UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE
https://archive-ouverte.unige.ch

Article scientifique | Article | 2025

Published version | Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Open Licensing and Data Trust for Personal and Non-Personal Data: A Blueprint to Support the Commons and Privacy

Benhamou, Yaniv; Dulong de Rosnay, Mélanie

How to cite

BENHAMOU, Yaniv, DULONG DE ROSNAY, Mélanie. Open Licensing and Data Trust for Personal and Non-Personal Data: A Blueprint to Support the Commons and Privacy. In: IIC, 2025. doi: 10.1007/s40319-025-01636-y

This publication URL: https://archive-ouverte.unige.ch/unige:187780
Publication DOI: 10.1007/s40319-025-01636-y

© The author(s). This work is licensed under a Creative Commons Attribution (CC BY 4.0)
https://creativecommons.org/licenses/by/4.0

IIC
https://doi.org/10.1007/s40319-025-01636-y

ARTICLE

Open Licensing and Data Trust for Personal and Non-Personal Data: A Blueprint to Support the Commons and Privacy

Yaniv Benhamou · Mélanie Dulong de Rosnay

Accepted: 21 August 2025
© The Author(s) 2025

Abstract The present contribution proposes a novel commons-based copyright licensing model that provides individuals better control over all their data (including copyrightable, personal and technical data) and that covers recent developments in AI technology. The licensing model also proposes restrictions and boundaries (e.g. to authorised users and groups) to protect the commons, allowing communities to define and maintain the political values they choose. Building on the practice of collective management of copyright, it also empowers data trusts to govern and monitor the use and re-use of the concerned data. The model is ultimately meant to address the power imbalance and information asymmetry that characterise today's economy of data as well as the "data winter" effect that restricts the accessibility of data for public interest, while accommodating and empowering individuals and communities that may have different political values and visions.

Keywords Open licenses · Open data · Data commons · Data trusts · Data trust · Licensing personal data

Melanie Dulong de Rosnay designed and wrote the license template, its principles and elements as part as the SUDACO (Sustainable Data Commons) collaboration convention between the Centre Internet et Société, CNRS and the Open Knowledge Foundation. https://okfn.org/en/projects/sustainable-data-commons/.

Y. Benhamou (✉)
Ph.D.; Associate Professor of Digital, Information and Media Law, Digital Law Center, University of Geneva, Geneva, Switzerland
e-mail: yaniv.benhamou@unige.ch

M. Dulong de Rosnay
Ph.D.; Research Director, Center for Internet and Society (CIS), National Centre for Scientific Research (CNRS), Paris, France
e-mail: melanie.dulong@cnrs.fr

Published online: 09 September 2025

1 Introduction

Data are subject to a multitude of rights and legal regimes (e.g. original works or personal information posted on social media, that may be subject to copyright, and data protection) and shared through different voluntary non-standardised, non-interoperable contractual terms.[1] This leads to fragmented, sometimes conflicting legal regimes and contractual terms. This has also become an even major challenge in the AI era, for example, when digital platforms and service providers set their own Terms of Services (ToS) in business-to-consumer relationships (B2C). Instead of letting platforms and service providers impose their own ToS to control data, and in order to give control back to rightholders, a solution would be to rely on open licenses. These are standard licenses,[2] according to which an object (e.g. an artistic work or software) is available to all but subject to their terms. Open licenses are *inter alia* used for public administration (Open Government Data)[3], research (Open Research Data), software (Free Libre and Open Source), data published on Wikipedia (Wikidata), collaborative creation (Creative Commons), cultural heritage, and the private sector.[4] Their benefits are obvious: (i) they lower transaction costs, by reducing the need to develop and read complicated contractual terms; and (ii) they support collaborative processes, by permitting everyone to use, innovate and improve the licensed object.[5]

However, existing open licensing schemes have shortcomings. In particular, they are insufficient to build data commons,[6] as they traditionally apply to copyrighted data only (i.e. images and software, thereby excluding personal and technical data); they furthermore lack modularity in their terms (e.g. as they do not foresee the licensor to restrict certain types of uses or users). Further, while some actors turn to standard open licenses, intermediary platforms and service providers continue to

[1] Voluntary data sharing can take the form of individual contracts between two or more actors (*e.g.* data transfer or exchange contract) or standardised contracts (*e.g.* general terms and conditions, or privacy policy). *Cf.* Swiss Federal Institute of Intellectual Property (IGE) (2021); Benhamou and Tran (2016), p. 572.

[2] These standardised licenses are offered by various organisations, *e.g.* the General Public License (GPL) by the Free Software Foundation www.gnu.org/licenses/gpl-3.0.en.html, accessed 4 December 2023, or Creative Commons licenses by the Creative Commons, an international non-profit organization https://creativecommons.org, accessed 4 December 2023.

[3] Open data often come from public data, *i.e.* data that can be used, edited, analysed and transmitted freely, without legal, financial or technical restrictions.

[4] Private companies are gradually using open data, in particular for economic reasons (the open data market being estimated at EUR 550 billion in Europe by 2025) but remains limited, especially for the purpose of protecting intellectual property rights and personal data. Data are then usually shared through publications (*e.g.* annual reports), on a website and/or an application programming interface (API) containing only specific types of data (*e.g.* master data and metadata). *cf.* IGE (2024), p. 26.

[5] On the benefits of Creative Commons for AI, *cf.* Vezina and Pierson (2021). On the benefits of open software for AI, *cf.* Theben et al. (2021), pp. 16 ff.

[6] We refer to "data commons", as any information and knowledge (hereinafter "objects") created and/or maintained online by communities that collectively and sustainably govern data and their relationships. *See* Van Maanen et al. (2024), p. 13.

impose their ToS on end users. These shortcomings ultimately reinforce the power imbalance and information asymmetry in a today's data economy,[7] and the "data winter" effect that restricts the accessibility of data for the public interest.[8] Finally, existing open data schemes are challenged by technical developments, such as the advent of complex AI models, which raise novel questions such as the copyrightability of their underlying training data and their output.

Our proposal addresses these shortcomings. This article analyses how to extend open licenses to all data types, including personal and technical data, while enabling granular control over data with modular clauses which allow licensors to restrict access outside of certain boundaries (e.g. authorised users and uses) – thereby ultimately fostering a commons-based approach. It also examines how to modulate open data so as to accommodate for different political values and visions, and so as to "future proof" its use in AI models (Sect. 2). Turning towards implementation, the present contribution then proposes to combine its open license model with collective data governance exercised through data trusts, in order to ensure their monitoring and enforcement (Sect. 3). Finally, the article offers a template for open licenses, embedding the broadest scope of commons-based values and future interoperability with other licensing systems (Sect. 4).

The following legal analysis does not focus on a specific jurisdiction and applicable law, as legal regimes vary from one jurisdiction to another and from one dataset to the other. Particular attention is given, however, to EU law,[9] notably in view of the global influence of the EU's General Data Protection Regulation (GDPR).[10] Finally, this contribution does not claim to be exhaustive or immediately implementable, but rather consists of an exploratory path to build a data common ecosystem with a bottom-up approach. Further work will require collaboration with data protection and IP scholars as well as other stakeholders.

2 Open License Spanning All Data Types: Towards Open Data Commons Licenses

In this section, we analyse how to extend open licenses traditionally designed for copyrighted works (e.g. text, images and software), to all kind of data, including personal data and technical data. In order to build the proposed Open Licenses, we need to revisit the paradigms of open licensing schemes (contractual chain, liability and moral rights), and define legal solutions to address their limits (consent validity and the right to erasure) since they were neither designed to cover personal nor technical data.

[7] Pasquale (2015); Beer (2017), p. 1; Kitchin (2017), p. 14.

[8] *See* Longpre et al. (2024), on the urgent need to preserve data access for the public interest.

[9] For an overview of the open data policies and strategies implemented by the Member States of the Union, see the Open Data Maturity Reports published yearly by the European Commission.

[10] Bradford (2020).

2.1 Copyrighted Data

2.1.1 Open Licenses and the Contractual Chain

Open licenses traditionally facilitate the sharing of copyrighted works, such as software with Free, Libre and Open Source licenses, and artistic works with Creative Commons licenses.[11] With these open licenses, the creator (licensor) allows any interested user (licensee) to access the licensed work, to modify it[12] and then make it freely available again under certain conditions.[13] It relies on copyright to apply it backwards and convert exclusivity to inclusivity.[14] Open licenses promote the circulation of the work and public access to it while limiting reuse under certain conditions, unlike the public domain which has no legal protection and leads to a risk of potential reappropriation. Open licenses facilitate collaboration and enrichment, aligned with the open source and free culture movements, through an *ex ante* permission to redistribute and make derivatives.[15]

With open licenses, a contractual chain exists between the author of the primary work ("primary author") and multiple downstream users.[16] A new license agreement is concluded between the primary author and each downstream user.[17] A distinction is made between a "simple" user of a copy and a user who creates a derivative work ("derivative user"), whose status is acquired by a creative contribution to the primary work (and not by simply accessing to the work, or by adding technical modification to the work).[18] A distinction is also made between a "simple" derivative work in the sense of copyright derivative,[19] and works that

[11] We define "copyrighted works" as any data subject to copyright, or related rights, under the applicable national law in accordance with international treaties (including the Berne Convention, the TRIPS Agreement, and WCT (WIPO Copyright Treaty) and WPPT (WIPO Performances and Phonograms Treaty), which can cover original intellectual creations, like software and original databases or other original arrangements.

[12] Except in the case of non-free licenses reserving derivative rights such as Creative Commons No Derivative.

[13] Examples of licenses for software include: MIT License, the General Public License (GNU/GPL), Apache License, BSD License, and CeCILL. For AI specifically, *see* the Open Source AI Definition 1.0. Examples of licenses for literary and artistic creations include Creative Commons License and the Free Art License. Despite variations in their level of openness, they all have the objective of encouraging the circulation of the object with or without the possibility of enriching it. *Cf.* Fröhlich-Bleuler (2012). On Creative Commons, *see* Dulong de Rosnay (2009), p. 7.

[14] *Cf.* Dusollier (2015), p. 50, in reference to the reversal of "copyright" to "copyleft"; "all rights reserved" to "some rights reserved".

[15] Theben et al. (2021), p. 16. *Supra* Sect. 1.

[16] Guadamuz (2004), p. 331.

[17] Jaccard and Cellina (2017), pp. 229 ff, 236.

[18] Creative Commons License, Arts. 1(a) and 2(a)(4) ("Simply making technical modifications […] never produces Adapted Material"). Bénabou and Farchy (2007), p. 37.

[19] We define "copyright derivatives" here as works (developed by a downstream user) based on an earlier work (developed by an upstream contributor) which remains recognizable and substantial in the new work.

have undergone derivations of such importance that they acquire a status of a new primary work distinct from the primary work.[20]

2.1.2 *Extension of Open Licenses to All Data, Including AI Training Data*

Some open licenses intend to propagate their open terms to all copies and copyright derivatives with a copyleft or Share Alike clause, by obliging to include the same terms in further distributions of the object (so-called propagating effect of copyleft clauses).[21] The terms become almost enforceable *erga omnes*, as opposed to contractual provisions that are normally enforceable *inter partes*.[22]

In the AI era, however, some argue that copyright does not extend to training data (and so neither do the exceptions), as training data are used for their informational content only, not for their copyrighted elements (e.g. tokenised fragments of texts for text generators and magma of pixels for image generators).[23] With this interpretation, training data under open licenses may not propagate their contractual terms to other elements within AI models (e.g. datasets that incorporate the training data or the output arising from the training data). However, we could argue the other way around, i.e. that the use of AI training data would propagate their open conditions to other elements within AI models.[24] This seems aligned with a strict interpretation of copyright applicable in most jurisdictions.[25] It also tracks with the recently published Open Source Artificial Intelligence Definition 1.0, which foresees that full access to training data is not required to meet the definition of

[20] Bénabou and Farchy (2007), p. 37.

[21] *E.g.* Art. 2 MIT License ("The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software"); art 3 Creative commons License ("If You Share the Licensed Material (including in modified form), You must […] indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License"). *Cf.* Dusollier (2015), p. 50, indicating that the circulation is thus ensured by the "physical" attachment of the text of the license to the digital file of the work, which makes it enforceable against any user like a property right.

[22] Dusollier (2015), p. 50.

[23] *See* Lazarova et al. (2021); Margoni and Kretschmer (2022), pp. 685–701.

[24] *See* Benhamou and Tran (2016), p. 579, giving as an example of propagating effect of Art. 5 of GNU GPL 3.0. All programs based on GNU GPL license are subject to the terms of this license, unless "identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves" (Art. 2 GNU GPLv2). On the propagation of open licenses to derivative in the AI context ("AI Derivatives"), *see* Benhamou (2024), considering that "AI Derivatives" deserve a different approach than software derivatives and that contractual terms may extend to AI derivatives.

[25] *See* EU law at Art. 2 of EU Directive 2001/29/EC that protects identical or partial reproduction, directly or indirectly, by any means, in whole or in part; recital 105 EU AI Act that confirms that the use of copyrighted works for AI training purposes has copyright relevance. *See* Benhamou (2021a), p. 398; Cherpillod (2023), p. 445; and Szkalej and Senftleben (2024), analysing that copyright relevance depends on the stage in the AI lifecycle, the initial stages (creating a training dataset or converting data into desirable formats) being copyright-relevant, while the last stages (creating an artefact) are copyright-relevant, as the artefact exists independently from the copyright-protected data that have been used for training purposes.

"open source", notably when the said data is private or legally restricted.[26] The model proposed in the Annex is in line with this second interpretation, i.e. that the use of data in an AI system triggers application of the copyleft clause to the system's output.

This approach, however, poses a number of challenges and limitations.[27] First, it assumes that elements within an AI system are considered as derivative of the training data under a copyleft clause, while it is not clear if there needs to be a minimum amount (or threshold) of the data for it to trigger the propagating effect of the copyleft clause. Secondly, the propagating effect is likely to apply in the training phase, less if at all, once the model is trained and to the generated output. For instance, in the training phase, the training dataset usually contains and reproduces training data. Consequently, it is an open question whether the copyleft terms[28] may or may not propagate to the training dataset. However, the trained model neither contains nor reproduces its training data (either fully or partially), and so neither does the generated output (except for the cases of memorisation of training data). Consequently, the copyleft terms arguably do not propagate to the trained model or to the generated data. Thirdly, the generated output is often not copyrightable when there is a lack of human authorship, so that the copyleft clause could apply on a contractual basis only, and not on the basis of exclusive copyright. This is what the model proposed in Annex intends to do: i.e. to apply on a contractual basis, even to non-copyrightable elements (see Sect. 2.3.1).

Finally, technical opt-out solutions exist to ensure enforcement of such terms, such as indicating the terms under which the data can be used (e.g. robots telling *data crawlers* and *data scrapers* whether or not to use certain data); this is a growing practice since the reservation of right in Art. 4 of the DSM Directive.[29]

2.1.3 Liability of Users and of Authors

Open licenses can raise user liability if they do not adhere to the license terms (e.g. omitting the mention of the author when it is required) or author liability if the object causes harm to users (e.g. a dysfunctional object causing damage or infringing a third party's rights).[30] Authors, users and third parties may then attempt

[26] *See* "The Open Source AI Definition – 1.0", https://opensource.org/ai/open-source-ai-definition, accessed 27 November 2024.

[27] On the propagating effect of copyleft clauses in the AI context Benhamou (2024), claiming that the AI context differs from the software context; Szkalej and Senftleben (2024), p. 313.

[28] Guadamuz (2023).

[29] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (DSM Directive), OJ 2019 L 130/92. *See* Keller (2024), discussing the different policies and ways to implement machine-readable opt-outs.

[30] *E.g.* an open source guidance software for autonomous cars that would cause the vehicle to lose control because of a design error. *See also* Benhamou (2021b), p. 422.

to claim liability based on contractual or tort liability, regardless of eventual contractual disclaimers in the license.[31]

Contractual liability presupposes the existence of a contract, which is the case with open licenses. The licensor makes a pre-formulated offer to an undefined number of users, which users tacitly accept when they use the object in a way that complies with the license terms.[32] There must be a wrongful conduct or negligence, which depends on the promised qualities, and may be admitted, for example, when the documentation provided with software reveals promised qualities but not when a work is admitted "*as is*".[33] Open licenses generally provide for an exclusion or limitation of liability clause,[34] which is not valid in all jurisdictions, in particular in civil law jurisdictions for cases of fraud or gross negligence.[35] Thus, the primary or derivative author who knowingly (or involuntarily) incorporates works in violation of copyright could incur liability towards the copyright infringement victim or the downstream involuntary infringer, notwithstanding the exclusion clause.

From the tort liability standpoint, it is conceivable that third parties who consider themselves harmed by the object of the license may wish to sue the author, especially since the exclusion of liability clause is not enforceable against them. However, the causality condition may be difficult to establish or may even be interrupted by the fact that the author does not address a particular user and may ignore whether the subject matter may cause harm, thus excluding the author's tort liability.[36]

[31] With regard to the liability of authors towards users of Open Source software, *cf.* Fröhlich-Bleuler (2012) p. 5; with regard to the liability of users towards authors of works licensed under Creative commons, *cf.* Jaccard and Cellina (2017), p. 236.

[32] Cellina (2020), p. 241 and references, considering that the user can easily know the content of the license and that, on the contrary, some consider that the simple use of the object is not sufficient to admit the existence of an acceptance of the conditions. This is a more general question of the integration of general conditions, the rules of which have been developed by the jurisprudence and for which we allow ourselves to refer to the references cited by Benhamou (2021b). To note is also that open licenses contain typical elements of contract (*e.g.* scope of the license and liability issues) and the validity of the licenses has been recognised by many courts, especially in France, Germany and the USA. For references, *cf.* Fröhlich-Bleuler (2012), p. 4. For a recent examples in case-law: in France, *see* TGI Paris 21 June 2019 3ème ch 3ème section, recognising the GNU-GPL license as a (membership) contract; in the USA, *see Jacobsen v. Katzer* 535 F.3d 1373 (Fed. Cir. 2008).

[33] With free software, the documentation sometimes contains an implicit guarantee that the software is *fit-for-purpose*. In the field of cryptocurrencies, the documentation even states that the codes have been audited by an independent firm.

[34] *E.g.* The Creative Commons license or the MIT license, stating that the license is offered as is, that it does not guarantee non-infringement of third-party IPRs and/or that it excludes liability for damages caused to third parties.

[35] This explains the Creative Commons license restriction (Art. 5(b) "Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You"). For references to comparative law, *cf.* Benhamou (2012), p. 101 ff., recalling other legislation preventing such exclusions, such as consumer protection and contract law.

[36] As an alternative to tort liability, one could imagine applying liability for the creation of a dangerous state of affairs when the object of the license is intended for specifically dangerous uses (*e.g.* software allowing the guidance of autonomous cars or the printing of 3D files, such as weapons or other dangerous objects) or liability for defects of the thing in the sense of the Product Liability Directive (PLD). For an analysis of liability under the PLD, *cf.* Benhamou and Ferland (2021)

Our model in the Annex proposes a liability clause corresponding to high European standards of consumer law since malpractice and infringement threatens the construction of safely reusable commons.

2.1.4 Limits Related to Moral Rights

Open licenses generally grant the right to modify the original work in advance to any users without having to ask permission from the licensor. They are also irrevocable, so that the author no longer has control over copies of the work once it is made available.[37] This sort of waiver of the author's right to exercise his/her rights, in particular the moral rights over the work, is not valid in all jurisdictions, especially those that enshrine the non-transferability of moral rights and the prohibition of excessive commitments.[38]

These difficulties led to creation of the Creative Commons licenses Version 4.0 so as to better suit some national legislations. A clause stating that authors "agree to waive or not assert any publicity, privacy, or personality rights that you hold in the material you are licensing, to the limited extent necessary for others to exercise the licensed right" minimises the effect of moral rights on the licensees' ability to use the work and ensure that the license operates as intended internationally.[39] The author also retains the right to be recognised as the copyright owner, thus ensuring that the attribution right is upheld.[40]

However, the transfer of moral rights and of all rights in advance to all future works is considered to be excessive so as to constitute a personality infringement.[41] Following this interpretation, we agree that authors should be able to reconsider their commitments at any time. This implies the offering of an *opt-out* right to authors, in particular so that they can object to an alteration contrary to their personality, if the use of the work is denaturing.[42]

[37] Jaccard and Cellina (2017), p. 236; Ginsburg (2013), pp. 3 ff.

[38] This is the case even where there is a foreign element, depending on the jurisdiction, as ruled in France by the Cour de Cassation in the case of "*Asphalt Jungle*" (Cour de Cassation Civ 1ère 28 May 1991 89-19.522 89-19.725); *cf.* Bénabou and Farchy (2007), p. 26.

[39] Creative Commons, "Frequently Asked Questions" (25 May 2023) https://creativecommons.org/faq/#how-are-publicity-privacy-and-personality-rights-affected-when-i-apply-a-cc-license, accessed 15 November 2023. Prior to that, there were national versions in order to comply with moral rights. *E.g.* the French CC 3.0 version stated that licensee had to respect licensor moral rights to the extend licensor could not renounce to them (see clause 4.d). Since CC 4.0, national versions have been abandoned, because the Share Alike clause allowed licensing of derivatives under another jurisdiction's license which may not have this clause and in order to reach compatibility among one another's versions.

[40] *Cf.* License CC BY 4.0, Art. 2(b)(1).

[41] Concerning the prohibition of excessive commitments (and criteria such as intensity, duration, and purpose), *cf.* Marchand (2010), CC 27 N 14; Benhamou and Tran (2016). Concerning the non-transferability of moral rights, *cf.* de Werra (2013), LDA 7 N, recalling that the principle and scope of this non-transferability is, however, controversial, some considering moral rights as non-transferable, others proposing to distinguish between the hard core of moral rights corresponding to the protection of the right of the personality (which is non-transferable) and the outer sphere of the moral right (which is transferable).

[42] *Cf.* Bénabou and Farchy (2007), p. 27.

The exercise of such right of revocation is, however, difficult to put into practice: The return or withdrawal of publicly available or commercialised copies seems unrealistic, if not impossible, in the digital context.[43] It also contradicts the logic of irrevocable open licenses,[44] in which licensors waive moral rights,[45] and which do not terminate even if the licensor ceases distribution.[46] As a solution, the exercise of the right of revocation could be only valid for the future (*ex nunc*) and would not oblige users to return or withdraw already circulated copies.[47]

In view of the above considerations, our model in the Annex supports the assertion of authorship and producer attribution. It also provides an *opt-out* right to licensors in order to comply with their moral and personality rights, however, with the limitation of being valid for the future (*ex nunc*).

2.2 Personal Data

2.2.1 Extension of Open Licenses to Personal Data

Open licenses traditionally apply to copyrighted data, but exclude personal data.[48] Yet, traces of personal data can be found almost everywhere,[49] whether they have been left voluntarily or incidentally, whether mobility data (e.g. data from smart cities or Google Maps), cultural data (e.g. photographs or metadata including personal data), or data collected by social media, browsing cookies or captcha to train algorithms. Such data are often transferred directly or by third parties (e.g. Data Brokers) through personal data licenses, either by consenting to the temporary

[43] *Cf.* Bénabou and Farchy (2007), p. 28, indicating that these modalities (transaction costs and compensation) are likely to neutralise the exercise of the right of revocation. Moreover, if the revocation comes from an author of a second work, this revocation cannot have any influence on the fate of the upstream works.

[44] Creative Commons, FAQs (2023), "What happens if the author decides to revoke the CC license to material I am using? The CC licenses are irrevocable. This means that once you receive material under a CC license, you will always have the right to use it under those license terms, even if the licensor changes his or her mind and stops distributing under the CC license terms. Of course, you may choose to respect the licensor's wishes and stop using the work".

[45] *Ibid*, Clause: "Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise".

[46] *Ibid*, The CC Licensor "may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License".

[47] Another solution is to consider that the right to revoke one's commitment contradicts the authorisation given to third parties to use, modify and distribute the work, and that the exercise of the right of revocation should be conditioned to a prior compensation of all those who would suffer a prejudice because of the revocation.

[48] *Cf.* IGE (2021), p. 25. *See also* the Creative Commons 4.0 license which indicates that it covers both copyright and *sui generis* database rights (*cf.* Art. 4) but not the integrity rights, or privacy rights, or any other personality or similar rights (Art. 2(b)). This is also the case with the Open Source AI Definition 1.0, which excludes training data insofar as it may contain personal data.

[49] We define "personal data" in the meaning of Art. 4(1) GDPR as all information relating to an individual.

use of the right, or by waiving the right to claim protection during the contractual period.[50]

The question then arises whether it is possible to apply open licensing mechanisms to personal data, thus allowing a better circulation, on the one hand, and a better control by individuals on the other hand, if data subjects may determine the conditions of access and use of their data. If so, open licenses could span both personal data and other data types, bearing in mind that it has become difficult and sometimes impossible to distinguish between personal data and other data types, both legally[51] and conceptually, since the scope of privacy laws seems to expand even further in a growing range of situations due to datafication, from data analytics to AI.[52]

Open licenses already exist in the area of copyright and of non-personal data with model contracts for the exchange of technical data between companies.[53] However, while initiatives start to emerge regarding personal data,[54] we are not aware of any finalised standardised licenses applicable to all types of personal data; thus, we explore this avenue, drawing inspiration from research towards data commons or collective data governance,[55] and personal data commons.[56]

Licenses covering personal data are generally proposed in a standardised, non-negotiable manner directly by platforms and service providers: the licensee (the provider) impose its conditions on the licensor (the data subject). Here, we propose to reverse the mechanism and would allow licensors (the data subjects) to specify their own conditions for access and use of their personal data. These licensors could use standardised licenses to indicate their terms before providing their personal data in exchange for services with third parties.

[50] It remains controversial as to which type of contract should be used for personal data (*e.g.* personal data license or usage agreement). Legal scholars seem to accept the qualification of a licensing agreement, *see* Rouvinez (2011) N 170. For an in-depth analysis of contracts for the commercialisation of personal data, *cf.* Cellina (2020), pp. 158–60. For a critical discussion on personal data as economic goods, *see* Purtova and van Maanen (2024), p. 1.

[51] Finck and Pallas (2020), p. 11.

[52] Purtova (2018), p. 40.

[53] *E.g.* IGE (2021), p. 6, and templates for transfer, or exchange contracts. *Cf.* also the Microsoft model contracts, Microsoft, "Removing Barriers to Data Innovation" https://news.microsoft.com/datainnovation, accessed 15 November 2023, and the "Montreal Data License", *cf.* Benjamin et al. (2019).

[54] *See* the 2012 Privacyleft proposal; Jurcys et al. (2020), p. 13; Saint-Aubin (2012). Standardisation efforts may be worth mentioning: (i) health data between patients and cooperatives like "Salus Common Good License" https://www.saluscoop.org/licencia, accessed 15 November 2023; (ii) between providers like Swiss Personalized Health Network, "Legal Agreement Templates" https://sphn.ch/services/dtua/, accessed 15 November 2023; (iii) urban data like Métropole de Lyon, "Licence de réutilisation des Données" https://download.data.grandlyon.com/files/grandlyon/LicenceEngagee.pdf, accessed 15 November 2023; and (iv) consumer data Customer Commons https://customercommons.org, accessed 15 November 2023.

[55] Zygmuntowski (2023); Kariotis et al (2020).

[56] Wong et al. (2020). More practically, *see* all the development of a data commons infrastructure by the DECODE project (https://tools.decodeproject.eu/, accessed 15 November 2023) including the Digital Data Commons Privacy Pledge of the DECODE EU project by Bassi et al. (2023); D'Intino and Roio (2019); Peugeot (2014); Casilli (2014), pp. 423–434.

We claim that open licenses designed for copyright could apply *mutatis mutandis* to personal data: in both cases, the object of the license is an absolute subjective right (copyright and personal data protection).[57] The terms are attached to the object which thus allows the object to be tracked and contractual rights to be extended to any user,[58] creating a contractual relationship between the data subject and any downstream user (who would then become a data controller, or a data recipient). This contractual chain would strengthen the position of data subjects who could determine their own terms of use, and, in the event of non-compliance with licensing terms, enforce them by asserting a tort or contractual violation.[59] One aspect, however, that distinguishes personal data from copyright is that, unlike copyright, the granting of access to personal data contains an inherent direct risk to the data subject to be reidentified, all the more with sensitive health or research data. So, the analogy does not hold entirely. While copyright is usually win-win; personal data is more complex to manage for the data rightsholder. This advocates for a robust application and monitoring of a personal data open license. This could be operated by combining this license with technical tools (e.g. robots.txt on a website)[60] and a collective governance structure such as cooperative-like data trusts (Sect. 3).[61]

2.2.2 *Limits Related to Consent Validity*

The question of consent validity remains, given that the open licenses relate to the use of the object by an undefined number of users and uses (a kind of "worldwide consent"). This would contradict the philosophy of strict privacy laws, such as the GDPR, according to which consent must relate to one or more specific processing activities (excluding the possibility of a blanket general consent[62]) and there must be a certain control over the data. Consequently, it cannot be excluded that courts may invalidate open licenses applied to personal data, based on a strict application of the GDPR consent requirements. Another potentially difficult issue would be the traceability of the data subject's consent; if the initial service provider ceases its activities, for instance in case of bankruptcy, then the initial consent would need to be evidenced by all downstream users of the contractual chain.

As a counterargument, the open license would be limited to one or more specific processing activities, precisely those defined by the data subject using our template for Open Data Commons Licences optional conditions. Also, the freedom of

[57] Benhamou and Tran (2016), p. 579.

[58] *Cf. supra* Sect. 2.1.1.

[59] *Cf.* supra Sect. 2.1.2. *See also* Jaccard and Cellina (2017), p. 236, about Creative Commons licenses ("*en cas de violation des conditions de la licence par l'utilisateur, l'auteur pourra faire valoir ses droits qui découlent non seulement du droit d'auteur, mais aussi des termes de la licence que l'utilisateur a accepté*").

[60] *See supra* Sect. 2.1.1.

[61] *Infra* Sect. 3.

[62] Meier and Tschumy (2023) LPD 6 N 91; Meier (2010), N 883.

disposal of data means that data can also be made freely available.[63] Strict privacy laws, such as the GDPR, focus on the data subject as the origin point of the license (rather than the controller or the processor), which should thus be able to determine the use of its data.

Finally, by analogy with copyright law – with a similar debate regarding a blanket assignment – it is not possible to contractually assign in advance all copyright for all future works, but it is possible to assign one's rights of use in a work.[64] Transposing to personal data, one could argue that open license relates to specific data, not to all data relating to a data subject.

This approach based on the consent raises, however, a major challenge. The fact is that, under several privacy laws (e.g. the GDPR), consent is not the only legal basis for processing data. For instance, in the AI context, AI developers tend to rely on an overriding legitimate interest instead of training their AI models based on publicly available data (e.g. social media conversations, photos and videos).[65] One solution to ensure that consent is the legal basis for processing, and therefore that the data subject can impose the contractual terms, would be that data subjects could object to the processing of their data by exercising their opt-out rights, a growing practice since the reservation of rights in Art. 4 DSM Directive, such as indicating the terms under which data can be used (e.g. robots telling *data crawlers* and *data scrapers* whether or not to use certain data) (see Sect. 2.1.2). Therefore, our proposed model includes a right to erasure clause by default, which practical implementation is to be solved (see Sect. 2.2.3).

2.2.3 Limits Related to the Self-Determination Right and the Right to Erasure

The question of irrevocability of personality rights also remains, given that the individual data subject has the right to withdraw consent at any time. More generally, the non-transferability of informational self-determination rights covers continuous control rights, such as the right to access, rectify, erase, object against harm to integrity, and withdrawal of consent at any time.[66]

This irrevocability is known under copyright law (the so-called non-transferability of moral rights and the prohibition of excessive commitments) and has been addressed by Creative Commons licenses.[67] Just as with copyright, where moral

[63] That is why some privacy laws provide certain exceptions to the protection, in particular to the consent and/or the safeguards to the international transfer: *see* Art. 49(1)(a) GDPR; Art. 17(1)(a) and Art. 30(3) Swiss Federal Act on Data Protection (LPD).

[64] De Werra (2013), LDA 16 N 15; OGer ZH 23 March 2006, in [2006] 12 sic! 2006 851 ("PMS").

[65] EDPB (2024).

[66] Cellina (2020), p. 257; Benhamou and Tran (2016), p. 579, recalling that this irrevocability was relativised in a ruling of the Swiss Federal Court (ATF 136 III 401) which admitted the irrevocable transfer of the image rights of a model in her photos, on the grounds that here images were not part of the intangible core of the human essence like name or voice given the importance of marketing these products over the last few decades. It should also be specified that, assuming that free licenses are considered a justification for a violation of the right to privacy, insofar as the data are made publicly available to any user (Art. 30(3) FADP), the opt-out would still seem to be possible, since the revocation would mean that the individual objects to the processing, and thus the justification would no longer apply.

[67] *Cf. supra* Sect. 2.1.4.

rights also allow authors in some jurisdictions to object to any use that denatures the work, in the context of personal data we foresee that the data subject may withdraw consent and request the deletion of data (*opt-out*). However, unlike copyright Creative Commons licenses, the personal data withdrawal right should apply retroactively (i.e. to use and copies done in the past) (*ex tunc*), and not only for the future (*ex nunc*).[68]

Therefore, our proposal in the Annex includes a right of withdrawal, revocation, rectification, to be forgotten (right to erasure), which must be complied with upon request from the data subjects. The challenge is then one of practicability, since data may be shared with a multitude of actors.[69] One solution is to incorporate a metadata anchor to trace data uses (a growing practice since the reservation of right in Art. 4 DSM Directive),[70] coupled with a data trust to enforce their members' rights and requests to erasure.

If such solution is neither available nor practical, it is worth remembering that the right to erasure has limits. Where the controller has made personal data public, the controller must only inform other controllers about the request to erase data "taking account of available technology and the cost of implementation" (Art. 17.2 GDPR). Moreover, in the AI era, from the output data standpoint, training data are generally inexistent in the output and, from an input data standpoint, it seems impossible to make an AI model "forget" the training data.[71]

2.3 Technical Data

2.3.1 Extension of Open Licenses to Technical Data

Another important question is to analyse if and how open licenses can be extended to technical data. We define "technical data" as non-personal data that do not relate to an individual (e.g. commercial or machine-generated data).[72] A distinction must be made between such technical data that is subject to an exclusive right and data with no exclusive right.

[68] Another solution is to consider that revoking consent contradicts the contractual commitment given to third parties to use, modify and distribute the data and could be conditional on prior compensation of all those who would suffer harm as a result of the revocation. *See supra* note 59.

[69] This practical challenge explains why, in medical research, medical institutions often continue to use data even after revocation of consent, which can be allowed in some jurisdictions. For instance, under Swiss medical law, in the event of revocation of consent, the data must be anonymised, unless it is clear from the beginning of the research project that anonymisation is not possible, *cf.* Art. 10 of the Swiss Federal Ordinance of 20 September 2013 on Human Research (HRO; RS 810.310); Swissethics, "Model for the Development of a Written Information Sheet to be Submitted for Studies Involving Persons in Accordance with the HRA/HRO" (Version 2.4 of 6 November 2018), p. 9.

[70] *See* Keller (2024).

[71] Li (2023), p. 479; Pastis (2023), indicating that machines do not "forget" due to their design, once an AI has learned something, unlearning proves complicated.

[72] Under EU law, such data are generally defined as non-personal data as opposed to personal data and includes such technical data (*e.g.* aggregated datasets used for big data analytics or weather conditions generated by sensors or data on maintenance needs for industrial machines) as well as anonymised data (*i.e.* initially personal data that were made anonymous). *See* the Regulation on a framework for the free flow of non-personal data in the European Union and its guidance.

2.3.2 Technical Data Subject to an Exclusive Right

Technical data may be subject to property-like rights in some jurisdictions, like the EU *sui generis* database right, which provides an exclusive database right in technical data or the datasets contained therein.[73] In practice, this means that the holder of technical data to which he/she has effective or legal control ("data holder") could prohibit the downloading of the data or datasets contained therein, even if the underlying data are not protected by an exclusive right like copyright. The EU *sui generis* database protection requires that substantial investments have been made (which excludes the protection of individual data) and extends to the collection only, not to the creation of data (which excludes machine-generated data that are created instead of "obtained").[74]

Under EU law (or other jurisdictions providing an exclusive database right), licenses covering technical data or datasets ("database licenses") are common and allow data holders to shape an ecosystem around their data aligned with their values and interests. In the case of B2B, there are efforts to standardise database licenses with model contracts for the exchange of technical data between companies.[75]

In the case of B2C, a few standardised licenses can be mentioned as well (e.g. for urban and map data).[76] Moreover, in an open ecosystem, both the Open Data Commons and the Creative Commons 4.0 licenses include the EU *sui generis* database right, because their definition of "Work" includes databases. They allow their optional conditions (attribution, non-commercial use, no derivative work and sharing alike) to also apply to the *sui generis* database right and not only to their copyrightable elements.[77] Our model in the Annex follows this trend and includes any such technical data through the definition of "Work".

[73] The EU *sui generis* database right was developed in order to protect data producers' investments and to prevent free-riding on somebody else's investment in creating the database.

[74] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of Databases (Database Directive), Art. 7(1)–(2). For the limited scope of the Database Directive, *see* Benhamou (2021a), p. 398.

[75] *See* IGE (2021), p. 6, and templates for transfer, or exchange contracts, available at https://www.ige.ch/fr/propriete-intellectuelle/pi-et-societe/traitement-et-securite-des-donnees, accessed on 15 November 2023.

[76] *See* the Open Database License (ODbL), used by OpenStreetMap. *See also* Lyon and Paris Open Data for urban data, a project constructed around map data, whose contributions came initially by volunteers that intend to map faraway places in need of maps (*e.g.* for humanitarian purposes). *See* Microsoft (SUPRA NOTE NUMBER?) and the "Montreal Data License". *Cf.* Benjamin et al. (2019).

[77] CC licenses prior to the 4.0 version excluded databases from the scope of the license. The rationale was to avoid exporting database *sui generis* rights to jurisdictions where they do not exist (the same argument could have applied to moral rights). However, this approach deprived database providers from deciding under which licensing terms their database would be licensed. To fil this gap, the UK-based Open Knowledge Foundation issued the Open Data Licence that includes database into the license followed by Creative Commons with its CC 4.0 in 2012 to correct a "market failure" and "to align with expectations of users". Vollmer (2015).

2.3.3 Technical Data Subject to No Exclusive Protection

In the absence of an exclusive right, licensing technical data can only be envisioned by contractual mechanisms. A distinction must be made between data provided by a data holder to pre-determined identified contractual partners and to an undefined number of third-party users (for the sake of simplification reference will be made to "one-to-one" and respectively "one-to-many").

In one-to-one relationships (typically as in the case of a B2B bilateral contract), such licensing mechanisms are common. Indeed, in practice, contractual partners often include "data", "databases", "confidential information" and "know-how" as part of the scope of the license. The definition chosen by the parties can therefore easily include technical data. Legally, this practice is, however, questionable under the principle that it is not possible to transfer more rights than one can have.[78] If there is no exclusive right in the data, then how can the data supplier or user claim ownership, and much less license it? Regardless this legal uncertainty, some courts have recognised the validity of such contracts, considering that the common intention of the parties is to give the licensee the benefit of an asset over which the licensor has *de facto* control and can therefore negotiate the exploitation.[79]

In one-to-many relationships (typically as in the case of a B2C unnegotiated contract), such licensing mechanisms may also be envisioned. Consequently, standardised licenses applicable to databases are beginning to emerge.[80] In EU law, such licenses appear valid, as it seems possible to restrict the use of databases through contractual mechanisms.[81] The CJEU notably confirmed that when a database is not protected by copyright or the *sui generis* database right, the website owner can rely on the online terms of use to prohibit data scraping by a lawful user.[82] Nevertheless, such licenses raise the delicate issue of their validity with respect to third parties that have no contractual relationship with the data supplier (e.g. e-commerce websites or social media that provide access to data to an undefined number of users, often via general terms of use). This implies an analysis of whether and how the conditions of the open license can be imposed on any third-party user.

Abroad where there is no exclusive database protection, these licenses may be challenged or even invalided. In US law for instance, while many companies continue using different techniques to bind website's visitors (e.g. browsewrap,[83]

[78] In civil law jurisdictions, this principle arises from the principle of "*nemo plus juris ad alium transferre potest quam ipse habet*", *see* Gilliéron (2021), p. 443.

[79] Gilliéron (2021), p. 443, quoting BGH, NJW 2012, 3512, spec. 3514.

[80] *See* the Community Data License Agreement v2 license by the Linux Foundation to address the "growing need to share and collaborate with data sets to analyse and use for AI training" and used notably by IBM. *See* MooreMoore and Puri (2023).

[81] *Cf.* Case C-30/14 *Ryanair v. PR Aviation* [2015] ECR-I 15 (EU:C:2015:10).

[82] The enforceability of such a clause contained in the online terms of use is then a matter of national law.

[83] Link on the screen with notice that using the site indicates acceptance of terms.

sign-in-wrap,[84] clickwrap[85] or scroll wrap agreements), some techniques such as browsewrap agreements are often unenforceable, in particular against the consumer due to insufficient notice. On the other hand, browsewrap agreements are enforceable when the website or app contains an explicit textual notice that continued use will act as a manifestation of the user's intent to be bound. Consequently, website owners can still rely on the enforceability of browsewrap agreements, provided that the consent to those agreements complies with case-law.

Our model in the Annex follows this trend and is intended to extend to such non-copyrighted data ("non-IP-data") through the definition of "Work". The validity, enforceability of this extension, of capturing/contractualising of non-IP-data will be ensured by data trusts, which will ensure that the delivery and consent mechanism of the model contract complies with the case-law, in particular on web and app design.

2.4 Restricting Openness to Support the Commons and Privacy

2.4.1 The Concept of "Open" According to Open Licenses and the Open Definition

Data Commons contends that the underlying data should be governed as a common, rather than by the state or surveillance capitalism. However, it is essential to draw a clearer distinction between open data, which is available to all, and data commons, where access to the data can be differentiated between community members and third parties. The establishment of data commons must also address the apparent contradiction between personal data and property, and between privacy and open access. A personal data commons would not entail the sharing of personal information but rather govern its reuse in accordance with the shared values of the digital commons.[86]

Consequently, our model in the Annex offers a set of optional elements meant to be sufficiently modular and adaptive to cover the needs of personal and non-personal data, to respect privacy and to accommodate a range of different political values to be defined and refined by individuals and communities,[87] from fully open data to options influenced by licensing schemes implementing reciprocity and peer

[84] Notify the user of the existence of the contract and advise to click on the button to proceed to the next screen.

[85] Also known as "click-accept" or "clickthrough".

[86] Dulong de Rosnay and Stalder (2020). Those values are defined in the Data Commons Manifesto and analysed in, also citing this fragment, Senabre Hidalgo et al. (2024), p. 45.

[87] For data commons design principles, *see* Tarkowski and Zygmuntowski (2022). Their recommendation to choose who can access and use data corresponds to the clause on authorised users and uses in the license template proposed in the Annex, and their call to a democratic management of data as a commons corresponds to Sect. 3 on data trusts.

production values to build sustainable data commons ecosystems.[88] Indeed, privacy and commons-based purposes may be achieved through setting further conditions applicable to the nature of the user or to the kind of use, even if these restrictions break the non-discrimination openness principle at the core of free software and open licenses.[89] Instead, an Open Definition revision could include the respect of the rights contained in a Privacy Pledge (Sect. 2.4.2)[90] and the openness could be achieved within a closed consortium or among certain users based on their status or type of usage as long as the purpose is commons-based (Sect. 2.4.3).[91] These restrictions enforce certain shared values and translate a certain vision of data altruism.[92] They would require governance and monitoring by a collective data trust (Sect. 3).

2.4.2 Restrictions Related to Privacy (Privacy Pledge)

A Privacy Pledge should be set by default and applies whenever personal data are involved. The Privacy Pledge is a guarantee for the licensor that the licensee will use the object in respect of personal data, e.g. by use of high technical standards of cybersecurity and human rights protections. Encryption protocols and privacy-by-design benchmarks can serve to enforce these standards. In case of breach of the Privacy Pledge, the license terminates, including for downstream users due to the propagating effect of the license. The breaching party may further held be liable for breaching the contract, which should be enforced by the data trust. Minimum privacy criteria are expressed in the Privacy Pledge, which both licensor and licensee promise to respect. The Pledge could be used as a compulsory standard, similar to other definitions in the FLOSS licensing world.[93] It could also be upgraded, and subsequent versions put through a certification process by issuer groups and serve as benchmark to certify other licenses as compatible. Finally, other

[88] Reciprocity and peer production values have been conceptualised in two commons-based licensing schemes: the Peer Production license is "allowing commercial commons-based reuse by cooperatives, and non-profit actors, but preventing "commercial entities intent on making profit through the commons without explicit reciprocity." The Copyfair license requires reciprocity in commercial activities, it "aims to subject commercialization of any such knowledge commons to some form of contribution to that commons". *See* Dulong de Rosnay (2021), pp. 109–122.

[89] https://opendefinition.org/, accessed on 15 November 2023; Romback et al. (2022), indicating that the OpenRAIL license already proposes use-based restrictions banning "social cost, harmful use, restricted scenarios".

[90] https://privacy-pledge.com/, accessed on 15 November 2023; https://zenroom.org/privacy/, accessed on 15 November 2023.

[91] Bassi et al. (2023).

[92] Data altruism can be defined as the voluntary sharing of data for general interest purposes *in the general interest*.", *see* Art. 2(16) DGA.

[93] The Open Definition https://opendefinition.org/od/2.1/en/, accessed 15 November 2023; Wikipedia, "The Open Source Definition" https://en.wikipedia.org/wiki/The_Open_Source_Definition, accessed 11 December 2023; Wikipedia, "The Free Software Definition" https://en.wikipedia.org/wiki/The_Free_Software_Definition, accessed 11 December 2023; Wikipedia, "Definition of Free Cultural Works" https://en.wikipedia.org/wiki/Definition_of_Free_Cultural_Works, accessed 11 December 2023.

declarations related to specific fields (e.g. patient data) or specific to the mission of a closed-door consortia (a project, a scientific association) could be added.

2.4.3 Restrictions Related to Authorised Users and Uses to Protect the Commons

The restrictions on shareability included in our licenses, which are used to support and enforce the shared values and purposes of commons, are based on the most recent generation of open licenses (e.g. the Peer Production license, the Copyfair, Copyfarleft, and Coopyright policies).[94] These conceptualised models restrict openness in order to protect the commons and develop a sort of Non-Commercial No Derivatives license outside of certain boundaries.[95] The commercialisation of the object remains free for those who contribute to the commons, while a fee must be paid by other stakeholders. This restriction is similar to traditional knowledge or other community governance benefit sharing mechanisms.[96] It also echoes the Open Source AI principle of the Open RAIL-M license[97] governing the use of Stable Diffusion which contains use-restrictions for ethical concerns.[98]

In order to accommodate the full spectrum of political economy models, from commercial to full openness, an option in our proposed licenses also accommodates compatibility with commercial reuse and, in turn, data monetisation. In order to comply with the EU's data privacy standards, defined also to support collective goals such as the contribution to the commons, the licenses can also be adapted in a more granular way, for instance by accepting the processing of personal data only for selected accredited public interest purposes and/or only under certain pre-defined conditions.

The granularity of the licenses is implemented through optional terms defining authorised users and usages in a manner similar to the different modulations of the Creative Commons licenses. Rights can be reserved or granted only to accredited users (e.g. non-profit education, open access research, climate change, public health research or any other high-level public interest or common purpose defined by the data trust) or usages (e.g. UN sustainability goals, ethical or responsible use, competition-friendly use, clinical data use, vulnerable persons-friendly use, environmental-friendly use, no surveillance use). The clauses defining authorised users and usages could be merged; however, since they correspond to different

[94] P2P Foundation, "Peer Production License" http://wiki.p2pfoundation.net/Peer_Production_License, accessed 15 November 2023; P2P Foundation, "Copyfair license" http://wiki.p2pfoundation.net/CopyFair_License, accessed 15 November 2023; and P2P Foundation, "Copyfair" http://wiki.p2pfoundation.net/Copyfair, accessed 15 November 2023; P2P Foundation, "Copyfarleft license" http://wiki.p2pfoundation.net/Copyfarleft, accessed 15 November 2023; La Coop des Communs, "Le "Coopyright" pour La Coop des Communs" (26 December 2017) https://coopdescommuns.org/fr/le-coopyright-pour-la-coop-des-communs/, accessed 6 December 2023.

[95] For instance, the Peer Production license, an adaptation version of the Creative Commons NonCommercial option. Only other commoners, cooperatives or non-profit can share and reuse, and commercial benefits must remain in the collective.

[96] Jeet Singh and Gurumurthy (2021).

[97] Romback et al. (2022).

[98] BigScience, "BigScience RAIL License v1.0" (19 May 2022) https://huggingface.co/spaces/bigscience/license, accessed 15 November 2023.

ontological categories, they are kept separate in the proposed model. This also facilitates the licenses' compatibility with existing Rights Expression Languages,[99] machine-readable metadata expressing licensing conditions, which may distinguish users and uses.

The last optional choice, Share Revenues, is meant to support financial sustainability for data commons. It can lead to revenue sharing for the benefit of data trusts and supporting the commons, the community, the infrastructure, fund data maintenance, or data care operations, through a fee collected and managed through collective management entities, like CMOs for radio broadcasting since the 1970s, and file-sharing legislative proposals in the 2000s.[100] The question of offering an individual remuneration in exchange of digital labour or the use of personal data is highly controversial, because introducing data ownership could lead to an unbalanced power relationship between users and platforms. This proposal is not meant to dispossess users of fundamental rights in their personal data, which could be transferred as assets for a couple of cents, but rather to dedicate revenues from reserved rights to support collective projects which can be local, commons- or values-based.

These optional terms are meant to foster the construction of sustainable data commons which are not necessarily fully open access. The status of Authorised User or Authorised Usage will require a certification, called Shared Values Network and Stamped Usage, based on characteristics of the re-user or of the reuse, or by reference to an existing standard. Such a process would be similar to a license compatibility assessment, a process already led by the Creative Commons and the Free Software Foundation as described in the next section.

If such modular terms can accommodate different political theories and values, they raise a major challenge. They could lead to a fragmentation of different open licenses, away from their goal of standardisation and simplification, and to a compatibility challenge among open licenses. A solution to mitigate this risk is to rely on collective management to ensure standardisation, effectiveness and enforceability of the template, as well as the definition of compatibility clauses with other open licenses (Sect. 3).

2.5 Compatibility Among Open Licenses

The current existing licensing frameworks have been created separately without resolving the question of their compatibility or interoperability. Attempts have been made to assert such compatibility between different licenses, but the fact remains

[99] Coyle (2004).

[100] Aigrain (2012).

that they sometimes follow radically different logics or traditions, making such exercises arduous.[101]

Legal compatibility is necessary when merging datasets under different licenses, in particular when data are subject to different, sometimes conflicting legal regimes and are inextricably linked (so-called mixed datasets).[102] One solution is the insertion of a downstream compatibility clause declaring that the object (and its derivatives) must be licensed under the same license, or under licenses deemed compatible.[103]

This can be achieved through three solutions. First, unilateral compatibility only requires a declaration by the source. It has notably been achieved in this way by Creative Commons BY SA (to GNU-GPL), and by Open Government License (to Creative Commons).[104] Second, reciprocal or bidirectional compatibility requires evaluation and bilateral negotiations. This has been achieved between Free Art Licence 1.3 and CC BY-SA 4.0.[105] Licensees may reuse works under both licenses and the product can be further licensed under any of these two licenses. Third, in the same vein, the Creative Commons compatibility recognition process refers to respecting the Free Cultural Work definition as "minimum compatibility criteria"; a solution is to use a standard as a proxy to help assess compatibility among the proposed Data Commons Licences and other licenses. This leads to the proposed solution of an Open Definition and a Digital Data Commons Privacy Pledge to define openness, commons principles, and personal data protection standard as minimum criteria, based on a Europeanist extra territorial effect, reflecting the imperative nature of data protection.

Our proposal invites to reflect on compatibility before releasing new licenses for public use in order to avoid data silos. The ideal solution would be (i) reciprocity; if

[101] Bénabou and Farchy (2007), p. 35. In addition to legal compatibility discussed in this section, technical compatibility could be envisioned via computable contracts, *i.e.* to give robots or other automated means a set of specifications, in order to search the relevant dataset with the relevant data under specific licenses, including the degree of openness (*see infra* 4.1). This practice (of computable contract) may be a growing practice since the reservation of rights in Art. 4 DSM Directive, *see supra* 2.1.1 and Keller (2024).

[102] Dulong de Rosnay (2009); EU Guidance on the Regulation on a framework for the free flow of non-personal data, COM(2019) 250 final, Brussels, 29.5.2019 (Guidance on the FFD), 9, defining "mixed datasets" as situations when data disentanglement is (i) not be technically feasible and (ii) results in depleting the dataset's value.

[103] Dulong de Rosnay (2009), p. 81. This is the case in Creative Commons Share Alike 4.0 licenses where BY-NC-SA and BY-SA "Compatible License means a license listed at creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License". The license applied by the licensee to adaptations "must be a Creative Commons license with the same License Elements, this version or later, or a (BY-NC-SA or BY-SA) Compatible License".

[104] This is the case for the UK Open Government Licence towards CC BY and Open Data Commons BY: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/, accessed on 15 November 2023. This is also the case for CC BY 4.0 towards GNU GPLv3: "Compatibility with the GPLv3 is one-way only, which means you may license your contributions to adaptations of BY-SA 4.0 materials under GPLv3, but you may not license your contributions to adaptations of GPLv3 projects under BY-SA 4.0".

[105] "The Free Art license 1.3" http://artlibre.org/licence/lal/en/ was declared a "BY-SA–Compatible License for version 4.0 on 21 October 2014". Creative Commons, "Compatible Licenses" https://creativecommons.org/share-your-work/licensing-considerations/compatible-licenses, accessed on 15 November 2023.

this is not achievable, then (ii) unilateral compatibility; and (iii) in the absence of declared compatible license by the licensing framework issuer, a long-term option is a process led by data trusts.

3 Collective Management Through Data Trusts

In this section, we explain how to set up the proposed implementation template of licenses (alternatively of similar open licensing schemes) with collective governance instruments, in particular in the form of data trusts (or trust-like institutions). This should ensure effectiveness and enforceability of the template, as well as compatibility with other open licenses, along with control or sovereignty of individuals and collectives over their data.[106] We also recommend drawing inspiration from copyright collective management organisations (CMOs) due to their similarities with data and long-standing experience in exercising rights on behalf of others.

3.1 Setting-Up Intermediaries with the Mandate to Subject Data Rights to the Model Contract

Open licensing should be combined with collective data governance to ensure its effectiveness and enforcement, including licensor opt-out, necessary for personal data and moral rights. Collective governance instruments are a growing trend of relying on a third party[107] – such as "data trusts" (DTs)[108] or "data rights intermediaries" (DRIs)[109] collectively referred to here as "data intermediaries" (DIs)[110] – allowing individuals to influence the terms by which their data may be

[106] There are a variety of theories and tools for data collective governance, *see* Micheli et al. (2020) mention "data sharing pools, data cooperatives, public data trusts and personal data sovereignty".

[107] This trend is fostered in the EU with the Data Governance Act (DGA), that encourages independent data management services via intermediaries, in particular commercial data intermediaries and altruistic organisations. EU Regulation on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), L 152/1.

[108] "Data trusts" refer here to an organisation, a trust, in which trustees make decisions about data use on their behalf. Normally, data trusts steward data on behalf of others (with independent stewardship, institutional safeguards, and collective action), and can take different legal forms, from real "trust institutions" in common law jurisdictions to "trust-like institutions" in civil law jurisdictions (*e.g.* via data agent-style contractual relationships), according to which individuals task an agent to exercise their data rights on their behalf under a fiduciary responsibility to act in the individual's best interests, *see* Data Trust Initiative, "Data Trusts: from theory to practice" (Working Paper, 26 November 2020) 4.

[109] "Data rights intermediaries" refer here to third parties that facilitate or enable the collective exercise of data rights under a mandate (by general-purpose, or specific *proxy*). They should be distinguished from data trusts, that steward data on behalf of others. *See* Giannopoulou et al. (2022), pp. 316–331.

[110] "Data intermediaries" are defined here as an umbrella term for organisations that manage data on behalf of others under a mandate (*supra* note 109) or under stewardship with fiduciary obligations (*supra* note 108), *see* Open Data Trust Initiative, "Data trusts: international perspectives on the development of data institutions" (Working Paper, 28 January 2021) 4; Delacroix and Lawrence (2019), pp. 236–252.

used and to redress the bargaining power imbalance and information asymmetry between individuals and businesses.[111]

Such collective governance instruments have the potential to allow licensors to transfer their rights according to specific terms (i.e. according to the mandate for DRIs, or to the fiduciary obligations for DTs). However, to be effective, they need to be further developed and adapted to the different types of data and goals.[112] Indeed, DIs can have different prerogatives, ranging from exercising portability and access rights on behalf of one or more data subjects or leveraging data rights to obtain better terms and conditions from service providers, to pooling, centrally storing, and granting access to data.[113]

For the purpose of this article, in terms of prerogatives, DIs should have, as a minimum, the mandate to leverage data rights to obtain terms and conditions from service providers, in particular to subject data rights (e.g. access and use of data) to our template (or similar open licensing models). This can be done by using our contract model with its compulsory elements as a default minimum standard contract (like the Creative Commons licenses can be used with the CC BY).[114] This does not necessarily require DIs to pool, centrally store and/or grant access to data. It is sufficient for the DI to have the mandate to negotiate, and to subject the data rights to compliance with the template.[115]

Moreover, DIs can have further prerogatives, such as pooling and directly granting access to the data, or the power to negotiate more granular conditions through optional elements (e.g. restricted users and uses, compatibility among licenses).[116] This could be particularly relevant for more complex datasets and sensitive data and/or communities willing to foster certain values. This requires the DIs to either pool or centrally store the data, in order to grant access or to grant stewardship in order to exercise extensive data rights (e.g. define and negotiate the optional elements, and the compatibility among the licenses). Effectively, DIs would become a DT, on top of having a DI status.

[111] *See supra* note 3 . *Cf.* also Element AI and Nesta, "Fiducies de Données – Un nouvel outil pour la gouvernance des données" (White Paper, 2019), p. 9 https://hello.elementai.com/rs/024-OAQ-547/images/Fiducies_de_Donnees_FR_201914.pdf, accessed on 15 November 2023, indicating that current models are flawed in many ways: (i) users have no power to negotiate the terms of use that provide for increased use of their data, to which they blindly consent without understanding its scope; and (ii) companies often do not know or prefer not to disclose all uses of data. According to several studies, 99% of privacy policies are not understandable to high school graduates and it would take the average user 244 hours to read the entire privacy policies of the sites they visit annually.

[112] Goals may vary according to the individual, or community's values and privacy preferences. *cf.* Element AI and Nesta (*supra* note 111), p. 15.

[113] For instance, in healthcare, there are often DIs that help the management of research data but with different uses and prerogatives. For instance, in Switzerland, some DIs pool and store research data of Swiss universities and hospitals in secure infrastructures (e.g. BioMedIT nodes) in order to grant access to others, while other DIs "just" redirect researchers and businesses to the holder of the research data and provide usage data template (*e.g.* SWISSUbase). Giannopoulou et al. (2022), p. 318.

[114] For the presentation of the Creative Commons licenses, *see supra* Sect. 2.1 (Copyrighted Data).

[115] For an analysis of the possibility to grant such mandate over personal data, *see* Giannopoulou et al. (2022).

[116] For the optional and compulsory elements, *see* the legal explanations of the template below in the Annex.

With either minimum or more extensive prerogatives, DIs would allow the management of complex datasets, increase data sharing, and improve trust and control over the data, ultimately correcting the power and information asymmetry (by offering the licensor the ability to decide the terms) and fostering a data common ecosystem (by propagating open licenses). In order to fully exercise its mission, in particular to enforce the license, its optional elements and compatibility among licenses, we believe that it is advisable to rely on DTs with stewardship and fiduciary obligations, which can grant extensive powers to exercise data rights under a fiduciary responsibility to act in the individuals' interests with *erga omnes* effect and institutional safeguards.[117] Individuals, or communities would delegate the power to manage their data to a DT and exercise their self-determination through their memberships.

The question of feasibility remains, in particular the adoption and scalability of such model, bearing in mind that licensors, in particular data subjects, sometimes lose interest in the protection and control over their data. This is the case in the privacy context with the concept of "consent fatigue", namely data subjects who are tempted to agree with only one click instead of reading lengthy and incomprehensible terms of use, or with the concept of "privacy resignation", namely data subjects who realise they are powerless, and simply give up.[118] Adoption and scalability requires first to set up universal, or thematic DIs (e.g. per sector, such as social media with social media data, music platforms with user data, smart cities or local cooperatives with urban data, health providers or patient organisations with health data). Then, for data already collected and in use by stakeholders, DIs would exercise mass access requests towards providers to retrieve the data and license them back under the license (or similar open licensing schemes).[119] For data not already collected, setting up universal or thematic DIs would incentivise platforms to voluntarily adhere to such terms (Fig. 1).

3.2 CMOs as Blueprint to Design Intermediaries for Open Data Commons

DTs may take different legal forms. The choice of the legal forms depends on the sector and the objectives pursued by the community (e.g. political economy values, sensitiveness of data). This being said, it seems interesting to combine both a corporate structure and a contractual architecture to impose internal rules among the members and external rules towards users, and the more complex the environment (e.g. changing membership and evolving goals), the more appropriate a corporate structure will be.[120] The Open Data Institute (ODI) has also identified the following success factors (or drivers of adherence) based on pilot-projects in different contexts (e.g. urban data, health data, online platforms): the importance of privacy, which

[117] For an analysis of the stewardship, fiduciary responsibility, and institutional safeguards, *see* Data Trust Initiative (*supra* note 108), p. 5.

[118] For an analysis on the consent fatigue and privacy resignation, *see* Draper and Turow (2019), p. 1824.

[119] For the ability to grant data rights to DIs, in particular the exercise of access requests, Giannopoulou et al. (2022), p. 318.

[120] Reed and Ng (2019).

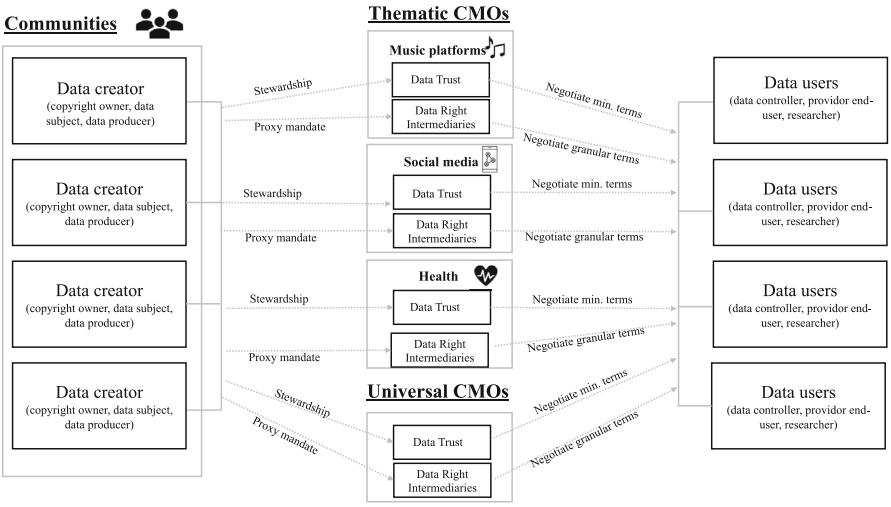

Fig. 1 Overview of the ecosystem with the different actors, and their relations

can vary according to personal preferences and sectors involved; the importance of social equalities and recognition of a plurality of interests, which can be enhanced through inclusion and representativeness on bodies; the importance of transparency, which can be fostered by subjecting the DT to transparency laws; the importance of the values conveyed by the DT.[121]

As model of inspiration that fulfils all the above functions, and success factors, we draw inspiration from CMOs due to their similarities with data, with data trusts and their long-standing experience in exercising rights on behalf of one or more right owners.[122]

Firstly, CMOs and DTs deal with similar rights (licensed objects). Copyrighted works and personal data are absolute subjective rights (with unwaivable rights in some jurisdictions), which can be transferred to a third party for management purposes.[123] Their exploitation is subject to a massification and out-of-control use: in the copyright context, this led to the emergence of CMOs, while in the data-driven economy, DIs are called on to manage the vulnerabilities to which

[121] *Cf.* Open Data Trust Initiative (*supra* note 108). These success factors echo the Ostrom's eight classical governance principles governing the commons. *See* Ostrom (1990). Ostrom principles are not easily transposed to intangible goods. However, there are efforts to port Ostrom design principles to knowledge commons, *see* Frischmann et al. (2014) and other books in this collective series, and to data commons, *see* Ruhaak et al. (2021).

[122] For a general analysis and history of CMOs, *see* Gervais (2010).

[123] With open licenses, the terms are attached to the resource which thus allows the resource to be tracked and contractual rights to be extended to any user, creating a contractual relationship between the data subject and any downstream user (who would become a data controller), *see supra* Sect. 2.1 (Copyrighted Data).

individuals and groups are exposed, and to empower individuals to influence the terms by which their data can be used.[124] Secondly, CMOs and DTs are similar in their legal mission and structure: CMOs are non-profit organisations, created by rightsholders, to administer copyright and monitor usages (namely to grant authorisation to use a copyrighted work to a multitude of users, by collecting the royalties due according to the copyright).[125] The legal structure depends on the jurisdiction, but they mostly take the form of corporate structures (e.g. associations, or cooperatives), with accessibility and equal treatment to all members (open door principle to all new members and equal vote principle of all members), appropriate participation in the society's decisions (internally) and clear transfer to the CMO which makes the right enforceable towards third parties (externally). Membership in the collective administration and the transfer of copyright is generally effected by the conclusion of a membership contract, supplemented by general management conditions (akin to general terms and conditions (GTCs)).[126] Thirdly, CMOs have long-standing experience in negotiating tariffs depending on the type of use (i.e. the "price" of using each copyrighted material), and of managing copyright, including that of non-represented owners thanks to extended collective licenses (ECLs).[127] This could be developed for massive use of both personal and non-personal data (including AI training data) as well, or to support and maintain the commons' infrastructure by charging certain categories of users or uses.

This echoes privacy DTs and their success factors identified by pilot projects (importance of social equalities, transparency, and values conveyed by the DT).[128] Consequently, DTs could take the form of CMOs, as non-profit associations or cooperatives, that guarantee equal treatment of all members, accessibility to all rightsholders, and appropriate participation in the society's decisions.[129] A DT would transfer, on a fiduciary basis, the rights of the owners it represent, all of whom would become members of the DT and take part in its activities. On the top of that, DTs could issue price recommendations to help individuals or companies determine the amount of remuneration they wish to charge for the use of the data.[130]

[124] Open Data Trust Initiative (*supra* note 108), p. 4.

[125] Gervais (2010), pp. 1 ff.; Cavalli (2005), Internet p. 181, recalling that, historically, they were the unions responsible for assisting their members, first in their negotiations with theaters, then in the field of musical and literary rights.

[126] Gervais (2010), pp. 1 ff.

[127] For a definition and references relating to ECLs, *see* Benhamou and Ferland (2022), p. 403.

[128] *E.g.* Nesta Foundation, Decidim civic participation and *op cit*. design principles for governing the commons, the knowledge commons and the data commons.

[129] It is specified that the non-profit condition refers to the corporate structure, not to the members who may want to monetise their data, and that it is not part of the identified success factors, but of the concept of altruistic organisation and trust law, so it could be removed depending on the DT models chosen.

[130] Benhamou and Ferland (2022), pp. 165 ff, suggesting that damages be based on tariffs per infringement, as is done in European copyright law (*e.g.* the royalty method) or in American copyright law (statutory damages).

3.3 Towards Legal and Political Compatibility Definition

DIs, DTs and CMOs play a role for the interoperability between datasets governed by different licenses, or managed by different intermediaries, and standardisation would facilitate transfer and portability from one trust to another. Legal compatibility can be achieved either upfront through certification by open license drafters' organisations, or later by DTs based on declaratory shared values.[131] A risk remains of multiplications of DTs and licensing proliferation, which is an inevitable caveat of the FLOSS world. Legal compatibility requires cooperation with other trusts and communities, all of which may differ in their sought political economy values.[132]

Members would delegate the power to manage their data to DTs and CMOs, and exercise their self-determination through their memberships. Members would select their DT like a political party, an association, or a trade union, and elect board members based on shared political values. Values, uses and users can be defined and standardised in Rights Expression Languages shared ontologies.[133] Values can also be published in declarations of principles, similar to the Open Definition or Privacy Pledges as in the proposed model. Further work requires collaboration between data protection scholars, open and commons actors. DTs could implement a commons-based governance, collect fee and share fees, towards redistribution to licensors, or to cover shared infrastructure maintenance costs. Certification protocols could be developed to negotiate and agree on behalf of licensors/data subjects whether a proposed use/user can be granted based on conditions of access and reuse.

4 Conclusion

This article suggests exploratory paths to develop an open licensing model applicable to all kinds of data (including personal and non-personal data, and mixed datasets covered by multiple intricated rights, preventing users from having to consider multiple conflicting legal and licensing frameworks), and to enable commons-based collective data governance, management of multiple datasets of different natures, while reinforcing the control of individuals over their data. Open licenses have been designed for collaborative copyrightable resources. We propose to extend them to personal data, to develop collective data governance models, and to interpret and enforce data commons licenses, through data trust structures, inspired by copyright management societies and commons-based pilot projects. We propose in the Annex a template for open licenses supporting privacy and the commons. To facilitate reading and future compatibility recognition, the template has drawn inspiration from the Creative Commons framework. This template has

[131] Such management data trusts would define rules based on their members' values, a process that is in line with Ostrom second design principle which recommend communities to define their own rules.

[132] Again, in line with Ostrom's principle No. 8, which recommends, as a success factor for commons, coordination with other groups.

[133] *See* W3C, "ODRL Information Model 2.2" https://www.w3.org/TR/odrl-model/, accessed on 15 November 2023; Abelson et al. (2012).

the triple view of: (1) implementing the findings of this article to apply open licensing to all kinds of data, including personal and non-personal data; (2) implementing both privacy and commons-based values; and (3) relying on the development of CMOs or data trusts for the application of these licenses, including for compatibility assessment, in the same way the Open Source Initiative (OSI) maintains a list of "approved licenses".

Our approach contains certain limitations. When it comes to extending the proposed model to the AI context and to technical data, the challenge is to enforce contractual mechanisms to non-copyrightable elements (e.g. elements of AI models that are not copyrightable or that do not contain or reproduce training data) as well as in one-to-many relationships (typically as in the case of a B2C unnegotiated contract). When it comes to extending the proposed model to personal data, the challenge is to enforce the contractual terms to processing activities based on other legal grounds such as an overriding legitimate interest. Another challenge is that the proposed model with modular terms can lead to a fragmentation of different open terms. Solutions to overcome these challenges include to enforce contractual mechanisms via opt-out rights, a growing practice in copyright, and to rely on a standardised contract like our proposed template (alternatively of similar open licensing schemes) with collective governance and negotiations instruments, in particular in the form of data trusts (or trust-like institutions).

This proposal is a first step towards the design of a revamped open licensing framework which would go beyond existing open licenses[134] and be compliant with copyright, privacy and other laws.[135] Regarding the applicable law, we use EU regulations as a high standard baseline due to their practical influence and their essential focus on privacy and self-determination. However, in order ensure compatibility with existing legal regimes, the licenses align with the strictest standards of data protection and moral rights. While we are aware that not all practical questions are solved, such as the enforceability or the technical implementation of withdrawal rights, the proposed template offers a first panorama of the rights and values for a data licensing scheme, which would be both applicable to the widest possible range of data, and compatible with other existing and future schemes (e.g. between full open-access and semi-closed commons).

The normative position regarding open data and the commons, i.e. the need to avoid the extraction and private appropriation of the value created by communities' as a side effect of open licensing,[136] prompted the development of this new scheme. Another design goal is interoperability, since the incompatibility between software components, works or datasets licensed under different licenses is a long-standing pitfall of existing open licensing systems. The proposed model thus endeavours to cover and accept the full scope of approaches, from more liberal open data to data commons and even stricter data governance flavours. The scepticism regarding the benefits of opening and sharing data, on the usefulness of open data, could be a

[134] Dulong de Rosnay (2021), pp. 109–122; Dulong de Rosnay (2016).

[135] As envisioned in the ODECO EU project towards a sustainable Open Data ECOsystem, *see* Skłodowska-Curie Grant Agreement No. 955569.

[136] Bodó (2019).

justification to provide a model which is less open, but representing the fully open version allows to have a model which is both conceptually complete and malleable. As capitalism has already taken advantage of openness, open data does not necessarily benefit citizens, as mentioned in the introduction; the model proposed thus does not ambition "to empower or help citizens to protect themselves against big tech",[137] but to enable communities to define the terms governing their data with more granularity than existing open licenses. Some may fear that such a legal approach towards complex data and capitalism-related problems with the development of tools may lead to those tools being abused and/or ignored by companies and overlooked by citizens. Such a concern echoes one of the earliest critiques[138] expressed towards Creative Commons relying on the master's tools (copyright licensing) to try to dismantle the master's house (as copyright extension was the power imbalance into question in the early 2000s). A consultation of public, private and commons actors as well as critical data, personal data, legal design and open licensing experts is ongoing to further refine and develop this template with the Open Knowledge Foundation.

5 Annex: Template for Open Licenses Supporting Privacy and the Commons

The proposed template is inspired by standard open licenses and the Creative Commons framework. It contains definitions, the scope of the license (optional and compulsory elements), and a selection of standard clauses such as liability and termination.

Optional elements are meant to provide a modular template sufficiently adaptive for communities and data trusts to accommodate different political theories and values.[139] They provide that data may be stored, copied and shared, may or may not be enhanced, correlated to create derivative data, that data must remain free or can be monetised, that data must be fully or partially anonymised, and, crucially, that terms may limit who may access the data and reserve the categories of downstream users or uses.

Compulsory elements are meant to cover the needs of copyrighted and personal data.[140] They contain a Privacy Pledge embedding minimum privacy criteria which both licensor and licensee promise to respect whenever personal data are involved. They also contain a right of withdrawal (or right to erasure), to respect the unwaivable moral rights and self-determination right recognised in certain jurisdictions and which must be complied on request of licensor. For copyrighted data, this right of withdrawal can be limited to the future (*ex nunc*) and, for personal data, must cover the past (*ex tunc*), as the right to erasure can be used retroactively.[141] Consequently, the template provides a reservation of personality

[137] Purtova and Van Maanen (2024), p. 1.

[138] Dusollier (2006), pp. 101–123.

[139] *See supra* Sect. 2.4 (Restricting openness to support commons and privacy).

[140] *See supra* Sects. 2.1.4, 2.2.2 and 2.2.3.

[141] *See supra* Sect. 2.2.3.

rights (following the reservation of moral rights in CC 3.0), an *opt-out* right if the data subject wants to withdraw consent and a scope of anonymity is provided: full anonymity means strict confidentiality with fully anonymised data; limited anonymity means semi-private with pseudonymity, or aggregation for research or statistical purposes.[142]

Regarding moral rights, the template supports the assertion of authorship and producer attribution,[143] but permits the waiving of attribution and authorship, for instance to meet the expectations common in big data science, which often involves reusing numerous databases and makes strict compliance with attribution and authorship challenging.[144]

The model also applies a Share Alike clause that enables to propagate its terms to derivatives, by requesting to keep the same license.[145] The notion of derivative is defined differently from the "copyright derivatives" of most open licenses (which is somehow far removed from the traditional notion of "copyright derivatives") in order to span all data derived from the initial data like aggregated and output data in the AI-context,[146] as well as (non-copyrighted) technical data ("non IP-data").[147] The right to make derivatives permits to reuse and modify the licensed object (*e.g.* extract, transform, adapt, aggregate, translate, update for scraping, as well as other kind of processing).[148] The option of controlling the access to data (not only the right to make derivatives) is also included in the model, as it can be a useful restriction, for instance to decide whether derivatives can be deployed for public use, or granted within a closed consortium only.[149] This allows to address privacy or to avoid that open data feeds closed platforms or surveillance models.[150]

Other standard clauses include a termination clause to ensure that the compulsory elements are respected (Privacy Pledge, Share Alike, right to withdraw) and a liability clause corresponding to high European standards of consumer law since malpractice and infringement threatens the construction of safely reusable commons.[151] This choice is at odds with free software and open licensing schemes, where the licensed object is offered "as is", without warranty that it does not contain infringing content or data. Putting a higher burden on the licensor is meant to ensure that data can be reused safely. This legal duty of care clause contributes to a virtuous propagation virality. For each downstream use, there is a new cascading license, and it will avoid C to turn against A because of B, or C to turn against B

[142] Guimarães Moraes et al. (2021), pp. 32–47.

[143] *Supra* Sect. 2.1.4.

[144] Open Knowledge, "Open Definition" https://opendefinition.org/od/2.1/en/, accessed 15 November 2023.

[145] *Supra* Sect. 2.1.1.

[146] *Supra* Sect. 2.1.2.

[147] *Supra* Sect. 2.3.3.

[148] *See* the Montreal Data Licence: Benjamin et al. (2019).

[149] For instance, some governmental data, personal data, scientific data, health data or certification of general interest data such as Open Food Fact should not be modified.

[150] Tarkowski and Warso (2023); Huang and Siddarth (2023).

[151] *Supra* Sect. 2.1.3.

because of A, as with the European Union Public License (EUPL) liability clause.[152]

Clause 1. Definitions

Consortium: Closed group authorised to exercise rights.

Collective Management Organisation: A body mandated to exercise rights on behalf of licensors and assess compatible licenses, compatible users and compatible uses. Similar to Data Trust.

Compatible license: Licenses recognised as compatible include future versions of the Open Data Commons Licences (revisions) and other licenses recognised as compatible by the license issuer or the Content Management Organisation and listed below.[153]

Compatible user: Rights are granted only to users accredited as compatible to become licensees.

Compatible use: Rights are granted only for uses accredited as compatible.

License elements: Optional and compulsory elements, illustrated by pictograms or icons. The scope of rights and the conditions can be selected by individuals or communities according to their needs.

Licensed material: Includes works, databases, personal data, technical data, mixed data, any sort of data and derivatives. For the sake of clarity, "derivatives" span, and are defined here as "any data derived from the initial data like, in the AI-context, aggregated and output data based on training data, no matter of the amount of training data used and of the recognizability into the output, and like technical data that may or may not be subject to copyright or database protection".

Licensee: Individual rightsholders or data subjects, or their representative, a collective or a community who accesses or makes derivatives of the material and accepts these conditions.

Licensors: Individual rightsholders or data subjects, or their representative, a collective or a community.

Clause 2. Scope of license. Optional conditions

This material is made available under the terms of this license, which combines certain compulsory and optional license elements.

Licensed rights are granted by the license and can be exercised by the licensee. These conditions apply to the exercise of rights granted.

[152] Commission, "European Union Public License" https://joinup.ec.europa.eu/collection/eupl/eupl-text-eupl-12, accessed 15 November 2023.

[153] Based on the Creative Commons clause: "means a license listed at creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License" https://creativecommons.org/share-your-work/licensing-considerations/compatible-licenses/, accessed 18 June 2023.

Condition 1: Request attribution, or not

- Authorship and producer attribution
- No attribution needed

Condition 2: Authorise the making of derivatives or not

- Right to make derivatives

This option grants the right of adaptation, modification, derived version, compilation, aggregation, reuse, or data processing, it being that any treatment should be considered as data processing.[154]

Derivatives may be licensed under this license or, if merged with resources licensed under compatible licenses, the derivative product may be licensed either under this License version, this License future versions, or under licenses recognised as compatible licenses and listed as such.

- No derivative

This option reserves data integrity or for selected users/uses

Condition 3: Request confidentiality, or partially, or not. A spectrum of three options is to be selected, from closed to open.

- Full anonymity[155]
- Limited anonymity
- Clear

Condition 4: Define the scope of use and users

- Public use (full public Open Access)

Data is made available in open standards to anyone (for any use, even for commercial purposes).

- No open access use (manually controlled use)

No open access use has options for licensors to control the data destination:

- The scope of the closed doors (consortium only, or accredited users and uses) and
- The financial condition (for free or for a fee).

Data are not made available publicly and subjected to choices 5, 6 and 7.

Condition 5: Determine the scope of compatible users or uses certified by a data trust, CMO, or pledges and external definitions to be referred to

- Rights can be exercised for free only within the consortium
- Rights can be exercised for free only by compatible users or trusts based on shared values

[154] *See* Schreiber (2020).

[155] Jurcys et al. (2020), p. 9.

– Rights can be exercised for free only for agreed purposes or by special categories of re-users

Condition 6: Financial conditions

– Rights can be exercised within the consortium or by compatible use or user for free
– Rights can be exercised by compatible use or user only for a fee and this fee will be dedicated to a common purpose and managed by the consortium or a designed CMO to share revenues)

Clause 3. Values. Compulsory elements

These conditions are compulsory and part of the grant of all licenses' baseline (in the same way Attribution is a compulsory element in all Creative Commons licenses). Rights can be exercised by licensees only if they respect certain standards or values. In some instances, the conditions apply to the licensor before licensing the material.

– Compulsory element 1. Privacy Pledge
– Compulsory element 2. Right to erasure
– Compulsory element 3. Share Alike

Other Standard Clauses

Term and Termination

The License applies for the term of the Rights licensed here and terminates automatically in case of breach of terms.

Notice and *license persistence*

The licensee must keep a link to this license, or a compatible license when sharing it even in an unmodified version. Copies and derivatives must carry a link to this license or a compatible license.

Enforcement and *jurisdiction*

Different mechanisms could be set up before judicial procedure: tracking through metadata, with the assumption reuse projects further carry the license, mediation, arbitration or alternative dispute resolution boards. The Creative Commons 3.0 International Governmental Organisations licenses have a clause addressing mediation and arbitration.[156]

Most open licenses usually do not mention any applicable law, and the practice of legal porting or national legal adaptations has been abandoned by Creative

[156] *See* clause 8.h. at the very end of CC BY 3.0 available at https://creativecommons.org/licenses/by/3.0/igo/legalcode, accessed 15 November 2023.

Commons, since it created internal incompatibilities. Some open licenses (such as the European Union Public Licence[157]) contain an applicable law clause, which does not preclude compatibility recognition.

Limited disclaimer of warranty if not fit for purpose

This excludes a waiver of liability for personal data or copyright infringement.

Liability

The licensor certifies to own or have obtained all the necessary rights to grant the license. The licensee, when redistributing, also certifies to have obtained all the necessary rights to grant the license.[158]

Funding Open access funding provided by University of Geneva.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Books

Aigrain P (2012) Sharing: Culture and the Economy in the Internet Age. Amsterdam University Press, Amsterdam

Bradford A (2020) The Brussels effect: how the European Union rules the world. Oxford Academic, Oxford

Cellina E (2020) La commercialisation des données personnelles—aspects de droit contractuel et de protection des données. Schulthess. JD thesis, University of Geneva

Frischmann B et al (eds) (2014) Governing Knowledge Commons. Oxford Academic, Oxford. https://doi.org/10.1093/acprof:oso/9780199972036.001.0001

Gervais D (2010) Collective management of copyright and related rights, 2nd edn. Kluwer Law International, Alphen Aan Den Rijn

Meier P (2010) Protection des données. Stämpfli, Berne

Ostrom E (1990) Governing the Commons: the Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge

[157] *See* Commission, "Implementing decision (EU) 2017/863 of 18 May 2017 updating the open source software license EUPL to further facilitate the sharing and reuse of software developed by public administrations" COM(2017) L 128/59 Art. 15.

[158] *See* EUPL (*supra* note 152), Art. 6: "The original Licensor warrants that the copyright in the Original Work granted hereunder is owned by him/her or licensed to him/her and that he/she has the power and authority to grant the Licence. Each Contributor warrants that the copyright in the modifications he/she brings to the Work are owned by him/her or licensed to him/her and that he/she has the power and authority to grant the Licence. Each time You accept the Licence, the original Licensor and subsequent Contributors grant You a licence to their contributions to the Work, under the terms of this Licence".

Pasquale F (2015) The black box society: the secret algorithms that control money and information. Harvard University Press
Rouvinez J (2011) La licence des droits de la personnalité. Schulthess. JD thesis, University of Lausanne

Book Chapters

Abelson H et al (2012) CC REL: The Creative Commons rights expression language. In: Dulong de Rosnay M, De Martin JC (eds) The digital public domain: foundations for an open culture. Open Book Publishers, Cambridge, p 149. http://books.openedition.org/obp/557
Benhamou Y, Ferland J (2021) Artificial intelligence & damages: assessing liability and calculating the damages. In: D'Agostino G, Gaon A, Piovesan C (eds) Leading legal disruption: artificial intelligence and a toolkit for lawyers and the law. Thomson Reuters-Yvon Blais, Montreal, pp 165–197
De Werra J (2013). In: de Werra J, Gilliéron P (eds) Commentaire Romand – Propriété intellectuelle. Basel, Helbing Lichtenhahn
Dulong de Rosnay M (2021) Open licensing peer production. In: O'Neil M, Pentzold C, Toupin S (eds) The Handbook of Peer Production. Wiley-Blackwell, Hoboken, pp 109–122
Ginsburg J (2013) Author's transfer and license contracts under US copyright law. In: de Werra J (ed) Research handbook on intellectual property licensing. Edward Elgar, Worcestershire, pp 33-ff
Marchand S (2010) In: Pichonnaz P, Foëx B (eds), Commentaire Romand – Code civil. Helbing Lichtenhahn, Basel. CC 27
Meier P, Tschumy N (2023) In: Meier P, Métille S (eds) Commentaire romand – Loi fédérale sur la protection des données, Helbing Lichtenhahn, Basel. LPD 6

Journal Articles

Beer D (2017) The social power of algorithms. Inf Commun Soc 20(1):1–13. https://doi.org/10.1080/1369118X.2016.1216147
Benhamou Y (2021a) Big data big data and the law: a holistic analysis based on a three-step approach—mapping property-like rights, their exceptions and licensing practices. Revue Suisse de droit des affaires et du marché financier 4:393–418
Benhamou Y (2021b) Intelligence artificielle: licence libre et gouvernance collective des données à travers l'altruisme des données et les data trusts. Revue Suisse de droit des affaires et du marché financier 4:419–434
Benhamou Y, Ferland J (2022) Digitization of GLAM collections and copyright: policy paper. GRUR Int. https://doi.org/10.1093/grurint/ikac024
Benhamou Y, Tran L (2016) Circulation des biens numériques: de la commercialisation à la portabilité. Revue du droit de la propriété intellectuelle, de l'information et de la concurrence
Cavalli J (2005) La gestion collective des droits d'auteur exploités sur Internet. Internet 2005: travaux des journées d'étude organisées à l'Université de Lausanne le 17 novembre 2004 et le 10 mars 2005.
Cherpillod I (2023) Intelligence artificielle et droit d'auteur. iusnet P 445
Delacroix S, Lawrence N (2019) Bottom-up data trusts: disturbing the "one size fits all" approach to data governance. IDPL 9(4):236–252. https://doi.org/10.1093/idpl/ipz014
Draper N, Turow J (2019) The corporate cultivation of digital resignation. New Media Soc. https://doi.org/10.1177/1461444481983333
Dulong de Rosnay M, Stalder F (2020) Digital commons. Internet Policy Review 9(4). https://doi.org/10.14763/2020.4.1530
Dusollier S (2006) The master's tools v. the master's house: creative commons v. copyright. Columbia J Law Arts 29(3):101–123
Dusollier S (2015) Le jeu du copyleft entre contrat et propriété. Cahiers du droit de l'entreprise 6:49–56
Finck M, Pallas F (2020) They who must not be identified—distinguishing personal from non-personal data under the GDPR. Int Data Privacy Law 10(1):11–36. https://doi.org/10.1093/idpl/ipz026
Giannopoulou A et al (2022) Intermediating data rights exercises: the role of legal mandates. IDPL 12(4):316–331. https://doi.org/10.1093/idpl/ipac017

Gilliéron P (2021) Intelligence artificielle: la titularité des données. RSDA 4:435 ff
Guadamuz A (2004) Viral contracts or unenforceable documents? Contractual validity of copyleft licenses. EIPR 26(8):331–339
Guimarães Moraes T et al (2021) Open data on the COVID-19 pandemic: anonymisation as a technical solution for transparency, privacy, and data protection. IDPL 11(1):32–47. https://doi.org/10.1093/idpl/ipaa025
Jaccard M, Cellina E (2017) Les Creative Commons, avenir du droit d'auteur. SJ II pp 229 ff
Jurcys P et al. (2020) My data, my terms: a proposal for personal data use licenses. Harv JL & Tech
Kariotis T et al (2020) Emerging health data platforms: From individual control to collective data governance. Data Policy. https://doi.org/10.1017/dap.2020.14
Kitchin R (2017) Thinking critically about and researching algorithms. Inf Commun Soc 20(1):14–29. https://doi.org/10.1080/1369118X.2016.1154087
Li T (2023) Algorithmic destruction. SMU L Rev 75(3):479 ff. https://doi.org/10.2139/ssrn.4066845
Margoni T, Kretschmer M (2022) A deeper look into the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. GRUR Int 71(8):685–701. https://doi.org/10.1093/grurint/ikac054
Micheli M, Ponti M, Craglia M, Berti Suman A (2020) Emerging models of data governance in the age of datafication. Big Data Soc 7(2). https://doi.org/10.1177/2053951720948087
Purtova N (2018) The law of everything. Broad concept of personal data and future of EU data protection law. Law Innov Technol 10(1):40–81. https://doi.org/10.1080/17579961.2018.1452176
Purtova N, van Maanen G (2024) Data as an economic good, data as a commons, and data governance. Law Innov Technol 16(1):1–42. https://doi.org/10.1080/17579961.2023.2265270
Schreiber A (2020) Mere access to personal data: is it processing? IDPL 10(3):269–277. https://doi.org/10.1093/idpl/ipaa005
Senabre Hidalgo E et al (2024) Co-creation of the digital democracy and data commons manifesto: alternative sociotechnical visions of data. Open Res Europe 4:45. https://doi.org/10.12688/openreseurope.17020.2
Szkalej K, Senftleben M (2024) Generative AI and Creative Commons licences, the application of share alike obligations to trained models, curated datasets and AI output. JIPITEC 15(3)
Van Maanen G et al (2024) Data commons. Internet Policy Rev. https://doi.org/10.14763/2024.2.1748
Wong J et al (2020) Data protection for the common good: developing a framework for a data protection-focused data commons. Data Policy. https://doi.org/10.5281/zenodo.3965670
Zygmuntowski JO (2023) Data governance in a trilemma: a qualitative analysis of rights, values, and goals in building data commons. DISO. https://doi.org/10.1007/s44206-023-00058-y

Online Documents

Bassi E et al (2023) Licensing of digital commons including personal data—update. DECODE. https://decodeproject.eu/publications/licensing-digital-commons-including-personal-data-update. Accessed 15 Nov 2023
Bénabou V-L, Farchy J (2007) La Mise à Disposition ouverte des œuvres de l'esprit. CSPLA. https://www.culture.gouv.fr/espace-documentation/rapports/Commission-du-CSPLA-sur-la-mise-a-disposition-ouverte-des-aeuvres. Accessed 15 Aug 2025
Benhamou Y (2012) Dommages-intérêts sous l'angle de la méthode des redevances suite à la violation de droit de propriété intellectuelle en droit suisse et comparé. (JD thesis, University of Geneva 2012) p 251
Benhamou Y (2024) Open source AI: does the copyleft clause propagate to proprietary AI models? Revisiting the Definition of derivatives in the AI-context. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4859623#. Accessed 16 Jun 2025
Benjamin M et al (2019) Towards standardization of data licenses: the Montreal data license. https://arxiv.org/abs/1903.12262. arXiv:1903.12262 [cs.CY]. Accessed 15 Nov 2023
Bodó B (2019) Was the open knowledge commons idea a curse in disguise? – Towards sovereign institutions of knowledge. https://doi.org/10.2139/ssrn.3502119
Casilli A (2014) Quatre thèses sur la surveillance numérique de masse et la négociation de la vie privée. In: Richard J, Cytermann L (eds) Etude annuelle du Conseil d'Etat, Le numérique et les droits

fondamentaux. La Documentation française. pp 423–434. https://shs.hal.science/halshs-01055503. Accessed 15 Aug 2025
Coyle K (2004) Rights expression languages: a report for the Library of Congress. Library of Congress. https://www.loc.gov/standards/relreport.pdf. Accessed 6 Dec 2023
Creative Commons (2023) Frequently asked questions. https://creativecommons.org/faq/#how-are-publicity-privacy-and-personality-rights-affected-when-i-apply-a-cc-license. Accessed 15 Nov 2023
D'Intino A, Roio D (2019) Smart contracts for data commons integrated with GDPR compliant legal rules and tested in pilots. DECODE. https://decodeproject.eu/file/697/download. Accessed 15 Nov 2023
Dulong de Rosnay M (2009) Creative Commons licenses legal pitfalls: incompatibilities and solutions. Creative Commons Nederland. https://www.creativecommons.nl/downloads/101220cc_incompatibilityfinal.pdf. Accessed 14 Aug 2025
Dulong de Rosnay M (2016) Alternative policies for alternative internets. J Peer Prod. http://peerproduction.net/issues/issue-9-alternative-internets/experimental-format/alternative-policies-for-alternative-internets/. Accessed 16 Jun 2025
EDPB (2024) Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en. Accessed 14 Aug 2025
Fröhlich-Bleuler G (2012) Open source compliance. Jusletter. https://jusletter.weblaw.ch/fr/juslissues/2012/686/_10750.html. Accessed 15 Nov 2023
Guadamuz A (2023) Creative Commons and AI training. Technollama blog. https://www.technollama.co.uk/creative-commons-and-ai-training. Accessed 9 Jun 2025
Huang S, Siddarth D (2023) Generative AI and the digital commons. cs.CY https://arxiv.org/abs/2303.11074. arXiv:2303.11074 [cs.CY]. Accessed 4 Dec 2023
Jeet Singh P, Gurumurthy A (2021) Economic governance of data: balancing individualist-property approaches with a community rights framework. Draft for discussion at Quarterly Roundtable of Data Governance Network. https://itforchange.net/sites/default/files/1880/Economic-governance-of-data.pdf. Accessed 14 Aug 2025
Keller P (2024) Considerations for implementing rightholder opt-out by AI model developers. Open Future. https://openfuture.eu/publication/considerations-for-implementing-rightholder-opt-outs-by-ai-model-developers/. Accessed 14 Aug 2025
Lazarova A et al (2021) Creative Commons statement on the opt-out exception regime / rights reservation regime for text and data mining under Article 4 of the EU Directive on Copyright in the Digital Single Market. https://www.researchgate.net/publication/358730256_Creative_Commons_Statement_on_the_OpO-Out_Exception_Regime_Rights_Reservation_Regime_for_Text_and_Data_Mining_under_ArticlA_4_of_the_EU_Directive_on_Copyright_in_the_Digital_Single_Market_Summary_Posi. Accessed 15 Aug 2025
Longpre et al. (2024) Consent in crisis: the rapid decline of the AI data commons. https://arxiv.org/abs/2407.14933. arXiv:2407.14933 [cs.CL]. Accessed 15 Aug 2025
Moore T, Puri R (2023) New collaborative data license agreement created to make sharing data sets easier. IBM Developer Blog. https://developer.ibm.com/blogs/new-collaborative-data-license-agreement-makes-sharing-data-sets-easier?mhsrc=ibmsearch_a&mhq=new%20collaborative%20data%20license%20agreement&mhp=0. Accessed 15 November 2023
Pastis S (2023) A.I.'s un-learning problem: researchers say it's virtually impossible to make an A.I. model "forget" the things it learns from private user data. Fortune. https://fortune.com/2023/08/30/researchers-impossible-remove-private-user-data-delete-trained-ai-models/. Accessed 15 Nov 2023
Peugeot V (2014) Données personnelles: sortir des injonctions contradictoires. Vecam. https://vecam.org/2002-2014/article1289.html. Accessed 15 Nov 2023
Reed C, Ng I (2019) Data trusts as an AI governance mechanism. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3334527. Accessed 15 Nov 2023
Romback R et al. (2022) CreativeML Open RAIL-M. Hugging Face. https://huggingface.co/spaces/CompVis/stable-diffusion-license. Accessed 15 Nov 2023
Ruhaak A et al. (2021) A practical framework for applying Ostrom's principles to data commons governance. Mozilla Foundation. https://www.mozillafoundation.org/en/blog/a-practical-framework-for-applying-ostroms-principles-to-data-commons-governance/. Accessed 15 Aug 2025
Saint-Aubin T (2012) Design your privacy: pour une licence de partage des données personnelles. InternetActu. http://www.internetactu.net/2012/06/22/design-your-privacy-pour-une-licence-de-partage-des-donnees-personnelles/. Accessed 15 November 2023

Swiss Federal Institute of Intellectual Property (IGE) (2021) Accès aux données non personnelles dans le secteur privé. https://www.ige.ch/fr/propriete-intellectuelle/pi-et-societe/acces-aux-donnees-non-personnelles-dans-le-secteur-prive. Accessed 15 Aug 2025

Swiss Federal Institute of Intellectual Property (IGE) (2024) Accès aux données non personnelles dans le secteur privé (Report, 1 March 2021)

Tarkowski A, Warso Z (2023) AI_COMMONS: filling the governance vacuum related to the use of information commons for AI training. Open Future.https://openfuture.eu/wp-content/uploads/2023/01/ai-commons-report.pdf. Accessed 15 Aug 2025

Tarkowski A, Zygmuntowski J (2022) Data commons primer: democratizing the information society. Open Future. https://openfuture.eu/wp-content/uploads/2022/07/220723data-commons-primer.pdf. Accessed 15 Aug 2025

Theben A et al (2021) Challenges and limits of an open source approach to artificial intelligence, study for the Special Committee on Artificial Intelligence in a Digital Age. European Parliament PE 662.908. https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)662908. Accessed 15 Aug 2025

Vezina B, Pierson S (2021) Artificial Intelligence and copyright in the cultural heritage sector: views from Creative Commons'. Europeana pro. https://pro.europeana.eu/post/artificial-intelligence-and-copyright-in-the-cultural-heritage-sector-views-from-creative-commons. Accessed 15 Nov 2023

Vollmer T (2015) CC licence version 4.0: helping meet the needs of open data publishers and users. Open Knowledge Foundation. https://blog.okfn.org/2012/08/15/cc-license-version-4-0-helping-meet-the-needs-of-open-data-publishers-and-users/. Accessed 15 Nov 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.