

La propriété des données dans les programmes de science participative : état des lieux des pratiques et du droit

La question de la propriété des données, de leur valorisation et de leur protection, a été posée de façon précoce dans les programmes de science participative et elle se pose d'ailleurs de façon plus générale pour nombre de programmes collaboratifs où des citoyens mutualisent et échangent des données, ou plus largement encore pour ceux qui se réclament de la contribution au bien commun que constitue le partage des connaissances. Ce travail exploratoire réalisé à la demande de l'association Planète Mer pour son programme Biolit est basé sur un bref état des lieux du statut juridique des données ainsi que sur l'expérience de quelques programmes de science participative.

1. Le statut juridique des données

1.1. Le statut juridique des données naturalistes non publiques

L'Office des données naturalistes d'Alsace (ODONAT), groupement associatif créé en 1995 à l'initiative d'Alsace Nature, du Conservatoire des Sites Alsaciens (CSA), de la LPO-Alsace et du Groupe d'Etude et de Protection des Mammifères d'Alsace (GEPMA) a publié en 2001 un Livre blanc², élaboré avec un juriste, qui rassemble des propositions concernant la propriété et l'échange des données naturalistes. Ce document part du principe de libre circulation et divulgation des données dans l'intérêt de la science et du public, mais également du droit, offert aux contributeurs de "définir" l'usage qui peut ou non être fait de leurs données afin notamment de se prémunir contre un usage dévoyé de ces données par des tiers.

En droit français, les données naturalistes privées peuvent relever de plusieurs régimes, notamment du droit de propriété intellectuelle (droit d'auteur régit par la loi), du droit des contrats (régit par un contrat, une convention, un règlement accepté par les parties), du droit commun des marques (régit

¹ Merci à ceux qui ont bien voulu me consacrer leur temps pour me faire partager leur connaissance de ce sujet : David Delon (Tela Botanica), Cédric Vanappelghem (GON), Olivier Pratte (Pôle Faune), Romain Julliard (Vigie Nature). Il va de soi que les approximations éventuelles restent de mon fait et n'engagent pas leur responsabilité. Document sous licence CC-BY-SA.

² Téléchargeable ici : http://www.odonat-alsace.org/documents/ODONAT_Livre_Blanc_Donnees_naturalistes_Alsace.pdf

par la loi) et encore du droit *sui generis* des bases de données (droit et jurisprudence en cours). Le livre blanc propose de distinguer plusieurs types de données :

- Les données primaires sont les informations non traitées, par exemple celles qui font état de la présence ou de l'absence de telle espèce à tel moment et à tel endroit y compris au moyen de protocoles scientifiques, ou les agrégats de données (tableaux récapitulatifs sans traitement). Elles ne peuvent pas être considérées comme des « œuvres originales de l'esprit » mais comme une description du réel. A ce titre, elles ne relèveraient pas du droit de propriété intellectuelle mais du droit des contrats.
- Les données secondaires sont les données agrégées et faiblement commentées, par exemple un tableau croisé dynamique sous excel (tableau avec données classées et traitement statistique) ou même une carte de synthèse ou un graphe. Elles relèvent du même statut.
- Les données élaborées sont les données interprétées, c'est-à-dire « les données agrégées selon un plan cohérent mais arbitraire, manifestant un choix de l'analyste, commentées et interprétées en portant l'empreinte de l'esprit et de la personnalité » de l'auteur. Elles relèvent du droit d'auteur. En effet, le droit d'auteur est défini par la jurisprudence par l'originalité, l'interprétation, la subjectivité. Les œuvres relevant du droit d'auteur doivent être nouvelles et porter la marque de leur créateur, notamment sa part de jugement personnel. Ces créations peuvent également relever du droit des marques si leur auteur dispose d'une marque déposée.
- La protection juridique est spécifique et plus complexe que celle des simples données. Elle a fait l'objet de mesures législatives en France depuis 1998³ et désormais en Europe, qui débouchent progressivement sur un droit *sui generis*⁴. « Le producteur d'une base de données, entendu comme la personne qui prend l'initiative et le risque des investissements correspondants, bénéficie d'une protection du contenu de la base lorsque la constitution, la vérification ou la présentation de celui-ci atteste d'un investissement financier, matériel ou humain substantiel.⁵ » La jurisprudence a précisé cette notion d'investissement substantiel.
- Les photographies peuvent généralement prétendre au droit de propriété intellectuelle.

Cette analyse juridique fait donc une distinction entre :

- Les données primaires ainsi que leur agrégation simple (qui relèvent du droit des contrats).
- Leur agrégation ou leur classement élaboré, ainsi que leur analyse, s'ils relèvent d'une création intellectuelle originale (droit de propriété intellectuelle, droit des marques).
- Les bases de données (droit *sui generis*)
- les photographies (droit de propriété intellectuelle).

L'auteur d'une donnée primaire peut s'en réserver l'usage en garantissant lui-même sa confidentialité. Dès qu'il la transmet sans l'assortir de réserves particulières, cette donnée perd son caractère privé et tombe dans le domaine public, sans qu'il soit nécessaire à celui qui l'utilise (du

³ Voir Loi n° 98-536 du 1 juillet 1998 portant transposition dans le code de la propriété intellectuelle de la directive 96/9/ CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données.
<http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000573438>

⁴ <http://www.jurisexpert.net/droit-sui-generis-des-bases-de-donnees/>

⁵ http://fr.wikipedia.org/wiki/Protection_juridique_des_bases_de_donnC3%A9es

moment qu'il y a eu accès sans fraude) de mentionner le nom du collecteur de cette donnée ou de lui proposer une rémunération. L'usage (scientifique notamment) veut que son nom soit cependant mentionné, mais la traçabilité peut se perdre avec les emprunts successifs.

Cependant, les données primaires relevant du droit des contrats, « *l'auteur de l'œuvre première [donnée primaire] peut déterminer contractuellement les conditions et le cadre de son utilisation et peut la limiter par exemple, à un certain type de supports, d'usage ou à une certaine durée* ».

Quant à l'auteur d'une base de données, il peut prétendre au droit de propriété intellectuelle pendant une durée de 15 ans s'il peut montrer que celle-ci a fait l'objet d'un investissement, matériel et intellectuel, important. Cependant, si cette base est mise à disposition du public (accès en ligne sans autorisation spécifique par exemple), son utilisation à des fins privées ne peut pas être interdite. Il peut également mobiliser le droit des marques s'il dispose lui-même d'une marque déposée à l'INPI mais cette protection étant limitée dans le temps, toutes les données ou bases de données qu'une marque souhaite ainsi protéger doivent être millésimées et ces données déposées chaque année à l'INPI. Enfin, il peut renoncer à ces droits en décidant de placer sa base de données sous un régime plus favorable à sa libre diffusion.

Afin d'encadrer l'utilisation des données collectées, certains programmes de science participative font appel à des licences standards (voir plus loin), qui font office de contrats auxquels l'utilisateur du programme de science participative doit adhérer en s'inscrivant. D'autres établissent des contrats ad hoc qui doivent être lus et validés par chaque contributeur lors de son inscription (voir plus loin).

1.2. Le statut juridique des données naturalistes publiques

Les données publiques sont les données détenues par des organismes ayant une mission de service public, notamment liée à l'environnement.

Le statut des données naturalistes publiques est encadré par différents textes spécifiques à l'environnement comme la convention d'Aarhus du 25 juin 1998⁶ sur le droit d'accès des citoyens aux informations relatives à l'environnement ou la directive européenne Inspire⁷ qui fait obligation aux Etats membres de rendre publiques les données de biodiversité géolocalisées.

En France, leur diffusion est également encadrée par des textes plus généraux concernant l'ensemble des données publiques comme la Loi n° 78-753 du 17 juillet 1978 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, dont l'application a été confiée à la Commission d'accès aux documents administratifs (CADA). La CADA a notamment rendu un arrêt indiquant qu'il n'était pas nécessaire d'établir une convention ou une licence pour diffuser des données primaires⁸. L'ordonnance du 6 juin 2005 a modifié cette loi du 17 juillet 1978 en consacrant notamment un droit à la réutilisation des données publiques dont la mise en œuvre a été précisée par un décret et une circulaire du Premier ministre. Les obligations des organismes publics ont été précisées dans la loi du 26 octobre 2005 et son décret d'application du 22 mai 2006, codifiés

⁶ <http://www.unece.org/fileadmin/DAM/env/pp/documents/cep43f.pdf>

⁷ <http://inspire.ign.fr/>

⁸ Source SINP : <http://www.naturefrance.fr/sinp/presentation-du-sinp/aspects-juridiques/donnees-publiques>

au chapitre IV du code de l'environnement (parties législative et réglementaire). La circulaire MEDAD du 18 octobre 2007 précise la portée et la mise en œuvre de ces textes.

Le ministère de l'Environnement a publié en 2010 un résumé de ces droits, qui est disponible sur le portail toutsurlenvironnement.fr⁹. Ce document rappelle que le principe général est celui de la libre diffusion des données publiques, sauf notamment si cela est contradictoire avec la protection de l'environnement, par exemple si la localisation d'espèces rares est « susceptible de faire l'objet d'un prélèvement autre que scientifique, ou d'une destruction volontaire » (p. 17).

2. Les outils des programmes de science participative

2.1. Les licences

Les licences libres et les licences ouvertes sont des contrats qui permettent à l'auteur d'une donnée de faciliter son utilisation tout en l'assortissant éventuellement de certaines restrictions (utilisation commerciale, modification...). Elles se veulent une alternative aux « droits réservés » (*copyright*) qui freinent l'exploitation des contenus. Ces droits offerts à l'utilisateur ne sont pas nécessairement assortis de la gratuité (cas de certains logiciels open source). Il existe de multiples licences, qui doivent toutes être compatibles avec le droit d'auteur pour être reconnues comme ayant valeur légale. Les plus importantes sont les suivantes :

- ***Creative Commons***

Les licences libres Creative Commons¹⁰ sont devenues des standards internationaux mises en place par l'organisation à but non lucratif du même nom elle-même inspirée par les logiciels libres. Elles ont été transposées en droit français par le CERSA, Centre d'Etudes et de Recherches en Sciences Administratives de l'Université Paris II. Il en existe 6, qui ont toutes en commun de reconnaître la paternité des données (l'auteur doit être cité), mais qui peuvent varier selon que cet auteur autorise ou non leur utilisation commerciale, leur modification et, dans cette hypothèse, la modification ou non - pour les œuvres dérivées - des droits que lui-même a choisi pour son œuvre originale. Elles sont décrites succinctement dans le tableau ci-dessous. Elles sont gratuites. Elles reposent sur un cadre juridique complet¹¹.

Il existe plusieurs versions successives de chacun des licences CC, légèrement différentes (les nuances ne sont pas explorées ici). Un des utilisateurs les plus connus de la licence CC est le site Wikipédia (licence CC-BY-SA version 3.0).

⁹ (http://www.toutsurlenvironnement.fr/files/Tout%20ce%20que%20vous%20avez%20toujours%20voulu%20savoir_2.pdf)

¹⁰ <http://creativecommons.org/>

¹¹ présenté sur cette page : <http://creativecommons.org/licenses/by-nc-nd/2.0/fr/legalcode>

Licence	Symboles			Droits associés
CC-BY				L'utilisation des données est libre : elles peuvent être valorisées dans un but commercial, modifiées et les auteurs des modifications peuvent appliquer à leurs produits les droits qu'ils désirent.
CC-BY-SA				L'utilisation des données est libre (valorisation commerciale et modifications autorisées) mais les produits dérivés doivent eux-mêmes bénéficier des mêmes droits.
CC-BY-ND				Les données peuvent être valorisées, y compris commercialement, mais pas modifiées.
CC-BY-NC				Les données peuvent être modifiées mais pas valorisées commercialement.
CC-BY-NC-SA				Les données peuvent être modifiées mais pas valorisées commercialement. Les produits dérivés doivent bénéficier des mêmes droits.
CC-BY-NC-ND				Les données ne peuvent être ni modifiées ni valorisées commercialement.

D'après le site Creative Commons

- **Open database (ODbL)**

La licence ODbL¹² est une licence spécifique aux bases de données, dont le droit est plus complexe que celui des simples données. Elle est utilisée par le projet OpenStreetMap (openstreetmap.fr), Toulouse métropole¹³ ou la ville de Paris¹⁴ pour la mise à disposition publique de certaines bases de données.

La licence ODbL offre les droits suivants (version résumée du code légal¹⁵ et adaptée par la ville de Paris):

Vous êtes libres :

- 
De partager : copier, distribuer et utiliser la base de données.
- 
De créer : produire des créations à partir de cette base de données.
- 
D'adapter : modifier, transformer et construire à partir de cette base de données.

¹² <http://opendatacommons.org/licenses/odbl/>

¹³ <http://data.grandtoulouse.fr/web/guest/la-licence>

¹⁴ http://opendata.paris.fr/opendata/jsp/site/Portal.jsp?page_id=10

¹⁵ <http://opendatacommons.org/licenses/odbl/1.0/>

Aussi longtemps que :



Vous mentionnez la paternité : Vous devez mentionner la source de la base de données pour toute utilisation publique de la base de données, ou pour toute création produite à partir de la base de données, de la manière indiquée dans l'ODbL. Pour toute utilisation ou redistribution de la base de données, ou création produite à partir de cette base de données, vous devez clairement mentionner aux tiers la licence de la base de données et garder intacte toute mention légale sur la base de données originaire.



Vous partagez aux conditions identiques : si vous utilisez publiquement une version adaptée de cette base de données, ou que vous produisiez une création à partir d'une base de données adaptée, vous devez aussi offrir cette base de données adaptée selon les termes de la licence ODbL.



Gardez ouvert : si vous redistribuez la base de données, ou une version modifiée de celle-ci, alors vous ne pouvez utiliser de mesure technique restreignant la création que si vous distribuez aussi une version sans ces restrictions.

http://opendata.paris.fr/opendata/jsp/site/Portal.jsp?page_id=10

- **Licence ouverte (Open Licence)**

La Licence Ouverte¹⁶ est une licence libre mise au point pour le gouvernement français pour assurer la diffusion de données publiques. Elle est compatible avec des standards internationaux, notamment la licence CC-BY.

Elle précise notamment à l'utilisateur qu'il est :

[...] libre de réutiliser « l'Information » :

- Reproduire, copier, publier et transmettre « l'Information » ;
- Diffuser et redistribuer « l'Information » ;
- Adapter, modifier, extraire et transformer à partir de « l'Information », notamment pour créer des « Informations dérivées » ;
- Exploiter « l'Information » à titre commercial, par exemple en la combinant avec d'autres « Informations », ou en l'incluant dans votre propre produit ou application.

Sous réserve de :

- Mentionner la paternité de « l'Information » : sa source (*a minima* le nom du « Producteur ») et la date de sa dernière mise à jour.

<http://www.data.gouv.fr/Licence-Ouverte-Open-Licence>

¹⁶ <http://www.data.gouv.fr/Licence-Ouverte-Open-Licence>

2.2. Les pratiques inspirées du statut des données publiques

Certains organismes associatifs financés plus ou moins largement par des subventions publiques (Etats, collectivités locales) estiment, même s'ils font également appel à des contributeurs bénévoles, que la diffusion de leurs données doit être soumise aux impératifs de libre diffusion des données publiques. Ils jugent souvent que la libre diffusion des données naturalistes est plus favorable à la protection de l'environnement qu'une diffusion restreinte, même si – pour certains d'entre eux - des exceptions doivent être aménagées à ce principe dans le cas par exemple d'espèces rares et potentiellement recherchées pour d'autres usages que la protection.

Dans l'avenir, les échanges de données, publiques et privées, militent effectivement dans le sens d'un alignement des pratiques de ces organismes vers celles du secteur public. Notamment, la mise en place du SINP (Système d'information sur la nature et les paysages, ministère de l'Environnement) invite les possesseurs de bases de données naturalistes à se positionner de façon claire sur la diffusion de leurs données.

Le principe de base des données publiques est la libre diffusion des données primaires et secondaires (non soumises au droit d'auteur), mais l'interdiction de leur rediffusion. Autrement dit, un utilisateur peut utiliser ces données pour son propre usage (y compris non gratuit, par exemple lorsqu'un bureau d'étude les intègre dans un travail de diagnostic qui va lui être rémunéré) mais il ne pourra pas les rediffuser à son tour à d'autres utilisateurs. L'origine de ces données doit être mentionnée.

La convention d'Aarhus et la Directive européenne Inspire sont les textes de référence qui encadrent ces pratiques, qui doivent évidemment être compatibles avec le droit français, notamment le droit d'auteur (pas de diffusion des données tertiaires sans consentement de l'auteur).

2.3. Les contrats ad hoc

Certains projets de science participative ne se réfèrent pas à des licences particulières ou au droit des données publiques mais proposent des contrats spécifiques auxquels les utilisateurs doivent souscrire dès lors qu'ils contribuent au projet (généralement en cochant une case lors de leur inscription).

Ces contrats doivent également respecter le droit français et notamment le droit d'auteur.

Ils sont souvent plus restrictifs que les licences en matière de diffusion des données. Celle-ci est soumise à l'accord express du gestionnaire de la base de données. Comme les licences, ils visent à la fois à préserver les droits des fournisseurs de données et à prévenir d'éventuels contentieux avec ceux d'entre eux qui pourraient s'opposer à l'exploitation de leurs données primaires, mais ils visent souvent aussi à préserver les droits de celui qui fournit un travail d'exploitation et même d'agrégation de ces données, c'est-à-dire le gestionnaire de la base qui est généralement un organisme gérant le programme de science participative. Lorsque cet organisme exploite lui-même commercialement de ces données (en les cédant contre rétribution à des tiers, par exemple) il peut en effet avoir intérêt à limiter la libre diffusion des données.

Les spécifications de ces contrats dépassent souvent le cadre strict de l'utilisation des données et abordent par exemple des questions liées à leur validation.

2.4. Les « bonnes pratiques »

Plusieurs responsables de programmes de science participative, quelque soit le régime choisi concernant leurs données, énoncent des principes qui relèvent de la déontologie plutôt que du droit, et qui sont considérés comme relevant de bonnes pratiques. Ceux-ci se rajoutent donc aux dispositions légales mais ne s'y substituent évidemment pas. Par exemple :

- Même si une donnée primaire ne relève pas du droit d'auteur, chaque contributeur se voit généralement réaffirmer la « propriété » de ses propres données, c'est-à-dire le droit de les transmettre librement à qui bon lui semble sans en rendre compte à quiconque, ou même éventuellement le droit de les retirer de la base de données au cas où un désaccord survient avec les orientations du programme, celles-ci pouvant évoluer ou être initialement mal comprises. Dans ce cas extrême (et rare), des données peuvent ainsi être effacées à la demande d'un contributeur lors d'une actualisation de la base de données (expérience de Tela Botanica), ce qui est sans effet notable si la part effacée est peu significative par rapport à l'ensemble des données, donc dans le cas de programmes importants ; noter que l'utilisation par d'autres des données antérieurement à cette suppression reste cependant irréversible. Pour le Muséum, le droit de propriété des données du contributeur est reconnu dans les programmes anciens (droit d'être cité) mais en revanche, dans les nouveaux programmes, le contributeur cède ses droits sur ses propres observations dès lors qu'il accepte qu'elles soient mutualisées.
- Plusieurs programmes offrent la possibilité aux contributeurs de maintenir « cachées » certaines données par convenance personnelle, c'est-à-dire de ne pas les rendre publiques ni les mutualiser au sein du réseau des contributeurs. Cela peut être le cas par exemple d'informations personnelles (adresse...), mais aussi naturalistes (certains relevés, certains commentaires...). La base fait ainsi office pour ces contributeurs de carnets de notes personnels, dont une partie seulement est mutualisée.
- La libre diffusion des données peut être limitée si elle peut provoquer des effets négatifs sur la biodiversité (révéler les sites de nidification de certains oiseaux ou la localisation de plantes rares, etc.). Un Comité de pilotage ou un comité scientifique est généralement chargé de définir des critères et de les rendre transparents. Pour le Muséum, cette crainte est rarement justifiée dans les faits et la libre diffusion de toutes les observations, même localisées, est préférable car globalement favorable à la connaissance de la biodiversité et donc à sa préservation.
- L'affirmation du droit ne suffisant pas à la prévention des malentendus, plusieurs programmes jugent qu'il est nécessaire de faire preuve de pédagogie sur la façon dont les données peuvent être traitées, cédées et utilisées. En effet, certains contributeurs bénévoles ne sont pas toujours bien au fait de l'utilisation qui peut être faite de leur travail : valorisation par des bureaux d'études y compris contre rémunération, communication à des organismes publics d'Etat ou des collectivités locales pour des études d'impact relatives à des projets d'aménagement, etc.

3. Quelques exemples

Les programmes de science participative font des choix différents en matière de droit des données. Les principales options sont les suivantes.

- ***Licences libres***

L'ONEM et Tela Botanica sont deux programmes qui ont en commun d'assurer la libre diffusion de leurs données en adoptant une licence. Inspirés par l'association Outils-Réseaux qui se réfère aux licences Creative Commons et à la « doctrine Odonat » (livre blanc de 2001), ils font le choix de la licence CC-BY-SA (libre diffusion sous réserve de mention de la source et de libre diffusion des produits dérivés). L'ONEM dispose à la fois d'une licence CC et d'une charte (en cours d'élaboration) qui encadre l'utilisation des données.

Voir : <http://www.tela-botanica.org/page:licence?langue=fr>
<http://www.onem-france.org/wakka.php?wiki=BDDProprio>

- ***Statut inspiré du droit des données publiques***

Le programme « Atlas des Odonates » mené par le Groupe ornithologique et naturaliste du Nord-Pas-de-Calais (GON) aligne ses pratiques sur celles des organismes publics. Les données primaires à l'échelle communale sont en libre accès sur le site du SIRF (Système d'information régionale sur la faune), mais leur rediffusion interdite. Les données géoréférencées plus précisément sont diffusées gratuitement (toujours à des utilisateurs finaux) mais sur demande écrite afin d'assurer une traçabilité des démarches dans ce sens.

Le Muséum national d'histoire naturelle est un organisme public dont les pratiques ont évolué avec le temps. Dans les premiers programmes (par exemple, le programme STOC), les données globales sont fournies sur demande pour usages scientifiques et les demandes sur des données précisément localisées sont renvoyées vers des coordinateurs locaux (des individus liés généralement à des associations). Dans les programmes plus récents (SPIOLL), les données sont en libre accès pour tous, y compris les non-contributeurs. Le Muséum ne fait pas de distinction entre données primaires et secondaires, estimant qu'une donnée individualisée sur la présence d'une espèce à tel endroit n'a de valeur que lorsqu'elle est insérée dans une base de données.

Voir :

<http://www.sirf.eu/index.php?cont=common&tpl=accueil>
<http://www.naturefrance.fr/sinp/presentation-du-sinp/aspects-juridiques/donnees-publiques>
<http://www.naturefrance.fr/sinp/presentation-du-sinp/aspects-juridiques/deontologie>
<http://www.naturefrance.fr/sinp/presentation-du-sinp/aspects-juridiques/mise-a-disposition-des-donnees>
http://www.naturefrance.fr/sites/default/files/documents/pdf/fiche_politique_de_mise_a_disposition.pdf

- **Contrats ad hoc**

Faune LR, projet du groupement Méridionalis, ou Faune Aquitaine, géré par la LPO, ont élaboré des contrats spécifiques, sous l'inspiration du même organisme : Biolovision. Ces contrats sont présentés sous forme de conditions générales ou de règlement posant des limites à l'utilisation des données (voir adresses ci-dessous).

La LPO met ses données en accès sur le site Visionature (www.ornitho.fr). Une inscription est nécessaire.

DORIS, projet de la Fédération Française d'Etudes et de Sports Sous-Marins, a élaboré sa propre licence (licence Doris), assez restrictive sous certains aspects, notamment concernant l'utilisation des photos postées sur son site.

Voir :

<http://files.biolovision.net/www.faune-lr.org/userfiles/docsaide/conditionsgeneralesdutilisationfaune-lr.pdf>

http://files.biolovision.net/www.faune-aquitaine.org/pdf/files/Code_FA_080508-8628.pdf

<http://doris.ffessm.fr/Copyright.asp>

Note de travail établie par Pierre-Yves Guihéneuf, DialTer pour l'association Planète Mer avec l'appui de la Fondation de France. www.dialter.fr. 12-2012