l'aide et

le soutien de la communauté ont atteint des sommets de productivité, démontrés dans les hackathons, des rencontres, qui peuvent durer plusieurs jours (et nuits), pendant lesquelles des spécialistes développent de manière collaborative des programmes informatiques.

Laissons la parole à Yves Caseau<sup>1</sup>, qui décrit la nouvelle façon de développer du logiciel :

Le premier élément de changement important, c'est la manière de coder. Avant, pour un problème complexe, on se lançait dans une analyse poussée et le développement d'un logiciel monolithique. Les jeunes actuels commencent par chercher avec Google si un bout de programme existe déjà quelque part qui résout une partie du problème. Du coup, avant on essayait d'avoir les développeurs les plus compétents. Maintenant, on préfère avoir des profils capables de manipuler avec brio des bouts de codes, de les combiner, et de vérifier ce qu'ils font, sans nécessairement comprendre leurs mécanismes intimes. C'est le rêve de l'approche composant où on réutilise sans arrêt des fragments. Le deuxième changement important, c'est que ça raccourcit le processus de développement. [...] Du code est fabriqué sans arrêt. Et la façon dont le code est fait importe autant que le résultat. On se retrouve comme à la belle époque de l'informatique. Il faut aimer le code ! Au début de l'informatique, les développeurs vivaient dans leur code et ils aimaient ça. Puis on a changé de modèle. Quelqu'un écrivait des spécifications d'une application. Ce modèle, très à la mode entre 1980 et 2000, est un modèle qui se meurt. En éloignant le concepteur de l'application du développeur, on avait perdu l'artisanat du développement de code. Aujourd'hui, avec le développement agile, on a retrouvé la proximité, le contact, l'amour du code<sup>2</sup>.

---

1. Yves Caseau, directeur des systèmes d'information de Michelin, a occupé des fonctions de responsabilités autour des logiciels chez Alcatel-Alstom, Bouygues et Axa. Il est membre de l'Académie des technologies depuis 2005 et a publié, en 2020, *L'Approche Lean pour la transformation digitale*, aux éditions Dunod.

2. Yves Caseau, « Comment le numérique a transformé l'informatique », Blog binaire, *Le Monde*, 2015, <https://www.lemonde.fr/blog/binaire/2015/05/22/yves-caseau-comment-le-numerique-a-transforme-linformatique/>.

Notons au passage qu'Yves Caseau évoque le développement d'un logiciel quelconque en entreprise, mais que ses propos sont particulièrement adaptés au développement de logiciel libre, en entreprise ou pas.

**RESSOURCES ÉDUCATIVES LIBRES.** On imagine facilement une manière moderne pour une enseignante de préparer son cours : elle identifie les ressources disponibles sur le sujet du cours, elle les consulte, elle interagit avec ses collègues qui ont proposé des contenus dans ce domaine. Puis elle assemble un certain nombre de ces ressources pour constituer de nouveaux contenus (textes, exercices et corrigés, MOOC, etc.), en y ajoutant des touches personnelles, des contributions qu'elle met tout ou partie à la disposition de la communauté. Oui, bien sûr, la réalité n'est pas toujours aussi simple. Il n'est pas simple de trouver des ressources, pas simple non plus de publier des ressources qu'on vient de réaliser, de les faire connaître. En particulier, la plateforme parfaite permettant la recherche des ressources, leur accès, leur dépôt et l'organisation de la communauté, tient encore de l'utopie. Mais on se dirige vers cela avec des plateformes comme le Réseau Canopé (Réseau de création et d'accompagnement pédagogiques) de l'Éducation nationale qui a pour mission, l'édition, la production et la diffusion de ressources pédagogiques et administratives destinées aux professionnels de l'enseignement.

**SCIENCE OUVERTE.** L'impact sur la méthode de travail est plus subtil à analyser dans ce contexte. Avec l'*open access*, le chercheur a typiquement accès à un plus large éventail de travaux, et surtout à des travaux plus récents, peut-être pas encore même officiellement vérifiés. Il a accès à une gamme plus large de matériels comme des données et des logiciels. La reproductibilité des travaux est simplifiée avec ses avantages : détecter les erreurs, apprendre des expérimentations. Évidemment, l'ouverture a un coût. Pour partager un corpus de données ou un logiciel, il faut le nettoyer, le documenter, le restructurer peut-être. Ces coûts sont largement compensés par les bénéfices de l'ouverture. On peut d'ailleurs observer que

la production de communs comme un logiciel libre ou un corpus ouvert, fait maintenant lentement, mais de plus en plus sûrement, partie du travail de chercheur.

**LES DONNÉES OUVERTES PUBLIQUES.** Les ressources consistent en des données qui existaient en général avant leur ouverture, et qui servent à l'État dans l'exercice de ses missions. Est-ce que leur ouverture change la façon dont elles sont produites ? Pas fondamentalement à l'origine. Mais la tendance est de construire des communautés autour de ces données, par exemple pour participer à leur maintenance. Le public peut participer au signalement des changements de la réalité que les données cherchent à capturer, voire des erreurs présentes dans les jeux de données publiés.

**L'OPEN DATA CROWDSOURCÉ.** Pour ces données, les communs ont véritablement inventé une nouvelle façon de travailler. Les milliers de bénévoles qui recueillent les données d'OpenStreetMap ou d'Open Food Facts, produisent des connaissances, de la valeur, et ils le font d'une façon totalement originale.

### *Un processus de distribution, lui aussi différent...*

Les produits des entreprises sont en général mis sur le marché par des équipes de vente, qui regroupent des spécialistes du marketing et des commerciaux. Les biens numériques de nature commerciale suivent en général la même logique : une force marketing pour mettre en avant le produit, le faire connaître, identifier des prospects potentiels, puis une force commerciale pour démarcher les prospects identifiés, les convaincre d'acheter le produit. La même force commerciale essaiera ensuite de vendre d'autres produits de la gamme à ces clients (*cross sale*) ou de vendre plus du même produit (*upsale*). Pour le logiciel, on évalue en général le coût récurrent de



commercialisation d'un produit à près de la moitié du coût du produit.

Dans la démarche des communs numériques, le produit est souvent gratuit, donc potentiellement, toute cette partie commerciale devrait disparaître, les utilisateurs venant « se servir » d'eux-mêmes du produit. Quand les communs numériques s'imposent dans un domaine, cela sonnerait-il la fin du marketing et du commerce ? C'est la réalité pour certains communs. C'est plus compliqué pour d'autres.

Si on revient à l'exemple de l'*Encyclopædia Britannica* remplacée par Wikipédia, le contraste est saisissant : l'*Encyclopædia Britannica* était vendue au porte-à-porte par une force commerciale de 2 000 vendeurs qui tiraient les sonnettes. Aujourd'hui, personne ne tire de sonnettes pour « vendre » Wikipédia : les utilisateurs se connectent et la consultent, c'est tout. Ce qui se rapproche le plus peut-être du commercial pour Wikipédia est le travail pour convaincre les utilisateurs de soutenir par leurs dons l'encyclopédie.

Pour les données ouvertes publiques, les jeux de données qui sont maintenant disponibles gratuitement n'ont besoin d'aucune force commerciale pour les distribuer et les services de l'État qui s'occupaient de la commercialisation de ces données ont disparu. D'un autre côté, on sait maintenant qu'il ne suffit pas de mettre des données en accès ouvert sur le Web pour qu'elles soient utilisées massivement. Une tâche de l'État reste de faire le *marketing* de l'utilisation de ces données ouvertes. Cela n'a évidemment rien à voir avec faire de la publicité pour vendre des données.

L'*open data crowdsourcé* comme celui d'OpenStreetMap et d'Open Food Facts est également distribué gratuitement sans intervention de la moindre force commerciale ; la promotion et la visibilité de l'offre sont prises en charge naturellement par la communauté, le plus souvent par un bouche-à-oreille viral. C'est aussi le cas pour les ressources éducatives libres. Bien sûr, cela n'empêche pas d'avoir

à faire le *marketing* de l'utilisation d'OpenStreetMap, Open Food Facts ou, par exemple, de Canopé.

Pour ce qui est de la science ouverte, les revues en accès ouvert offrent une bonne diversité de situations. Quand ce sont de grands clients qui paient pour l'accès, il faut recruter ces clients. Quand ce sont les auteurs qui paient, on peut imaginer que l'attrait de la renommée de la revue est une publicité suffisante. Mais les éditeurs continuent à faire la publicité de leurs revues, à les faire connaître, à mettre en avant leur renommée, etc. On peut imaginer que pour des revues directement gérées par les scientifiques ou leurs institutions comme le CNRS, marketing et publicité deviennent inutiles, ou pour le moins, tiennent une place moins importante, perdent considérablement en coût.

On pourrait croire qu'avec le logiciel libre, toute forme de commerce disparaît : les produits sont disponibles, les utilisateurs potentiels peuvent les essayer et éventuellement les adopter. La réalité est plus nuancée.

Des logiciels et de nombreux composants *open source* échappent totalement à la logique commerciale et sont simplement disponibles sur le Net. C'est généralement le cas pour certains logiciels en début de vie, quand les moyens d'en faire la publicité manquent, ou quand le bouche-à-oreille est suffisant pour en faire la promotion. Quand le projet prend de l'ampleur, comme Scikit-learn, sa promotion est indispensable ne serait-ce que pour trouver des leviers de croissance.

Quand les liens avec l'industrie du logiciel s'intensifient, la situation change. Des éditeurs de logiciels *open source* opèrent alors suivant un modèle entrepreneurial classique et ont des forces de vente relativement classiques. La version Red Hat de Linux est commercialisée par Red Hat avec des commerciaux de façon assez classique. C'est encore plus le cas d'entreprises comme IBM et Microsoft notamment qui ont mis en *open source* une partie de leurs logiciels. La situation est encore plus complexe avec des logiciels comme MySQL vendus par des commerciaux sous double licence.

Il faut aussi mentionner les nombreuses entreprises du marché du service autour du logiciel libre. En France, on les appelle des SSSL, sociétés de services du logiciel libre. Elles fournissent du support, de la formation, de l'installation et de l'hébergement de logiciels libres. Elles sont souvent structurées de façon classique avec des équipes commerciales.

### *... et des produits bien sûr différents !*

Pour clore ce tour d'horizon sur ces transformations apportées par les communs, on se doit de mentionner l'impact sur le produit lui-même. Un produit élaboré dans le cadre d'un commun numérique est-il différent de son équivalent dans un système propriétaire ?

Une fois vendu, un bien numérique n'est plus la responsabilité de son fabricant (sauf garanties contractuelles). L'utilisateur est encouragé à acheter la prochaine version. Au contraire, dans les communs, la maintenance du bien et sa capacité de continuer à servir vont de soi, sauf disparition du commun. Dans le cadre d'un monde numérique qui évolue très vite, où l'adaptation permanente est l'enjeu majeur, cette différence est essentielle.

Pour ce qui est des produits, c'est peut-être dans le contexte de l'encyclopédie que les différences sautent le mieux aux yeux. Une entrée Wikipédia, contrairement à une entrée dans une encyclopédie classique, est un travail en cours (*work in progress*). Dans un premier temps, ce n'est qu'une ébauche comme une bannière le proclame. Ensuite, les éditeurs de Wikipédia vont mettre un point d'honneur à améliorer l'entrée. Comme le souligne Florence Devouard, une grande wikipédienne : « On vise la perfection, par exemple, le label "Article de qualité"<sup>1</sup>, qui est un label très difficile à obtenir. Certains travaillent comme des dingues sur un article pour y arriver. C'est de

---

1. Ce label est décerné à la suite d'un vote de contributeurs enregistrés de Wikipédia.

cette quête de perfection qu'ils tirent leur fierté. » On peut d'ailleurs voir l'historique des modifications et comprendre comment la version actuelle a été construite. Un article d'une encyclopédie classique assène une vérité ; il est figé. Une entrée de Wikipédia décrit la connaissance en train de se construire. Une erreur dans une encyclopédie attendra la prochaine édition pour disparaître. Son sort peut être fixé dans l'instant sur Wikipédia.

On retrouve un contraste semblable dans le domaine de l'éducation. Un manuel éducatif privé est figé, statique. Il faudra attendre la prochaine édition pour le voir évoluer. Par contre, nous verrons que les ressources éducatives sont guidées par les 5 R : retenir, réutiliser, réviser, remixer, redistribuer. Plusieurs d'entre elles insistent sur le droit de faire vivre les ressources en les modifiant, les combinant. Donc, d'un côté, un produit fixe et immuable, de l'autre, quelque chose qui ressemble plus à de la pâte à modeler, que l'on peut utiliser comme outil personnalisable et malléable. Cela montre bien une différence essentielle de nature des deux produits.

Un commun éducatif, en conséquence, est le résultat d'un travail plus collaboratif. Comme nous l'avons vu, la notion d'auteur s'estompe, remplacée par celle de contributeurs. Il devient plus compliqué de dire pour un élément du tout, qui a contribué, plus complexe de définir la propriété intellectuelle. De fait, cette propriété perd en importance devant l'objectif : réaliser un produit de qualité au service de la collectivité. Pour un cours développé collectivement, la notion d'auteurs n'est pas essentielle.

La situation du logiciel libre est également assez semblable. Un logiciel propriétaire est fermé, ne peut être utilisé que pour faire exécuter le code dans un cadre très précis. On ne sait pas comment il est fait à l'intérieur, on ne peut vérifier la façon dont il fonctionne, la seule documentation dont on dispose est celle de l'utilisateur. Dans le cas d'un logiciel libre, on dispose du code source et de sa documentation, on peut en faire des usages différents : le comprendre, l'analyser, le corriger, l'adapter à ses besoins, l'améliorer, le porter

dans un environnement différent. Là encore, la notion de contributeur change. On peut contribuer à un logiciel libre en se contentant d'en lire le code, en détectant par exemple un trou de sécurité, sans même écrire une seule ligne de code.

## CHAPITRE 15

# La création et le développement des communs

La création d'un commun, rappelons-le, passe par le choix d'une ressource ou d'un ensemble de ressources, puis par la constitution d'une communauté qui va mettre en place des règles de production et de gestion du commun, enfin par la mise à disposition des ressources du commun, soit à l'ensemble de la population, soit à une communauté d'utilisateurs plus restreinte. Le tout s'appuie sur un modèle économique garant de la pérennité et de la durabilité du commun. La vie du commun passe, dans de nombreux cas, par la croissance de la communauté de producteurs et de celle des consommateurs.

Une telle organisation peut conduire à un véritable écosystème : un ensemble de personnes et d'organismes, qui vivent au sein d'un environnement défini par le commun en question, et interagissent entre eux. Cet écosystème va se mettre en place, fonctionner et se développer parfois en coexistence avec d'autres systèmes propriétaires ou communs. L'existence d'un écosystème autour d'un commun, et sa coexistence avec d'autres systèmes (coopération, concurrence, destruction...) sont au cœur de ce chapitre.

## *Création ex nihilo*

La mise en place d'un commun numérique peut être l'occasion de la création d'un nouvel ensemble de ressources qui jusqu'à ce jour n'était ni produit ni diffusé. Il peut en résulter un nouvel acteur, qui interagit avec des acteurs existants, voire est à l'origine de nouveaux acteurs. C'est ce qui s'est produit par exemple avec Open Food Facts.

Open Food Facts, un succès français devenu international a été créé en 2012 par Stéphane Gigandet (un centralien) rejoint ensuite par Pierre Slamich (un de Science Po). Le domaine de la santé et de la nutrition saine et équilibrée est un sujet important du début de ce siècle. Pour y répondre et donner aux consommateurs les informations leur permettant de faire des choix éclairés, Open Food Facts a constitué et maintient une base de données des informations disponibles sur les produits alimentaires et la met à disposition de tous.

Le modèle initial de Stéphane Gigandet est celui d'OpenStreetMap et de Wikipédia. Les données sont recueillies par des contributeurs volontaires qui prennent des photos de l'emballage des produits et soumettent les données. Elles sont ensuite traitées et analysées par l'équipe pour les valider et les intégrer dans la base. La communauté est donc constituée de bénévoles (comme pour Wikipédia ou OpenStreetMap) en mode *crowdsourcing*. Chaque produit est identifié de façon unique par son code-barres. Pour chaque produit référencé on retrouve, notamment, une dénomination générique du produit, sa quantité, le conditionnement (carton, surgelé, etc.), la ou les marques du produit, la catégorie d'aliment du produit afin de permettre des comparaisons, les lieux de fabrication ou de transformation, les magasins et pays où le produit est en vente, la liste des ingrédients et de traces éventuelles (pour les allergies, interdits alimentaires ou tout régime spécifique), les additifs alimentaires

détectés à partir de cette dernière, et les informations nutritionnelles. Une taxonomie des ingrédients permet de structurer la base.

La base est mondiale et multilingue, 180 pays sont couverts. La barre du million de produits a été passée en 2019. La langue de travail est l'anglais (au grand dam des contributeurs québécois) et des contributeurs traduisent dans de nombreuses langues. La base de données est ouverte (sous licence ODbL<sup>1</sup>), accessible à tous. Le logiciel est sur GitHub, développé et maintenu par une équipe technique propre, même s'ils ont reçu l'aide de quelques contributeurs.

Une application Open Food Facts permet de scanner le produit et d'avoir instantanément ses données sur le produit, dont souvent un Nutri-Score<sup>2</sup>. Les produits alimentaires évoluent dans le temps (leur composition évolue). Open Food Facts les identifie par leur code-barres et mémorise cette évolution. Certains produits changent d'un pays à l'autre : par exemple, le Fanta Orange est très différent d'un pays à l'autre.

Le modèle économique est basé sur le bénévolat et sur l'appel aux dons. Jusqu'en 2018, l'association n'a fonctionné qu'avec des bénévoles. En 2018, une subvention de Santé Publique France a permis de recruter et l'association compte à ce jour quatre salariés permanents. L'association s'interdit tout financement de la part des industriels pour garder une totale indépendance. Mais la plateforme leur est accessible *via* une interface dédiée, ils peuvent notamment envoyer les informations sur leurs produits. Certains soumettent jusqu'à 5 000 produits et peuvent bénéficier d'analyses d'ensemble.

Open Food Facts est à l'origine d'un véritable écosystème. Les données sont utilisées par de nombreuses applications (plus de 200) en France et dans le monde. Certaines de ces applications gèrent des

---

1. La licence ODbL est une des licences *open data* classique, elle est notamment virale.

2. Le Nutri-Score, proposé en 2014 par des scientifiques et mis en place à l'initiative du gouvernement français en 2016, est ensuite repris dans d'autres pays comme la Belgique, l'Espagne, l'Allemagne et les Pays-Bas, et son utilisation est recommandée par l'OMS. Il donne une note à cinq niveaux de A à E pour évaluer la valeur nutritionnelle d'un aliment.



sujets spécialisés, par exemple l'alimentation pour les diabétiques ou pour les personnes souffrant d'allergies.

En 2017, la start-up Yuka a réalisé une application similaire à celle d'Open Food Facts, basée sur un modèle commercial. Pendant un temps, Yuka a utilisé les données d'Open Food Facts, mais a arrêté en 2018. En revanche, Yuka est un contributeur important à la base d'Open Food Facts. Yuka a une base propriétaire et ne la met pas à disposition du public. Son modèle économique s'appuie sur une application gratuite avec une version premium par abonnement payant. Il est intéressant d'observer qu'aux yeux du grand public, Yuka est plus visible qu'Open Food Facts. Mais il faut se rappeler que les objectifs d'Open Food Facts dépassent largement ceux de la seule information du public sur les produits alimentaires et visent notamment un impact plus général de santé publique et d'influence sur l'industrie agroalimentaire.

Leurs données servent par exemple à l'action publique : ainsi, l'Assemblée nationale dispose de données et d'une vision d'ensemble.

La communauté est très active et fonctionne de façon assez informelle. Elle utilise pour communiquer un *slack* avec plus de 300 canaux.

### *Perturbation d'un marché existant*

La rencontre entre un produit privé et un commun peut conduire à une confrontation entre deux façons de produire, d'utiliser, de diffuser des ressources. Le point de départ du travail d'Elinor Ostrom était le mouvement des *enclosures* qui privatisait des pâturages communaux jusqu'alors gérés en commun. Dans ce cas, des communs préexistants se trouvaient sous la pression d'une production privée. Parfois, des communs s'installent au contraire sur des marchés gouvernés par la propriété privée. Dans ce cas, la mise en

place du commun et son développement se heurtent à un système et des acteurs bien établis. Deux scénarios sont alors possibles :

- Le grand remplacement : le commun, en s'appuyant peut-être sur un écosystème, prend tout l'espace et fait disparaître les acteurs existants (au moins dans leur rôle initial), c'est l'exemple de Wikipédia et la disparition de l'*Encyclopædia Britannica*, qui a dû se réinventer sur de nouveaux types de produits.
- La coexistence : l'écosystème du commun coexiste avec les acteurs existants pour se partager le marché. C'est le cas par exemple du logiciel libre qui partage le marché avec le logiciel propriétaire.

Plus que le modèle du chemin de fer remplaçant les diligences, le modèle dominant est celui de la coexistence, de celle entre logiciel libre et logiciel propriétaire, entre données ouvertes et données propriétaires ou celle de divers modèles de publications scientifiques. Trois exemples vont nous permettre d'illustrer la diversité des situations : le cas Wikipédia contre l'*Encyclopædia Britannica* présente un remplacement, MySQL contre Oracle, une coexistence ; enfin le cas de l'*open source* face à Microsoft Office illustre la difficulté que les communs peuvent rencontrer quand ils cherchent à s'installer sur un marché bien fermé par une entreprise privée puissante.

### *Wikipédia vs. Encyclopædia Britannica*

La rédaction de la première édition de l'*Encyclopædia Britannica* dura trois ans. Elle fut publiée entre 1768 et 1771 à Édimbourg en trois volumes. Dans sa dernière version, l'ouvrage faisait 32 volumes et pesait 58 kilos. Pendant deux cents ans, les nouvelles versions de l'encyclopédie étaient réalisées manuellement et avec des moyens mécaniques, chaque nouvelle version prenait plusieurs années, jamais moins d'un an.

Dans les années 1970, l'éditeur commence à utiliser des moyens informatiques : le contenu de l'encyclopédie est chargé sur un serveur pour améliorer le processus des mises à jour annuelles. Dans les années 1980, l'entreprise produit l'édition électroniquement ce qui améliore la productivité. Le modèle économique de l'entreprise est la vente par des commerciaux à prix coûtant, le profit étant généré par les abonnements aux mises à jour annuelles sous forme d'un livret.

Au sommet de son histoire, le démarchage au porte-à-porte de l'*Encyclopædia Britannica* est effectué par plus de 2 000 commerciaux. En 1990, elle est vendue à plus de 100 000 exemplaires. Mais cette date marque le pic de la distribution de l'encyclopédie sous forme papier. Le modèle porte à porte marche moins bien et l'arrivée des PC avec des lecteurs de CD-ROM est une concurrence forte. Les ventes chutent brutalement : 51 000 en 1994, 3 000 en 1996.

En 1993, Microsoft lance *Encarta*, une encyclopédie disponible sur CD-ROM et en ligne. Originellement vendue 395 dollars, son prix baisse rapidement à 99 dollars et le CD-ROM est souvent disponible gratuitement avec tout PC Wintel. *Encarta* sera développé jusqu'en 2009, date à laquelle Microsoft abandonne le produit.

En 1994, l'*Encyclopædia Britannica* réagit et produit son premier CD-ROM au prix de 1 200 dollars, à peu près le même prix que la version papier. En même temps que la version CD-ROM, la société introduit une version Web, un concept original pour l'époque. Ces deux nouveaux canaux de vente cannibalisent bien entendu la version papier, et ce à un rythme plus rapide que l'entreprise ne l'avait anticipé.

En 1996, à court de cash, l'entreprise est rachetée par Jacob E. Safra, un investisseur suisse. Comme souvent dans ces cas-là, un consultant extérieur, Jorge Cauz, est nommé à la tête pour définir et mettre en œuvre une nouvelle stratégie pour sauver l'entreprise avec les inévitables coupes dans les coûts. Dès les premiers mois, Cauz démantèle le réseau de vente des vendeurs porte à porte, et le remplace par une approche de marketing direct, puis il réajuste le

prix du CD-ROM, non à ce que le produit coûte à produire mais à ce que le marché est prêt à payer : de 1 200 dollars à moins de 1 000, puis à 150, enfin à moins de 100.

Alors que l'entreprise se bat pour survivre sur le marché du CD-ROM, celui-ci est disrupté par un nouveau phénomène : le passage sur Internet. La société peut basculer alors sur un nouveau modèle de vente en ligne.

Deux essais malheureux, l'un, avec un modèle gratuit en ligne supporté par la publicité, et l'autre, un site pour les scolaires mal conçu, se soldent par des échecs, mais permettent à l'entreprise de mieux comprendre le marché Internet et de pivoter sur un marché différent, celui de l'éducation.

En 2001, Wikipédia est lancée, à la suite de l'essai avorté de Nupedia (une première version d'une encyclopédie en ligne). L'*Encyclopædia Britannica* regarde avec curiosité et scepticisme ce nouvel acteur. Après les premiers succès de ce produit original dont l'usage croît très rapidement, l'*Encyclopædia Britannica* se positionne par rapport à ce concurrent sur la qualité de son contenu.

En 2003, ayant montré qu'il peut sauver l'entreprise, le consultant coupeur de coûts devient président. En 2005, la revue *Nature* publie un article comparant la qualité de l'*Encyclopædia Britannica* et de Wikipédia. Le travail met en avant 50 entrées des deux encyclopédies sur les mêmes sujets et un panel de scientifiques évalue la qualité en comptant le nombre d'erreurs. Il trouve 2,9 erreurs en moyenne par article pour l'*Encyclopædia Britannica* et 3,9 pour Wikipédia.

En 2006, Cauz déclare que Wikipédia deviendra « une masse énorme d'articles médiocres, inégaux, peu fiables et la plupart du temps illisibles ». Cela n'empêche pas l'*Encyclopædia Britannica* de sortir effectivement du marché de l'encyclopédie pour se focaliser sur celui de l'éducation en fournissant des outils pour les élèves et les enseignants.

En 2012, Cauz annonce la fin de la publication de la version papier. La nouvelle est un scoop, mais elle ne fait que confirmer un virage déjà pris depuis longtemps par l'entreprise.

Cela désespère certains des vieux lecteurs de l'*Encyclopædia Britannica*. Sur Twitter on peut lire des commentaires du type « désolé de vous avoir été infidèle, l'*Encyclopædia Britannica*. Wikipédia était juste là et si pratique, mais ça ne voulait rien dire. Revenez, s'il vous plaît ! » Mais l'*Encyclopædia Britannica* a depuis longtemps déjà changé de métier.

En 2015, une nouvelle étude, réalisée à l'initiative de Wikipédia par une équipe indépendante, constituée du groupe EPIC et d'une équipe l'Université d'Oxford, montrait que Wikipédia était supérieur à l'*Encyclopædia Britannica* en anglais, et à Enciclonet en espagnol.

Depuis cinq ans, la nouvelle société, avec son nouveau modèle économique a vu une croissance annuelle à deux chiffres, mais elle est effectivement sortie du marché de l'encyclopédie, l'abandonnant (pour le moment du moins) à Wikipédia.

Cet exemple illustre le basculement complet d'un domaine dans le modèle des communs, la solution propriétaire disparaissant totalement.

## *MySQL vs Oracle*

Dans certains cas, en particulier pour des logiciels, le produit dominant bénéficie d'un énorme effet réseau. On dit que le « *winner takes all* » : « le vainqueur emporte la mise. » Il devient extrêmement compliqué de détrôner les produits dominants. C'est ce qui s'est passé dans le monde des bases de données jusqu'à ce qu'un commun numérique vienne rebattre les cartes d'un marché que l'on croyait quasi fermé.

Une base de données est un corpus de données dans un domaine particulier. Un système de gestion de bases de données est le logiciel

permettant de gérer ces données, en particulier les stocker et les interroger. Le marché mondial annuel des systèmes de gestion de base de données pèse aujourd'hui de l'ordre de 85 milliards de dollars. Il a fait, depuis sa naissance, l'objet de nombreuses convoitises.

Les bases de données relationnelles (qui stockent les données sous forme de tableaux avec des lignes et des colonnes, et sont interrogeables avec le langage SQL) naissent à la fin des années 1970, à la suite des travaux de chercheurs académiques et industriels. De grands industriels, comme IBM ou Microsoft, et une poignée de start-up californiennes (Oracle, Ingres, Sybase et Informix) se lancent sur ce créneau.

Après une lutte acharnée entre ces acteurs pendant les années 1980 et 1990, le marché se structure et se consolide : IBM est un acteur établi avec DB2, Oracle taille des croupières à tout le monde, Microsoft entre sur le marché avec SQL Server en 1989. Ingres est racheté par Computer Associate en 1994, Informix par IBM en 2001, et Sybase par SAP en 2010.

Donc, au milieu des années 1990, on peut dire que la messe est dite : le marché est stabilisé et Oracle est le grand gagnant, « *the 800-pound gorilla* », et il paraît très compliqué, voire impossible, pour un nouvel acteur de se faire une place sur ce marché d'un produit qui est pratiquement devenu une commodité.

Mais en 1995, en Suède, Michael Widenius lance, avec d'autres, la start-up MySQL (My est le prénom de sa première fille). MySQL, tout d'abord positionnée comme un système de bases de données personnel, devient rapidement un système multiutilisateur. Le système est diffusé sous une double licence, *open source* et propriétaire. La version *open source* et gratuite devient vite très populaire, là où personne n'attendait un nouvel acteur. Les avantages de l'*open source* jouent ici à plein. Les entreprises peuvent inspecter le code et développer des logiciels compatibles, ce qui conduit à la création d'un écosystème autour de MySQL.

Linux, Apache, MySQL et PHP, forment la « pile » *open source* LAMP utilisée par un très grand nombre de développeurs, donc d'applications. (Une pile de logiciels, *software stack* en anglais, est un ensemble de logiciels qui permet de développer des sites et des applications Web.) La pile LAMP se généralise sur la plupart des serveurs Web.

Après plusieurs tours de financement par des capitaux-risqueurs, MySQL est rachetée par Sun, le constructeur informatique californien en 2008 pour 1 milliard de dollars, ce qui en fait à l'époque la plus grosse acquisition dans le monde de l'*open source*. Selon Enlyft, avec 190 000 clients uniques, MySQL serait le système le plus utilisé au monde, avec plus de 20 % de parts de marché des bases de données transactionnelles. Ironie de l'histoire, Sun sera racheté en 2010 par Oracle.

Pour la petite histoire, on notera qu'en 2009 Michael Widenius réalise un *fork*<sup>1</sup> de MySQL, sous le nom de MariaDB (Maria est le prénom de sa deuxième fille). MariaDB est gérée par une fondation, mise en place par Widenius. Cette initiative vient en réaction à l'achat de MySQL par Sun, pour garantir que le logiciel restera libre et *open source*. MariaDB va vivre sa vie indépendamment de MySQL canal historique, et est aujourd'hui installé par défaut sur certaines distributions Linux à la place de MySQL. Si MySQL est encore très populaire, MariaDB, plus performant, est de plus en plus utilisé.

Cet exemple illustre comment un commun peut arriver à défier des entreprises dominantes bien installées, renverser des situations apparemment bien établies. Mais c'est tout sauf simple comme le montre un autre exemple.

---

1. Le *fork* est une scission de la communauté résultant en deux logiciels concurrents. Voir la définition exacte au chapitre XXX.

## *LibreOffice et OpenOffice vs Microsoft Office*

Le marché des suites bureautiques est énorme parce que tous les utilisateurs ont besoin d'une suite bureautique pour produire avec leurs PC des documents, des tableurs et des présentations. Il a donc été, depuis toujours, l'objet de convoitises. Dans les années 1980, la concurrence a été rude entre des acteurs comme Lotus 123, Wordperfect et Microsoft Office. Microsoft a profité intelligemment de sa position de quasi-monopole sur les systèmes d'exploitation de PC pour imposer son logiciel Office comme le standard de fait. À ce jour, plus d'un milliard de personnes utilisent MS Office dans le monde.

De nombreuses tentatives ont été faites pour essayer de remettre en cause cette domination. En 1985, la start-up Star Division, en Allemagne, créée par un développeur de 16 ans, Marco Börries, développe le logiciel StarWrite, concurrent de Microsoft Word. Le produit, initialement vendu comme un logiciel propriétaire, est un succès et sera utilisé sur 25 millions de machines. Sun rachète Star Division en 1999 et remet le produit sur le marché sous le nom de StarOffice, cette fois-ci en *open source*, puis sous le nom d'OpenOffice. La suite logicielle est compatible Microsoft, c'est-à-dire qu'elle peut ouvrir et mettre à jour un document Microsoft Office et que tout document OpenOffice peut être exporté sous un format Microsoft.

En mars 2009, une étude montre qu'OpenOffice a un taux de pénétration de 3 % dans le marché de l'entreprise, un résultat mitigé pour Sun. Sun est rachetée par Oracle en 2010 qui ne sait pas trop quoi faire d'OpenOffice, alors que la communauté OpenOffice n'est pas ravie de se retrouver dans la maison Oracle.

Oracle se débarrasse du bébé en le refilant à la Fondation Apache, qui met le logiciel en licence Apache (une licence légèrement



différente de GPL sous laquelle se trouvait OpenOffice). La communauté OpenOffice, trouvant la fondation Apache pas assez active sur le produit, crée une nouvelle fondation (The Document Foundation) et fait un *fork* d'OpenOffice qui donne naissance à LibreOffice.

Aujourd'hui, OpenOffice continue sous le parapluie de la Fondation Apache, LibreOffice sous le parapluie de The Document Foundation. Mais Microsoft Office reste un standard de fait. Pourtant, la messe est-elle dite ? Pas certain. Des acteurs publics ont basculé des dizaines ou centaines de milliers de PC vers LibreOffice, comme l'administration de la région de Valence ou le ministère de la défense italien. Si ce mouvement semble perdurer, il n'est pas assez massif pour remettre en question la domination de Microsoft.

Donc le miracle de MySQL ne s'est pas (encore) reproduit, et cela parce que tout le monde reçoit des documents MS Office (et doit pouvoir les traiter) et doit pouvoir envoyer des documents MS Office.

Si cet exemple illustre bien la puissance de l'effet de réseau qui a permis à Microsoft Office de se maintenir en position de monopole, on peut aussi en tirer deux leçons : qu'un *fork* par une partie de la communauté peut permettre de redynamiser un commun à la peine, et que les pouvoirs publics peuvent s'avérer des appuis puissants des communs.

### *Microsoft et l'open source*

Dans certains cas, une entreprise privée peut être convertie, en dépit de son plein gré à l'*open source*.

En 1976, des années avant que le logiciel libre ait trouvé un nom et une pratique, un jeune chef d'entreprise écrit une lettre ouverte dénonçant les programmeurs qui utilisent du logiciel sans le payer. « C'est du vol ! » explique-t-il. Son nom : Bill Gates. C'est bien l'esprit que défendra l'entreprise qu'il a créée, Microsoft, qui vit

de la vente de logiciels. Bill Gates et Steve Ballmer, un autre dirigeant de Microsoft, avancent que les partisans de l'*open source* et les développeurs du noyau Linux sont des communistes (une insulte suprême aux États-Unis !), et que Linux est un cancer qui va détruire la propriété privée.

Dans le documentaire *Revolution OS*, de 2001, Eric Raymond, l'une des figures clés du logiciel *open source*, raconte sa rencontre avec Craig Mundie, *chief research and strategy officer* de Microsoft, qui lui demande qui il est. Raymond répond : « *I'm your worst nightmare* » (« Je suis votre pire cauchemar »). C'est dire si les rapports entre Microsoft et la communauté du logiciel libre ne sont pas bons.

On peut être profondément ancré dans une culture et en changer par intérêt économique. Le vent tourne et Microsoft va progressivement se mettre au logiciel *open source*.

C'est Satya Nadella, le nouveau CEO nommé en 2014 à la place de Steve Ballmer, qui va confirmer le virage vers l'*open source*. Il met en *open source* une partie du code de .net, un environnement de développement de logiciel central pour Microsoft, pour les systèmes d'exploitation Windows, Linux et macOS. Il investit dans le développement de Linux, et adhère à la Fondation Linux et à l'Open Source Initiative. Il conclut l'acquisition de GitHub en 2018, l'infrastructure d'hébergement la plus populaire des projets *open source*. En 2017, Microsoft, l'ennemi de toujours du logiciel libre, devient le plus gros contributeur mondial aux projets *open source* (mesuré par la présence de ses développeurs sur GitHub). Par exemple, un produit phare de Microsoft est Visual Studio Code, qui a été publié en *open source*. Ce système est un environnement de développement de logiciel très populaire sous Windows, mais également sous Linux.

Ces évolutions de Microsoft illustrent comment l'arrivée de communs et leur succès peuvent amener des acteurs privés à se convertir eux-mêmes à l'arrivée des communs, à mettre en place des modèles hybrides, propriétaires- communs.

## *Elastic vs Amazon*

Elasticsearch est un logiciel de moteur de recherche, créé en 2010 par Shay Banon. En 2012, la société Elastic NV est créée à Amsterdam et fournit services et produits additionnels autour d'Elasticsearch. Le logiciel, sous licence libre, est très largement utilisé par de nombreux projets et produits qui veulent installer un moteur de recherche. La société devient publique en 2018 sur le New York Stock Exchange. Sa capitalisation est à ce jour de l'ordre de 8 milliards de dollars. Les revenus d'Elastic viennent des services et des produits complémentaires à son moteur de recherche.

Le produit est initialement disponible sous licence Apache 2.0, une licence reconnue comme libre. En 2021, la société annonce un changement de licence et passe à une double licence sous Server Side Public License (non reconnue comme licence libre) et sous une licence Elastic propriétaire. C'est la fin de ce logiciel libre.

Les raisons de ce changement ? Un conflit avec Amazon. L'offre *cloud* d'Amazon Web Services inclut un nombre de services, dont Elasticsearch, et Elastic considère qu'Amazon ne respecte pas les règles et lui reproche notamment une violation de la marque (en ayant une offre appelée Amazon Elasticsearch Services) et une annonce incorrecte de partenariat avec Elastic. D'où le changement de licence pour forcer Amazon à une meilleure conduite. Le sujet est clairement financier : Elastic, fournissant sa technologie, veut une part du gâteau qu'Amazon reçoit en vendant son *cloud* qui inclut Elastic.

La réponse d'Amazon : un *fork* ! Donc, Amazon crée une nouvelle version, appelée Opensearch en avril 2021. Cette version reste, elle, sous licence Apache 2.0. La bagarre se complique quand Elastic fait son possible pour rendre incompatibles ses connecteurs avec l'utilisation d'Opensearch.

Qui va gagner ? Le temps le dira, mais la communauté y perd certainement en se fragmentant. Les deux produits sont maintenant en compétition pour les utilisateurs, mais aussi pour la communauté des contributeurs, et donc pour savoir qui attirera le plus de développeurs. On se retrouve avec deux produits distincts et concurrents qui doivent tous les deux être maintenus et améliorés.

Qui a raison ? La situation est assez classique. Une entreprise innove, ici Elastic, une grande entreprise profite de cette innovation, et elles ne sont pas d'accord sur le partage du gâteau. Elastic n'est pas seule dans cette situation, d'autres entreprises du monde *open source* comme MongoDB ont changé leur licence pour bloquer Amazon dans leur stratégie de vente. En général, on essaie de se mettre d'accord parce que personne n'a à gagner d'une guerre. La grande entreprise participe au développement et compense la plus petite. Elastic dit d'ailleurs avoir des relations sans problèmes avec d'autres fournisseurs de *cloud* (Microsoft, Google, Alibaba, etc.).

Nous n'essaierons pas de dire qui est le bon de cette histoire et qui en est le méchant. Ce qui nous intéresse surtout c'est qu'elle illustre les difficultés et les limites de l'utilisation de l'*open source* par des entreprises.

### *Le GFII vs les données ouvertes françaises*

Terminons ce tour d'horizon avec un cas où la mise en place du commun oblige l'ensemble des acteurs à se repositionner et à faire évoluer leur métier.

Le GFII est le Groupement français de l'industrie de l'information. Il regroupe des industriels français ou internationaux qui vendent des données et de l'information. À la fin des années 2000, le petit monde de la donnée publique française ronronne paisiblement sur un modèle bien rodé. L'État produit des données et les vend. Les industriels du GFII achètent ces données à l'État en gros et les

commercialisent au détail après les avoir plus ou moins reformatées ou restructurées.

Des exemples de ces services publics sont la DILA (Direction de l'information légale et administrative), Infogreffe et l'Insee.

La DILA est installée dans le 15<sup>e</sup> arrondissement de Paris. Dans son hall d'entrée trônent deux énormes machines d'imprimerie datant probablement du début du xx<sup>e</sup> siècle, souvenir d'une époque où cette direction imprimait de nombreuses revues papier (dont le célèbre *Journal officiel*). La DILA vend des licences d'accès au *Bulletin officiel des annonces civiles et commerciales* (Bodacc), qui assure la publicité des actes enregistrés au Registre du commerce et des sociétés (RCS), notamment les ventes et cessions, les immatriculations, les créations d'établissements, les procédures collectives, les radiations, et les annonces de dépôt des comptes des sociétés. Elle vend des licences au *Bulletin officiel des annonces de marchés publics* (BOAMP), dans lequel sont regroupées toutes les annonces de marchés publics. Elle vend enfin des licences d'accès au *Bulletin des annonces légales obligatoires* (BALO), où les sociétés cotées sont tenues de publier leurs résultats.

L'Insee, bien connu de tous les Français, regroupe avec ses services détachés dans les ministères, pas loin de 4 500 personnes. L'Insee gère et maintient notamment le fichier Sirene qui liste les quelque 10 millions d'entreprises françaises. Elle vend un accès à l'ensemble du fichier Sirene pour plusieurs centaines de milliers d'euros, le mettant ainsi hors de portée de toute PME qui désire faire des travaux statistiques sur les millions d'entités légales françaises. Et, cerise sur le gâteau, l'Insee ajoute une redevance à payer chaque fois que l'entreprise licenciée utilise un numéro Insee, un mécanisme complexe à mettre en place et à vérifier et se prêtant bien sûr à diverses violations.

Infogreffe n'est pas un établissement public, mais un GIE (groupement d'intérêt économique) rassemblant les greffiers des tribunaux de commerce, qui gère le RNCS (Registre national des commerces

et des sociétés) par délégation de service public. Le RNCS contient les informations disponibles sur des millions d'entreprises françaises. Comprenons bien la mécanique des années 2010 : les organismes publics ou assimilés sont alimentés en données par les entreprises, lesquelles ont une obligation de le faire et paient pour le dépôt des informations. Les mêmes organismes publics ou assimilés qui ont été payés pour recueillir et maintenir ces informations, les revendent à des entreprises grossistes, lesquelles les revendent à nouveau au détail aux acteurs qui en ont besoin, qui, parfois même, sont obligés d'y avoir accès. Pour prendre un exemple plus précis, le Kbis est la carte d'identité d'une entreprise, il est nécessaire à la vie de l'entreprise et exigé par l'État pour de nombreuses démarches. Il est constitué à partir de données fournies par l'entreprise à Infogreffe, l'entreprise payant alors un droit à Infogreffe. Ensuite, chaque fois que l'entreprise veut un Kbis, elle doit l'acheter à Infogreffe. À l'époque, le Kbis en version électronique était fourni par Infogreffe à un prix plus élevé que le Kbis papier envoyé par la poste !

Qui sont les « grossistes » du GFII ? Des entreprises de l'information juridique comme Lexis Nexis, ou Francis Lefebvre, des entreprises de l'information économique comme Altarès, Ellisphère, Société.com ou Bureau van Dick.

C'est cette merveilleuse rente de situation que va dynamiter partiellement l'ouverture des données en France au début des années 2010.

Entre 2010 et 2020, sous l'effet des efforts de lobbying de la communauté des données ouvertes, et sous l'action d'Etalab, un grand nombre de ces sources vont devenir gratuites et disponibles à tous. Dans cette décennie, le Sirene, le Bodacc, le BALO, le RNCS, le BOAMP passent en données ouvertes et deviennent disponibles gratuitement pour tous : entreprises, citoyens, journalistes, etc.

Les entreprises qui achetaient les données aux organismes publics et les revendaient à la découpe n'ont évidemment pas été ravies : leur quasi-monopole disparaissait et les données qu'elles faisaient payer

devenaient tout à coup accessibles à tous et gratuitement. Certains utilisateurs pouvaient maintenant se servir directement à la source et de nouveaux acteurs offrir des services sur ce marché, les barrières d'entrée ayant été levées. Ces entreprises ont bien sûr ferrailé contre le passage à l'ouvert, des organismes publics également en expliquant que le passage au gratuit se traduirait pour eux par un désastre, leur finance ne s'en remettant pas. Le plus bel exemple de cette bataille a été le cas du RNCS, l'État ayant longuement hésité pour décider qui allait être responsable de l'ouverture de ce fichier. Après une lutte acharnée en coulisse entre Infogreffe et l'INPI (Institut national de la propriété intellectuelle, en charge des brevets), c'est l'INPI qui a hérité du bébé. De fait, il a fallu plusieurs années pour que le transfert des données à l'INPI soit effectif et on peut même dire que le combat d'arrière-garde d'Infogreffe n'est pas complètement terminé.

Dans ce contexte, la mise en place de communs numériques a bouleversé des rentes de situation et leur a substitué un système plus efficace, plus équitable et profitant plus à l'écosystème et à la société.

## CHAPITRE 16

# Le patrimoine

Quelles sont les interactions entre patrimoine et communs numériques ? Nous distinguons dans ce chapitre deux facettes : 1) l'utilisation de communs pour préserver le patrimoine ; 2) la préservation des communs numériques considérés eux-mêmes comme valeurs patrimoniales.

### *La préservation du patrimoine*

Les monuments, les œuvres d'art, les livres et toutes formes de document, constituent entre autres notre patrimoine. Ils méritent donc à ce titre d'être préservés pour être disponibles à de futures générations. Cette préservation est indépendante de la nature de la propriété du bien : qu'il soit public ou privé, nous voyons un superbe moulin à eau du XVIII<sup>e</sup> siècle comme un bien commun à protéger. La préservation du patrimoine trouve son origine dans la création de la Bibliothèque nationale de France par François I<sup>er</sup> en 1537. Ce qu'on appelle alors la Bibliothèque royale doit servir de « dépôt légal » pour les copies de tous les livres et documents officiels publiés dans le royaume. La préservation des monuments existe aussi avec l'inscription ou le classement comme monument historique. De tels



mécanismes protecteurs de patrimoine matériel nous rapprochent du cadre des communs.

Le numérique participe à la préservation du patrimoine. Des États numérisent d'ailleurs déjà des pans entiers de leur patrimoine historique et culturel dans des archives comme Europeana, la bibliothèque numérique européenne. S'ils le font, c'est d'abord pour en faciliter l'accès à tous, mais la numérisation offre également une forme particulière d'immortalité. Une fois numérisé et placé dans une archive ouverte, un vieux parchemin ne disparaîtra pas, même si la bibliothèque qui l'héberge brûle. Le *data center* peut lui brûler, des mécanismes de réplication dans d'autres *data centers* sont alors supposés garantir la pérennité du parchemin.

Au-delà de l'archivage de telles informations numériques, il existe de nombreuses autres données qui méritent d'être préservées. Les algorithmes complexes et les programmes informatiques sont des produits de l'ingéniosité et de l'expérience humaines. Ils font partie de notre mémoire collective, de notre patrimoine et de notre histoire. La prise de conscience de la nécessité de la préservation du patrimoine logiciel est récente. Le projet Software Heritage a été lancé en 2015 sous l'égide d'Inria, maintenant avec le soutien de l'Unesco. Son objectif est de collecter, stocker et partager les codes logiciels afin de constituer une archive universelle et libre d'accès. En quelques années, ces archives ont déjà accumulé des milliards de fichiers, représentant un vaste trésor de l'ingéniosité humaine et de la résolution de problèmes. Software Heritage fonctionne un peu comme un commun avec sa communauté de contributeurs de logiciels, des scientifiques, des industriels, des sociétés savantes, des fondations, des organisations indépendantes et institutionnelles. Ce n'est pas par hasard qu'on trouve à son initiative Roberto di Cosmo, une figure du logiciel libre.

## *La préservation des communs numériques*

La pérennité des ressources construites par des communs numériques peut elle-même être questionnée. Une communauté peut s'épuiser, disparaître. Les ressources qu'elle construit peuvent alors se trouver en péril. Les informations, les logiciels, les données, etc., produits aujourd'hui par des communs numériques seront-ils disponibles dans cinquante ans, quand des chercheurs voudront étudier le monde dans lequel nous vivons aujourd'hui ?

Dans certains cas, les solutions globales de préservation de patrimoine, comme Europeana pour les livres, ou Software Heritage pour les logiciels, existent. Mais dans les autres cas ?

La question se pose pour le Web. En 1994, la Bibliothèque nationale du Canada a été l'un des pionniers de son archivage. Puis, en 1996, la fondation Internet Archive a lancé un système d'archivage mondial des pages Web. Le site Internet archive.org donne accès à la Wayback Machine. Le site est conçu pour remonter le temps, en suivant l'évolution des différentes versions de pages Internet. La création de telles archives nécessite des moyens informatiques importants. Ce droit d'archiver des connaissances peut se trouver en conflit d'autres droits, comme le droit d'auteur. Pour avoir scanné et prêté des copies numériques de livres protégés par le droit d'auteur, Internet Archive a été accusé par des éditeurs, Hachette Book Group, HarperCollins, John Wiley & Sons et Penguin Random House, de « violation massive des droits d'auteur ».

La question se pose aussi pour la pérennité d'autres communs numériques. Dans le cadre d'un projet de recherche, un groupe de chercheurs peut développer un commun, par exemple une base de données sur un moulin à eau du XVIII<sup>e</sup> siècle. Ils réunissent des photos, des plans, peut-être des données sur des simulations de son

fonctionnement à l'époque. Quand le projet s'achève, qui va stocker ces données, les préserver ? Le contrat de recherche qui a financé les différentes équipes a pris fin. On ne sait pas en général qui va payer pour leur préservation.

De telles questions se posent pour nombre de communs, y compris les plus importants. Qu'est-ce qui garantit, par exemple, que nous disposerons dans cinquante ans des données actuelles d'Open Food Facts ou d'OpenStreetMap ?



PARTIE D

# Les usages



Après avoir étudié les différents communs numériques, leur nature et leur fonctionnement, nous nous intéressons maintenant aux domaines dans lesquels ils sont utilisés. Il sera ici question, successivement, de science, de gouvernement, d'éducation, de santé, de justice, d'agriculture et d'entreprises.





## CHAPITRE 17

# La science

La déclinaison des communs pour ce qui concerne les sciences est la « science ouverte », un mouvement qui vise à rendre tous les produits de la recherche scientifique – publications, données, logiciels, protocoles, cahiers de laboratoires, etc. –, accessibles à tous, aux professionnels de la recherche comme aux amateurs du grand public. La science s’entend évidemment ici au sens le plus large, incluant donc les sciences sociales et les sciences humaines. La science s’est beaucoup transformée récemment. Elle est devenue plus numérique et plus collaborative dans des changements de cultures souvent liés au développement de communs numériques.

Quels sont les buts de la science ? Notamment, de mieux comprendre le monde dans lequel nous vivons et d’améliorer la vie des humains. Partager les fruits de leur travail avec toute la société semble donc bien naturel aux scientifiques. C’est en particulier vrai quand ils sont payés par l’État pour enseigner et faire de la recherche, le cas d’un grand nombre d’entre eux aujourd’hui en France. Mais pourquoi les scientifiques publient-ils leurs résultats ?

Ils le font d’abord pour apporter leur contribution, même modeste, au progrès de l’humanité. En s’appuyant sur les résultats d’un scientifique ou d’une équipe, d’autres scientifiques peuvent faire à leur tour avancer la science, peut-être développer des techniques

qui vont améliorer l'agriculture, l'industrie, les transports, la vie de chacun. Quand Isaac Newton déclare : « Si j'ai vu si loin, c'est que j'étais monté sur des épaules de géants », il résume bien le fait que les progrès de la science se basent sur la science passée. En rendant publics des résultats positifs de son travail de recherche, on permet à d'autres de les vérifier, de les améliorer, voire de se servir des idées exposées pour innover peut-être dans d'autres domaines. En rendant public un échec, on évite à d'autres de faire les mêmes erreurs. Dans tous les cas, on permet aux sciences d'aller plus loin avec les mêmes ressources. C'est pour cela qu'on « ouvre » la science, qu'on la rend plus accessible, pour optimiser les effets des ressources qu'on lui consacre.

Ces idées de partage, de collaboration, dans les sciences peuvent d'ailleurs se prolonger dans l'innovation, avec l'innovation ouverte, qui consiste pour une entreprise à intégrer dans sa démarche d'innovation des contributions extérieures venant de la recherche publique ou d'autres partenaires industriels.

Les scientifiques publient également leurs résultats pour obtenir la reconnaissance de leurs pairs, peut-être la gloire. Si Pasteur est si populaire, c'est parce qu'il a expliqué ses résultats et convaincu ses pairs de leur justesse et de leur importance. Un scientifique plus modeste publie également ses résultats pour faire reconnaître leur valeur. Il est difficile d'accorder à quelqu'un du crédit sur des résultats restés confidentiels. Par exemple, en 1976, W. Diffie et M. Hellman ont introduit la notion de cryptographie à clé publique, et, en 1978, R. Rivest, A. Shamir et L. Adleman ont proposé une réalisation pratique du système de Diffie et Hellman. Tous resteront dans l'histoire comme les inventeurs d'une technologie massivement utilisée dans les échanges sur Internet, la cryptographie à clé publique. Pourtant, d'autres chercheurs, J. Ellis, C. Cocks, et M. J. Williamson auraient obtenu au préalable des résultats semblables. Mais comme ils travaillaient pour les services du chiffre britannique, leur travail resta secret et n'eut que peu d'impact.

## *L'histoire de la science ouverte*

L'idée de faire de la recherche scientifique au service de l'humanité, évidente aujourd'hui pour une chercheuse du CNRS ou un professeur d'université, l'était moins pour un scientifique du Moyen Âge. À cette époque, les sciences relevaient davantage du hobby et les scientifiques devaient être riches de naissance. Sinon, il leur fallait rentabiliser leurs résultats pour vivre, par exemple, en se trouvant de riches mécènes. Les mécènes finançaient des artistes, des architectes ou des scientifiques pour leur propre gloire. Ce sont eux qui décidaient du secret des résultats, ou pas. Encore aujourd'hui, une entreprise qui subventionne des recherches peut vouloir pour un temps garder certains résultats confidentiels afin d'en tirer des profits. Les États eux-mêmes peuvent préférer que les résultats de la recherche publique servent en priorité aux entreprises nationales.

Le monde scientifique se transforme au XVII<sup>e</sup> siècle avec la naissance des académies et des premiers journaux scientifiques. C'est à cette époque que naît la science moderne, basée sur plus de partage, peut-être parce que le corpus scientifique est devenu si massif qu'il faut repenser les transferts de connaissances. L'ouverture est au départ exclusivement à destination d'autres scientifiques. Au XVIII<sup>e</sup> siècle, les échanges s'intensifient entre les savants européens. En 1872, aux États-Unis, *Popular Science*, la première revue scientifique à destination des non-professionnels éclairés, voit le jour. Au XX<sup>e</sup> siècle, le nombre de scientifiques explose et rend encore plus indispensable le partage. Les bibliothèques deviennent numériques et mondiales. Vers la fin du XIX<sup>e</sup> siècle, et surtout au début du XX<sup>e</sup>, le concept moderne de « science ouverte » prend son essor même s'il s'inscrit en véritable continuité de l'idée de diffusion des connaissances qui avait conduit à la création de journaux scientifiques.

En décembre 2001, un groupe de 16 scientifiques se rencontre à Budapest à l'invitation de l'Open Society Institute (qui deviendra Open Society Foundations). Les discussions sur les publications scientifiques sont passionnées et souvent divergentes. Mais le groupe se met d'accord sur l'idée d'un communiqué conjoint, qui paraîtra en février 2002. L'initiative de Budapest pour l'accès ouvert déclare la nécessité de faciliter l'échange et la valorisation d'archives numériques. On voit se mettre en place des bases de données bibliographiques. Pour les publications ouvertes, la publication devient à portée de clic. Deux stratégies complémentaires émergent :

- l'auteur publie dans des journaux en libre accès ;
- l'auteur met en ligne ses propres articles dans des archives. Quand c'est avant leur publication dans un journal, on parle de *preprint* ou prépublication en bon français ; *a posteriori* ce sont des *post-prints*.

Aujourd'hui, le site Sci-Hub (et des sites miroirs) mis en place et animé par Alexandra Elbakian donne librement accès à la quasi-totalité des articles scientifiques. Même s'il est dans l'esprit de Budapest, ce site est considéré par certains, surtout des éditeurs scientifiques qui lui font des procès, comme pirate.

Au XXI<sup>e</sup> siècle, la science ouverte s'invite dans les agendas politiques. Les gouvernements commencent à questionner les modèles des éditeurs scientifiques. Comme le déclare le biologiste Michael Eisen, fervent soutien de la science ouverte : les contribuables qui ont déjà payé pour la recherche ne devraient pas payer à nouveau pour lire les résultats. Le mouvement pour la science ouverte est de plus en plus actif aux États-Unis. L'Europe définit son programme en 2018. C'est à l'origine de l'European Open Science Cloud (EOSC), un environnement pour héberger, traiter et partager les programmes et les données de la recherche. Des programmes ambitieux, Horizon 2020 (2014-2020) et Horizon Europe (2021-2027), structurent le monde scientifique européen. Le premier concernait

principalement l'accès ouvert aux articles scientifiques. Le second définit une véritable vision de la science ouverte pour l'Europe. La notion de données FAIR est adoptée. L'acronyme vient d'exigence sur les données qui se doivent d'être « faciles à trouver, accessibles, interopérables et réutilisables », en anglais *findable, accessible, interoperable, and reusable*.

Dans ce contexte, la France est motrice. La science ouverte est inscrite par exemple dans la loi « pour une république numérique » d'Axelle Lemaire en 2017. Les chercheurs financés sur fonds publics peuvent mettre leurs articles en accès libre, même s'ils ne sont pas encore tenus de le faire. Cette obligation sera introduite par la suite au niveau européen pour ceux qui bénéficient de subventions européennes. Depuis, les différents ministères français de l'Enseignement supérieur, de la Recherche et de l'Innovation ont poursuivi une politique ambitieuse de science ouverte. La science ouverte est devenue un socle essentiel de la recherche française.

### *L'accès ouvert aux publications scientifiques*

Le premier combat pour la science ouverte s'est joué sur les publications. Au bout du chemin, le but était clair : l'accès ouvert et gratuit de tous à toutes les publications scientifiques n'importe où dans le monde.

Cet accès ouvert était le rêve de nombreuses personnes. Aaron Swartz a ouvert la voie, ce qui l'a conduit à la mort. Alexandra Elbakyan a en partie réalisé ce rêve même si cela l'a entraînée dans des démêlés avec certaines justices.

**Aaron Swartz** (1986-2013) était un informaticien et hacktiviste américain qui a contribué au développement du format de flux Web RSS. Il est aussi l'un des fondateurs de l'organisation Creative Commons. Il est connu pour ses prises de position militantes en faveur de la liberté sur Internet, et pour son combat contre le *Stop Online Piracy Act*. Il a été poursuivi par la justice pour avoir téléchargé des articles de revues universitaires. En 2013, Aaron Swartz est retrouvé mort par suicide dans son appartement de Brooklyn. Il n'a pas trente ans. Son destin tragique fait de lui un des héros des libertés sur Internet.

**Alexandra Elbakyan** est une informaticienne kazakhe, créatrice du site Web Sci-Hub, qui, au mépris de tout copyright, offre un accès ouvert et gratuit à une énorme masse de documents scientifiques. Elle a démarré ce projet parce qu'elle trouvait inacceptable qu'un chercheur ne puisse avoir accès à des publications qui auraient permis de faire avancer la science.

Les chiffres suivants (fournis par elle) donneront une idée de son succès. Sci-Hub compte environ 1 million d'utilisateurs par mois depuis la Chine, 1 million d'Europe, et 250 000 depuis l'Afrique. Et c'est sans compter les sites « miroirs » qui ont suivi son mouvement. Elle est parfois surnommée « la reine des pirates de la science ». Admirée par certains, attaquée en justice par d'autres, Alexandra Elbakyan a clairement fait bouger les lignes. Son action a permis de prendre conscience qu'il faut trouver des solutions pour ouvrir la science à tous.

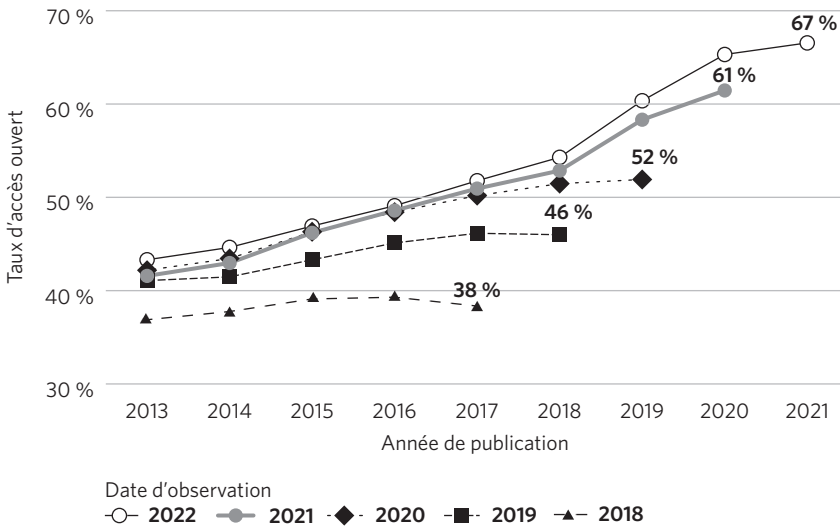
L'action radicale d'Alexandra Elbakyan conduit à s'interroger sur le modèle classique des publications, un héritage d'une autre époque. Dans le temps, le scientifique produisait un manuscrit, le plus souvent rédigé à la main, puis tapé à la machine. Le rôle de l'éditeur était important : le mettre en forme, l'imprimer et le diffuser. Aujourd'hui, la mise en forme est souvent réalisée par les auteurs

eux-mêmes, et la diffusion se fait le plus souvent numériquement sur le Web, rendant caduque la question de l'impression. On voit bien comment le travail de l'éditeur de journaux scientifiques s'est considérablement amoindri. Les scientifiques font des recherches, rédigent les articles, les mettent en forme, les sélectionnent (avec le mécanisme d'évaluation par les pairs) et, en bout de chaîne, les lisent. Les éditeurs se contentent parfois de les mettre sur le Web, de garantir leur pérennité (un rôle non négligeable), et de faire payer pour leur accès. On aurait pu penser que ce changement de contexte ferait diminuer les prix des revues et les revenus des éditeurs. C'est le contraire qui s'est produit : leur marge et leurs revenus ont augmenté.

L'Académie des sciences, dans ses recommandations pour une mise en pratique des principes de la science ouverte en 2022, souligne les questions que soulève le modèle économique actuel. Quatre éditeurs se partagent encore, en 2018, plus de 50 % du marché de l'édition scientifique. Il est aberrant de voir l'importance des marges d'éditeurs, près de 40 %, au regard de leur modeste contribution. Ce système fait porter des coûts énormes aux laboratoires de recherche.

Sous la pression de scientifiques promoteurs de la science ouverte, et de laboratoires qui veulent à la fois contrôler leurs coûts et garantir l'accès de toutes les publications à leurs scientifiques, le domaine des publications scientifiques est en train de bouger malgré les résistances d'éditeurs qui veulent continuer à faire des profits importants. Nous avons vu en partie C différents modèles pour la publication scientifique – doré, diamant, platine, vert –, qui se distinguent principalement sur la question de décider qui doit payer pour le fonctionnement du système.

Le mouvement de l'accès ouvert aux publications scientifiques semble irrésistible, au moins en France et en Europe. Pour ce qui est de la France, le Baromètre de la science ouverte montre une évolution du taux d'accès impressionnante : le pourcentage de publications disponible en accès ouvert passe de 47 % en 2013 à 67 % en 2021, comme montré dans le graphique ci-dessous.



Source : Baromètre français de la science ouverte (CC BY MERS).

**Figure 4.** Évolution du taux d'accès ouvert des publications scientifiques françaises, par année d'observation.

Ce taux dépend beaucoup de la discipline. S'il est à près de 75 % en 2020 pour les mathématiques qui ont une longue tradition d'accès ouvert, il est bien plus bas en sciences humaines et sociales où l'importance des maisons d'éditions reste prépondérante. Tout en célébrant une nette amélioration, on peut quand même regretter que la France soit moins en pointe que d'autres pays d'Europe comme l'Allemagne ou l'Angleterre même si elle semble rattraper son retard.

De plus en plus, avant la date de publication dans une revue, les articles sont publiés sur des sites spécialisés sous forme de *preprint*. Si ce mouvement a commencé bien plus tôt, il est devenu massif dans les années 1990 avec des bases de données comme arXiv et HAL. Grâce aux *preprints*, les résultats scientifiques sont ainsi partagés beaucoup plus rapidement. On a pu assister de ce fait à un considérable échange de connaissances dans le cas de la pandémie du Covid-19 qui a peut-être fait gagner du temps à la science, mais



qui a aussi malheureusement ajouté des pièces dans la machine à *fake news*. Le problème est qu'un *preprint* n'est pas encore passé à travers le mécanisme de l'évaluation par les pairs. Cette évaluation peut prendre des mois, voire des années pour améliorer un article, le transformer en « vraie » publication, quand il n'est pas tout simplement refusé. Le *preprint* peut inclure des erreurs ; il se peut même qu'il ne soit qu'un outil de désinformation. S'ils contribuent énormément à la transmission des connaissances, les *preprints* présentent des risques avec lesquels il faut apprendre à vivre.

### *Publish or perish*

Les scientifiques publient leurs résultats sous forme d'articles publiés dans des journaux, des actes de conférences, des chapitres de livres, etc. Les coutumes varient énormément d'une discipline à l'autre. Dans certaines, seules les publications dans des revues « comptent » ; pour d'autres, des conférences peuvent être plus sélectives que des journaux.

Ces publications sont une pièce essentielle de la carrière du ou de la scientifique. Elles servent de base à son évaluation pour obtenir un doctorat, des crédits de recherche, un poste temporaire ou permanent, une promotion. Cette évaluation avait, dans le temps, un côté artisanal. On prenait le temps de se familiariser avec le travail de la personne évaluée. Mais le nombre de scientifiques a explosé, la compétition est devenue féroce et les choix si complexes que les évaluateurs s'en remettent trop souvent au simplisme de comptages bibliométriques comme l'indice  $h^1$  dont les limites sont connues. Le scientifique ambitieux doit se plier au jeu et publier énormément, dans des revues dites de prestige. Cette frénésie de publications

---

1. Ou  $h$ -index en anglais. D'après son inventeur, Jorge Hirsch : « Un scientifique a un indice  $h$  si  $h$  de [ses]  $N$  articles ont chacun au moins  $h$  citations, et les autres ( $N - h$ ) articles ont au plus  $h$  citations chacun. »

est-elle au bénéfice de la science ? C'est discutable parce qu'elle se focalise plus sur la quantité que sur la qualité.

Favorise-t-elle au moins la science ouverte ? Pas vraiment, parce que les revues de prestige sont historiquement en accès fermé. Et puis, même si les choses sont en train de changer, les évaluateurs ne tiennent pas encore assez compte dans leur évaluation des autres produits de la recherche comme les logiciels développés ou les corpus de données construits. Avec la déclaration DORA (Declaration on Research Assessment), des chercheurs et des institutions de recherche s'efforcent de reprendre la main sur l'évaluation des chercheurs et de la recherche, en évitant de donner trop de poids à des métriques calculées automatiquement comme le « facteur d'impact » des journaux scientifiques et en analysant en profondeur la qualité des publications. (Le facteur d'impact d'une revue est un indicateur de sa notoriété ; il est calculé par exemple en divisant le nombre de citations reçues par une revue dans une année et le nombre d'articles publiés par cette revue au cours des deux années précédentes.)

La méthode d'évaluation classique consiste en une évaluation par des pairs, des experts du domaine, en double aveugle (*double blind peer reviewing*). Un éditeur choisit en fonction du contenu d'un article des évaluateurs, des « relecteurs », qui lui remettent en un ou plusieurs tours, des évaluations, avec des décisions de publication ou de rejet, et d'éventuelles demandes de modification, parfois substantielles. Les relecteurs ne connaissent pas l'identité des auteurs, et ces derniers ignorent qui effectue l'évaluation, d'où le terme « double aveugle ». Ces anonymats croisés sont là pour protéger le processus d'évaluation des biais de copinage ou d'inimitié, pour éviter de trop pénaliser les jeunes chercheurs sans réseaux, avec une réputation à construire. Le processus peut prendre du temps, des mois, voire des années. Pendant ce temps, les résultats de l'article ne sont accessibles qu'à de rares privilégiés.

Ce type d'évaluation peut conduire à des frustrations du fait de la lourdeur du dialogue entre évaluateurs et évalués, ou entre évaluateurs qui peuvent avoir des avis divergents, voire à des agressions verbales des évalués sous la protection de l'anonymat des évaluateurs. Un des auteurs de cet essai a vu dans un rapport le commentaire suivant : « Il faut être vraiment débile pour proposer cet algorithme. » Une telle « opinion » peut être dévastatrice pour un jeune chercheur. Des efforts sont réalisés pour corriger ces abus heureusement rares.

Une proposition plus radicale consiste à ouvrir l'évaluation. L'article est immédiatement disponible avec une « évaluation par les pairs ouverte ». Tout devient transparent : n'importe qui peut se proposer pour évaluer un article, les auteurs dialoguent publiquement avec les évaluateurs.

Une initiative de la Commission européenne, Open Research Europe, permet aux chercheurs de certains programmes européens de publier les résultats de leur recherche qu'ils souhaitent partager, favorisant la reproductibilité, la transparence et l'impact. Les articles sont très rapidement disponibles en accès ouvert. L'évaluation par des pairs est ouverte. On peut imaginer le confort scientifique d'une telle approche. Les résultats sont immédiatement disponibles même si le site n'a pas le prestige des revues bien établies.

Une limite de ce modèle est que la publication est restreinte aux travaux de ces programmes. On peut aussi questionner l'évaluation ouverte. Ne risque-t-on pas de réintroduire des travers considérables que l'évaluation en double aveugle a cherché à éliminer : les rapports de complaisance et ceux qui servent de règlements de compte ?

On peut pousser cette idée d'évaluation par les pairs ouverte à tous types de contenus. C'est ce que propose un logiciel libre, Hypothes.is. L'idée est de pouvoir accueillir des commentaires sur tout contenu du Web. Les internautes donnent un avis, et le système va jusqu'à permettre d'évaluer les commentaires. Le Web propose une masse de connaissances considérable, mais c'est au milieu d'une

masse de contenus-poubelle tout aussi considérable. En s'appuyant sur de tels outils, en définissant ses propres règles, une communauté d'internautes ne pourrait-elle nous aider à séparer le bon grain de l'ivraie ? Cela constituerait la base d'une véritable alternative aux systèmes de recommandation actuels. N'est-ce pas un beau défi pour les communs numériques ?

### *Les données ouvertes*

Les données ont de tout temps tenu une place centrale dans la science. Les premiers astronomes se servaient des données d'observation accumulées pour décider, par exemple, quand planter des cultures. La sophistication des sciences a accéléré ce besoin de données, notamment, de nouvelles techniques comme l'analyse de données massives (*big data*) et l'apprentissage automatique (*machine learning*). Les données ont pris une importance considérable dans de nombreux domaines scientifiques, au point que le simple fait de construire un corpus de données peut devenir de fait un projet scientifique.

Le Human Genome Project a été construit sur le principe des données ouvertes : « Toutes les informations sur les séquences génomiques humaines devraient être librement disponibles et dans le domaine public afin d'encourager la recherche et le développement et de maximiser ses avantages pour la société. » On est pile-poil dans la science ouverte.

L'impossibilité d'accéder à des données propriétaires peut interdire à un scientifique d'arriver à des résultats. Dans un tel cas, la fermeture des données est un frein au progrès scientifique. Les astronomes ont compris cela assez tôt, sans doute parce que le partage était plus essentiel dans leur discipline que dans d'autres. Les physiciens aussi, et ce n'est pas un hasard si le Web est né dans un laboratoire de physique, le CERN. La volonté d'échanger des

connaissances était la première motivation. Du coup, ces communautés ont pris rapidement l'habitude de l'accès ouvert. Dans leurs disciplines, on a l'habitude de mieux documenter, mieux structurer ses données, de les partager, dans l'optique que d'autres puissent les utiliser.

De fait, de nombreux organismes, universités ou labos de recherche dans différents domaines, ont commencé à s'organiser pour mettre en place des centres de stockage de corpus de données. Il leur fallait aussi jouer l'accompagnement des scientifiques dans des tâches dont ces derniers n'étaient pas familiers. En s'appuyant sur une plateforme développée par l'Inrae, le ministère de l'Enseignement supérieur et de la Recherche a lancé la plateforme Recherche Data Gouv en juillet 2022 qui offre à tous les scientifiques un cadre pour partager des corpus de données validés. Cette plateforme ouverte à tous devrait encourager le partage d'expérience, d'outils logiciels, d'améliorer l'accompagnement des scientifiques, de leur faciliter la vie quand ils changent d'organisme.

### *Accès ouvert à tous les produits de la science*

Le numérique ne s'est pas contenté de transformer la production et le partage de publications scientifiques et de données, il a aussi rendu possible le partage de *tous* les produits de la recherche : les cahiers d'expérience, les protocoles, jusqu'à ces nouveaux objets de la science que sont les logiciels. Si le scientifique rend accessible les détails de ses travaux, sa recherche peut devenir si transparente qu'elle devient « reproductible ». Un autre scientifique peut s'emparer d'un processus particulier, décider de le reproduire pour en vérifier les résultats, ou peut-être juste pour apprendre de l'exercice.

Le résultat d'une équipe de recherche peut en particulier être du logiciel. L'Inria est à l'origine de nombreux projets majeurs de logiciels libres, entre autres : OCaml, une version très populaire du

langage de programmation caml, et Scikit-learn, une bibliothèque logicielle de référence en *machine learning*. Dans de tels cas, le logiciel devient lui-même un outil de transmission des connaissances. Mais il est également un lieu de collaboration. Le projet OCaml, par exemple, inclut des collaborateurs de plusieurs équipes de recherche qui constituent ainsi une communauté contribuant collectivement au projet. C'est véritablement un commun numérique en action avec son produit, sa communauté, sa gouvernance, ses règles.

La collaboration autour de logiciels est évidente pour une équipe de recherche en informatique. Mais elle est tout aussi essentielle en physique, biologie, etc. De fait, on peut embarquer des années de résultats de recherche dans du code informatique. Le code sert alors à expliquer de manière précise les résultats, et permet à une communauté de se les approprier si elle le souhaite. En particulier, dans de nombreux domaines, les logiciels sont à la base de simulations qui font avancer la science : des simulations du climat, d'un organe comme le cœur, d'un véhicule, etc. De plus en plus d'articles scientifiques dans toutes les sciences mentionnent des logiciels développés par des physiciens, des biologistes, des géographes, des sociologues, etc.

Quand on parle de tels partages, on pense d'abord aux succès pour que d'autres puissent les utiliser, construire à partir d'eux. Mais il est important de partager aussi les échecs. En partageant une expérience qui n'a pas donné les résultats escomptés, on évite peut-être à d'autres d'investir du temps et des ressources pour rencontrer ce même échec. Il est légitime d'avoir l'idée d'expérimenter un traitement à base d'hydroxychloroquine pour lutter contre le Covid 19. Le fait que cela ne donne pas de résultats probants est une information utile qu'il faut diffuser pour éviter que d'autres se lancent dans la même direction. Aux États-Unis comme en Europe, on se dirige vers l'obligation de rendre public tout essai clinique dès son lancement.

L'ouverture permet aussi de découvrir plus rapidement des erreurs dans un raisonnement, une preuve, un logiciel, des données, et en cela, elle est également indispensable au progrès scientifique. Et là, on ne vise pas seulement les escrocs qui existent dans le monde scientifique comme ailleurs, mais aussi le scientifique qui s'est tout simplement trompé, qui a commis une erreur.

### *Les freins*

Il existe un environnement qui peut être défavorable à l'ouverture de la science : la recherche financée par le privé. Le scientifique peut se trouver dans le laboratoire d'une entreprise privée, ou dans un laboratoire public sous contrat avec une telle entreprise. Sa liberté de publier ses résultats peut se voir alors limitée par les intérêts de l'entreprise.

Quelles peuvent bien être les raisons pour une entreprise de subventionner de la recherche ? Elle peut le faire pour des raisons de responsabilité sociale ou pour sa réputation. Elle aura alors sans doute à cœur de rendre public les résultats. Mais elle le fait souvent, ce qui est dans son ADN, pour développer un nouveau service ou produit et va exiger, au moins pour un temps, un embargo sur les résultats pour garder ainsi un avantage technique sur la concurrence.

Pour ce qui est du cas des données, un frein spécifique à leur ouverture est clairement la protection de la vie privée. Certaines données médicales et biologiques sont par exemple extrêmement sensibles. On peut évidemment utiliser des techniques de « pseudo-nymisation », mais leurs limites sont connues. Pour de telles données, on préfère à l'ouverture la mise en place de centres de données protégés où des analyses de données peuvent être réalisées sans faire courir de risque à leur caractère confidentiel.

Un dernier frein enfin concerne le scientifique lui-même. En ouvrant sa recherche, on la rend publique et on en tire la reconnaissance de ses pairs qui vont pouvoir peut-être s'en servir comme socle pour de futurs travaux. Mais, on court aussi le risque de se faire « doubler ». En médecine, on peut déployer énormément d'efforts à construire un corpus de données cliniques et, en les mettant en accès ouvert, voir d'autres, plus rapides, en tirer des résultats. On touche ici à la difficile tâche d'attribuer correctement les résultats scientifiques. Francis Crick, James Watson et Maurice Wilkins ont obtenu en 1962 le prix Nobel de physiologie ou médecine pour leur découverte de la structure à double hélice de l'ADN. Or leurs travaux devaient sans doute beaucoup à Rosalind Franklin qui n'a pas eu la reconnaissance qu'elle méritait.

Certains chercheurs développent donc l'habitude de cacher leurs données et leurs résultats, et ce pour éviter que d'autres qu'eux n'en tirent le fruit. Pour l'humanité, qu'un résultat soit attribué à Pierre ou Jeanne, cela n'a guère d'importance. Pour Pierre ou Jeanne, cela en a et cela conduit à une moins grande ouverture des produits de la recherche.

### *La science pour tous*

La science ouverte est d'abord destinée à la communauté scientifique. Il s'agit de mettre les produits de la science à disposition de tous les chercheurs. Mais la science ouverte est aussi l'occasion de partager les résultats avec l'ensemble de la population. Le monde moderne est technique, complexe. Les citoyens n'y comprennent souvent pas grand-chose. Est-ce que les apports de la 5G sont suffisants pour accepter son empreinte environnementale ? Est-ce que les avantages des vaccins pour le Covid-19 contrebalancent des effets peut-être négatifs ? Les questions sont tout sauf simples et pour



avoir des avis éclairés, les citoyens doivent monter en compétences scientifiques. Pour cela, il faut leur ouvrir les sciences.

On ouvre l'accès des revues scientifiques aux amateurs qui se passionnent pour un sujet et sont suffisamment éclairés pour comprendre les articles qui le traitent. Mais ces revues sont le plus souvent hors de portée du grand public. C'est pourquoi il appartient aux scientifiques d'expliquer à tous leurs résultats, de jouer le rôle de « passeur de sciences » à destination de ce grand public, comme cela se fait, par exemple, avec la Fondation La main à la pâte ou la Fondation Blaise Pascal.

La participation d'amateurs aux sciences prend des formes surprenantes.

- Le projet participatif Galaxy Zoo invite les personnes intéressées à donner un coup de main à la classification morphologique d'un grand nombre de galaxies. L'amateur devient chercheur et le résultat du travail de tous est en accès ouvert bien sûr.
- Pl@ntnet permet d'identifier des plantes en utilisant une banque d'images annotées construite avec l'aide de la foule.
- Dans un autre projet, Folding@home, on peut participer à la recherche en mettant la puissance de calcul de son ordinateur personnel à la disposition d'un projet de recherche médicale visant à simuler le repliement des protéines dans diverses configurations de température et de pression.

On imagine (on aimerait être convaincu) que le mouvement de la science ouverte est inéluctable, que bientôt 100 % des publications scientifiques seront en accès ouvert, que les autres produits de la recherche scientifique suivront. Les canaux de publications de la science, notre façon de faire la science, de partager cette science avec toute la société, devraient en être profondément transformés.

Le sujet du prochain chapitre, l'éducation ouverte, pourrait être considéré comme une autre facette de la science ouverte. Nous avons choisi, compte tenu de son importance, d'en faire un chapitre distinct.

**Human Genome Project (Projet Génome humain)** a établi un séquençage de 92 % de l'ADN du génome humain entre 1988 et 2003. Le résultat est disponible sur Internet. Le séquençage complet (à 0,3 % près) attendra 2021.

Une entreprise privée, Celera Genomics, s'est donné à l'époque le même défi en s'appuyant sur des techniques plus robotisées. Celera visait à vendre ses résultats à l'industrie pharmaceutique. La compétition s'est terminée par un match nul. Le projet de Celera pose la question de savoir si le patrimoine génétique humain ne devrait pas être considéré comme un commun. Si c'est le cas, une entreprise privée ne devrait pas avoir le droit de garder pour elle des données comme le séquençage du génome humain. Celera a aussi reconnu avoir utilisé des données du projet Génome humain pour compléter son séquençage. Encore un exemple d'entreprise privée utilisant un commun sans y contribuer !

## CHAPITRE 18

# L'éducation

Le premier rôle d'un enseignant est de partager son savoir avec ses élèves ou ses étudiants. Le partage est véritablement à la base de l'éducation. Ce savoir pourrait paraître une ressource non rivale : le fait qu'un enseignant expose son savoir à un groupe d'élèves ne lui enlève rien de ce savoir. Pourtant, à l'époque prénumérique, la rivalité existait du fait du temps dont disposait l'enseignant. Il pouvait doubler la taille de l'amphi ou de la classe, répéter un même cours ou le même TD plusieurs fois, son temps était une ressource limitée, donc rivale. En outre, les manuels scolaires et autres objets pédagogiques étaient alors des biens rivaux.

La puissance du numérique permet aujourd'hui de dispenser le savoir massivement en bénéficiant de la non-rivalité des ressources numériques. En transformant potentiellement la transmission du savoir en un bien non rival, le numérique a véritablement révolutionné l'éducation. Rien ne s'oppose à des amphis virtuels de milliers d'élèves partageant les mêmes ressources éducatives numériques (et d'ailleurs cela existe). Rien ne s'oppose à ce qu'élèves et enseignants se retrouvent sur Internet dans des forums ou dans de nouvelles interactions, certaines encore à inventer. Si cela ne peut se substituer bien sûr au contact direct entre maître et élève, cela peut le compléter, voire l'enrichir.

Avec le numérique, nous disposons d'une nouvelle technologie hyperpuissante à mettre au service de l'éducation. Cette technologie arrive au bon moment dans un monde où la demande pour plus d'éducation est massive, en particulier, dans des pays en développement dont le personnel éducatif et les moyens disponibles pour l'enseignement sont limités. Avec le numérique, ces pays pourraient avoir accès à un enseignement riche et varié de leur population, pour un coût raisonnable. Ce n'est pas un hasard si l'expression « ressources éducatives libres » a été adoptée lors du Forum de l'Unesco de 2002 sur l'impact des didacticiels libres pour l'enseignement supérieur dans les pays en développement.

Comme nous l'avons vu dans le chapitre sur la science ouverte, des chercheurs sortent de leurs labos pour expliquer leurs résultats aux étudiants, mais aussi au grand public. C'est une forme d'éducation ouverte. L'idée de mettre dans l'éducation, en général, l'accent sur l'ouverture est naturel. Mais le mot « ouverture » est polysémique. En particulier, dans le contexte de l'éducation, on peut l'entendre de deux manières :

- l'éducation peut être ouverte à tous, les classes le plus accessibles possibles à tous points de vue : sociétal, géographique, générationnel, etc.
- des ressources éducatives peuvent être ouvertes ou libres, les contenus sont produits de manière collaborative, et leur accès est garanti pour tous.

Ce sont ces deux aspects que nous considérons dans la suite de ce chapitre.

## *L'éducation ouverte et les MOOC*

L'ouverture ou la liberté dans l'éducation peuvent qualifier tout un ensemble de pratiques pour enseigner autrement, et une culture orientée autour de la liberté de l'enseignant. Dans ce modèle,

l'éducation sort de l'école et des programmes pour se retrouver dans des lieux comme les fablabs ou les tiers lieux. À travers des projets, enseignants et élèves décloisonnent les disciplines.

**Le projet MineStory** est une frise immersive d'histoire des arts vise à la modélisation du patrimoine architectural français.

Des enseignants du primaire modélisent avec leurs élèves des édifices du patrimoine architectural proche de leurs écoles. Grâce à des logiciels, ils les recréent virtuellement. Ils le font avec l'aide des services de la mairie et d'architectes. Les élèves apprennent en s'amusant mathématiques et histoire de l'art. Ils sortent de l'école, physiquement, mais ils sortent aussi des programmes. Surtout, ils collaborent pour atteindre un objectif concret, ils partagent. Avec MineTest, un clone libre de MineCraft, ils créent une ressource libre, une superbe frise immersive d'histoire de l'art, dans une démarche qui tient elle-même de la liberté de l'éducation. Génial !

Pour certains, l'ouverture peut aussi signifier une autre façon d'enseigner plus orientée sur la participation des élèves, les projets, le dialogue. Cela peut impliquer d'être particulièrement clair avec les étudiants sur les objectifs, les méthodes éducatives mises en avant. Cela peut aussi inclure des schémas d'évaluation plus transparents.

Avec l'éducation ouverte, on insiste aussi pour que l'accès à l'éducation ne soit pas limité aux étudiants inscrits au cours, mais soit ouvert à tout public qui pourrait être intéressé. On supprime également l'exigence de qualification préalable pour suivre l'enseignement, et, plus généralement, toutes les barrières à l'entrée de l'éducation. Ce mouvement a démarré avec des établissements comme l'Open University en Grande-Bretagne. Les universités du temps libre ou du troisième âge tiennent d'une philosophie assez proche. Ce sont des associations dont les membres cherchent à

partager, à acquérir des savoirs sans que cela ne débouche forcément sur un diplôme.

L'éducation ouverte a trouvé récemment un essor considérable avec les MOOC, les *massive open online course*, qu'on traduit parfois en français par FLOT, pour « Formation en ligne ouverte à tous ». Ce sont des formations sur Internet ouvertes à tous. Elles se situent dans la continuité de l'enseignement à distance dont un fleuron en France est le Centre national d'enseignement à distance (CNED).

Un MOOC est diffusé sur Internet. Des élèves, d'une poignée à des dizaines de milliers, le suivent en ligne, peut-être partout dans le monde. L'inscription au MOOC est gratuite. Parfois, l'élève peut avoir à payer pour recevoir un certificat. Un MOOC offre des lectures, des quiz, des exercices, parfois des projets. Des interactions entre les élèves, les enseignants, des assistants d'enseignements sont proposées via des forums ou des réseaux sociaux.

Les précurseurs états-uniens ont construit des entreprises comme Coursera et Udacity ou des organisations à but non lucratif comme edX. Depuis 2013, le gouvernement français propose des MOOC sur France université numérique (FUN). La plateforme tourne sur le logiciel *open source*, *open edX*, développé par un consortium fondé par MIT et l'Université Harvard. Fin 2021, la plateforme FUN avait déjà proposé plus de 700 cours, le plus souvent en français. D'autres acteurs comme Coursera, OpenClassRoom ou Sillages offrent aussi des MOOC francophones.

Deux MOOC sur les bases de données sont proposés par le premier auteur de ce livre, Serge Abiteboul, dans le cadre de FUN :

- bases de données relationnelles : apprendre pour utiliser, avec Philippe Rigaux et Benjamin Nguyen ;
- bases de données relationnelles : comprendre pour maîtriser, avec Philippe Rigaux.

La préparation de chacun de ces MOOC a demandé un volume de travail largement supérieur à celui d'un cours classique. L'animation du forum à chaque session annuelle demande aussi du travail. Ces cours ont été suivis par des milliers d'élèves, ce que leurs auteurs n'avaient jamais connu avant. Et puis, et c'est une particularité de FUN du fait de la francophonie, le cours a été beaucoup suivi en Afrique. Il est difficile, par contre, avec cette plateforme, d'évaluer le niveau atteint par les élèves.

La pédagogie ouverte est poussée encore plus loin par les cMOOC où le « c » insiste sur la connectivité entre des participants qui construisent et partagent collectivement des connaissances.

En allant plus loin, on peut se passer, en partie, d'enseignants. Cela permet de s'attaquer à l'un des plus grands problèmes de l'éducation, le manque d'enseignants dans les régions où ils seraient le plus indispensables. L'une des réalisations les plus impressionnantes dans cette direction est celle du « Trou dans le mur » de Sugata Mitra. Dans ces expériences commencées dans un bidonville de New Delhi, des ordinateurs avec accès au Web sont proposés en accès libre à des enfants de milieux défavorisés. Les enfants sont amenés à apprendre de manière totalement autonome à se servir d'un ordinateur ; ils apprennent aussi le calcul, l'anglais et d'autres matières.

L'éducation est également ouverte à tous à travers un nombre incroyable de tutoriels sur la toile. Vous voulez savoir comment nouer une cravate, construire une cheminée, installer la dernière PlayStation VR, apprendre à parler en public, à tricoter, ou vous maquiller ? Quelqu'un a probablement préparé un tutoriel pour vous. Toutes ces personnes partagent généreusement leur connaissance, vous donnent envie d'apprendre parfois en vous amusant. Vous n'avez pas besoin d'aller à l'école, toutes les connaissances théoriques ou pratiques sont à portée de clic. Certes, toutes ne le font pas uniquement par altruisme et certaines essaient d'en tirer des profits.

Avec les MOOC et les tutoriels sur la Toile, va-t-on pouvoir se passer de l'école ? Pas vraiment. Il reste encore à l'école le rôle considérable de permettre un dialogue direct entre élèves et professeurs, d'amener tous les élèves à un niveau de connaissance choisi collectivement, des surdoués aux élèves plus lents, des enfants des milieux favorisés, à ceux des milieux difficiles. Il lui reste également celui de valider les compétences.

### *Les ressources éducatives libres*

Les ressources éducatives libres (REL pour faire court) sont des supports pédagogiques librement accessibles et sous licence ouverte tels que des textes, du multimédia, et d'autres objets numériques utiles pour l'enseignement, l'apprentissage ou à des fins de recherche. Des enseignants les développent pour que tous puissent les utiliser pour se former librement. Dans le cadre de l'éducation, cette liberté se décline avec 5 « R<sup>1</sup> » :

- retenir/conserver : le droit de créer, de garder des copies de la ressource, de contrôler ses usages ;
- réutiliser : le droit d'utiliser ces ressources, en particulier dans ses cours, mais aussi sur un site Web, à l'intérieur d'une vidéo ;
- réviser : le droit d'adapter ou de modifier le contenu, en particulier le droit de traduction dans une autre langue ;
- remixer : le droit de créer une nouvelle ressource en mélangeant des morceaux de ressources existantes, en les modifiant ;
- redistribuer : le droit de partager avec d'autres des copies du matériel original, modifié ou remixé.

Les REL visent à élargir le nombre de personnes qui peuvent accéder à un contenu, en éliminant les barrières à son utilisation. Mais elles cherchent également à faciliter la création de nouveaux contenus

---

1. En anglais : *retain, reuse, revise, remix, redistribute*.



pédagogiques à partir de matériel existant. Un peu à la manière du logiciel libre, on imagine un effort collectif pour construire des contenus pédagogiques. C'est une différence avec un MOOC qui est en général un produit éducatif plus figé.

Un argument massif pour les REL est leur coût. Une fois la ressource développée, on peut la diffuser massivement sans coût supplémentaire.

Les REL présentent de nombreux avantages, mais pour en tirer pleinement profit certains travers sont à éviter. D'abord, en se privant des éditeurs, on sacrifie le contrôle-qualité qu'ils réalisent. Des élèves peuvent donc se trouver confrontés à des contenus médiocres. Il est indispensable que la communauté instaure ses propres contrôles en combinant des avis de spécialistes, et ceux de la foule *via* le *crowd-sourcing*. On est également parfois confronté à la frilosité de certains enseignants qui rechignent à partager leur matériel éducatif. Partager demande des efforts de présentation des matériels. Et puis, les enseignants peuvent avoir l'impression d'être dépossédés de leur travail. En l'absence de copyright, les producteurs de contenus sont mal ou pas du tout rétribués. La récompense énorme pour un éducateur de participer à la création d'une ressource éducative de qualité qui est utilisée par de nombreux enseignants et élèves compense bien sûr. Mais il faudrait trouver des moyens de valoriser plus explicitement le partage.

Un dernier risque des MOOC est que, dans un souci d'économie, on profite de leur existence pour limiter le suivi des élèves. Des outils algorithmiques de plus en plus sophistiqués permettent de personnaliser l'expérience de chaque élève. Un suivi humain et personnalisé reste pourtant indispensable.

Pour les besoins de la présentation, nous avons distingué entre ressources éducatives libres et éducation ouverte, mais évidemment tout cela se rejoint.

Pour conclure, soulignons la place importante du logiciel libre dans ce contexte. Un logiciel *open source* comme Scratch permet de

former des jeunes et des moins jeunes à l'informatique. S'il est utilisé par l'Éducation nationale, il est aussi à la base de communautés sur le Web. De nombreux aspects des communs numériques sont à la fois à la manœuvre.

## CHAPITRE 19

# L'État

Les possibilités d'interactions entre la puissance publique et les communs numériques sont nombreuses. Après la science et l'éducation ouverte, nous évoquerons dans les chapitres à venir les questions de santé, de justice, d'agriculture. Ce sont des domaines où l'État est un acteur essentiel et où les communs trouvent toute leur place. Dans ce chapitre, nous considérons l'utilisation directe par l'État de logiciels libres et de données ouvertes. Nous discutons aussi de la mise des communs au service de la souveraineté numérique.

### *L'accélération de l'histoire*

En France, l'accès des citoyens aux documents administratifs est encouragé depuis des années, au moins depuis la création de la Commission d'accès aux documents administratifs (CADA) en 1978. Cette dernière était un premier pas, modeste, vers l'ouverture. En 2005, l'ordonnance sur la réutilisation de l'information publique transpose la directive européenne de 2003 sur le sujet, c'est le premier texte qui officialise l'ouverture des données en France. En 2011, Etalab est créé avec pour première mission de mettre en place un portail unique recensant les données publiques ouvertes. Depuis

2011, la plateforme [data.gouv.fr](http://data.gouv.fr) référence les données ouvertes de l'État français. Pour les données ouvertes états-uniennes, [data.gov](http://data.gov) existe depuis 2009. Plusieurs centaines de telles plateformes dans le monde sont listées dans [DataPortals.org](http://DataPortals.org).

En 2015, la loi relative à la gratuité et aux modalités de la réutilisation des informations du secteur public est promulguée. Enfin, en 2016, la loi « pour une république numérique », proposée par Axelle Lemaire, va plus loin en affirmant la mise en ligne par défaut des données publiques, et en encourageant l'utilisation de logiciels libres et de formats ouverts dans les systèmes d'information des administrations publiques. Avec cette loi, les codes sources deviennent des documents administratifs communicables et réutilisables.

**Etalab** (<https://www.etalab.gouv.fr/>) est un service de l'État organisé autour des données ouvertes publiques. Son but est d'améliorer l'ouverture des données publiques pour encourager leur réutilisation. Pour que des utilisateurs accèdent à une donnée, il faut d'abord qu'ils sachent qu'elle existe. C'est la fonction assurée par les portails de données ouvertes publiques comme [data.gouv.fr](http://data.gouv.fr). Le portail liste les données disponibles, les moyens d'y accéder, que ce soit pour une information particulière ou pour télécharger l'ensemble d'un jeu de données. Il fournit aussi des informations sur certaines des utilisations qui en sont faites. Etalab est à l'origine d'une licence pour les données publiques (la Licence Ouverte). Son périmètre a été élargi aux logiciels libres, aux algorithmes de l'État, à l'intelligence artificielle et à la science des données. Enfin, Etalab cherche à ouvrir l'administration à des talents du numérique avec le programme « Entrepreneurs d'intérêt général. »

L'ouverture comme moyen d'améliorer le fonctionnement et les services de l'État est régulièrement rappelée par des textes législatifs et des circulaires, tout comme le fait que les données doivent être ouvertes par défaut et que des justifications doivent être fournies

quand elles ne le sont pas. Ces rappels montrent bien qu'il n'est pas si évident d'en convaincre l'administration. L'ouverture est aussi considérée comme un levier pour la création de services innovants tant publics que privés. Comme nous le verrons, elle est enfin vue comme un moyen de transformer la nature même de la démocratie.

Récemment, cette recherche d'ouverture a été réaffirmée dans le rapport Bothorel en 2020, et dans une circulaire du Premier ministre, Jean Castex, en 2021. Une mission « dédiée à l'animation et la promotion interministérielles en matière de logiciel libre et de communs numériques » est créée au sein de la Direction interministérielle du numérique (Dinum). Il s'agit d'adapter l'État à une société de plus en plus numérique. Chaque ministère doit maintenant avoir son Administrateur ministériel des données, des algorithmes et des codes sources (son Amdac).

### *Les données de référence*

Parmi les données ouvertes de l'État, le décret du 14 mars 2017 définit certaines données comme étant « des données de référence ». Il s'agit de données qui présentent un intérêt stratégique et qui peuvent intéresser de nombreux acteurs, citoyens ou entreprises. Ces données doivent être traitées avec une attention particulière et on leur attribue quatre rôles :

- Les producteurs de ces données traitent les remontées des utilisateurs, et gèrent les mises à jour des données. Ils désignent les diffuseurs pour chacun des jeux de données de référence.
- Les diffuseurs mettent à disposition les données avec un haut niveau de qualité, des engagements sur des niveaux de performance et de disponibilité.
- Les utilisateurs utilisent les données de référence pour produire de nouveaux services et créer de la valeur économique et sociale.

Ils participent à la montée en qualité des données de référence en signalant des erreurs, et en proposant des améliorations.

- Enfin, Etalab recense pour l'État les jeux de données de référence. En cas de défaillance des diffuseurs, ou de non-respect des engagements notamment en matière de performance et de disponibilité, Etalab est habilité à se substituer au diffuseur désigné par le producteur.

L'État prend divers engagements sur ces données, notamment de disponibilité (une disponibilité de 99,5 % est annoncée) et de qualité (non précisément définie). Parmi ces données de référence, on peut citer en particulier :

- la BAN, Base Adresse Nationale des adresses postales ;
- le Plan cadastral informatisé et ses 600 000 « feuilles » ;
- le Référentiel géographique à grande échelle (RGE), qui décrit le territoire national et l'occupation de son sol ;
- la base Sirene des quelque 10 millions d'entreprises françaises et de leurs établissements ;
- le Registre parcellaire graphique (RPG) qui contient les données graphiques des parcelles (depuis 2015) et îlots (éditions 2014 et antérieures) agricoles et leur culture principale ;
- le Répertoire national des associations (RNA) ;
- le code officiel géographique, qui rassemble les codes et libellés des communes, des cantons, des arrondissements, des départements, des régions et des pays et territoires étrangers ;
- le référentiel de l'organisation administrative de l'État, qui liste toutes les institutions régies par la Constitution de la V<sup>e</sup> République et les administrations qui en dépendent, soit environ 6 000 organismes ;
- le Répertoire opérationnel des métiers et des emplois (ROME).

L'enthousiasme de 2017 pour étendre cette liste à d'autres jeux de données semble être un peu retombé depuis et la liste n'a que peu évolué pendant les cinq dernières années.

## *Un État plus efficace*

Un premier but de l'ouverture est de permettre d'améliorer et d'optimiser les services publics. Prenons n'importe quel service de l'État, comme la gestion des cimetières, distribué sur tout le territoire. Une autorité locale (la région, le département, la commune, etc.) va chercher à se numériser, c'est-à-dire à financer le développement d'un logiciel, ou acheter un logiciel propriétaire. Pour les plus de 40 000 cimetières en France, cela peut finir par coûter cher. De plus, si chaque autorité définit localement ses propres formats et ferme l'accès à ses données, l'État a beaucoup de mal à obtenir une vision consolidée à l'échelle du pays.

Un choix de partage de logiciels et de formats ouverts, ainsi que l'ouverture des données vont faciliter l'intégration de toutes ces données. L'existence de ces données disponibles encourage également le développement de nouveaux services. Le résultat : un service tant localement que globalement moins coûteux et plus performant qui tire peut-être tout un écosystème. On remarquera que cet exemple combine logiciel et données, souvent la réalité des services informatiques sur le terrain.

**LES DONNÉES.** L'ouverture des données est un moyen d'améliorer la qualité des données de l'État dont tous vont bénéficier : les citoyens et les entreprises qui pourront mieux bénéficier de ces services, mais également la presse et les chercheurs qui étudient justement ces services de l'État. À partir du moment où les données sont ouvertes, tout le monde peut les voir et signaler les erreurs, améliorant ainsi leur qualité, en supposant bien sûr qu'un mécanisme de retour utilisateur soit mis en place, ainsi qu'un processus permettant de traiter ces retours. Cela est aussi vrai pour les informations sur les adresses dans la Base Adresse Nationale que pour la géolocalisation des arrêts de bus d'une ville. Cette participation citoyenne à l'amélioration de la qualité des données est parfois un

premier pas vers la mise en place d'une communauté, donc d'un commun numérique.

L'ouverture des données publiques est largement encadrée par la loi, et par des licences associées aux données en question qui définissent comment elles peuvent être utilisées. Il existe bien entendu des limitations à cette ouverture et en particulier des lois comme le RGPD et le respect de la vie privée. La publication de données peut aussi entrer en conflit avec d'autres lois comme celle du secret-défense ou le droit des entreprises.

L'ouverture prônée des données publiques rencontre régulièrement des résistances dans les services de l'État. Certains agents ressentent le partage d'information comme une forme de perte de pouvoir. Le mot ouverture prend pour eux un peu le sens d'ouverture dans « ville ouverte », une ville qui choisit de se rendre sans combat. L'État perdrait sa souveraineté en ayant à s'expliquer, à justifier ses actions. Plus concrètement, l'ouverture peut être comprise par les services de l'État comme encourageant les critiques des citoyens, voire comme facilitant les recours juridiques contre ces services. Enfin, on a tant répété que « les données étaient le nouveau pétrole de l'économie », que tout responsable de données auquel on demande de les ouvrir se sent dépossédé d'un trésor. Il est vrai que l'ouverture des données pour certains services de l'État qui avaient l'habitude de les monnayer s'est parfois traduite par une perte de revenus, et comme c'était prévisible, ces services ont parfois longtemps résisté à l'ouverture. C'était le cas de la DILA qui commercialisait le BALO, le Bodacc et le BOAMP, de l'Insee qui commercialisait le fichier Sirene, du Cadastre qui commercialisait ses plans et de l'IGN qui commercialisait ses cartes.

L'absence de données ouvertes peut être un frein à de nombreuses activités. Par exemple, un système unifié définissant précisément les adresses a longtemps manqué en France. On peut rencontrer des problèmes pour se faire installer la fibre parce que son adresse n'est pas reconnue par les opérateurs télécoms. Plus grave, une équipe d'urgentistes peut perdre un temps précieux pour trouver le logement



d'une personne en détresse, et cela à cause de l'absence d'un système d'adressage global et ouvert. Le remède est la Base Adresse Nationale proposée par un consortium incluant deux services de l'État, une association et une entreprise : l'IGN, Etalab, OpenStreetMap et La Poste. Ensemble, ils ont créé un système qui permet de relier des adresses postales et des coordonnées géographiques (dans les deux directions). Par exemple, le 12, rue de l'Écu à Romorantin correspond aux coordonnées GPS 47.360335, 1.74283.

Avant cet effort, il existait plusieurs bases adresses, toutes incomplètes, toutes imparfaites, à La Poste, chez EDF, à l'IGN, à la DGfip, etc. L'ambition de la BAN est de remplacer toutes ces bases de données d'adresses, de proposer une référence unique et ouverte pour tous les lieux pourvus d'une adresse en France.

Pendant un temps, les adresses ont été l'objet d'une compétition entre l'État qui dispose massivement de cartes, et d'associations comme OpenStreetMap qui s'appuient sur la foule pour obtenir des données cartographiques. Et d'un certain point de vue, un randonneur a pu avoir à choisir entre les cartes de l'IGN et celles d'OpenStreetMap. Mais nous avons dépassé cette époque de compétition, quand les cartes IGN étaient une source de revenu pour l'État, et qu'OpenStreetMap se développait lentement. Les deux coopèrent aujourd'hui notamment dans le projet Street View, pour visualiser des lieux situés sur des voies publiques. Ensemble, ils veulent offrir une vraie alternative au service privé Google Street View.

**LES LOGICIELS LIBRES.** Le logiciel libre présente pour l'État la possibilité d'une mutualisation. Par exemple, pendant la crise du Covid-19, l'Éducation nationale s'est trouvée piégée avec un logiciel propriétaire pour les cours en vidéo : chaque nouveau déploiement coûtait cher et le logiciel n'était pas vraiment adapté aux besoins. Depuis, elle a choisi de basculer vers BigBlueButton, un logiciel libre. Les spécifications des besoins ont été définies avec l'aide des enseignants. Le développement par une entreprise privée est financé par l'Éducation nationale. Le code reste libre et sert à toute la collectivité.

Ce n'est pas encore la solution idéale. Il faut s'accommoder un temps d'un logiciel encore imparfait. Il faut installer et dimensionner des serveurs. C'est le prix à payer pour garder une forme d'indépendance, et faire, on l'espère, de sérieuses économies.

L'État français a bien réalisé les avantages du logiciel libre. Aujourd'hui, quand un service a besoin d'un logiciel, il est encouragé à considérer toutes les options, propriétaires comme libres. Etalab publie le SILL, le socle interministériel de logiciels libres, une liste de logiciels libres recommandés pour les administrations publiques. Il est intéressant de voir que cette liste est elle-même coconstruite par des agents des services publics.

L'Adullact a été créée en 2002. Cette association réunit des administrations et des collectivités locales autour du logiciel libre. Elle regroupe aujourd'hui 230 entités. L'idée simple est que beaucoup d'organismes publics ont des besoins similaires, si ce n'est identiques. Donc plutôt que chacun fasse son application spécifique dans son coin, une administration peut réaliser ou faire réaliser l'application, la mettre en logiciel libre et permettre à tous de l'utiliser. Ainsi est né par exemple Opencimetière, une application de gestion de cimetière que de nombreuses communes utilisent. Pour chacune de ces applications, une communauté se crée qui va la faire vivre et évoluer.

**PARLE-T-ON DE COMMUNS ?** Peut-on parler de communs pour ces ressources ouvertes publiques que nous avons évoquées ? Dans le cas de la BAN, le triptyque ressources-communauté-gouvernance est à l'œuvre. C'est un bel exemple de commun. Mais, ce n'est pas le cas pour tous les corpus de données publiques. Chaque entreprise dispose d'un numéro SIREN (Système d'identification du répertoire des entreprises). Ce numéro permet d'identifier l'entreprise auprès des administrations. Le fichier Sirene est en accès ouvert. On peut interroger le fait que ce soit un commun, car le fichier est totalement géré par l'État et aucune communauté ne participe à sa maintenance.

De nombreux services publics français ont pris le pli d'ouvrir leurs données. L'ouverture des données marque un premier changement

de culture. Mais c'est une autre affaire que de collaborer avec une communauté, de faire vivre collectivement ces données, et peut-être de les enrichir. Ce passage aux communs est un autre changement de culture que seuls quelques services ont engagé.

Comme dans tout changement de culture, tout ne se passe pas toujours facilement. Certains cherchent des excuses pour ralentir le passage aux données ouvertes. Et comme ce n'est pas facile d'ouvrir les données, de nombreuses excuses peuvent être trouvées pour traîner les pieds : la qualité des données n'est pas parfaite, on n'arrive pas à suivre les mises à jour assez vite, leur publication est peut-être attentatoire à la protection de la vie privée, cela va coûter très cher, etc. Il faut en revenir à une règle simple : « Les données utilisées par l'État dans l'exercice de ses missions doivent être disponibles pour tous. » On peut bien sûr accepter que tout ne soit pas tout de suite parfait, mais l'ouverture ne doit pas attendre plus que nécessaire.

Pour le logiciel, la question de savoir si on se situe ou pas dans le cadre de communs est encore plus complexe. D'abord, les services de l'État peuvent utiliser un logiciel libre. Ils font alors partie au même titre que d'autres de la communauté des utilisateurs. Le logiciel en question peut être un commun, mais l'État n'y est pas pour grand-chose. Par contre, un service de l'État peut être moteur dans le développement d'un logiciel libre, notamment en le finançant. Si une vraie concertation avec les utilisateurs est mise en place, voire avec des mécanismes de coconstruction, cela devient alors un commun dont l'État, en plus d'être un bénéficiaire, est un partenaire essentiel.

### *Un État plus ouvert*

Henri Verdier, qui a été le deuxième dirigeant d'Etalab et qui lui a donné sa véritable impulsion, retrouve l'origine de l'idée de l'ouverture des données dans la Déclaration des droits de l'homme et du citoyen, de 1789 à l'article 15 : « La Société a le droit de

demander compte à tout agent public de son administration. » Historiquement, c'est l'administration Obama qui a accéléré le mouvement en 2009 en donnant trois semaines à chaque administration fédérale pour ouvrir trois jeux de données « de haute valeur » qui n'étaient pas ouverts à ce jour, puis à l'issue de ces trois semaines de recommencer. Une grande partie du monde a suivi avec plus ou moins d'enthousiasme.

L'idée de « gouvernement ouvert » est proposée pour réduire la distance entre le citoyen et l'État. Pour y arriver, on dispose de deux angles :

- La transparence. Des données et des algorithmes ouverts permettent au citoyen de mieux comprendre les politiques publiques, ce qui participe à construire la confiance. Comme leurs comportements peuvent être observés, les services sont valorisés quand ils réussissent et critiqués sinon ; ils sont ainsi encouragés à mieux faire.
- La participation du citoyen et de la société civile. Le citoyen est consulté pour les décisions le concernant, et peut participer à la vie démocratique. La société civile à travers les associations, labos de recherche, etc., est également consultée. De nouveaux canaux de communications basés sur les nouvelles technologies enrichissent les liens entre les citoyens et l'État. S'ils ne décident pas (c'est le rôle des élus), au moins on leur demande leur avis.

Prenons la sélection de la filière après le bac avec le processus Parcoursup. Dans le temps, cette sélection se faisait en toute opacité, favorisant les passe-droits et occultant les inefficacités. C'est maintenant réalisé par un logiciel, ce qui conduit à plus de justice. De plus, le logiciel est libre, et les choix opérés peuvent maintenant être débattus sur la place publique. Donc le logiciel libre offre un mécanisme de transparence, qui permet à tous (du moins tous ceux qui peuvent lire un code source) de comprendre, de vérifier les règles qu'il applique, voire de les contester. Nous ne portons pas de jugement sur ce que fait ce logiciel, nous nous contentons de

saluer le fait que les choix de ses orientations sont lisibles et maintenant publiquement débattus. OpenFisca, dont nous parlons dans le chapitre sur les communs dans la justice, est un autre exemple de la transparence.

Depuis 2014, la France fait partie des États du Partenariat pour un gouvernement ouvert<sup>1</sup> (PGO) où se retrouvent 78 pays (janvier 2022) et de nombreux organismes de la société civile. En plus de la transparence et de la participation, le PGO promeut l'intégrité de la vie publique et de ses agents, et l'utilisation des nouvelles technologies en faveur de l'ouverture et de la redevabilité. Avec le numérique, peut-être avons-nous enfin les moyens de réaliser une vieille utopie, celle du citoyen éclairé qui décide de son destin ? Devant certaines dérives comme les *fake news*, il faut une certaine dose d'optimisme pour penser cela. Le chemin est sûrement long, mais pourquoi ne pas au moins essayer ?

### *Un État plus souverain*

La « souveraineté numérique » est la déclinaison du principe de souveraineté nationale dans le cadre de l'informatique. L'idée sous-jacente est que dans un monde numérique, la maîtrise de son destin passe par la maîtrise des technologies numériques, pour éviter de dépendre des technologies numériques des autres. Le sujet n'est pas nouveau. Déjà en 1966, le général de Gaulle lançait le Plan Calcul pour construire les bases d'une informatique nationale – on ne parlait pas encore de souveraineté numérique. Il le faisait parce que les États-Unis venaient de s'opposer à l'acquisition par la France d'un supercalculateur nécessaire au développement de la force nucléaire. Ses successeurs n'ont pas fait prospérer l'idée, ce qui a conduit, au début du XXI<sup>e</sup> siècle, Pierre Bellanger parmi de

---

1. <https://www.opengovpartnership.org/>.

nombreux autres à militer pour la souveraineté numérique qu'il définissait comme : « La maîtrise de notre présent et de notre destin tels qu'ils se manifestent et s'orientent par l'usage des technologies et des réseaux informatiques. »

Le sujet revient dans l'actualité avec de grands projets comme la French Tech, ou des slogans comme la start-up nation. Dans un monde numérique, dominé par les États-Unis et la Chine, et par leur rivalité, la question de la souveraineté est maintenant reprise au niveau européen.

Un exemple des risques que nous fait courir l'absence de souveraineté a été mis en évidence par la guerre en Ukraine. La Russie a été confrontée à d'énormes difficultés quand les États-Unis ont restreint son accès aux composants de haute technologie, matériels et logiciels, fabriqués avec des pièces ou des propriétés intellectuelles états-uniennes. L'inimitié états-unienne se paie cher, comme l'a aussi découvert l'entreprise chinoise Huawei qui a dû abandonner des pans entiers de son marché parce qu'elle n'avait plus accès aux logiciels et aux matériels états-uniens. Dans un monde globalisé, une indépendance numérique totale n'a bien sûr pas beaucoup de sens. Par contre, il est devenu évident qu'il est indispensable d'occuper le terrain de l'industrie du logiciel et du matériel informatique pour préserver une relative indépendance économique et politique.

Face à des États qui cherchent à transposer leur domination technique et économique en une domination politique, face aux entreprises géantes du numérique, on peut essayer de défendre nos citoyens et protéger nos entreprises à coups de régulations comme le RGPD, le DMA, ou le DSA déjà évoqués. On peut aussi taxer ces géantes. Ce sont des moyens purement défensifs. De manière plus offensive, on peut essayer de regagner des espaces sur les terrains économiques et techniques. Cela passe par les entreprises privées, mais ce n'est pas simple. Les communs numériques proposent une autre voie très prometteuse.

Avec les communs numériques, on peut créer des communautés, des écosystèmes qui offrent des alternatives crédibles aux monopoles ou oligopoles existants, comme Linux pour Windows OS, ou OpenStreetMap pour Google Maps. Ces communs présentent l'avantage de remettre en question les stratégies d'enfermement. En 2022, en prenant la présidence de l'Europe, la France annonce vouloir promouvoir les communs numériques avec des références explicites à Wikipédia, Linux, OpenStreetMap et Open Food Facts, dont les trois derniers sont européens à l'origine. La France propose un investissement massif pour soutenir des start-up européennes, par le biais de Scale-Up Europe. Elle propose aussi de soutenir au niveau européen les logiciels libres dans les domaines stratégiques comme Internet, et d'augmenter leur part dans les appels d'offres, et de manière générale de soutenir le développement de communs numériques.

Les domaines stratégiques dans lesquels il serait bon de ne pas être pieds et poings liés à des entreprises privées étrangères sont légion : les logiciels d'Internet, le *cloud*, les puces, les moteurs de recherche, les systèmes d'exploitation, les bases de données, le commerce électronique, le métavers, etc. À côté de la maîtrise du matériel et du logiciel se pose la question de la gestion des données. Cela a conduit, en 2020, au démarrage par la France et l'Allemagne de Gaia-X, un projet de cloud souverain, puis, en 2021, de Euclidia, plus indépendant des grands états-unis.

Nous gagnerions un peu de notre souveraineté numérique en construisant des communs dans ces domaines. Les États européens et l'Europe, ont les moyens de réaliser des investissements considérables tant d'un point de vue technique, que marketing et vente, pour réaliser de tels objectifs. Mais il ne faut pas sous-estimer la difficulté de la tâche. L'exemple de Huawei est ici édifiant. Privée de logiciels et de matériels par le gouvernement américain, l'entreprise qui était alors un leader mondial de smartphone a dégringolé sur ce marché. Elle a eu brutalement à développer son propre OS

pour smartphone, HarmonyOS et sa version ouverte OpenHarmony. L'histoire dira si elle arrivera à se réimplanter sur ce marché. Le cas de Huawei a aussi mis en évidence l'imbrication entre eux des produits numériques tant logiciels que matériels, et la complexité d'atteindre une forme d'autonomie technologique.

L'angle « souveraineté numérique » des communs peut irriter certains qui voient dans les communs d'abord des lieux de partage, de collaboration ; c'est vrai qu'on est bien loin de ces idéaux. Les communs pour l'Europe pourraient pourtant réduire les dépendances à de grandes entreprises privées typiquement états-uniennes et chinoises. L'indépendance est une dimension indispensable à notre développement économique et à la défense de nos valeurs et de nos libertés de choix géostratégiques.



## CHAPITRE 20

# La santé

Le partage de données de santé s'impose. Il est important à la fois pour favoriser les progrès de la recherche, mais aussi pour piloter le système de santé. Le caractère personnel des données est naturellement un frein à ce partage.

Les dépenses de santé ont représenté en France 12,4 % du PIB en 2020, c'est dire leur importance dans l'économie ; comme point de comparaison, celles de l'éducation représentent 7 % du PIB. Il s'agit donc bien là d'un secteur essentiel de l'économie et de la vie des Français. Ce secteur est en crise. Pour s'en convaincre, il suffit de consulter les médias ou de parler à des professionnels. C'est un domaine où les données sont massivement produites : essais cliniques, dépenses de santé, tableaux de bord sur la pandémie, etc. Le partage des données pourrait-il contribuer à améliorer les choses ? Certainement, mais ce n'est pas simple. De nombreuses données de santé sont des données personnelles, classifiées comme sensibles par la CNIL. Et puis, le domaine de la santé a une longue histoire de résistance à la transformation numérique. L'usage de la carte vitale si longue à s'implanter et encore refusée par certains professionnels de santé est un bel exemple de cette résistance bien gauloise.

## *Le Système national des données de santé*

Les sources de données de santé sont multiples : les professionnels de santé, les établissements de santé, les autorités dans l'accomplissement de leur mission de service public, les industriels et entreprises du secteur de la santé, les chercheurs et enfin les patients.

Depuis longtemps, l'État se préoccupe de ces données, de leur recueil, de leur protection et leur exploitation. Pour se concentrer uniquement sur les derniers épisodes d'une longue histoire, un rapport sur les données dans la santé est remis par Pierre-Louis Bras en octobre 2013, avec pour objectif de permettre de partager davantage les données de santé<sup>1</sup>. Il est suivi par la mise en place d'une commission sur le même sujet qui remet le fruit de ses travaux en juillet 2014. Cet autre rapport sur les données en santé fait l'état des lieux et recommande la mise en place d'un système de données intégrant diverses sources et l'accès sous conditions offert aux acteurs du domaine.

Ces deux rapports insistent sur l'importance d'une stratégie gouvernementale volontariste en faveur des données de santé, présentent les résultats d'une réflexion approfondie menée par le Ministère de la santé, en vue d'ouvrir davantage l'accès aux données de santé, dans le respect de l'anonymat des patients.

Ces travaux aboutissent à la mise en place du Système national des données de santé, le SNDS. Unique en Europe, le SNDS est créé en 2016 pour développer l'usage de ces données. Il est géré par la Caisse nationale de l'assurance-maladie (CNAM). Il regroupe essentiellement les données du SNIIRAM (Système national d'information inter-régimes de l'assurance-maladie) et celles du

---

1. Pierre-Louis Bras, *Rapport sur la gouvernance et l'utilisation des données de santé*, 2013, <https://drees.solidarites-sante.gouv.fr/publications/rapports/rapport-sur-la-gouvernance-et-lutilisation-des-donnees-de-sante>.

PMSI (Programme de médicalisation des systèmes d'information). Eh oui, c'est un peu compliqué...

- Le SNIIRAM réunit les données sur les soins et les remboursements pour les soins du secteur libéral (médecine de ville). Les données sont anonymisées. Pour chaque patient, la base donne l'âge, le sexe et la commune de résidence, et pour chaque professionnel de santé, l'âge, le sexe, la commune d'exercice et la spécialité. Pour chaque soin, elle donne la date et le montant remboursé et liste les médicaments délivrés, les actes techniques et les prélèvements biologiques réalisés.
- Le PMSI joue un rôle similaire pour la médecine hospitalière, y compris les hospitalisations à domicile. On y trouve en outre les bases de données sur les causes médicales de décès et les données relatives au handicap.

Tout cela constitue une mine de données d'une valeur considérable.

Le SNDS définit à quoi doivent servir ces données, à quoi elles ne doivent pas servir, et de quelle manière elles doivent être traitées. Les finalités visées par l'utilisation des données du SNDS concernent notamment la mise en œuvre et le suivi des politiques de santé, la connaissance des dépenses de santé, la surveillance, la veille et la sécurité sanitaire. Est en revanche exclu tout ce qui pourrait être utilisé pour prendre une décision sur une personne physique identifiée, ou qui viserait à faire la promotion de produits ou services auprès des personnels ou des établissements de santé. Est aussi interdite toute utilisation par un assureur de ces données pour modifier les primes d'un individu ou d'un groupe d'individus. Enfin, les accès aux données du SNDS et leurs utilisations doivent garantir la confidentialité et l'intégrité des données et la traçabilité des accès et autres traitements.

Cette mine de données est relativement accessible : une liste de plusieurs dizaines d'organismes ayant un accès permanent à tout ou

partie des données du SNDS est fournie dans le décret d'application de la création du SNDS. En outre, tout organisme public ou privé, à but lucratif ou non, peut accéder aux données du SNDS après avoir reçu une autorisation de la CNIL. Cette autorisation nécessite de respecter un certain nombre de procédures pour assurer la confidentialité des données et d'expliquer en quoi on ne va pas faire un usage interdit des données.

### *Vers le Health Data Hub*

L'étape suivante pour le SNDS est le Health Data Hub (HDH). Le HDH a été créé en 2019 pour faciliter un partage sécurisé des données de santé afin de favoriser la recherche, améliorer la qualité des soins, et l'accompagnement des patients. L'objectif est de mutualiser les données et notamment de permettre l'utilisation des techniques d'intelligence artificielle sur toutes ces données. Le Hub doit récupérer les données du SNDS et les augmenter d'une quarantaine de bases de données supplémentaires.

Le Hub a déposé une demande d'agrément auprès de la CNIL, puis a annoncé en juin 2022 qu'il retirait sa demande. La raison ? Il faut aller la chercher dans la bronca provoquée par le fait que le Hub avait choisi Microsoft comme fournisseur de technologie *cloud*. Cela posait problème parce que les lois extraterritoriales des États-Unis donnaient l'accès à ces données de santé, donc sensibles, aux autorités états-uniennes. Même si l'hébergement est européen, est-il acceptable d'utiliser des logiciels états-uniens sur des ordinateurs non européens localisés ? Ou doit-on insister pour que les logiciels soient eux-mêmes européens ? Devrait-on exiger leur ouverture pour plus de transparence ? Il semblerait que le projet soit totalement repensé. Le chemin vers l'intégration des données médicales et leur partage est long et sinueux.

## *Les associations de patients*

Tout ce que nous avons décrit jusqu'ici demeure sous le contrôle de l'État et des autorités médicales, même s'il est fait largement appel aux principes du partage de données. Y a-t-il place dans cet écosystème pour des organisations relevant plus clairement du principe des communs numériques ? Oui, et les pistes pour cela sont à trouver du côté du grand nombre d'associations de patients (près de 3 000 suivant certaines sources) qui réunissent les patients atteints d'une maladie spécifique.

Une telle association pose les premiers éléments d'un commun :

- une communauté, constituée des patients, ceux qui les accompagnent, familles et soignants, et éventuellement de chercheurs ;
- un intérêt commun clair : échanger et partager des informations, pour faire bénéficier les malades de meilleurs soins, mieux les accompagner, et faire progresser la connaissance de leur maladie.

Ces associations de patients peuvent bien entendu accéder au SNDS en suivant le processus d'accès défini par la plateforme. Avec ces associations, la logique de la fourniture des données est inversée : ce n'est pas un organisme médical qui recueille les données sur les patients et les utilise pour des besoins de recherche, ce sont les patients eux-mêmes qui prennent l'initiative de fournir leurs données pour faire progresser la recherche. L'association « Moi Patient », par exemple, a pour objet de développer une base de données pilotée et renseignée par les patients pour faire notamment de la recherche sur des sujets comme l'égalité d'accès aux soins, l'éducation thérapeutique, la qualité des soins, l'expérience patient et l'adhérence aux traitements.

Plusieurs études ont été conduites par diverses associations de patients sur des sujets comme les migraines, les médicaments anti-rejet pour les greffés, la sclérose en plaques et son interaction avec le Covid-19.

L'état de l'ouverture des données de santé, ainsi que la mise en place de certains communs, illustre bien les caractéristiques de ces données : le besoin est important, la progression est lente mais régulière et devrait se poursuivre, les freins et les difficultés sont nombreux.

## CHAPITRE 21

# La justice

Dans le domaine de la justice, le numérique est bien sûr porteur de grandes promesses puisque le travail de juriste consiste principalement à acquérir des connaissances et produire des arguments et des décisions qui sont autant de nouvelles connaissances partageables. Tout cela, le numérique peut le faciliter. Mais, quels peuvent être les apports des communs numériques dans ce domaine ? Le dynamisme d'une initiative devenue association – Open Law, Le droit ouvert – l'illustre. Ses missions principales sont : « Mettre en œuvre l'application des principes de l'innovation légale au monde du droit, en adoptant une démarche ouverte, afin de promouvoir la mise à disposition sous un format électronique et à titre gratuit (*open data*) des données juridiques et de favoriser l'émergence d'une économie numérique du droit. » Les références aux communs sont bien présentes.

Dans le cadre de la justice, nous allons d'abord considérer les données ouvertes, puis les logiciels *open source* et conclure avec des approches collaboratives envisagées dans les préparations de procès.

## *Des données ouvertes : la loi et la jurisprudence*

La première étape de l'ouverture des données légales a été le remplacement du *Journal officiel* (qui n'est plus édité sous forme papier depuis 2016) par un site Web, Légifrance, mis en place en 1999. C'est le site du gouvernement français dont la mission est la diffusion des textes législatifs et réglementaires et des décisions de justice des cours suprêmes et d'appel de droit français. Il inclut une partie importante de la jurisprudence.

Comme « nul n'est censé ignorer la loi », l'accessibilité de telles données semble bien un préalable à la protection des droits de chacun. La justice bénéficie notamment de la transparence apportée par la publicité des décisions de justice, qui rend les règles plus accessibles, plus compréhensibles. C'est en particulier une promesse de plus de prévisibilité des décisions de justice pour le justiciable et ses défenseurs.

La loi de 2016 « pour une république numérique » déclarait l'ouverture des données des décisions des juridictions judiciaires et administratives. Il faudra attendre 2020 pour que son application soit précisée dans des décrets d'application. Aujourd'hui, le judiciaire est pris en charge par la base de données Judilibre, et l'administratif par le service Décisions de la justice administrative. Les décisions sont disponibles gratuitement. Si toutes les décisions ne sont pas encore disponibles, de plus en plus le sont. Chaque cour de justice organise son propre calendrier d'ouverture pour les données nouvellement produites, et pour le rattrapage du passé.

L'ouverture prend du temps du fait de la multiplicité des acteurs et de l'éclatement des sources de données ; ce n'est pas simple d'intégrer les données d'un grand nombre de tribunaux équipés de logiciels souvent hétérogènes. Mais cela tient aussi à des freins. Lesquels ?



On peut d'abord noter les frilosités de juristes souvent peu friands de numérique. Il faut aussi tenir compte des réticences d'éditeurs juridiques qui contrôlaient jusque-là les données de jurisprudence et voient avec regret s'évanouir l'oligopole que cela représentait.

Enfin un verrou sérieux était et reste la protection de la vie privée. Les données de justice contiennent de nombreuses données personnelles, protégées par des lois et règlements, notamment le RGPD. Leur publication pourrait parfois porter tort à des personnes. Il a fallu des arbitrages pour clarifier ce point. La Cour de cassation a décidé quelles informations seront anonymisées avant d'être ouvertes. Elle a décidé par exemple qu'il fallait occulter les noms des justiciables, mais pas ceux des magistrats. C'est un choix français. D'autres pays ont fait d'autres choix.

Évidemment, il est difficile de faire disparaître tout élément permettant l'identification d'une personne. On peut même s'attendre à ce que, le plus souvent, il soit impossible de garantir l'anonymat contre un « attaquant » qui s'appuierait sur toutes les données disponibles en dehors des jugements. À l'impossible nul n'est tenu, les autorités n'anonymisent pas vraiment les données, mais se contentent de les « pseudonymiser », en remplaçant les noms, prénoms, adresses, et autres champs directement porteurs d'information par des identifiants sans sémantique.

On voit bien la tension entre la mise en accès ouvert de données de justice pour la transparence de cette justice et la protection des données personnelles. La numérisation des données de justice peut conduire à un paradoxe : quand on numérise et publie des données qui sont physiquement accessibles si on se rend dans les greffes des tribunaux, on cache certaines données personnelles.

## *Des logiciels ouverts*

Les données de justice requièrent des outils qui aident les spécialistes comme les avocats à faire leur travail, mais qui permettent aussi aux justiciables (avec l'aide d'acteurs de médiation) de comprendre une loi exprimée dans le jargon de la profession, voire d'interagir avec des structures spécialisées parfois opaques.

Cela conduit à une industrie des logiciels pour le droit et à la *legaltech* (en français, « technologie appliquée au juridique »), un tissu d'entreprises, dont des start-up, qui s'intéressent à ce marché. Nous avons vu comment, avec les logiciels ouverts, des développeurs pouvaient être collectivement plus performants que ceux d'entreprises isolées avec des logiciels propriétaires. De la même manière, dans le droit, les logiciels ouverts et interopérables sont porteurs de promesses. Leur ouverture semble d'ailleurs en adéquation avec les idéaux de transparence de la justice. C'est cette *legaltech* ouverte qui nous intéresse ici.

De plus en plus de logiciels *open source* sont disponibles dans le monde de la justice. En voici quelques exemples :

- Un logiciel d'anonymisation des décisions de justice a été développé en *open source* dans une collaboration entre des acteurs privés et publics, s'appuyant sur des logiciels *open source* développés dans d'autres secteurs (LABEL pour « logiciel d'anonymisation d'une base enrichie labellisée »).
- ClinicCases est un logiciel de gestion des données d'affaires juridiques ciblant notamment l'enseignement du droit. Il est *open source* et s'installe facilement sur son propre serveur. On garde donc le contrôle des données des différentes affaires auxquelles on s'intéresse.
- Le site [opensource.legal](https://opensource.legal) propose toute une gamme de logiciels *open source*, de produits avec des API ouvertes, des formats de données ouverts, pour développer des applications de *legaltech*.

Les logiciels peuvent servir à offrir des services, comme la recherche de jurisprudence. Ils peuvent également participer à préciser le droit. Sans aller jusqu'à une formalisation complète des lois sous forme algorithmique ou mathématique, on peut se servir des logiciels pour préciser la loi. C'est ce que s'est efforcé de faire OpenFisca, un logiciel libre de microsimulation du système sociofiscal, créé en 2011. Ce logiciel de simulation permet notamment de simuler l'impact d'une loi ou d'un règlement sur un individu, ou de simuler l'impact global à l'échelle du pays que peut avoir la mise en place éventuelle d'une nouvelle mesure. Avant OpenFisca, chaque ministère faisait son simulateur dans son coin. Comme OpenFisca est un logiciel libre, tous les services peuvent contribuer à un logiciel commun qui fait le boulot, en couvrant un champ bien plus large de lois. L'objectif est que toute la législation des impôts et toutes les règles des prestations sociales soient codées.

Les usages d'OpenFisca sont multiples :

- Pour les particuliers : chacun peut calculer les impôts qu'il va devoir payer et les prestations sociales auxquelles il a droit. On conçoit l'intérêt d'un tel logiciel quand on prend conscience, notamment, que des prestations sociales ne sont pas exigées par un nombre important de personnes qui ne savent pas qu'elles y auraient droit. Elles ne veulent pas se lancer dans un dossier compliqué sans savoir si ça va marcher. Le site OpenFisca demande des informations simples à l'utilisateur (cela peut être saisi *via* un médiateur numérique, un ami, un bénévole d'une association) et propose de calculer si la personne en question a droit à telle ou telle prestation.
- Pour les décideurs : ils peuvent tester ce qui se passerait si on changeait, par exemple, tel seuil ou tel taux d'imposition.
- Pour la presse et les politiques : en cas de réforme, les politiques dans leurs débats, et les journalistes dans leurs articles, n'ont plus d'excuses si leurs exemples et leurs affirmations dans ce domaine sont faux.

Il faut aussi signaler le site « la fabrique de la loi », une application mise en place par Regards Citoyens, une association de bénévoles qui aide à comprendre le fonctionnement des institutions démocratiques. La fabrique de la loi permet de suivre le processus de production de 1 042 lois par le parlement français.

### *Préparation ouverte d'arguments juridiques*

Dans le cadre du projet Openlaw, le Berkman Klein Center de l'Université Harvard expérimente l'élaboration collaborative ouverte d'arguments juridiques. Sur le Web, des avocats et des non-juristes participent à l'élaboration de ces textes, jusqu'aux plaidoiries. Chacun peut apporter ses idées de jurisprudence ou d'arguments. On s'appuie sur la puissance collective avec l'idée que le citoyen peut contribuer.

On imagine comment cela pourrait fonctionner dans le cas de désaccords entre une grande entreprise et des particuliers. La grande entreprise peut convoquer des moyens légaux considérables, loin des possibilités du particulier. Par contre, dans le cadre peut-être d'un recours collectif (*class action* en anglais), les particuliers peuvent préparer collectivement leurs arguments dans un mode ouvert.

Les données ouvertes, les logiciels *open source*, les approches collaboratives, un mouvement autour des communs semble bel et bien engagé dans la justice. On imagine mal qu'il puisse être remis en cause.

## L'agriculture et l'environnement

Avec des pâturages partagés par des paysans, le monde agricole connaît des communs depuis des siècles. Qu'en est-il aujourd'hui, d'abord pour les communs matériels, puis pour les communs numériques ?

### *Les communs matériels*

L'agriculture est aujourd'hui confrontée à une véritable transition, en particulier vers le bio. Il lui faut concilier d'une part les fonctions productives des biens alimentaires, et d'autre part, la préservation de l'environnement. L'agriculture intensive s'appuie sur l'utilisation massive d'implants. La réglementation tente de contenir les atteintes les plus graves à l'environnement. Il en résulte des tensions entre d'un côté des agriculteurs et de grands groupes qui cherchent à développer la production et de l'autre, des défenseurs de l'environnement, une tension classique entre biens privés et biens communs.

Comme l'explique Christian Huyghe, directeur agriculture de l'Inrae, la particularité de cette tension est d'être entre des biens privés d'aujourd'hui et des communs de demain. Pour produire plus aujourd'hui, on accepte de mettre en péril l'environnement de demain.

Si ce n'est ce décalage temporel, on se retrouve donc dans une situation typique de la « tragédie des communs » : les participants pensent à leur intérêt propre en ignorant l'impact sur l'intérêt collectif.

Par exemple, un agriculteur confronté à un problème immédiat, comme des insectes qui dégradent sa production de colza, ne va pas hésiter à choisir une solution avec des effets négatifs à longs termes : utiliser du Phosmet<sup>1</sup>, par exemple. Pourtant cet agriculteur connaît la nocivité de ce produit et il avait probablement arrêté de l'utiliser pour cette raison. Mais comme les autres pesticides ont été interdits et pas celui-là (parce que personne ne l'utilisait plus), il va de nouveau y recourir et... tant pis pour demain !

De telles tensions peuvent conduire à de véritables imbroglios politico-judiciaires, voire à des drames. C'est ce qui s'est passé au barrage de Sivens, un projet de barrage sur le cours du Tescou, un affluent du Tarn, dans le bassin de la Garonne. Pour les uns, la construction du barrage était indispensable à la survie de petites exploitations agricoles, aux productions variées. Pour les autres, il conduisait à la destruction d'une zone humide qui héberge de nombreuses espèces protégées. En 2014, un manifestant meurt lors d'affrontements avec la police. Le projet est abandonné par la suite. On peut s'attendre à la multiplication de ce genre de conflits entre biens privés d'aujourd'hui et communs indispensables à l'environnement de demain.

Ces problèmes débordent les frontières. Les États se montrent incapables de prendre les mesures qui s'imposeraient, comme une taxe sur l'extraction d'énergie fossile. Ils sont confrontés à des législations sur les biens publics et les biens privés, quand il s'agit de défendre des biens communs. Si tout le monde est concerné, personne n'est vraiment responsable du climat ou de la préservation de l'eau au niveau de la planète. La mondialisation s'est jusqu'ici surtout concentrée sur la défense des marchés et a oublié les dimensions environnementales. Les communs semblent être une des pistes de solutions. Pour citer

---

1. Le phosmet est un insecticide organophosphoré.

David Bolier, « le paradigme des communs recèle un potentiel inestimable pour réinventer des États en panne et réformer des marchés prédateurs ».

L'air, l'eau, les forêts, les rivières, etc., sont des biens communs (et clairement transfrontaliers). Leur gestion s'organise mal autour de la propriété privée ou même publique. Elle nécessite la mise en place d'une culture de partage, de nouvelles formes de gouvernance par une communauté de personnes, pour protéger ces ressources naturelles, ce que les États ou le marché n'ont pas su faire. Paradoxalement, c'est bien la responsabilité des États aujourd'hui, de s'approprier ces questions en encourageant le développement de communs, d'avoir le courage politique de visions à long terme ambitieuses, comme le Green Deal<sup>1</sup> européen.

Mais tout cela relève des communs matériels, revenons aux communs numériques.

### *Les communs numériques*

Dans le domaine des communs numériques, l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (l'Inrae) est en pointe de la promotion de la science ouverte. Cela a tenu d'abord à la personne d'un ancien P-DG de l'Inrae, François Houlier, qui a notamment réalisé en 2016 un rapport *Les Sciences participatives en France*. L'accent y est mis en particulier sur l'idée que la science doit s'ouvrir à des non-scientifiques, notamment les agriculteurs, pour qu'ils participent à la définition des objectifs, mais aussi au travail de recherche et d'innovation, et qu'ils s'approprient naturellement les résultats.

La science est appelée à sortir des laboratoires pour être partagée avec toutes les personnes concernées. Et quand on parle de partage,

---

1. Le Pacte Vert européen adopté en 2023 constitue un ensemble de mesures visant à faire de l'Europe, entre autres, un continent « climatiquement neutre » d'ici à 2050.

on tombe immédiatement sur l'accès ouvert aux articles scientifiques et aux données de la science.

Ce partage d'information et de données est également important pour une autre raison : la place du local en agriculture. La production dépend considérablement des conditions locales comme la nature du climat ou des sols. Le contexte économique dépend aussi beaucoup de la concurrence locale, des marchés locaux. Un pilotage au plus près de la production agricole doit donc s'appuyer sur des informations locales qui sont typiquement aux mains d'un grand nombre d'acteurs. Leur partage est indispensable, tout comme le partage de la capacité d'analyser ces données.

Qu'en est-il sur le terrain des communs numériques pour l'agriculture ? Pour qu'il y ait des communs numériques appliqués à l'agriculture, il faut d'abord qu'il y ait usage du numérique pertinent pour l'agriculture, puis que ces usages se prêtent à la mise en place de communs et enfin qu'il y ait des initiatives dans ce sens de la part des parties concernées. Cela conduit à nombre de questions comme : les usages du numérique dans l'agriculture font-ils sens ? Où en est leur développement ? Quelles sont les initiatives en cours pour la mise en place de communs ?

La réponse à la première question paraît positive et est mentionnée par de nombreux acteurs du domaine, comme OpenDataSoft, un acteur clé des données ouvertes en France :

Les données géospatiales, par exemple, permettent aux agriculteurs de repérer les parcelles les plus fertiles et de répandre les pesticides de manière sélective, tandis que les données météorologiques et tarifaires les aident à choisir les cultures les plus rentables à planter au cours d'une saison donnée. De même, les données sur les sols éclairent les décisions sur le type et la quantité d'engrais à utiliser, et les données fournies par les capteurs permettent de surveiller la croissance et la santé des plantes et du bétail<sup>1</sup>.

---

1. « 5 façons d'utiliser l'open data dans l'agriculture », Opendatasoft, 31 mars 2021, <https://www.opendatasoft.com/fr/blog/5-facons-dutiliser-lopen-data-dans-lagriculture/>.



Si on regarde maintenant les communs, on trouve quelques exemples d'initiatives qui vont dans cette direction.

L'ONG britannique Global Open Data for Agriculture and Nutrition soutient les efforts globaux pour rendre les données agricoles et nutritionnelles disponibles, accessibles et utilisables sans restriction à l'échelle mondiale. Concrètement, elle supporte les initiatives qui visent à avoir un impact positif sur l'agriculture et la nutrition grâce aux données ouvertes. L'initiative a été lancée en 2013, un an après le sommet du G8 en 2012, au cours duquel les dirigeants du G8 se sont engagés dans la Nouvelle Alliance pour la sécurité alimentaire et la nutrition en tant que prochaine phase d'un engagement commun pour parvenir à la sécurité alimentaire mondiale. Elle regroupe de nombreux partenaires, dont des gouvernements de divers pays.

« Frais et local » est une plateforme née d'un partenariat entre les chambres d'agriculture et le ministère de l'Agriculture et de l'Alimentation. Elle est ouverte à toutes les plateformes de vente directe de produits agricoles. L'idée est de fournir un point d'entrée unique aux consommateurs facilitant la recherche de produits fermiers de proximité et une visibilité renforcée aux agricultrices et agriculteurs adhérents des plateformes partenaires. Le consommateur y trouve la liste et l'accès géolocalisé aux sites de vente directe de l'agriculteur au consommateur.

Que fait l'État, et en particulier le ministère de l'Agriculture en France ? Comme souvent dans les situations où l'on pressent un besoin d'agir, un rapport est demandé à un spécialiste du domaine. En l'occurrence, c'est Jean-Marc Bournigal, président de l'Irstea (Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture), qui a été chargé du rapport. Sa mission était de préparer la mise en place opérationnelle d'un portail de données agricoles, et de proposer un plan d'action pour favoriser l'émergence de nouveaux services à partir des données du portail. Le rapport a été remis en janvier 2017 et a conduit à la création du Agdatahub, le Health Data Hub de la filière.

**Agdatahub.** La start-up Agdatahub propose la plateforme API-Agro, un catalogue sécurisé de données agricoles qualifiées. Les agriculteurs sont de plus en plus connectés. Leurs exploitations génèrent des masses considérables de données produites par des machines agricoles, des logiciels, des objets connectés, des stations météo, etc. Ces données ont une valeur considérable pour optimiser la production, la prévision, l'information à destination des pouvoirs publics ou du public. Basée sur la technologie Dawex et le *cloud* d'Orange, la plateforme facilite l'échange de données entre des agriculteurs, des éleveurs, leurs coopératives, des industriels, des instituts de recherche, des collectivités territoriales, etc. Les producteurs de données décident de leurs statuts, ouverts à tous ou plus ou moins limités. L'Agdatahub est le nœud français pour une mise en commun de données agricoles au niveau européen.

En résumé de ce rapide tour d'horizon, les communs numériques semblent faire sens et sont appelés à se développer dans l'agriculture, mais ils n'en sont pour le moment qu'au stade embryonnaire malgré des efforts réels des précurseurs pour les mettre en place.

Pour conclure, prenons un exemple qui tient de la conception ouverte. Quand on développe une semence qui a des caractéristiques très intéressantes, d'autres doivent pouvoir s'en servir. Il faut bien sûr rétribuer l'invention mais il ne faut pas la geler. Il faut laisser à d'autres la possibilité de l'utiliser et l'améliorer. On est bien dans les communs.

## CHAPITRE 23

# L'entreprise

Nous avons présenté les communs numériques comme une alternative au modèle économique libéral classique. On pourrait donc penser, *a priori*, que l'entreprise n'est pas véritablement concernée. Pourtant, nous avons vu que, dans différents usages comme la justice ou l'agriculture, les entreprises étaient bien présentes. Dans ce chapitre, nous revenons sur deux facettes de l'interaction des entreprises et des communs. La première est le logiciel *open source*, où les entreprises jouent un rôle considérable. Comme nous avons déjà abondamment traité le sujet, nous le mentionnons ici très rapidement. La seconde est l'ouverture des données dans des entreprises. Si cela nous conduit parfois hors du monde des communs au sens strict, avec des entreprises organisant des clubs de partage de données, nous verrons que cela reste très proche de l'esprit des communs.

### *Le logiciel open source*

Nous avons déjà abondamment traité le sujet dans les chapitres précédents. Nous avons vu comment les entreprises se sont approprié le modèle d'organisation des communs pour développer collectivement des logiciels. Nous en avons décrit différents aspects : le partage

des intérêts, la démographie de la communauté, et le modèle économique aux chapitres 6, 8 et 12. Nous avons aussi vu que dans certains communs de logiciels *open source*, on accepte bien volontiers la participation d'industriels, mais on se protège d'influences indues en spécifiant que les membres de la communauté sont là à titre individuel et que leur poids dans l'organisation est dû à leur apport et non à celui de l'entreprise à laquelle ils appartiennent. Pourtant, la participation de l'entreprise est bien réelle au-delà du simple financement, suivant les modes établis par la gouvernance à un commun numérique.

### *Les données et leur ouverture*

Si nous avons surtout rencontré la production de données par des organismes publics et par la foule (*crowdsourcing*), les entreprises privées produisent pourtant, elles aussi, des masses de données numériques. Leur arrive-t-il de les ouvrir ? Eh bien, oui.

Le mouvement est de bien moindre ampleur que dans le cas des organismes publics, au point que quand on parle de données ouvertes, on pense souvent implicitement à des données publiques. Cependant, la place de la mise en commun de données dans les entreprises est loin d'être négligeable. C'est ce que nous considérons maintenant.

Les sociétés cotées en Bourse (publiques, au sens anglo-saxon du terme) sont tenues à une certaine transparence qui exige qu'elles publient leurs comptes suivant des règles précises. L'idée est que, de par leur statut, elles sont « un peu » publiques et on exige d'elles la transparence des données qui permet à leurs actionnaires individuels de faire des choix financiers éclairés. En France, et dans d'autres pays européens (le Royaume-Uni par exemple), même les entreprises non cotées sont tenues à une certaine publicité : des informations administratives sont obligatoirement publiées (nom des responsables, changement de statut, adresse, etc.) et, suivant la taille de l'entreprise,

les comptes financiers doivent être publiés et sont disponibles dans le RNCS. Au-delà de ces obligations légales, il peut y avoir diverses raisons, pour une entreprise privée, de mettre en commun des données :

- 1) parce que l'entreprise peut mieux les valoriser en les partageant ;
- 2) pour le bien commun, celui de la société.

### *Valoriser les données par le partage*

Pour une entreprise, le partage des données peut aussi être un levier de recherche d'efficacité voire d'innovation.

Un consortium peut se créer pour échanger des données à l'intérieur d'un groupe d'entreprises. C'est ce que fait le consortium B3i, pour Blockchain Insurance Industry Initiative. Des assureurs et réassureurs se sont réunis pour créer une *blockchain* dans laquelle ils publient leurs contrats de réassurance, c'est-à-dire les contrats par lesquels ils partagent les risques. Cela leur permet d'avoir une vision consolidée sur ces risques, sur les primes, les sinistres et leurs remboursements, etc. Les entreprises participantes gardent le contrôle sur leurs données, et elles seules ont accès à la plateforme.

Cette recherche d'efficacité peut même parfois conduire des entreprises à ouvrir complètement les données. C'est le cas de la plateforme de données ouvertes d'Airbus, Skywise. Au départ, il y a le principe que, pour faire de l'analyse de données massives (du *big data*), il faut un maximum de données. Les données sont fournies par plus d'une centaine de compagnies et traitent de milliers d'avions. Ce sont des données massives recueillies par les pilotes, les ingénieurs, les ordinateurs de bord, etc. sur les vols, les avions, leur fabrication, leur maintenance, les incidents, etc. Ces données sont utiles à tous et peuvent servir, par exemple, à économiser du carburant ou à améliorer la maintenance des flottes.

Certaines entreprises ouvrent leurs données pour favoriser le développement d'autres entreprises qui pourraient contribuer à enrichir

leur offre. Elles encouragent ainsi l'existence d'un écosystème qui va favoriser leur croissance. C'est le cas de Twitter (bien avant Elon Musk). Quand le site de *microblogging* a ouvert ses données, on a vu fleurir des entreprises qui proposaient des outils d'engagement pour aider à publier du contenu sur Twitter, des outils d'analyse pour aider les entreprises à apprendre des connaissances sur un marché à partir du contenu Twitter, d'autres pour aider leurs clients à être plus visible sur la plateforme. Parfois, cette ouverture arrive à une certaine étape du développement de l'entreprise. Quand l'entreprise grandit, il est tentant pour elle de chercher à contrôler cet écosystème, voire à concurrencer ces mêmes entreprises dont elle a favorisé la création et dont le marché a démontré l'intérêt. C'est ce qu'on a pu observer avec Twitter autour de 2012.

L'ouverture de données par des entreprises rend possible des collaborations autour de ces données. Une telle coopération peut être génératrice d'idées comme le montre le club Datacraft, un lieu d'échanges entre *data scientists*. Dans de tels clubs, des ateliers, des hackathons, dans de nombreuses occasions de se former ensemble, d'innover ensemble, la force du collectif est à l'œuvre, un élément essentiel des communs. Et les données ouvertes facilitent ce travail.

Pour une entreprise, ses données peuvent constituer un patrimoine, un avantage industriel ou commercial. Donc, il est naturel qu'elle hésite à les confier à d'autres. Pour conclure ce chapitre, nous nous interrogeons sur les motivations pour une entreprise de partager ses données.

## *Le bien commun*

Ces données peuvent représenter pour la société une valeur inexploitée.

**AU SERVICE DU PUBLIC.** Il semble d'abord peu acceptable que des entreprises qui exercent des missions de service public comme

de favoriser les mobilités douces avec Vélib' cachent leurs données à l'État ou la collectivité qui leur a confié la mission. Même si ce n'est pas prévu par le contrat qui régit le service, l'ouverture des données paraît indispensable si ces données peuvent être utiles aux services publics. Et puis des entreprises peuvent disposer de masses de données qui pourraient permettre aux services de l'État ou de collectivité de mieux fonctionner. Cela peut être par exemple des données sur la mobilité dans telle ville, qu'une entreprise comme Uber accumule. Pour le bien collectif, les services publics devraient avoir accès à ces données.

**LA RECHERCHE ET L'INNOVATION.** Le cas de la recherche est particulièrement parlant : il est difficile d'accepter que les progrès des connaissances soient bridés par le fait que les chercheurs n'ont pas accès à des données dont ils ont besoin, qui existent dans des entreprises. Pour ce qui est de l'innovation, il serait également dommage de priver la société de nouveaux services parce que les entreprises qui possèdent les données nécessaires n'y ont pas pensé, ou parce qu'elles ne possèdent pas la technologie nécessaire. Il semble utile de favoriser le partage de leurs données avec d'autres entreprises, notamment des start-up, au bénéfice de services innovants.

**LA CONCURRENCE.** On réalise aussi aujourd'hui que les données peuvent engendrer des distorsions énormes de concurrence. Une entreprise peut se retrouver de fait en position de monopole parce qu'elle dispose de données que d'autres n'ont pas. En mobilisant le droit à la concurrence, la justice a tranché par le passé en faveur d'une ouverture de données. Par exemple, GDF-Suez a été obligé d'accorder à ses concurrents un accès à une partie des données de son fichier client pour limiter son monopole.

**LES DONNÉES D'APPRENTISSAGE.** Le partage de données est devenu particulièrement important pour l'entraînement des algorithmes d'apprentissage automatique. Dans des domaines comme les véhicules autonomes ou les outils de modération des réseaux sociaux, les données qui servent à entraîner les algorithmes sont indispensables.

Si on ne partage pas de telles données, on risque de plonger des pans entiers de l'économie dans des monopoles ou des oligopoles, de nombreuses entreprises de taille modeste ne disposant pas des moyens de construire ces corpus.

Tout cela constitue un éventail de raisons pour encourager le partage de données par les entreprises au profit du bien commun. Le terme consacré pour de telles données est : « données d'intérêt général » ; l'expression est explicitement mise en avant par la loi Lemaire en 2016. La loi Lemaire listait certaines données qui devraient *a minima* tomber dans cette catégorie : les données des délégations de service public (dans les transports, l'eau, la gestion des déchets, etc.), les données relatives aux subventions publiques supérieures à 23 000 euros ou encore les données de consommation d'énergie. Pourtant, il n'existe pas à ce jour de définition opérationnelle de ce que seraient les données d'intérêt général, ni de liste précise de ces données.

Les domaines d'application où on imagine facilement l'existence de tels corpus sont multiples : la santé, l'environnement, l'énergie, le transport, les villes intelligentes, ou la gestion des catastrophes naturelles. Le nombre de rapports officiels sur le sujet comme celui du CNNum en 2020 ou de Cytermann en 2015 montre l'intérêt général pour la notion. Le rapport Jutand en 2015 donne une bonne idée de ce que pourrait être une approche sectorielle du sujet dans un secteur spécifique, le transport.

Si le partage de ces données serait positif pour la société, on peut se demander pourquoi une entreprise qui les posséderait les ouvrirait à d'autres. Une première réponse : par altruisme.

**L'ALTRUISME.** C'est le moteur du *Governance Data Act* adopté par l'Union européenne en mai 2022. Pour une telle loi, le niveau européen est clairement le bon niveau, notamment parce que cela évite les distorsions de concurrence entre pays européens. Le partage de données est encouragé à des fins d'intérêt général en facilitant la mise à disposition volontaire de données privées. C'est l'entreprise



qui décide de déclarer des données comme d'intérêt général. Quand elle fait cela, elle bénéficie d'un cadre juridique protecteur, et peut préciser le contrôle qu'elle entend garder sur les données et comment ces données pourront être utilisées par d'autres.

L'altruisme peut rencontrer l'intérêt individuel d'une entreprise et la conduire à ouvrir des données, par exemple, dans le cadre de sa responsabilité sociétale, pour l'image positive que cela procure.

Prenons le cas d'Uber. De la bagarre entre chauffeurs d'Uber Pop et de taxis à Paris, jusqu'aux accusations d'exploitation de ses chauffeurs en passant par l'ubérisation de la société, cette entreprise est rarement présentée positivement dans la presse. Avec Uber Movement, Uber essaie de redorer son blason. Uber collecte et publie des données sur des milliards de trajets obtenues par l'app de ses chauffeurs ou celle de ses clients (en vélo ou scooter). Ces données contiennent des mines d'information sur la mobilité urbaine. Elles peuvent être utiles pour les villes concernées, par exemple pour améliorer des infrastructures de transport existantes, ou définir de nouvelles infrastructures. Des chercheurs peuvent aussi s'en servir pour étudier des phénomènes comme la propagation par effet domino des bouchons résultant du blocage d'une autoroute. Officiellement, Uber publie ces données pour le bénéfice des villes où l'entreprise est installée. On imagine bien qu'Uber cherche aussi, par ce biais, à approfondir ses partenariats avec ces villes, et à se rendre indispensable.

Les données d'Uber Movement sont sous une licence Creative Commons qui interdit les utilisations commerciales. Pourtant d'autres entreprises, par exemple, dans le domaine du tourisme (hôtels, restaurants, etc.) peuvent bénéficier de la transparence de la plateforme d'Uber comme elles peuvent bénéficier de celle de la plateforme DataVille de Airbnb. Ces entreprises peuvent ainsi mieux comprendre le marché, optimiser leur gestion et leur implantation.

**DES INCITATIONS PLUS FORTES.** L'altruisme, au cœur de l'approche des communs, a ses limites dans le cadre des entreprises. L'altruisme

ne peut être l'objectif principal pour une entreprise dont la raison d'être est le profit. En particulier, on peut craindre que les vellétés d'altruisme ne pèsent pas assez lourd pour convaincre les décideurs de l'entreprise face aux exigences de compétitivité. On imagine bien qu'il faudra un jour considérer des incitations plus fortes, voire imposer le partage, au minimum quand il s'agit de recherche ou dans des domaines où le partage de données est essentiel : la médecine, la mobilité verte, les territoires intelligents.

En conclusion, si les données d'intérêt général représentent une idée qui fait son chemin, il lui en reste encore pas mal à parcourir.

Nous avons donc vu que si l'entreprise privée n'est bien entendu pas le domaine de prédilection des communs numériques et de l'ouverture des données, ces deux notions y sont présentes, soit parce qu'elles servent l'intérêt économique des entreprises, soit parce que des incitations légales peuvent les y pousser.

PARTIE E

# Le temps et l'espace



Comment les communs numériques se développent-ils dans le temps et dans l'espace ? Une étude exhaustive nous conduirait sans doute trop loin. C'est pourquoi nous choisissons de discuter rapidement de certains aspects qui mettent en évidence le fait qu'ils se déploient de plus en plus, partout dans le monde.



## CHAPITRE 24

# L'histoire en bref

Les communs matériels remontent au moins jusqu'au Moyen Âge. Les communs numériques sont un phénomène relativement récent, une quarantaine d'années. C'est beaucoup dans l'histoire de l'informatique, mais peu à l'échelle de l'histoire de l'humanité.

### *La naissance*

Le premier commun numérique est le logiciel libre. Ses fondements se mettent en place avec la naissance d'Internet (autour de 1983) et du Web (en 1991) qui vont permettre à tout l'édifice de se construire. La séparation entre la branche *open source* et la branche logiciel libre s'officialise en 1998. Et c'est cette séparation qui accélère le mouvement à partir des années 2000. Le logiciel libre est la mère de tous les communs numériques. On lui reconnaît volontairement le rôle de précurseur et de *leader*. Il sert souvent de modèle pour la plupart de communs qui utilisent d'ailleurs souvent largement ce mode de logiciel.

Elinor Ostrom formalise la notion de communs à partir de 1992. (Elle obtient le prix Nobel d'économie en 2009.) Au début des années 2000, on voit naître Wikipédia et quasiment en même temps,

émerge le concept de science ouverte. Puis, l'idée des ressources éducatives libres émerge.

Plus tard, à la fin de cette décennie, vers 2010, on assiste au démarrage des données ouvertes publiques. Plus tardives que l'*open source*, les données ouvertes sont aussi largement adoptées dans l'ensemble des communs numériques.

Une chronologie de communs est donnée dans le tableau suivant (une chronologie plus complète est proposée en annexe, p. 285). Le choix de certaines dates est relativement arbitraire. Pour chaque commun, quand nous avons pu, nous avons choisi une personne clé. Si personne ne conteste à Richard Stallman la paternité de l'idée du logiciel libre ou à Jimmy Wales celle de Wikipédia, d'autres choix sont bien sûr plus discutables.

**Tableau 2.** Dates clé des principaux communs numériques.

Événement	Date	Lieu	Personne clé
Internet	1983	États-Unis	Vinton Cerf
Logiciel libre	1983	États-Unis	Richard Stallman
Formalisation du concept de commun	1990	États-Unis	Elinor Ostrom
Web	1991	Suisse	Tim Berners-Lee
Open Source	1998	États-Unis	Eric Raymond
Science ouverte	2001	États-Unis	
Wikipédia	2001	États-Unis	Jimmy Wales
Ressources éducatives libres	2002	États-Unis	
OpenStreetMap	2004	Royaume Uni	Steve Coast
Données ouvertes	2009	États-Unis	Barack Obama
Open Food Facts	2012	France	Stéphane Gigandet

Les licences associées suivent leur propre chronologie.



**Tableau 3.** Dates clé des principales licences de communs numériques.

Date	Acteurs	Licences
1989	Stallman	GPL
1991	Stallman	GPL V2, LGPL
1998	Université de Berkeley	BSD
2001	Lawrence Lessig, Aaron Swartz	Créative Commons
2007	Stallman	GPL V3
2009	OKFN	ODbL
2011	Etalab	Licence Ouverte
2013	CEA, CNRS, Inria	CeCILL

Voilà pour les naissances, qu'en est-il maintenant de l'évolution dans le temps ?

### *La croissance*

Les chiffres montrent une croissance régulière des communs numériques pour lesquels nous avons pu trouver des chiffres. Et dans le cas où le commun s'inscrit dans une activité économique identifiée, cette croissance se fait à une vitesse clairement supérieure à celle de l'ensemble du secteur économique.

Pour l'*open source*, si l'on mesure la croissance du marché associé en France, on voit un marché multiplié par 40 en moins de vingt ans passant de 200 millions d'euros en 2004 à 5,9 milliards en 2022. Le tableau ci-dessous montre l'évolution et la prévision d'évolution du marché de l'*open source* en France sur la période 2019-2027, avec une variation positive de la part de l'*open source* par rapport à l'ensemble du marché.

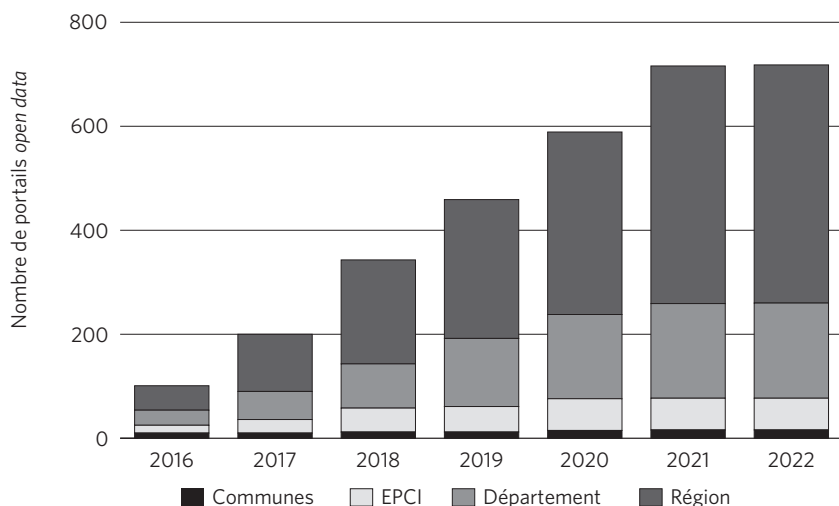
**Tableau 4.** Marché de l'open source en France.

	2019	2022	2027 (prévision)
Logiciels et services (en M€)	48 780	53 148	72 243
Dont <i>open source</i>	5 320	5 935	8 660
Pourcentage d' <i>open source</i>	10,9 %	11,2 %	12,0 %

Données : CNLL.

Pour la science ouverte, un indicateur significatif est le nombre de publications scientifiques en accès ouvert. Le chapitre 17 décrit la croissance régulière de cet indicateur pour la France.

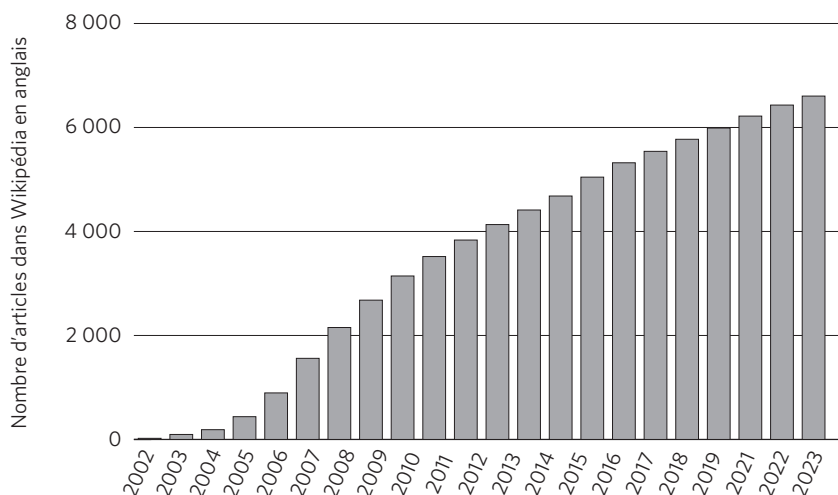
Pour les données ouvertes publiques en France, fions-nous à l'observatoire des données ouvertes des territoires publié par Open Data France, pour voir à quel rythme s'est poursuivie l'ouverture des données par les collectivités locales.



Données : Open Data France.

**Figure 5.** Nombre de portails *open data* en France par an et par type de collectivité.

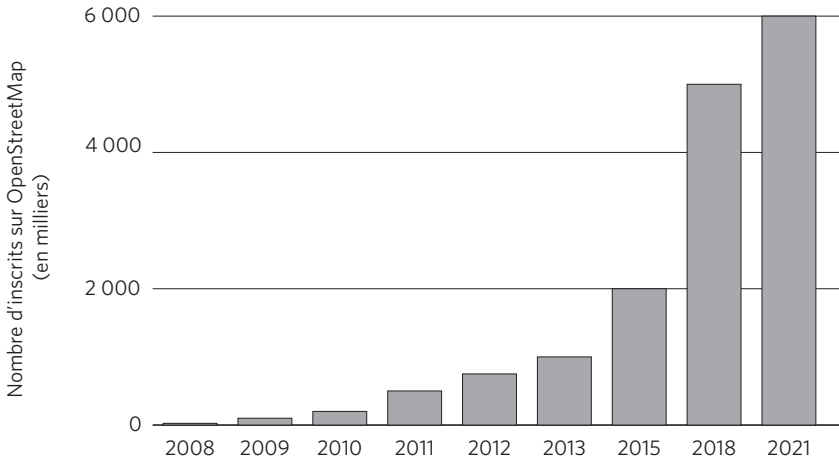
Une mesure possible de Wikipédia est le nombre d'articles dans l'encyclopédie. En novembre 2022, il est de 6,5 millions dans l'édition anglaise. Il continue de croître à un rythme solide : 17 000 par mois. Ce gain a été à son maximum en 2006 et a décliné régulièrement depuis. On en trouve l'illustration dans le graphique ci-dessous. Par contre, si l'on mesure la croissance de la taille de l'encyclopédie, elle reste autour de 1 gigaoctet par an, ce qui indique que la taille des articles augmente.



Données : Wikipédia.

**Figure 6.** Nombre d'articles dans Wikipédia en anglais par année.

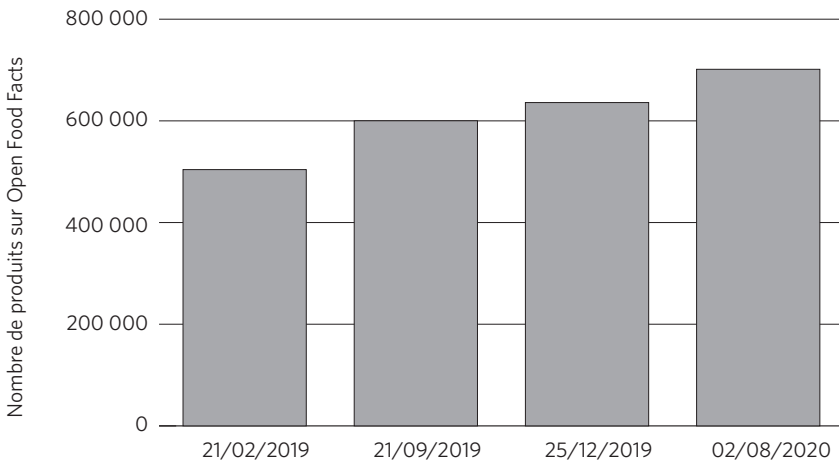
Chez OpenStreetMap, si l'on prend comme mesure de l'activité le nombre d'inscrits sur la plateforme, on constate une croissance forte, même si elle ralentit dans les dernières années, comme montré dans le graphique ci-dessous.



Données OpenStreetMap.

**Figure 7.** Nombre d'inscrits sur OpenStreetMap par an.

Chez Open Food Facts, si l'on prend comme mesure le nombre de produits référencés, la croissance est régulière comme montrée par le graphique ci-dessus.



Données : Open Food Facts.

**Figure 8.** Nombre de produits sur Open Food Facts par date.

Pour terminer ce rapide tour de l'histoire des communs numériques, le lecteur pourra se reporter à l'annexe, où les dates et événements clés de l'histoire des communs sont listés en ordre chronologique.

## CHAPITRE 25

# Géographie des données ouvertes publiques et de l'*open source*

L'origine des communs numériques est souvent états-unienne. C'est par exemple le cas pour Wikipédia, l'*open source* et les données ouvertes publiques. Il existe cependant des exceptions notables comme le Web, européen à l'origine, OpenStreetMap britannique, et Open Food Facts français.

Si les communs numériques se sont d'abord surtout répandus dans les pays occidentaux économiquement avancés, c'est qu'ils ont suivi la plupart du temps le taux de maturité numérique des pays. Sont-ils limités à ces pays ? Non. Pour éviter une analyse systématique par commun qui risquerait d'être fastidieuse, et sans doute aussi pour économiser le temps de nos lecteurs (et le nôtre), nous avons choisi de nous focaliser dans ce chapitre uniquement sur deux types de communs particulièrement incontournables, les données ouvertes publiques d'une part, l'*open source* et le logiciel libre, de l'autre.

## *Les données ouvertes publiques*

Plusieurs pays ont mis en place des outils de mesure de leur propre stade d'avancement des données ouvertes publiques. Comment mesurer cet avancement de façon objective ? On peut le faire de plusieurs façons : en comptant le nombre de jeux de données disponibles, de sujets traités par les données disponibles, de portails de données publiques, en constatant l'existence d'un portail central, la qualité des données, l'importance des données disponibles, ou l'usage qui en a été fait, etc. Il existe donc de nombreux indicateurs, souvent disponibles publiquement. Mais, de par la multiplicité des indicateurs, l'exercice, consistant à comparer les États en termes de données ouvertes publiques, n'est pas simple et comporte certainement une part d'arbitraire.

Un certain nombre d'organisations ont proposé des baromètres et les ont mis en œuvre de façon régulière dans le temps, ce qui nous permet des comparaisons et des mesures de l'évolution. Chaque organisation a sa propre vision de la bonne mesure, si bien que différents baromètres ont été proposés, permettant de comparer de façon plus ou moins objective les pays analysés. Ces baromètres diffèrent par leurs objectifs et les points sur lesquels ils mettent l'accent. À ce jour, une petite dizaine de *benchmarks* multipays est disponible, tous focalisés sur les données ouvertes publiques mis en œuvre par les gouvernements et leurs usages. On peut mentionner notamment :

- Open Data Barometer de l'Open Data Institute et la WWW Foundation ;
- Open Data Maturity in Europe de l'European Data Portal ;
- OURData Index de l'OCDE ;
- Global Open Data Index de l'OKFN (Open Knowledge Foundation).

Nous avons choisi de reprendre ci-dessous des résultats de l'Open Data Barometer, édités par l'Open Data Institute (un *think tank* britannique) et la WWW Foundation, qui nous ont semblé particulièrement intéressants. Le baromètre fournit des chiffres sur les 17 pays qui ont adopté la Charte de l'Open Data et les pays du G20, soit 30 pays en tout. Le baromètre se compose de trois notes qui mesurent le niveau de préparation, le niveau de mise en œuvre et l'impact de la publication des données.

Pour des raisons de lisibilité, nous nous sommes limités aux 15 premiers du classement.

**Tableau 5.** Classement des pays suivant l'Open Data Barometer.

Pays	Rang	Note globale	Préparation	Implémentation	Impact
Canada	1	76	86	87	55
Royaume-Uni	2	76	83	89	57
Australie	3	75	79	84	62
France	4	72	84	77	55
Corée du Sud	5	72	82	67	67
Mexique	6	69	79	67	62
Japon	7	68	78	68	58
Nouvelle-Zélande	8	68	79	72	52
États-Unis	9	64	79	76	37
Allemagne	10	58	76	72	27
Uruguay	11	56	71	70	28
Colombie	12	52	69	60	28
Russie	13	51	62	59	32
Brésil	14	50	63	56	30
Italie	15	50	61	61	27

Source : Open Data Barometer.



Le score de ces 15 *leaders* montre que la portée géographique de ce commun numérique est très large, et couvre l'Europe, les deux Amériques et la Russie.

Si l'on étudie les variations de ces notes dans le temps, on constate que 21 pays progressent, 8 restent à peu près stables et un seul régresse. Donc le mouvement d'ouverture des données se poursuit et s'amplifie. On voit avec plaisir que la France est dans le peloton de tête. Les États-Unis qui ont été précurseurs sur le sujet ne sont plus nécessairement *leader*. La Russie n'est pas mal placée. La Chine est en queue de peloton (24<sup>e</sup>), mais nous reviendrons sur ces deux pays.

## *L'open source et le logiciel libre*

Comment se développent l'*open source* et le logiciel libre dans le monde et qui en sont les principaux contributeurs ? Dans le cas des données ouvertes publiques, nous disposons de nombreux baromètres parmi lesquels choisir. Pour l'*open source*, nous ne disposons pas de tels baromètres, mais une méthode différente et relativement efficace s'impose : regarder les chiffres de GitHub, la plateforme au cœur du développement *open source*. GitHub fournit une solution très utilisée d'hébergement et de gestion de sources de programmes aux développeurs de logiciels. La plateforme est utilisée à ce jour par 73 millions de développeurs ! Fondée en 2008, la société GitHub a été initialement financée par ses trois fondateurs, puis après plusieurs injections de capital-risque, elle a été acquise par Microsoft en 2018 pour 7,5 milliards de dollars. Curieusement, le code de GitHub n'est pas lui-même en *open source*. GitHub est gratuit pour les projets *open source* et payant pour les autres.

Est-ce qu'un tel indicateur pourrait introduire un biais ? Github détient 84 % du marché de la gestion de codes sources<sup>1</sup>. Les développeurs du monde entier l'utilisent. S'il est possible que des pays spécifiques l'utilisent moins que d'autres, pour dégager les grands traits du déploiement des logiciels *open source* dans le monde, on peut raisonnablement se fier à ces chiffres. L'activité que l'on mesure sur GitHub est plus la contribution que l'utilisation. On parle donc plus de ceux qui donnent, que de ceux qui reçoivent même si on peut raisonnablement supposer que, pour chaque pays, il existe une corrélation entre développement et contribution. Donc, oui, cet indicateur est biaisé (en particulier si on s'intéresse à l'utilisation d'*open source*) mais il donne quand même des indications intéressantes.

Une façon de mesurer cette activité est de compter le nombre de *push* sur GitHub, un *push* correspondant à une mise à jour d'une version distante d'un logiciel à partir d'une version locale. Le nombre de *push* est une bonne façon de mesurer l'activité de contribution.

Nous nous limiterons là encore au top 15.

**Tableau 6.** Classement des pays suivant leur activité de push sur GitHub.

Pays	Nombre de <i>pushs</i>
États-Unis	1 709 062
Hongrie	428 786
Allemagne	319 787
Royaume-Uni	304 752
Chine	250 708
France	151 529
Canada	143 815
Inde	109 823

1. <https://www.slintel.com/tech/source-code-management/github-market-share>. URL NON VALIDE

Pays	Nombre de <i>pushs</i>
Japon	108 934
Brésil	104 802
Russie	101 331
Australie	90,174
Pays-Bas	75 198
Espagne	60 553
Suisse	53 652

Source : Felipe Hoffa, « Github Top Countries », Medium.

Si on normalise par rapport à la population, donc en comptant le nombre de *push* par habitant, on ne peut qu'être impressionné par les scores de la Hongrie et de l'Estonie. Ce sont des pays avec une forte culture informatique. Le passé communiste est-il en partie responsable ? Les États-Unis, premiers en volume absolu comme contributeurs, restent parmi les *leaders* pour les logiciels libres. Les pays Européens sont très présents ; d'autres ne sont pas loin dans la liste. La position de la France (24<sup>e</sup>) est un peu décevante ; il nous a fallu aller plus loin dans la liste pour qu'elle figure dans le palmarès. La Chine, le Japon, l'Inde et la Russie bien représentés en volume absolu, le sont moins quand on normalise par la population. On peut imaginer qu'ils sont beaucoup d'utilisateurs.

**Tableau 7.** Classement des pays suivant le nombre de push par habitant.

Pays	Nombre de push par habitant (en millième)
Hongrie	43
Estonie	28
Suisse	7

Pays	Nombre de push par habitant (en millième)
Norvège	6
Nouvelle-Zélande	6
États-Unis	6
Luxembourg	5
Suède	5
Islande	5
Royaume-Uni	5
Danemark	5
Pays-Bas	5
Canada	4
Finlande	4
Australie	4
Allemagne	4
Irlande	3
Belgique	3
Autriche	3
Grèce	3
République tchèque	2
Lettonie	2
France	2
Slovénie	2
Israël	1

Source : [www.whitesourcesoftware.com](http://www.whitesourcesoftware.com).

Les grandes entreprises de logiciel sont des utilisateurs de logiciel libre, mais elles sont aussi très actives en tant que contributeurs. Donc, intéressons-nous à la contribution de ces entreprises à l'*open*

*source*. En se basant sur le nombre de leurs employés qui contribuent, on obtient le classement suivant :

**Tableau 8.** Classement des entreprises par nombre de contributeurs.

Entreprise	Nombre de contributeurs
Microsoft	4 550
Google	2 267
Red Hat	2 027
IBM	1 813
Intel	1 314
Amazon	881
SAP	747
Facebook	710
GitHub	676

*Source : Statistics & Data.*

On trouve les GAFAM à la notable exception d'Apple ; il faut descendre à la 24<sup>e</sup> place pour la trouver. Cela confirme bien l'intuition qu'Apple n'est pas un fervent adepte de l'ouverture. On trouve 9 entreprises états-uniennes et une seule entreprise européenne (SAP), ce qui reflète la forte domination états-unienne dans ce domaine. S'il n'y a aucune entreprise chinoise à ce niveau, on les trouve en descendant un peu plus loin : Huawei (11), Tencent (13), Alibaba (16) et Baidu (28). En tenant compte de la taille des entreprises, notamment les BATX (Baidu, Alibaba, Tencent, et Xiaomi), l'équivalent chinois des GAFAM, on peut observer que l'*open source* est moins développé en Chine qu'en Occident. Et aucune entreprise russe ne figure dans les trente premières. Il s'agit ici de contributions, on ne mesure pas vraiment l'utilisation.

En résumé, l'*open source* est largement répandu dans le monde, les pays industrialisés et les grandes entreprises du numérique étant les premiers contributeurs et utilisateurs. On peut dire que ce mode de développement a été largement adopté par les industriels mondiaux.

## CHAPITRE 26

# La Chine et la Russie

Si une analyse particulière de l'appétence de chaque pays du monde pour les communs numériques serait fastidieuse, elle pourrait permettre de mieux comprendre, entre autres, l'influence du régime politique du pays sur cette appétence. On peut penser que les régimes qui placent le bien collectif avant celui de l'individu sont plus susceptibles de mettre en commun des données, de l'information, des connaissances. On peut penser également que des régimes totalitaires peuvent craindre des communs que les communautés apprennent à s'organiser avec moins d'État, et surtout que la transparence, notamment des données, ne leur soit préjudiciable. Il serait intéressant de disposer d'études poussées pour confirmer ou infirmer ces impressions. Pour apporter des débuts de réponses, nous considérons dans ce chapitre deux pays stratégiquement importants avec une bonne maîtrise de l'informatique : la Chine et la Fédération de Russie.

La Chine a maintenant ses propres Big Tech, des sociétés chinoises, leaders dans les technologies numériques dans certains domaines, grandes par leur revenu, leur nombre d'employés, leur capitalisation et leur part de marché national et mondial. Les BATX sont les équivalents chinois des GAFAM : Baidu (moteur de recherche), Alibaba (e-commerce), Tencent (jeux vidéo et réseaux sociaux) et Xiaomi (smartphone). On peut y ajouter d'autres grandes entreprises

chinoises comme ByteDance (partage de vidéos), Didi Chuxing (véhicule de transport avec chauffeur), Huawei (communications) et Meituan Dianping (vente en ligne).

Si la Russie n'en est pas là, elle cherche également à développer son industrie du logiciel. Comme la Chine, elle a su développer des alternatives à certains des GAFAM, notamment Yandex et Vkontakte, les équivalents de Google et Facebook, même si leur taille est bien plus modeste.

Pour chacun de ces deux pays, nous focalisons notre attention sur les logiciels *open source* et les données ouvertes publiques.

### *Le logiciel open source en Chine*

Comme beaucoup de pays, la Chine a une tradition de piratage massif de logiciels propriétaires venus notamment des États-Unis, et la pression sur le pays est continue pour se mettre en ordre vis-à-vis de ce problème. Une des solutions possibles pour les Chinois est bien sûr le remplacement de logiciels propriétaires piratés par des logiciels *open source*.

La Chine est devenue aujourd'hui un gros consommateur de logiciel *open source* et devient progressivement un des plus gros producteurs du monde. La communauté *open source* chinoise est en croissance organique depuis le début des années 2000 : la Chine est deuxième (derrière les États-Unis) en termes du nombre d'utilisateurs et du nombre d'usages (mesuré par le nombre de *forks* et de *clones*) sur GitHub. Parmi les cinq comptes GitHub les plus populaires, deux sont chinois (Ruan Yfeng et Evan You).

Cette conversion progressive à l'*open source* s'est produite sous la triple influence de l'État, des Big Tech et des start-up chinoises du numérique.

Le sujet du logiciel libre est déclaré important pour l'État. Par exemple, le MIIT (Ministry of Industry and Information Technology,



chargé de l'industrie et de la technologie) a déclaré son soutien pour OpenStack (une infrastructure logicielle pour le *cloud*) en 2014 et l'a promu auprès des entreprises publiques chinoises. Le MIIT a mis en place une fondation pour l'*open source* (OpenAtom Foundation), avec pour objectif d'avoir deux ou trois communautés avec une influence mondiale d'ici 2025.

Les Big Tech chinoises utilisent l'*open source* pour se développer et développer leur technologie. Leurs motivations sont celles que nous avons déjà mentionnées pour l'*open source* comme la gratuité ou la flexibilité. Les développeurs chinois comme ceux du monde entier préfèrent travailler avec de l'*open source* si leur employeur le permet. Dans certains domaines, des entreprises chinoises ont été des précurseurs dans l'utilisation de l'*open source*. Par exemple, JD.com a adopté Kubernetes en 2016, moins d'un an après la mise en *open source* du projet par Google. Pour Didi Chuxing (l'Uber chinois), l'*open source* est essentiel pour permettre de se développer rapidement.

On peut dater un basculement important vers l'*open source* de la « guerre de 3Q ». Cette guerre était le conflit légal entre Tencent et Qihoo360 (une société de sécurité Internet), commencée en 2010. Le sujet était de savoir si Tencent avait abusé de sa position dominante de marché en « empaquetant » (*bundling*, en anglais) son logiciel antivirus dans ses produits pour exclure Qihoo. Une dispute semblable avait opposé Microsoft et Netscape aux États-Unis à la fin du siècle dernier. Tencent a gagné, mais au prix de dommages considérables à sa réputation. Pour restaurer son image publique, l'entreprise a été conduite à libérer une partie de son code. Depuis, la plupart des Big Tech chinoises ont soit mis en *open source* de la technologie interne, soit créé de nouveaux projets *open source*. Par exemple, la plupart contribuent significativement financièrement au CNCF, un projet de la fondation Linux sur le *cloud*. En termes de contributions à ce projet, la Chine est troisième derrière les États-Unis et l'Allemagne.

Parmi les candidats à devenir les grands projets internationaux en 2025, on peut noter du côté des Big Tech chinoises : Apollo de Baidu (la voiture autonome), Open Harmony de Huawei (système d'exploitation) et OpenXuantie d'Alibaba (conception de semi-conducteurs RISC V). Et plus du côté start-up : TiDB/TiKV (base de données distribuée), OpenResty (API) et Océan Base (base de données distribuée).

D'autre part, Gitee a construit une alternative à GitHub et a reçu un investissement de Baidu en 2019 pour ce faire. Ce sujet est important, car à ce jour le lieu de développement international de l'*open source* est GitHub. La Chine peut-elle réussir à s'imposer internationalement sur ce créneau également ? Ce serait cohérent avec sa stratégie qui l'a vue développer avec succès ses propres plateformes équivalentes aux GAFAM. L'avenir le dira. Et nous rencontrerons la même tentation chez les Russes.

Doit-on parler d'*open source* (la recherche de l'efficacité économique dans le développement logiciel) ou de logiciel libre (la recherche de la liberté de création et d'utilisation dans le monde logiciel) ? De multiples rencontres avec des acteurs de cet espace nous amènent à penser qu'on peut parler des deux. Si les efforts chinois pour pousser l'*open source* relèvent clairement de la recherche d'un optimum économique, et si la majorité de nos interlocuteurs étaient de l'école *open source*, nous avons aussi eu l'occasion de rencontrer dans ce pays de vrais libristes que Richard Stallman aurait clairement reconnus comme ses enfants.

En résumé, la Chine a largement adopté, comme les États-Unis, l'*open source* comme mode de développement du logiciel, cette approche étant soutenue par l'État et les industriels dans le cadre d'une politique de souveraineté numérique.

## *Le logiciel open source en Russie*

La Russie bénéficie historiquement d'un système éducatif de qualité pour les sciences notamment en mathématiques, d'une importante force de bons développeurs logiciels dont la compétence est mondialement reconnue. Ces deux facteurs sont l'héritage de l'ancien système soviétique.

Comme la Chine, la Russie est coutumière d'un fort degré de piratage en matière de logiciel. La Business Software Alliance estimait en 2011 que 67 % du logiciel utilisé en Russie était piraté. Et là encore, le pays est sous pression internationale (américaine surtout) d'agir contre cet état de fait. Le logiciel *open source* se présente donc également comme une bonne réponse au problème.

En 2010, Vladimir Poutine a décrété que les agences gouvernementales devaient mettre leur logiciel en *open source* d'ici à 2015. Le plan initial était d'abandonner les produits Microsoft en faveur d'un OS *open source* russe basé sur Linux. La transition devait débuter en 2011. Les premiers essais de faire migrer certains secteurs du gouvernement vers l'*open source* ont jusqu'à maintenant échoué, essentiellement par manque de soutien politique. La guerre en Ukraine et l'interdiction de l'accès à la technologie américaine pourraient changer la donne.

Le gouvernement russe pousse au développement de logiciels alternatifs aux logiciels notamment de Microsoft, par exemple MyOffice de Kaspersky Lab comme alternative à Microsoft Office. Il est prévu de rendre obligatoire dans les écoles, les hôpitaux et les administrations publiques l'utilisation d'une plateforme de communication incluant e-mail et visio. L'objectif est de se passer de WhatsApp, Zoom et leurs équivalents. En 2019, Yandex a lancé son propre service vidéo et la société Rutube vient de se relancer avec un financement de Gazprom Media comme alternative à YouTube. Là où les alternatives russes ne sont pas disponibles, le Kremlin se satisfait

de l'approche *open source*. Par exemple, les administrations ont été poussées à abandonner Oracle pour PostgreSQL.

Le gouvernement russe prévoit aussi de développer une plateforme pour héberger le développement de logiciel libre, donc un concurrent de GitHub, suivant une stratégie similaire à celle des Chinois.

La tentation de couper les ponts avec l'Internet mondial reste une constante. L'ancien président Dmitri Medvedev avait déclaré que la Russie était prête juridiquement et techniquement à se couper du reste de l'Internet mondial et a fait la même déclaration sur le réseau SWIFT des paiements internationaux. La loi sur l'Internet souverain, votée en 2019, donne à la Russie la capacité de se couper du reste du Web, même si les experts techniques restent sceptiques sur la capacité technique de le faire.

Les solutions *open source* sont présentes dans à peu près tous les domaines. L'utilisation de l'*open source* dans l'infrastructure informatique, du support et de l'analytique. Les grands groupes ont fait le choix de l'*open source*, les barrières administratives ont été supprimées et les avantages compétitifs de l'*open source* sont maintenant visibles, un écosystème est en train de se mettre en place. Les utilisateurs ont par ailleurs commencé à contribuer et la frontière entre producteurs et consommateurs est de plus en plus floue. Un fond pour le développement de l'*open source* en Russie a été créé en 2021, la Fondation Russe de l'*open source*.

De même que pour la Chine, parle-t-on ici d'*open source* ou de logiciel libre ? Là encore, nous avons eu l'opportunité de rencontrer des contributeurs des deux écoles de pensée (avec un écart de philosophie important entre les deux).

Donc, comme pour la Chine, la lutte contre le piratage et la recherche de la souveraineté numérique, la recherche d'indépendance à l'égard des technologies logicielles de l'occident, poussent naturellement vers le logiciel libre, et ce avec la volonté de l'État, ce qui place la Russie parmi les nations actives en la matière.

## *Les données ouvertes publiques en Chine*

Les données ouvertes publiques en Chine, cela peut sembler à certains un oxymore. Mais il y a eu récemment des signes positifs de certaines agences gouvernementales, notamment le bureau national des statistiques qui maintenant publie des données sous forme lisible par machine et il y a surtout eu de nombreuses initiatives au niveau local. Le baromètre Open Data produit par l'OKFN place la Chine en position 34 sur 70, donc au milieu du peloton. L'index montre que le gouvernement chinois a fourni des données dans la plupart des grandes catégories, à l'exception notable des élections et des dépenses publiques.

La Chine n'a en revanche à ce jour pris aucune position nationale officielle en matière de données ouvertes publiques : ni le PCC, ni le gouvernement ne sont impliqués. Le mouvement est plutôt piloté par des gouvernements locaux, des citoyens, et des ONG.

En 2012, Shanghai et Beijing ont mis en ligne leur portail de données ouvertes avec des centaines de jeux de données. Les données publiées sur ces portails couvrent les mêmes sujets que la plupart des portails de données ouvertes publiques dans le monde : météo, qualité de l'air et de l'eau, commerce, transport, cartographie, etc. Ces villes sont aussi le siège d'expériences faites par des ONG et des chercheurs académiques.

À Beijing, dans le district de Xicheng, qui couvre la majorité de la vieille ville, un laboratoire fournit des données démographiques, socio-économiques et de cartographie pour la recherche. Pour éviter de violer les règles du gouvernement sur la publication des données, le laboratoire utilise le modèle de quasi-ouverture : il publie un catalogue de ses données pour que les personnes intéressées puissent venir au laboratoire et utiliser les données sur place.

À Shanghai, une ONG fournit des données ouvertes sur la qualité de l'air, l'eau et le sol. L'ouverture des données environnementales en Chine a ses limites et ces données soulèvent des doutes : les données centrales peuvent ne pas correspondre aux données locales et l'historique des données n'est pas toujours disponible. Pour faire face à ce manque de fiabilité, les ONG se débrouillent et combinent plusieurs approches : elles utilisent les données publiques disponibles, elles collectent des données supplémentaires via des opérations de *crowdsourcing*, elles émettent des demandes pour des données supplémentaires et font pression sur le gouvernement pour obtenir plus de données *via* la publication de rapports. Alibaba supporte en partie ce travail de fourmi via sa fondation.

Écoutons, Feng Gao, directeur d'Open Data China, une association fondée en 2014 en Chine qui fait le bilan des dix dernières années : « Les données ouvertes publiques existent en Chine mais ne fonctionnent que partiellement. Les données sont techniquement (donc lisible par machine), mais pas légalement (il n'y a pas de licence ouverte). Et ce parce que la licence ouverte est en conflit avec la norme sociale chinoise. Le partage des données est souvent limité : la licence n'est pas disponible, seuls certains éléments en sont fournis dans les conditions générales d'utilisation du site. Les données ouvertes publiques se sont diffusées en Chine de façon ascendante : du local au national. Il y a à ce jour 150 portails locaux, mais pas de portail national. Cela a commencé à Shanghai et à Beijing puis s'est diffusé dans les autres régions, surtout celles qui sont avancées économiquement. »

Le développement des données ouvertes publiques est pollué par le conflit entre l'ouverture des données et l'économie de la donnée. Chaque détenteur de données publiques est focalisé sur le retour sur investissement qu'il peut en tirer plutôt que sur l'ouverture et la transparence. Par exemple, Shanghai a défini sa stratégie de données ouvertes en séparant les données inconditionnellement ouvertes (les vraies données ouvertes) et les données conditionnellement ouvertes

(contrôlées et distribuées sous licence particulière). Et cette approche est en train d'être suivie par la plupart des portails locaux.

Si la proportion de sujets couverts par les données ouvertes nationales a augmenté régulièrement depuis 2018, la proportion de données régulièrement mise à jour a décliné.

Les situations sont souvent contrastées comme l'illustre l'exemple suivant de la publication de la qualité de l'air. La qualité de l'air en Chine a été mesurée et publiée par l'indice de pollution de l'air depuis 2000. En 2013, la mesure a basculé sur un autre indice, l'indice de qualité de l'air. Ce changement a provoqué une baisse de la publication des données dans ce domaine cette année-là, qui a été suivie par une augmentation importante en 2014 qui a fait plus que rattraper la baisse. L'Université Columbia qui compare les États-Unis, le Royaume-Uni et la Chine, pour ce qui est de l'ouverture des données environnementales, met le Royaume-Uni à 95/100, les États-Unis à 80/100 et la Chine à 50/100. Les raisons du 50 sont le fait que les données ne sont pas lisibles par des machines et la licence qui n'est pas libre. Les deux problèmes seraient faciles à corriger, mais pour ce qui est de la licence, le blocage est politique.

En résumé, les données ouvertes publiques existent en Chine, se développent progressivement, mais lentement, et butent clairement sur le problème de l'adoption d'une licence ouverte.

### *Les données ouvertes publiques en Russie*

Le mouvement pour les données ouvertes en Russie démarre en 2011 avec le président Medvedev qui lance une initiative en ce sens. En 2012, un groupe de travail est mis en place pour voir comment développer effectivement les données ouvertes et un ministre du gouvernement ouvert, Mikhail Abyzov, est nommé. La Fédération

de Russie, avec les membres du G8 a approuvé l'Open Data Charter pour faciliter le développement des données ouvertes publiques. Les principes de cette charte sont : données ouvertes par défaut, à jour et complètes, accessibles et utilisables, comparables et interopérables. Cette charte a constitué l'un des points de départ des données ouvertes dans la Fédération de Russie. La ville de Moscou a lancé sa propre plateforme (data.mos.ru) et le gouvernement régional de Ulyanovsk a lancé la première initiative « données ouvertes » avec le soutien de la Banque mondiale.

En 2013, 27 projets sont lancés. Mais la même année, la Russie quitte le partenariat pour le gouvernement ouvert qu'elle avait rejoint en 2012 et choisit de définir sa propre stratégie de gouvernement ouvert.

En 2014, le portail « Open Data » de la Fédération de Russie est mis en service et plus de 7 500 jeux de données sont disponibles. La structure du portail suit les principes de la charte et couvre un large spectre : gouvernement, économie, éducation, santé, culture, écologie, transport, sport, construction, divertissement, tourisme, électronique, commerce, cartographie sécurité et climat. Les données ouvertes sur les dépenses de l'État sont classées en 3<sup>e</sup> position dans le monde par l'Open Government Data Initiative. Le ministère des Finances travaille avec Infoculture pour organiser une compétition pour développer une application BudgetApps utilisant les données du ministère des Finances.

En 2015, un certain nombre de lois sont votées et de règlements publiés pour soutenir le partage d'informations publiques. Un hackathon attire 32 équipes de développeurs qui travaillent sur les données du budget.

Sur les 85 régions que compte la Fédération de Russie, 13 ont des portails de *données ouvertes* et 29 ont des données *ouvertes*. Le tout avec une synchronisation automatique avec le portail central de data.gov.ru.



À ce jour, data.gov.ru affiche plus de 26 000 jeux de données, dont 15 000 sur le thème du gouvernement et 1 500 sur celui de l'économie. Une école des données ouvertes, opendataschool.ru, a été créée. Le site Web clearspending.ru suit la dépense du gouvernement et sert à gérer plus de 12 millions de contrats, 270 000 sous-traitants et 900 000 fournisseurs.

Les acteurs clés de l'écosystème russe des données ouvertes publiques sont : 1) le gouvernement, responsable de l'analyse, de la planification et du développement des initiatives *données ouvertes publiques* ; 2) le personnel des gouvernements fédéraux et régionaux (analystes, managers et développeurs de portails) chargés du recueil de données et de la mise en place des sites ; 3) des entreprises de logiciels qui développent des services en utilisant des données publiques ; et enfin 4) des ONG comme Infoculture et Open City Foundation, des acteurs de la société civile, et du monde académique.

En conclusion, si l'on exclut le pas de deux autour de l'Open Government Partnership, la Fédération de Russie a bien avancé en matière d'ouverture des données, et de nombreux acteurs, y compris gouvernementaux, jouent le jeu.



# Conclusion

Ainsi s'achève notre parcours au sein des communs numériques. Nous avons essayé de partager un peu de notre passion pour ce nouveau modèle de production, de distribution et d'usage de ressources numériques, fondamentalement différent du modèle de l'entreprise privée ou du modèle étatique. Nous espérons avoir fait découvrir aux lecteurs pour qui ce domaine était déjà familier des aspects qui leur étaient moins familiers. Et à ceux pour qui la notion était relativement nouvelle, nous espérons être arrivés à montrer une partie de sa richesse, de sa complexité.

Que retenir de ce parcours ?

## *Autour des communs*

Les communs sont basés sur le triptyque ressources-communauté-gouvernance : la ressource est celle des objets numériques qu'il faut produire, distribuer et utiliser ; la communauté est l'ensemble des acteurs qui contribuent à la production de la ressource et à son utilisation ; et la gouvernance s'appuie sur des règles qui définissent le fonctionnement de cette communauté. La grande majorité des communs partagent d'autres traits comme la gratuité (des ressources

distribuées) et la transparence (des règles de gouvernance et de leur application).

Nous n'avons pas cherché à définir de façon stricte ce qu'est un commun. Certains systèmes de production de ressources numériques présentent un intérêt certain même s'ils ne collent pas à ce triptyque, s'ils ne collent pas parfaitement à l'idée qu'on pourrait se faire de communs. Nous avons choisi d'en parler aussi parce qu'ils sont parfois des étapes sur le chemin des communs, ou parce qu'ils partagent avec les communs l'esprit, les objectifs.

### *Une portée universelle*

Avec leur palette de formes, des logiciels aux données, avec leur belle variété de modèles économiques, leur grande diversité de modes de gouvernance, les communs numériques se portent bien. Ils sont en général dans une grande dynamique tant en nombre de nouveaux communs, qu'en croissance des communautés et des ressources. Nous avons également constaté un développement des communs dans de nombreux pays. Nous avons enfin montré qu'ils s'installaient dans de nouveaux domaines d'application.

### *Les menaces*

Nous avons parlé de formes modernes d'enclosures par des gouvernements qui cherchent à bétonner des morceaux du cyberspace, des entreprises qui essaient de s'approprier des communs numériques. Nous avons vu que les communs numériques étaient parfois fragiles, du fait de leurs modèles économiques, ou de leur gouvernance. Souvent, un problème fondamental est celui de la motivation des contributeurs, quand chacun peut profiter de la ressource gratuitement.

Malgré tous ces obstacles, les communs numériques se développent. Nous sommes convaincus qu'ils continueront à le faire parce qu'ils correspondent aux aspirations de partage de grandes communautés, parce qu'ils représentent aussi des outils exceptionnels pour des entreprises en quête de meilleures performances, pour des États en quête de plus d'efficacité – mais aussi de plus de participation des citoyens – et de souveraineté. Comme par le passé, cela ne se fera pas sans résistance, sans combats. Nous comptons sur l'action des États pour poursuivre et étendre leur soutien aux communs numériques, et sur l'adoption de nouvelles lois pour mieux les protéger. Nous comptons surtout sur l'implication de leurs membres et le dynamisme de leurs communautés.

Face aux défis écologiques, sanitaires, démographiques, nous aurons de plus en plus besoin de solidarité. Les communs en général ont énormément à offrir. Avec la place centrale du numérique dans les efforts qui nous attendent, les communs numériques nous paraissent donc particulièrement indispensables.



# Lexique

*Ce lexique explicite des sigles et acronymes du livre. Il a été principalement rédigé à partir de Wikipédia, en adaptant les définitions au cadre et objectifs de ce livre. Nous en profitons pour témoigner de notre reconnaissance à ce fabuleux commun numérique.*

**AMDAC (Administrateur ministériel des données, des algorithmes et des codes sources).** La circulaire du 27 avril 2021 du Premier ministre sur la politique publique de la donnée, des algorithmes et des codes sources met en place un tel administrateur par ministère.

**API (application programmer interface).** Façade clairement définie par laquelle un logiciel offre des services (fonctions, actions ou données) à d'autres logiciels afin que des données ou des fonctionnalités soient échangées.

**Arcep (Autorité de régulation des communications électroniques, des postes et de la distribution de la presse).** Autorité administrative indépendante française chargée de réguler les communications électroniques et postales et la distribution de la presse (entre autres).

**Arpanet (Advanced Research Projects Agency NETWORK).** Réseau à transfert de paquets de données conçu aux États-Unis par la

Defense Advanced Research Projects Agency (DARPA). Le projet fut lancé en 1966, mais Arpanet ne vit le jour qu'en 1969. Sa première démonstration officielle date d'octobre 1972.

**ASCII (American Standard Code for Information Interchange).** Norme informatique de codage de caractères apparue dans les années 1960. C'est la norme de codage de caractères la plus influente à ce jour.

**AWS (Amazon Web Services).** Division du groupe américain de commerce électronique Amazon, spécialisée dans les services de *cloud computing* à la demande pour les entreprises et particuliers.

**BALO (*Bulletin des annonces légales obligatoires*).** Publication éditée par la DILA, le BALO a été créé, par la loi du 30 janvier 1907 qui dispose que « l'émission, l'exposition, la mise en vente, l'introduction sur le marché en France d'actions, d'obligations ou de titres [...] devront faire insérer dans un bulletin annexe au *Journal officiel* une notice ».

**BAN (Base Adresse Nationale).** Base de données, mise en place en 2015, contenant la correspondance entre adresse postale (non nominative) et position géographique de plus de 25 millions d'adresses sur le territoire français. Développée et maintenue aujourd'hui par l'IGN, La Poste, OpenStreetMap et Etalab.

**BATX (Baidu, Alibaba, Tencent, Xiaomi).** Équivalent chinois des GAFAM : Baidu est principalement un moteur de recherche ; Alibaba une plateforme de commerce électronique et de paiement ; Tencent une plateforme de services Internet et de publicité en ligne ; Xiaomi un acteur de la téléphonie mobile et des smartphones.

**BNF (Bibliothèque nationale de France).** Organisme chargé de la collecte, l'archivage et l'entretien (conservation, restauration), de tout



ce qui se publie ou s'édite en France, ainsi que du patrimoine hérité des collections antérieures et reçu par d'autres voies (dons, legs, achats).

**BOAMP** (*Bulletin officiel des annonces de marchés publics*). Publié par la DILA, il contient les avis d'appels publics à la concurrence (AAPC) nationaux et européens, les avis de concession, les avis d'attribution, les contrats de partenariat public-privé lancés par l'État, l'armée, les collectivités territoriales et leurs établissements publics.

**Bodacc** (*Bulletin officiel des annonces civiles et commerciales*). Publié par la DILA, dans le cadre de sa mission de contribuer à la garantie de la transparence de la vie économique et financière, il assure la publicité des actes enregistrés au registre du commerce et des sociétés : les avis de ventes et cessions, les immatriculations, les créations d'établissement, les modifications et radiations d'une personne physique ou d'une personne morale.

**BSD (Berkeley Software Distribution)**. Ensemble de logiciels du système Unix distribué sous la licence BSD, licence de logiciel libre.

**CC (Creative Commons)**. Association à but non lucratif dont la finalité est de proposer une solution alternative légale aux personnes souhaitant libérer leurs œuvres des droits de propriété intellectuelle standard de leur pays.

**CEA (Commissariat à l'énergie atomique et aux énergies alternatives)**. Organisme divers d'administration centrale (ODAC) de recherche scientifique français dans les domaines de l'énergie, de la défense, des technologies de l'information et de la communication, des sciences de la matière, des sciences de la vie et de la santé.

**CeCILL (CEA CNRS Inria logiciel libre)**. Licence de logiciel libre compatible avec le droit français et définie en 2004 par les trois organismes mentionnés.

**CERN (Conseil européen pour la recherche nucléaire).** Centre de recherche européen de physique des particules situé sur la frontière franco-suisse.

**CNAM (Caisse nationale d'assurance-maladie).** Organisme qui gère, au niveau national, les branches maladie (risques maladie, maternité, invalidité, décès) et accidents du travail et maladies professionnelles du régime général de la Sécurité sociale, et pilote les organismes chargés de la mettre en œuvre.

**CNED (Centre national d'enseignement à distance).** Établissement public à caractère administratif du ministère de l'Éducation nationale proposant des formations à distance.

**CNIL (Commission nationale de l'informatique et des libertés).** Autorité administrative indépendante française chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

**CNRS (Centre national de la recherche scientifique).** Organisme public français de recherche scientifique exerçant ses activités dans tous les domaines de la connaissance.

**CSV (*comma separated values*).** Format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules.

**DAO (*decentralized autonomous organization*).** Organisation qui se fixe une mission et qui fonctionne grâce à un ensemble de *smart contracts*, permettant d'établir et de fournir des règles de gouvernance.

**DARPA (Defense Advanced Research Projects Agency).** Agence du département de la Défense des États-Unis chargée de la recherche et développement des nouvelles technologies à un usage militaire.

**DFX (*design for X* ou *design for excellence*).** Approche systématique, ensemble des règles à observer, dans la conception d'un produit, d'un service ou d'un processus pour atteindre un objectif dénoté par X. Il peut s'agir de la conception d'un circuit intégré ou d'un système électronique. Par exemple, DFA, *design for assembly* (en français, « conception de produits favorisant l'assemblage »).

**DILA (Direction de l'information légale et administrative).** Administration publique française chargée de la diffusion du droit, à son application dans la vie quotidienne des Français et l'accès à l'information.

**DINUM (Direction interministérielle du numérique).** Administration publique française chargée de coordonner les actions des administrations en matière de systèmes d'information.

**DMA (*Digital Market Act*).** Le DMA est un règlement de l'Union européenne qui vise à rendre l'économie numérique plus équitable et plus contestable en définissant des règles pour les grandes plateformes en ligne.

**DORA (San Francisco Declaration on Research Assessment).** Déclaration publiée par des membres de la Société américaine de biologie cellulaire et un groupe d'éditeurs de revues scientifiques qui remet en cause l'usage croissant du classement bibliométrique comme indicateur de qualité dans l'évaluation de la recherche et des chercheurs. Il a été adopté par un grand nombre d'associations scientifiques.

**DRM (*digital right management*).** Dispositifs de contrôle de l'utilisation des œuvres numériques par leur chiffrement.

**DSA (*Digital Service Act*).** Le DSA est un règlement proposé par la Commission européenne qui vise à réguler les contenus en ligne, notamment les contenus illégaux, et la désinformation.

**DWG (DraWinG).** Format natif des fichiers de dessins AutoCAD, standard de facto de l'industrie CAO et DAO.

**EOSC (European Open Science Cloud).** Initiative de la Commission européenne qui vise à développer un dispositif fournissant à ses utilisateurs des services de *cloud computing* pour les pratiques de science ouverte.

**ePub (*electronic publication*).** Format ouvert standardisé pour les livres numériques et fondé sur le XML.

**ERCIM (European Research Consortium for Informatics and Mathematics).** Fondé en 1989 par trois instituts de recherche : Inria (France), GMD (Allemagne) et CWI (Pays-Bas), il a pour mission de stimuler la collaboration à l'intérieur de la communauté européenne de la recherche et d'augmenter la coopération avec l'industrie européenne. Il héberge la branche européenne du W3C.

**ESN (*entreprise de services du numérique*).** Société de services experte dans le domaine des nouvelles technologies et de l'informatique, anciennement appelée SSII.

**FAIR (*findability, accessibility, interoperability, and reuse*).** Critères auxquels doivent obéir les données ouvertes : faciles à trouver, accessibles, interopérables et réutilisables.

**FCC (Federal Communications Commission).** Agence indépendante du gouvernement fédéral des États-Unis qui réglemente les communications par radio, télévision, fil, satellite et câble à travers les États-Unis.

**FSF (Free Software Foundation).** Organisation américaine à but non lucratif fondée par Richard Stallman en 1985, dont la mission est la promotion du logiciel libre et la défense des utilisateurs.

**FUN (France université numérique).** Plateforme française visant à promouvoir l'utilisation des MOOC supportée par le logiciel open source américain open edX.

**GAFAM (Google Apple Facebook Amazon Microsoft).** Les cinq grands groupes états-unis qui dominent le marché du numérique : Google (moteur de recherche), Apple (produits électroniques grand public), Facebook (réseaux sociaux), Amazon (commerce électronique), Microsoft (logiciels de base et grand public).

**GFII (Groupement français de l'industrie de l'information).** Association d'acteurs publics et privés de l'information numérique professionnelle regroupant une centaine d'organisations professionnelles adhérentes.

**GIF (*graphics interchange format*).** Format d'image numérique couramment utilisé sur Internet.

**GNU (*Gnu is not Unix*).** Le projet GNU est un projet logiciel débuté en 1984 par Richard Stallman pour développer le système d'exploitation GNU.

**GPL (*general public license*).** Licence qui fixe les conditions légales de distribution d'un logiciel libre du projet GNU.

**HDH (*health data hub*).** Groupement d'intérêt public, gérant une plateforme de croisement des données de santé à des fins de recherches médicales et visant l'amélioration de la qualité des soins.

**HTML (*HyperText Markup Language*).** Langage de balisage conçu pour représenter les pages Web. Les balises permettent de formater et de structurer la page.

**HTTP (*Hypertext Transfer Protocol*), HTTPS (*Hypertext Transfer Protocol Secure*).** Protocole de communication client-serveur

développé pour le World Wide Web, et sa variante sécurisée par le chiffrement.

**ICANN (Internet Corporation for Assigned Names and Numbers).** Société de droit californien à but non lucratif qui administre les ressources numériques d'Internet, gère le système de noms de domaine de premier niveau (génériques et nationaux).

**IEEE (Institute of Electrical and Electronics Engineers).** Organisation à but non lucratif de droit américain, association professionnelle de plus de 400 000 membres avec différentes branches dans le monde entier.

**IETF (Internet Engineering Task Force).** Organisme de standardisation ouvert qui élabore et promeut des standards Internet, en particulier les standards qui composent la suite de protocoles Internet TCP/IP.

**IGN (Institut national de l'information géographique et forestière).** Établissement public à caractère administratif ayant pour mission d'assurer la production, l'entretien et la diffusion de l'information géographique de référence en France.

**IMDb (Internet Movie Database).** Base de données en ligne sur le cinéma mondial et sur la télévision, rachetée en 1998 par Amazon.

**INPI (Institut national de la propriété intellectuelle).** Établissement public chargé de la gestion de la propriété industrielle, recevant les dépôts et délivrant les titres (brevets, marques, dessins et modèles), centralisant en outre le RNCS.

**Inrae (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement).** Établissement public à caractère scientifique et technique issu de la fusion en 2020 de l'INRA et

de l'IRSTEA, chargé des travaux de recherche scientifique et technologique dans les domaines de l'agriculture, de l'alimentation, de la forêt, de l'environnement, de l'eau, de la biodiversité, de la bioéconomie, de l'économie circulaire, de la gestion durable des territoires et des risques dans les champs de compétence précités.

**Inria (Institut national de recherche en informatique et automatique).** Établissement public à caractère scientifique et technique chargé du développement de la recherche et de la valorisation en informatique et mathématiques appliquées, placé sous la double tutelle du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, et du ministère de l'Économie et des Finances.

**IP (*Internet protocol*).** Famille de protocoles de communication de réseaux informatiques conçus pour être utilisés sur Internet. IPv4 et IPv6 sont deux versions de ces protocoles, le second étant appelé à remplacer progressivement le premier.

**IPX (*Internetwork packet exchange*).** Protocole de communication qui transmet des paquets à travers un réseau local.

**JPG, JPEG (Joint Photographic Expert Group).** Norme qui définit le format d'enregistrement et l'algorithme de décodage pour une représentation numérique compressée d'une image fixe.

**JSON (JavaScript Object Notation).** Format de données textuel dérivé de la notation des objets du langage JavaScript.

**KML (Keyhole Markup Language).** Langage fondé sur le formalisme XML et destiné à la gestion de l'affichage de données géospatiales dans les logiciels de SIG.

**LAMP (Linux Apache MySQL PHP).** Ensemble de logiciels libres permettant de construire des serveurs de sites Web et composé de

Linux (un système d'exploitation), Apache (un serveur Web), MySQL (un serveur de base de données) et PHP (un langage de script).

**LAN (*local area network*).** Réseau informatique, de taille restreinte, où les terminaux qui y participent s'envoient des messages sans utiliser l'accès à Internet.

**LGPL (*lesser general public licence*).** Version dérivée de la licence GPL qui permet de lier un programme A à un programme B en LGPL sans que le programme A soit assujéti aux contraintes de la LGPL.

**MIDI (*Musical Instrument Digital Interface*).** Protocole de communication et de commande pour les instruments de musique.

**MIT (*Massachusetts Institute of Technology*).** Université et institut de recherche américains privés situés à Cambridge près de Boston dans la Massachusetts, spécialisés dans les domaines de la science et de la technologie.

**MOOC (*massive open online course*).** Cours ouvert de formation à distance capables d'accueillir un grand nombre de participants.

**NFT (*non fungible token*).** Donnée valorisée composée d'un type de jeton cryptographique qui représente un objet, auquel est rattachée une identité numérique (reliée à au moins un propriétaire) et authentifiée grâce à une *blockchain*.

**NSA (*National Security Agency*).** Organisme gouvernemental du département de la Défense des États-Unis, chargé de la surveillance, de la collecte et du traitement d'informations et de données à l'échelle mondiale à des fins de renseignement et de contre-espionnage national et international. La NSA est également chargée de la protection des réseaux de communication et des systèmes d'information américains.



**Odt (*open document text*)**. Format ouvert de données pour les applications bureautiques : traitements de texte, tableurs, présentations, etc.

**OKFN (Open Knowledge Foundation)**. Association à but non lucratif de droit britannique, créée en 2004 à Cambridge au Royaume-Uni, promouvant la culture libre, en particulier les contenus libres et les données ouvertes.

**Open RAN (*open radio access network*)**. Réseau d'accès sans fil ouvert est basé sur l'interopérabilité et la standardisation d'éléments de réseau d'accès sans fil, notamment sur une norme unifiée d'interconnexion pour des éléments matériels et des éléments logiciels *open source* issus de différents fournisseurs.

**OSI (Open Systems Interconnection)**. Normes de communication, en réseau, de tous les systèmes informatiques, proposées par l'ISO.

**OWL (Web Ontology Language)**. Langage de représentation de connaissances construit sur le modèle de données de RDF. Il fournit les moyens de définir des ontologies structurées.

**PDF (Portable Document Format)**. Langage de description de page présenté par la société Adobe Systems et devenu une norme ISO en 2008.

**PGO (Partenariat pour gouvernement ouvert)**. Partenariat multilatéral visant à promouvoir un gouvernement ouvert, en faisant notamment levier sur le numérique et les nouvelles technologies.

**PMSI (Programme de médicalisation des systèmes d'information)**. Regroupe les données des patients dans le cadre de la médecine hospitalière.

**PNG (*portable network graphics*)**. Format ouvert d'images numériques.

**RCS (Registre du commerce et des sociétés), RNCS (Registre national du commerce et des sociétés).** Registre recensant l'ensemble des sociétés et les informations les concernant et géré par un tribunal de commerce. Le RNCS regroupe l'ensemble des données des différents RCS et est géré par l'INPI.

**RDF (Resource Description Framework).** Standard, développé par le W3C, destiné à décrire formellement des descriptions de graphes, en particulier, les ressources Web et leurs métadonnées, afin de permettre le traitement automatique de telles descriptions.

**REL (ressources éducatives libres).** Matériaux d'enseignement, d'apprentissage ou de recherche appartenant au domaine public ou publiés avec une licence de propriété intellectuelle permettant leur utilisation, adaptation et distribution à titre gratuit.

**RGE (référentiel à grande échelle).** Référentiel géographique pour la France produit par l'IGN.

**RGPD (Règlement général sur la protection des données).** Règlement de l'Union européenne relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, texte de référence en matière de protection des données à caractère personnel.

**RNA (Répertoire national des associations).** Registre national des associations en France, développé et maintenu par le ministère de l'Intérieur, disponible en données ouvertes.

**ROME (Répertoire opérationnel des métiers et des emplois).** Répertoire créé en 1989 par l'ANPE, identifiant des métiers et des emplois et comprenant plus de 10 000 appellations de métiers et d'emplois.

**RPG (Registre parcellaire graphique).** Système d'information géographique permettant l'identification des parcelles agricoles.

**RSS (Really Simple Syndication).** Famille de formats de données utilisés pour la syndication de contenu Web.

**SHP (*shapefile*).** Format de fichier pour les systèmes d'information géographique (SIG). Initialement développé par la société ESRI pour ses logiciels commerciaux, désormais un standard *de facto*, aux spécifications ouvertes.

**SIG (Système d'information géographique).** Logiciel informatique permettant d'utiliser des données géographiques.

**SNDS (Système national des données de santé).** Système informatique qui unifie des bases de données de santé existantes (notamment le SNIIRAM et le PMSI), géré par la Caisse nationale de l'assurance-maladie des travailleurs salariés.

**SNIIRAM (Système national d'information inter-régimes de l'assurance-maladie).** Base de données sur les soins et les remboursements du secteur libéral (médecine de ville).

**SQL (Structured Query Language).** Langage informatique pour interroger les bases de données relationnelles, c'est-à-dire de données organisées sous forme de tables.

**SRI (Stanford Research Institute).** À l'origine centre de recherche basé sur le campus de l'Université de Stanford, maintenant le SRI est devenue une entreprise privée faisant de la recherche sous contrat et basée à Menlo Park, Californie.

**TCP/IP (Transmission Control Protocol/Internet Protocol).** Ensemble de protocoles utilisés pour le transfert de données sur Internet.

**UCLA (University of California at Los Angeles).** Université de Californie à Los Angeles. Son sigle est couramment rencontré sur les T-shirts de jeunes et de moins jeunes.

**W3C (World Wide Web Consortium).** Organisme de standardisation à but non lucratif, fondé en 1994 chargé de promouvoir la compatibilité des technologies du World Wide Web telles que HTML5, HTML, XHTML, XML, RDF, SPARQL, CSS, XSL, PNG, SVG, MathML et SOAP.

**WAN (*wide area network*).** Définit un réseau informatique étendu couvrant une grande zone géographique (par opposition au LAN), typiquement à l'échelle d'un pays, d'un continent, ou de la planète entière. Le plus grand WAN est le réseau Internet.

**XML (Extensible Markup Language).** Langage informatique de balisage permettant d'ajouter dans un texte des marqueurs (appelées « balises ») donnant des informations sur le contenu du texte, de formatage par exemple. Sa syntaxe est dite « extensible » en ce qu'elle permet de définir différents langages avec pour chacun son vocabulaire et sa grammaire.

# Annexe

Date	Commun	Événement
1969	<i>Internet</i>	Premier échange de message entre UCLA et SRI
1978	<i>Données ouvertes</i>	Loi CADA en France
1983	<i>Logiciel libre</i>	Richard Stallman publie les plans de GNU
1985	<i>Logiciel libre</i>	Création de la Free Software Foundation
1990	<i>Web</i>	Tim Berners Lee diffuse le logiciel dans le CERN
1991	<i>Logiciel libre</i>	Première diffusion de Linux par Linus Torvalds
1993	<i>Web</i>	Diffusion de Mosaic et usage public du Web
1994	<i>Web</i>	Fondation du W3C, 2 700 sites Web
1998	<i>Open source</i>	Tim O'reilly organise l'Open Source Summit
	<i>Web</i>	Fondation de Google
2000	<i>Wikipédia</i>	Fondation de Nupedia
2001	<i>Creative Commons</i>	Fondation par Lawrence Lessing <i>et al.</i>
	<i>Science ouverte</i>	Initiative de Budapest, mise en ligne de HAL
	<i>Wikipédia</i>	Fondation de Wikipédia par Jimmy Wales
	<i>Éducation ouverte</i>	MIT Open Courseware Project

Date	Commun	Événement
2002	<i>Creative Commons</i>	Publication de la version 1.0
	<i>Données ouvertes</i>	Ouverture des données juridiques en France
	<i>Éducation ouverte</i>	L'Unesco adopte le terme « <i>open educational resources</i> »
2003	<i>Données ouvertes</i>	Directive européenne sur les données publiques
2004	<i>Open source</i>	Le CEA, le CNRS et l'Inria publient la licence CeCILL
	<i>Wikipédia</i>	Bataille de l'endive
	<i>OpenStreetMap</i>	Steve Coast créé le projet au Royaume-Uni
2005	<i>Données ouvertes</i>	Ordonnance sur la réutilisation de l'information publique en France
	<i>Wikipédia</i>	<i>Nature</i> publie un article comparant Wikipédia et l' <i>Encyclopædia Britannica</i>
2007	<i>Open source</i>	Création du GTLL du pôle de compétitivité Systematic en France
	<i>Données ouvertes</i>	Sebastopol Meeting (O'Reilly et Lessig)
2008	<i>Open source</i>	Microsoft rejoint la fondation Apache & Lancement de Github
2009	<i>Données ouvertes</i>	Barack Obama lance l'Open Government Initiative et data.gov
	<i>Données ouvertes</i>	Claire Gallon crée Libertic en France
2011	<i>Données ouvertes</i>	Création d'Etalab en France
	<i>Science ouverte</i>	Mise en ligne de Sci-Hub par Alexandra Elbakyan
2012	<i>Wikipédia</i>	Fin de la version papier de l' <i>Encyclopædia Britannica</i>
	<i>Open Food Facts</i>	Création par Stéphane Gigandet
2013	<i>Open source</i>	Publication du Socle interministériel des logiciels libres en France

<b>Date</b>	<b>Commun</b>	<b>Événement</b>
<b>2014</b>	<i>Open source</i>	Satya Nadella déclare : « <i>Microsoft loves Linux.</i> »
	<i>Données ouvertes</i>	La France rejoint le partenariat pour un gouvernement ouvert
<b>2016</b>	<i>Web</i>	1 milliard de sites
<b>2016</b>	<i>Données ouvertes</i>	Lancement de la BAN par La Poste, l'IGN et OpenStreetMap
<b>2017</b>	<i>Open Data France</i>	Ouverture du fichier Sirene,
	<i>Science ouverte</i>	Loi Lemaire « pour une république numérique »
<b>2018</b>	<i>Open source</i>	Microsoft rachète GitHub
	<i>Science ouverte</i>	Programme européen pour la science ouverte
	<i>Open Food Facts</i>	Première subvention (Santé publique France)
<b>2019</b>	<i>Science ouverte</i>	Mise en place du baromètre de la science ouverte en France
	<i>Open Food Facts</i>	1 million de produits référencés





# Bibliographie et repères

## CHAPITRE 1. LES COMMUNS MATÉRIELS

Elinor Ostrom, *Governing the Commons : The Evolution of Institutions for Collective Action*, Cambridge University Press, 2015.

Elinor Ostrom, *Discours de Stockholm, en réception du Nobel d'économie 2009*, préface de B. Coriat, traduit de l'anglais par J. Demazière et H. Le Crosnier, C&F Éditions, 2009.

Garett Hardin, « The tragedy of the commons », *Science*, 1968, 162 (3859), p. 1243-1248.

Marie Cornu, Fabienne Orsi, Judith Rochfeld (dir.), *Dictionnaire des biens communs*, PUF, 2021.

## CHAPITRE 2. LES COMMUNS NUMÉRIQUES

Charlotte Hess, Elinor Ostrom (dir.), *Understanding Knowledge as a Commons : From Theory to Practice*, MIT Press, 2006.

Hervé Le Crosnier, « Une introduction aux communs de la connaissance », *Tic & Société*, 2018, 12 (1), p. 13-41, <https://doi.org/10.4000/ticetsociete.2481>.

Hervé Le Crosnier, *En communs. Une introduction aux communs de la connaissance*, C&F Éditions, 2015.

« Les communs numériques », blog Binaire, *Le Monde*, <https://www.lemonde.fr/blog/binaire/page-les-communs-numeriques/>.

Judith Rochfeld, « Les communs numériques », blog Binaire, *Le Monde*, 28 septembre 2021, <https://www.lemonde.fr/blog/binaire/2021/09/28/les-communs-numeriques/>.

## CHAPITRE 7. LES CONNAISSANCES

Chris Anderson, *Makers : The New Industrial Revolution*, Crown Business, 2012 ; traduction française, *Makers. La nouvelle révolution industrielle*, Pearson, 2017.

## CHAPITRE 8. LES RÉSEAUX COMME BIEN COMMUN

Lawrence Lessig, *L'Avenir des idées. Le sort des biens communs à l'heure des réseaux numériques*, PUF, 2005

Félix Tréguer, *L'Utopie déçue. Une contre-histoire d'Internet XV<sup>e</sup>-XXI<sup>e</sup> siècle*, Fayard, 2019.

Serge Abiteboul, Valérie Peugeot, *Terra Data. Qu'allons-nous faire des données numériques ?*, Le Pommier, 2017.

Marc Augier, *Les Logiciels*, PUF, « Que sais-je ? », 1998.

## CHAPITRE 9. LA COMMUNAUTÉ

Nadia Eghbal, *Roads and Bridges : The Unseen Labor Behind Our Digital Infrastructure*, Fondation Ford, <https://www.fordfoundation.org/media/2976/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure.pdf>.

## CHAPITRE 10. LES LICENCES

François Pellegrini, Sébastien Canevet, *Droit des logiciels. Logiciels privatifs et logiciels libres*, PUF, 2013.

Lawrence Lessig, « The Creative Commons », *Florida Law Review*, 2003, 55, p. 763-777.

« Présentation générale. Une géopolitique de la propriété intellectuelle », dans Mélanie Dulong de Rosnay, Hervé Le Crosnier (dir.), *Propriété intellectuelle. Géopolitique et mondialisation*, CNRS Éditions, 2013, p. 9-35.

## CHAPITRE 11. GOUVERNANCE DES COMMUNS

« Linux – Governance », P2P Foundation, [https://wiki.p2pfoundation.net/Linux\\_-\\_Governance](https://wiki.p2pfoundation.net/Linux_-_Governance).

Gilles Sahut, « La gouvernance de Wikipédia : élaboration de règles et théorie d'Ostrom », *Tic & société*, 2018, 2 (1), p. 167-200, <https://journals.openedition.org/ticetsociete/2426#tocto2n2>.

## CHAPITRE 17. LA SCIENCE

Bernard Rentier, *Science Ouverte, le défi de la transparence*, Université de Liège, 2018, [https://orbi.uliege.be/bitstream/2268/230014/1/rentier\\_science\\_ouverte\\_pour\\_ORBi.pdf](https://orbi.uliege.be/bitstream/2268/230014/1/rentier_science_ouverte_pour_ORBi.pdf).

## CHAPITRE 19. L'ÉTAT

*Partenariat pour un gouvernement ouvert : quels engagements autour des logiciels libres ?*, Etalab, 2022.

Éric Bothorel, *Pour une nouvelle ère de la politique publique de la donnée, des algorithmes et des codes sources*, rapport au Premier ministre, 2021.

*Des barbelés sur la prairie Internet : contre les nouvelles enclosures, les communs numériques comme leviers de souveraineté*, France Diplomatie, ministère de l'Europe et des Affaires étrangères, 2020, <https://www.diplomatie.gouv.fr/fr/politique-etrangere-de-la-france/diplomatie-numerique/blog-de-l-equipe/article/des-barbelés-sur-la-prairie-internet-contre-les-nouvelles-enclosures-les>.

## CHAPITRE 21. LA JUSTICE

Marie Gosset, Antonin Yvon, « L'*open data* de la jurisprudence : intérêt et protection des droits fondamentaux des justiciables », *Affiches parisiennes*, 19 mai 2022.

La technologie Open Source Legal, <https://opensource.legal/home>.

L'association Open Law, Le droit ouvert, <http://openlaw.fr/>.

## CHAPITRE 24. L'HISTOIRE EN BREF

Étude sur le marché de l'open source France et Europe, Systematic Paris Region, 2022, <https://systematic-paris-region.org/download/etude-open-source-2022/>.



# Remerciements

Merci à Anne Alombert, Jean-François Abramatic, Christine Balagué, Jean Cattan, Pascale Cossart, Sophie Gamerman, Hervé Le Crosnier, Stéphane Grumbach, Valérie Peugeot, Brigitte Plateau, Pierre Senellart, pour leurs remarques très pertinentes sur une première version de ce texte.

Merci à toutes les personnes qui ont répondu à nos questions dans le cadre de la rubrique « communs numériques » du blog Binaire du *Monde*. Elles ont nourri notre réflexion : Jean-François Abramatic, Mehdi Benchoufi, Primavera de Filipi, Florence Devouard, Alexandra Elbakyan, Stéfane Fermigier, Stéphane Grumbach, Colin de la Higuerra, Christian Huyghe, Benjamin Jean, Charles Letailleur, Laure Lucchesi, Sylvain Massip, Judith Rochfeld, Sébastien Soriano, Gaël Varoquaux.

Partageant l'ambition de questionner et d'éclairer la réflexion autour des enjeux du numérique, le Conseil national du numérique s'est associé au blog Binaire le 2 juin 2023 pour construire une collection d'entretiens sur les « communs numériques » sur le site du CNNum. Merci au CNNum !

Nous remercions enfin Marin Dacos, Stéphane Gigandet, Jean-Marc Lazard, Charles Schultz, Henri Verdier, et Emmanuel Weisenburger pour des discussions qui ont également nourri notre réflexion.



# Table

Préface, par Gérard Berry .....	7
Mise en bouche .....	11

## PARTIE A

### Les communs numériques

CHAPITRE 1 – Les communs matériels.....	21
CHAPITRE 2 – Les communs numériques.....	27
<i>Des biens non rivaux</i> .....	28
<i>Des communs de la connaissance</i> .....	30
<i>Une terre de conflits</i> .....	31
CHAPITRE 3 – Aux frontières des communs numériques.....	33
<i>Ouverts et gratuits ?</i> .....	34
<i>Un commun doit-il être éthique ?</i> .....	35
<i>Le moteur de recherche de Google est-il un commun ?</i> .....	36
<i>Pour conclure sur la définition des communs</i> .....	39

## PARTIE B

## Les objets des communs numériques

CHAPITRE 4 – Les données .....	47
<i>L'ouverture des données</i> .....	48
<i>La production par la multitude de données ouvertes</i> .....	49
CHAPITRE 5 – L'information .....	52
<i>Le droit d'auteur</i> .....	53
<i>L'œuvre libre</i> .....	54
<i>Militantisme : l'art et la culture libre</i> .....	56
CHAPITRE 6 – Le logiciel .....	57
<i>Logiciel libre et logiciel open source</i> .....	60
CHAPITRE 7 – Les connaissances .....	64
<i>Les connaissances sous forme de texte</i> .....	64
<i>Les connaissances formalisées</i> .....	65
<i>La conception ouverte</i> .....	71
CHAPITRE 8 – Les réseaux comme bien commun .....	75
<i>Ces réseaux sont-ils des communs ?</i> .....	77
<i>Les réseaux télécoms mobiles</i> .....	79
<i>Internet et le Web</i> .....	82
<i>Les enclosures par les entreprises</i> .....	87
<i>Les enclosures par les États</i> .....	88
<i>Des réseaux au service des humains</i> .....	89
<i>Et demain les métavers...</i> .....	91



## PARTIE C

## Le fonctionnement

CHAPITRE 9 – La communauté.....	97
<i>L'esprit de la communauté</i> .....	97
<i>Quelle communauté ?</i> .....	98
<i>Pourquoi contribuer ?</i> .....	102
CHAPITRE 10 – Les licences.....	108
CHAPITRE 11 – Gouvernance des communs .....	113
<i>Le crowdsourcing</i> .....	114
<i>Les logiciels</i> .....	117
CHAPITRE 12 – L'organisation autonome décentralisée.....	121
CHAPITRE 13 – Le modèle économique.....	125
<i>Le financement public</i> .....	126
<i>Le crowdsourcing</i> .....	128
<i>Logiciel libre et open source</i> .....	129
CHAPITRE 14 – communs <i>vs</i> entreprises.....	132
<i>Un processus de production différent</i> .....	132
<i>Un processus de distribution, lui aussi différent...</i> .....	136
<i>... et des produits bien sûr différents !</i> .....	139
CHAPITRE 15 – La création et le développement des communs.....	142
<i>Création ex nihilo</i> .....	143
<i>Perturbation d'un marché existant</i> .....	145
<i>Wikipédia vs. Encyclopædia Britannica</i> .....	146
<i>MySQL vs Oracle</i> .....	149
<i>LibreOffice et OpenOffice vs Microsoft Office</i> .....	152
<i>Microsoft et l'open source</i> .....	153
<i>Elastic vs Amazon</i> .....	155
<i>Le GFII vs les données ouvertes françaises</i> .....	156

CHAPITRE 16 – Le patrimoine .....	160
<i>La préservation du patrimoine</i> .....	160
<i>La préservation des communs numériques</i> .....	162

## PARTIE D

## Les usages

CHAPITRE 17 – La science .....	169
<i>L'histoire de la science ouverte</i> .....	171
<i>L'accès ouvert aux publications scientifiques</i> .....	173
<i>Publish or perish</i> .....	177
<i>Les données ouvertes</i> .....	180
<i>Accès ouvert à tous les produits de la science</i> .....	181
<i>Les freins</i> .....	183
<i>La science pour tous</i> .....	184
CHAPITRE 18 – L'éducation .....	187
<i>L'éducation ouverte et les MOOC</i> .....	188
<i>Les ressources éducatives libres</i> .....	192
CHAPITRE 19 – L'État .....	195
<i>L'accélération de l'histoire</i> .....	195
<i>Les données de référence</i> .....	197
<i>Un État plus efficace</i> .....	199
<i>Un État plus ouvert</i> .....	203
<i>Un État plus souverain</i> .....	205
CHAPITRE 20 – La santé .....	209
<i>Le Système national des données de santé</i> .....	210
<i>Vers le Health Data Hub</i> .....	212
<i>Les associations de patients</i> .....	213

TABLE	299
CHAPITRE 21 – La justice .....	215
<i>Des données ouvertes : la loi et la jurisprudence</i> .....	216
<i>Des logiciels ouverts</i> .....	218
<i>Préparation ouverte d'arguments juridiques</i> .....	220
CHAPITRE 22 – L'agriculture et l'environnement .....	221
<i>Les communs matériels</i> .....	221
<i>Les communs numériques</i> .....	223
CHAPITRE 23 – L'entreprise.....	227
<i>Le logiciel open source</i> .....	227
<i>Les données et leur ouverture</i> .....	228
<i>Valoriser les données par le partage</i> .....	229
<i>Le bien commun</i> .....	230

## PARTIE E

### Le temps et l'espace

CHAPITRE 24 – L'histoire en bref.....	239
<i>La naissance</i> .....	239
<i>La croissance</i> .....	241
CHAPITRE 25 – Géographie des données ouvertes publiques et de l' <i>open source</i> .....	246
<i>Les données ouvertes publiques</i> .....	247
<i>L'open source et le logiciel libre</i> .....	249
CHAPITRE 26 – La Chine et la Russie.....	255
<i>Le logiciel open source en Chine</i> .....	256
<i>Le logiciel open source en Russie</i> .....	259
<i>Les données ouvertes publiques en Chine</i> .....	261
<i>Les données ouvertes publiques en Russie</i> .....	263

Conclusion.....	267
Lexique .....	271
Annexe.....	285
Bibliographie et repères.....	289
Remerciements.....	293

OUVRAGE DE SERGE ABITEBOUL  
CHEZ ODILE JACOB

*Nous sommes les réseaux sociaux* (avec Jean Cattani), 2022.



Cet ouvrage a été composé  
en Adobe Garamond Pro  
par Nord Compo  
à Villeneuve-d'Ascq (Nord).

N° d'édition : 4150-0798-X – N° d'impression : 00000  
Dépôt légal : mars 2024

*Imprimé en France*

Inscrivez-vous à notre newsletter !

Vous serez ainsi régulièrement informé(e)  
de nos nouvelles parutions et de nos actualités :

<https://www.odilejacob.fr/newsletter>

