



HAL
open science

Les “ non-consumptive research uses ” des ressources numériques

Julien Baudry

► **To cite this version:**

| Julien Baudry. Les “ non-consumptive research uses ” des ressources numériques. 2023. hal-04289382

HAL Id: hal-04289382

<https://amu.hal.science/hal-04289382>

Preprint submitted on 16 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Les « *non-consumptive research uses* » des ressources numériques

Cette note n'est pas un avis juridique mais une analyse de situation légale autour de la question des *non-consumptive research uses*.

Résumé

La *non-consumptive research* est la recherche où le seul usage que le chercheur fait d'une ressource numérique est d'y appliquer une analyse informatique (computationnelle), et non d'en lire et comprendre « humainement » des parties substantielles afin d'assimiler intellectuellement son contenu.

Exemples de *non-consumptive research use* : extraction de texte, analyse automatisée de texte, traduction automatique, synthèse et rapport statistique généré automatiquement, indexation automatique...

Conséquences légales de cette notion dans le droit anglo-saxon : depuis 2009, la *non-consumptive research* est assimilée à un *fair use*, c'est-à-dire un usage ne nécessitant pas d'autorisation spécifique. Ce point est contesté par les représentants des ayants-droits, mais largement utilisé par les bibliothèques universitaires pour autoriser leurs chercheurs à utiliser les ressources numériques à des fins de *non-consumptive research*.

Adaptation de la notion en contexte de droit d'auteur français/européen : les *non-consumptive research use rights* sont dépendant de la notion anglo-saxonne de *fair use* (= certains usages de documents sous droits sont autorisés par défaut si considérés « *fair* ») et non-applicables tel quel en droit français. Toutefois, depuis 2021, une ordonnance sur le Text and Data Mining ajoutée au Code de la Propriété Intellectuelle fait de cet usage une exception au droit d'auteur ne nécessitant pas d'accord spécifique tant que l'accès au contenu est licite. Par ailleurs, les contrats entre éditeurs et bibliothèques comprennent de plus en plus des clauses TDM.

Limites et points de vigilance : pour l'avenir deux points nous semblent mériter une vigilance accrue :

- Le développement conjoint des Intelligences Artificielles, d'une part, et de l'ouverture des publications et données scientifiques (Science Ouverte) d'autre part, invite à penser qu'un durcissement des conditions d'application de cet usage pourrait advenir par réaction des éditeurs.
- Sur ce sujet la bibliothèque, en ce qu'elle est la pourvoyeuse aux chercheurs du contenu licite sous droits est un acteur central de l'application des *non-consumptive use rights*. Une montée en compétences des acteurs impliqués dans les bibliothèques semble nécessaire. Dans l'avenir, si les bibliothèques sont amenées à constituer des corpus et bases de données en vue de TDM, ces questions juridiques devront être posées.

Définition et principes des *non-consumptive research use*

La définition primaire est celle de « *non-consumptive use* », terme courant en anglais pour désigner des usages d'un bien **qui ne sont pas liés à la « consommation », au sens alimentaire** (comme le lavage du linge par opposition à l'eau tiré du robinet pour être bu, ou la photographie animalière par opposition à la chasse). Dans ces domaines, la notion est importante car elle suppose que la ressource (l'eau, la faune) n'est pas « consommée », donc peut être réutilisée. En droit français on parle de « caractère non-consomptible » d'une ressource dans un sens identique.

Depuis 2010¹ environ cette notion est appliquée à la recherche (« *non-consumptive research* ») pour distinguer l'usage de « consommation » de la ressource (le chercheur lit et assimile intellectuellement le contenu de la ressource) et l'usage non-consomptible (**le chercheur interroge la ressource par l'intermédiaire d'un programme informatique ou plus généralement d'une machine**). Le terme est utilisé et défini dans des accords entre, initialement, les syndicats d'auteurs/éditeurs d'un côté et Google Books et les bibliothèques participantes au programme de numérisation de l'autre, comme présenté ici : <https://www.lawinsider.com/dictionary/non-consumptive-research>. Durant les années 2010, les éditeurs, fondations, entreprises, bibliothèques, impliquées dans la numérisation de contenus ont mis en place des *non-consumptive research use policies* pour clarifier juridiquement l'usage pratiqué. La définition suivante a été arrêtée, par exemple pour la Hathi Trust Digital Library (HTDL) :

« *Research in which computational analysis is performed on one or more volumes or textual objects in the HTDL, but not research in which a researcher reads or displays substantial portions of an in-copyright or rights-restricted work to understand the expressive content presented within that work.* »

Avec une insistance sur le terme « *substantial* », qui indique bien que le chercheur peut être amené à lire des extraits de la ressource, mais qu'ils doivent rester limités. On ne doit pas pouvoir reconstituer la ressource à partir de ce qui est présenté à l'œil humain. Parmi les usages cités, on peut trouver : extraction de texte, analyse automatisée de texte, traduction automatique, synthèse et rapport statistique généré automatiquement, indexation automatique...

Rôle des bibliothèques

Le rôle des bibliothèques dans l'application de cet usage est souvent mis en avant pour une raison simple : ce sont principalement les bibliothèques qui gèrent l'accès licite aux contenus que les chercheurs souhaitent interroger².

Certaines bibliothèques, comme celles du MIT³, précisent bien que l'un de leurs rôles est de s'assurer que les accès fournis aux ressources soient en adéquation avec l'évolution des pratiques scientifiques, et notamment du développement de la *non-consumptive research*. Un des rôles des bibliothèques pourrait être de guider les chercheurs dans ce nouveau type d'usage scientifique⁴.

Contexte anglo-saxon et mise en débat

Attention cependant : cette définition est intimement liée au contexte juridique anglo-saxon du *fair use*. Le *non-consumptive research use* est considéré comme un *fair use*, c'est-à-dire un usage permis sans autorisation explicite des ayants-droits de la ressource, ni paiement supplémentaire. Il est cité dans le *Code of Best Practices in Fair Use for Academic and Research Libraries*⁵. Il est par exemple considéré *fair use* pour une bibliothèque universitaire de constituer une base de données de ressources sous droits pour qu'un chercheur puisse y appliquer de la *non-consumptive research*.

Cette insistance à faire de la *non-consumptive research* un *fair use* est crucial dans le cas du développement des IA⁶, puisque ces applications ont précisément un usage non-consomptif des données et ressources. A l'inverse, les lobbys des ayants-droit insistent sur le fait que cet usage, bien que « non-consommé » par l'œil humain, reste un usage qui doit faire l'objet d'autorisation et de frais. Est-ce que les IA pourront « juridiquement » accéder gratuitement aux ressources sous droits, ou devra-t-on en passer par des accords et lois pour définir un cadre d'application ? C'est bien là tout l'enjeu. Plusieurs procès sont en cours pour trancher cette question⁷.

Application au contexte français

Fair use vs droit d'auteur

Le contexte juridique français (ou d'Europe continentale en général) n'applique pas le *fair use*. Le principe du droit d'auteur continental est simple : tout usage d'une ressource est interdit sans autorisation de l'auteur, sauf si une

exception légale existe pour l'autoriser. Pour cette raison la notion anglo-saxonne de *non-consumptive use rights* comme d'un droit d'usage nécessairement différent ne paraît pas applicable en droit d'auteur où tout usage nouveau non-couvert par exception rentre par défaut dans le cadre commun de l'autorisation préalable (ce qui n'est pas autorisé est interdit).

La question à poser en termes d'application de la notion de *non-consumptive research* au droit français est donc double :

- Les contrats signés avec les éditeurs de ressources numériques permettent-ils (ou mentionnent-ils) ce type d'usage ; y a-t-il des litiges possibles sur ce point ?
- L'évolution de la législation nationale ou européenne peut-elle aider à avancer sur cette question ?

Sur le premier point, on a vu en 2022 des cas d'avertissements envoyés par certains éditeurs à des universités sur l'usage de leurs ressources pour fouille de texte. Ceux-ci arguant que les chercheurs devaient passer par leurs applications de TDM pour réaliser la fouille, là où les chercheurs voulaient utiliser leur propre application. Ces cas ont été traités par l'ABES en tant que porteur des négociations. Dans tous les cas, il est important que les contrats négociés comprennent des clauses liées à ces usages, ce qui est en partie traité à travers l'ajout des clauses TDM, comme précisé dans la lettre de cadrage Couperin⁸.

TDM et IA Act

Dans le contexte français et européen, la question des *non-consumptive research use* a pour l'instant été traité par le biais du Text and Data Mining (TDM). Pour rappel, une ordonnance de 2021⁹ a transposé dans le Code de la Propriété Intellectuelle une directive européenne sur le TDM pour en faire un usage autorisé à des fins de recherche. Cela signifie qu'un éditeur/ayant droit ne peut pas s'opposer à un usage de TDM par un chercheur tant que le chercheur a pu accéder légalement à la ressource numérique en question (par exemple via sa bibliothèque). Dans ce contexte législatif, la bibliothèque est au cœur de la pratique de TDM du chercheur.

Cette ordonnance est un premier pas qui répond à la question légale soulevée par la *non-consumptive research*. Toutefois, la notion de TDM ne recouvre pas totalement celle de *non-consumptive research*, elle n'est qu'un aspect d'une notion qui se veut plus générale. Elle date d'une période d'avant la généralisation des IA « domestiques », type ChatGPT. On peut se demander si elle couvre correctement l'ensemble des usages considérés comme *non-consumptive*. Une veille sur les usages de recherche semble alors indispensable pour anticiper les litiges pouvant survenir avec les éditeurs.

De plus, le fait que de plus en plus de contenu scientifiques soit en libre accès peut, paradoxalement, être un frein au développement de la *non-consumptive research*, dans la mesure où ces contenus ouverts échappent (en partie) à une contractualisation avec les universités. C'est l'éditeur qui peut en dicter les conditions d'usage. La clé ici est la généralisation des licences libres de type CC BY qui autorisent par défaut tout usage.

Un projet de régulation de l'usage des IA est en cours à l'échelle européennes (IA Act¹⁰), qui comprend notamment des clauses liées au droit d'auteur. Un des buts sera de rendre obligatoire la déclaration des contenus et ressources ayant servi à entraîner l'IA afin de respecter, a minima, le droit moral.

Date : 27/09/2023

Rédacteur : Julien Baudry (Université Bordeaux Montaigne) pour le GTSO Juridique de Couperin

Summary

Non-consumptive research is research in which the researcher's only use of a digital resource is to apply computer (computational) analysis to it, rather than to 'humanly' read and understand substantial parts of it in order to intellectually assimilate its content.

Examples of *non-consumptive research* uses are text extraction, automated text analysis, automated translation, automatically generated synthesis and statistical reports, automated indexing, etc.

Legal consequences of this concept in Anglo-Saxon law are that since 2009, *non-consumptive research* has been treated as *fair use*, i.e. use that does not require specific permission. This point is contested by representatives of copyright holders, but is widely used by university libraries to authorise their researchers to use digital resources for *non-consumptive research* purposes.

Adapting the concept to the French/European copyright context seems complex, since *non-consumptive research* use rights are closely linked to the Anglo-Saxon concept of *fair use* (= certain uses of copyrighted documents are allowed by default if they are considered "fair") and cannot be applied as such under French law. However, as of 2021, an ordinance on text and data mining has been added to the French Intellectual Property Code, making this use an exception to copyright that does not require a specific agreement, as long as access to the content is licit. In addition, TDM clauses are increasingly included in contracts between publishers and libraries.

Limitations and points to be aware of: In the future, two points seem to us to deserve greater attention:

☒ The simultaneous development of AI, on the one hand, and the increasing openness of scientific publications and data (Open Science), on the other, suggest that publishers may react by tightening the conditions of use.

☒ In this respect, the library is a key player in the application of non-consumptive use rights, as it provides researchers with licit content under copyright. However, library staff will need to develop their skills if they are to be able to meet the needs of researchers in these areas, and if libraries are asked to build corpora and databases for TDM in the future, these legal issues will need to be addressed.

Références

¹ Eleanor F. Dickson, *et alii*, « Creating a Policy Framework for Analytic Access to In-Copyright Works for Non-Consumptive Research (non-consumptive research, text analysis, access) », abstract d'une intervention à *Digital Humanities 2017*, 8 au 11 août 2017, url : <https://dh2017.adho.org/program-2/abstracts/>

² Brandon Butler, « For Text and Data Mining, Fair Use Is Powerful, but Possession Is Still 9/10 of the Law », *SPARC*, 18 février 2018, url : <https://sparcopen.org/news/2018/possession-is-still-9-10-of-the-law/>

³ MIT Libraries, « MIT Framework for Publishers Contract », mise à jour le 19 mai 2020, url : <https://libraries.mit.edu/scholarly/publishing/framework/>

⁴ R.G. Samberg et C. Hennesy, « Law and Literacy in Non-Consumptive Text Mining: Guiding Researchers Through the Landscape of Computational Text Analysis ». In *Copyright Conversations: Rights Literacy in a Digital World*. UC Berkeley, 2019, url : <https://escholarship.org/uc/item/55j0h74gh>

⁵ <https://www.arl.org/code-of-best-practices-in-fair-use-for-academic-and-research-libraries/>, p.28-29

⁶ Carlo Scollo Lavizzari, « AI in publishing and licensing – publishing and licensing in AI », In WIPO Conversation on IP and AI: Second Session. Day 3, Draft 29 June 2020

⁷ Blake Brittain, « Getty images lawsuit says Stability AI misused photos to train AI », *Reuters*, 6 février 2023, url : <https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/>

⁸ Couperin, « Lettre de cadrage – négociations 2023 », *Couperin.org*, 21 juin 2023, url : <https://www.couperin.org/negociations/ressources/lettre-de-cadrage/>

⁹ Lien vers le texte de l'ordonnance n° 2021-1518 du 24 novembre 2021 : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044362034>

¹⁰ Matthieu Eugene, « IA Act : 5 choses à savoir sur le projet européen de régulation de l'IA », *Le blog du modérateur*, 12 mai 2023, url : <https://www.blogdumoderateur.com/ia-act-projet-regulation-intelligence-artificielle/>