



HAL
open science

No title

Yaniv Benhamou, Melanie Dulong de Rosnay

► **To cite this version:**

| Yaniv Benhamou, Melanie Dulong de Rosnay. No title. 2023. halshs-04343942

HAL Id: halshs-04343942

<https://shs.hal.science/halshs-04343942>

Preprint submitted on 14 Dec 2023

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

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

**OPEN DATA COMMONS LICENCES (ODCL): LICENSING PERSONAL
AND NON PERSONAL DATA SUPPORTING THE COMMONS AND PRIVACY**

Yaniv Benhamou and Melanie Dulong de Rosnay*

I.	Introduction.....	3
II.	Open licenses for all kinds of data: towards Open Data Commons Licenses	5
1.	Copyrighted data	5
1.1	Extension of open licenses to all copyrighted data, including AI training data	5
1.2	Contractual chain and the right to make derivatives.....	6
1.3	Liability of the users and of the authors.....	7
1.4	Limits of open licenses related to moral rights	8
2.	Personal data	10
2.1	Extension of open licenses to personal data	10
2.2	Limits related to consent validity.....	11
2.3	Limits related to the right to self-determination and the right to erasure	12
2.4	The case of mixed datasets	13
3.	Technical data.....	14
3.1	Extension of open licenses to technical data.....	14
3.1	Technical data subjected to an exclusive right	14
3.2	Technical data subjected to no exclusive protection.....	15
4.	Restricting openness to protect the commons and privacy	16
4.1	In general	16
4.2	Restrictions relating to privacy (Privacy Pledge)	17
4.3	Restrictions related to authorized users and uses to protect the commons	17
4.4	Compability among open licences	18
III.	Collective management through <i>Data Trusts</i>	19
1.	Introduction and attempt to define	19
2.	Success factors of a data trust.....	20
3.	Legal structures inspired by copyright collective management	21
4.	Towards legal and political compatibility definition	23
IV.	Conclusion	23
V.	Annex: Template for Open Data Commons Licenses (ODCL).....	24
1.	Table of elements	24
2.	Clauses to be combined.....	25

Abstract: Data are often subject to a multitude of rights (e.g. original works or personal data posted on social media, or collected through captcha, subject to copyright, database and data protection) and voluntarily shared through non standardized, non interoperable contractual terms. This leads to fragmented legal regimes and has become an even major challenge in the AI-era, for example when online platforms set their own Terms of Services (ToS), in business-to-consumer relationship (B2C).

This article proposes standard terms that may apply to all kind of data (including personal and mixed datasets subject to different legal regimes) based on the open data philosophy initially developed for Free and Open Source software and Creative Commons licenses for artistic and other copyrighted works. In a first part, we analyse how to extend open standard terms to all kinds of data (II). In a second part, we suggest to combine these open standard terms with collective governance instruments, in particular data trust, inspired by commons-based projects and by the centennial collective management of copyright (III). In a last part, after few concluding remarks (IV), we propose a template “Open Data Commons Licences“ (ODCL) combining compulsory and optional elements to be selected by licensors, illustrated by pictograms and icons inspired by the bricks of Creative Commons licences and legal design techniques (V).

This proposal addresses the bargaining power imbalance and information asymmetry (by offering the licensor the ability to decide the terms), and conceptualises contract law differently. It reverses the current logic of contract: instead of letting companies (licensees) impose their own ToS to the users (licensors, being the copyright owner, data subject, data producer), licensors will reclaim the ability to set their own terms for access and use of data, by selecting standard terms. This should also allow the management of complex datasets, increase data sharing, and improve trust and control over the data. Like previous open licencing standards, the model is expected to lower the transaction costs by reducing the need to develop and read new complicated contractual terms. It can also spread the virality of open data to all data in an AI-era, if any input data under such terms used for AI training purposes propagates its conditions to all aggregated and output data. In other words, any data distributed under our ODCL template will turn all outcome into more or less open data and foster a data common ecosystem. Finally, instead of full openness, our model allows for restrictions outside of certain boundaries (e.g. authorized users and uses), in order to protect the commons and certain values. The model would require to be governed and monitored by a collective data trust.

Keywords: open licenses, open data, data commons, data trusts

I. Introduction

Data access and reuse is key in a data-driven society, which explains why it is the subject of numerous legislative developments in Europe and abroad.¹ However, because data can come from multiple actors and a variety of sources, this legal framework is fragmented. Hence, data are subject to sometimes conflicting legal regimes (e.g. copyright, trade secrets, data protection). When the access is voluntary², actors have been turning to contractual mechanisms to regulate in a flexible way the different types of data and the interactions between participants in the ecosystem.³ However, platforms and service providers tend to impose their Terms of Services (ToS), for instance by permitting or prohibiting data scraping to train AI models. Instead of letting platforms and service providers impose their ToS to control data, this article proposes to give control back to rightholder users, by using open standard terms and collective governance schemes applicable to all data, including personal and non personal data.

Open licenses are standardized licenses, applicable broadly, including to public administration (Open Government Data), research (Open Research Data), software (Free and Open Source), data published on Wikipedia (Wikidata), collaborative creation (Creative Commons), cultural heritage, smart cities,⁴ the private sector.⁵ Open licenses make the resource they cover available to all or under certain conditions⁶. While most open data projects provide a Share Alike or a No Commercial clause, the ethos of the commons has yet to be fully performed and refined for data. The Share Alike clause is not sufficient to build data commons, as it applies to derivatives only, and the No Commercial clause is not fine-grained enough, as it could exclude social economy and solidarity companies. Our article offers a set of options meant to be sufficiently sophisticated and adaptive to cover the needs of personal and non-personal data, and to respect privacy and commons values, influenced by recent licensing schemes

* **Yaniv Benhamou** is Professor at the University of Geneva and attorney-at-law (IP & Data Protection). **Melanie Dulong de Rosnay** is Research Professor at Center for Internet and Society of CNRS (French National Center for Scientific Research). Authors wish to thank Mr. Michel José Reymond for his comments and Mr. Ilya Skurikhim for his help in finalizing the manuscript.

¹ In European law, one can think of sectoral or horizontal regulations, such as the General Data Protection Regulation (GDPR) for personal data, the Regulation on the protection of non-personal data for non-personal data, the Open Data Directive for public data, the Data Governance Act for governance as well as legislative proposals, such as the "Data Act" for data law. In US law, one can think of the Executive Order of 20 October 2023 in the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. In Swiss law, there are several initiatives that aim to promote access to personal and non-personal data, cf. Swiss Federal Institute of Intellectual Property (IGE), *Accès aux données non personnelles dans le secteur privé*, Report of March 1, 2021, 4 ff.

² Voluntary access can take the form of an individual contract between two or more actors (e.g. data transfer or exchange contract, subscription contract) or a standardized contract in the form of general terms and conditions or private policy for Open Data or Shared Data. Cf. IGE, 7.

³ Cf. YANIV BENHAMOU/LAURENT TRAN, *Circulation des biens numériques: de la commercialisation à la portabilité*, in: *sic!* 2016, 572 ff; JACQUES DE WERRA, *Perspective "Inside-Out". Défis du droit d'auteur dans un monde connecté*, in: *sic!* 2014, 194 ff, 206.

⁴ These standardized licenses are offered by various organizations, e.g. the General Public License (GPL) by the Free Software Foundation, <www.gnu.org/licenses/gpl-3.0.en.html> (29.06.2021) or Creative Commons license by <<https://creativecommons.org/>> (29.06.2021).

⁵ The sharing of open data by private companies is gradually emerging, in particular for economic reasons (the economic value of open data being estimated at USD 5,000 billion in 2013) but remains limited, especially for the purpose of protecting intellectual property rights and personal data. Data is then usually shared through publications (e.g. annual report), mentioned on a website and/or an application programming interface (API) containing only specific types of data (e.g. master data and metadata). Cf. IGE, 26.

⁶ In many jurisdictions, open data mostly come from public data, i.e. data that can be used, edited, analyzed and transmitted freely, without any particular legal, financial or technical restrictions. This free access to public data may be questionable, since it may lead to liability risks (cf. II.1.3) and since it is often reused by economic actors (e.g. Apple, which uses the open data of the Federal Office of Topography, Swisstopo, for the "Map" application on the iPhone and iPad, cf. ANOUCH SEYDTAGHIA, *Apple utilise les cartes suisses...mais sans les payer*, 2021, <<https://www.letemps.ch/economie/apple-utilise-cartes-suisse-payer>> (29.06.2021), published in *le Temps*). In Switzerland, the principle of free access of public data was recently reiterated by the Federal Council, on the basis in particular that it is not possible to define a pricing structure, cf. Swiss Federal Council Report in implementation of Postulate 19.3342 Badran of March 22, 2019, 2021. This argument may be surprising given that there is a market for data and efforts to price data. For example, in the case of urban data from the city of Lyon, cf. <https://download.data.grandlyon.com/files/grandlyon/Annexe%203_redevances_LicenceAssociee_Ch.pdf> (29.06.2021).

implementing reciprocity and peer production values and with the aim to extend this trend to all data in order to build data common ecosystems.

Traditionally, open licences apply to data of different nature, but exclude personal data⁷. However, traces of personal data can be found almost everywhere, whether they have been left voluntarily or incidentally, whether it is mobility data (eg. 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. Therefore, it seems useful to propose broader open licensing models applicable to both personal and non-personal data, which seems particularly relevant for complex technologies, such as Artificial Intelligence (AI). First because of the multitude of actors involved and the contractual relationships, which lead to a risk of interdependence between the people and institutions involved⁸. Also to facilitate interoperability of datasets licensed under different conditions and to set contractual fair terms. Finally, to access to training data for AI and open software, allowing others to verify, enrich and improve the quality of products and services⁹.

Collective governance models consist in producing, pooling and sharing resources through a structure, a community, which can be commons-based (e.g. Wikipedia and its community of contributors and users). They can be supported by Data Trust, an emerging concept which supports complex data management, circulation, trust and the control or sovereignty of individuals and collectives over their data. Collective governance aims at enforcing clauses values and virality to the output, and support a interoperable ecosystem.

This paper explores how to apply open licenses to both personal and non-personal data (II) with the support of a collective data governance (III). After few concluding remarks (IV), the annex of the paper suggests a template for Open Data Commons Licences (ODCL) embedding commons-based values and future interoperability with other licensing systems (V). The model is meant to apply to the broadest scope of data as inputs, including personal data and technical or mixed data, going beyond copyright and sui generis database scope of existing open licenses.

Our proposal intends to address the bargaining power imbalance¹⁰ and information asymmetry (by offering the licensor the ability to decide the terms) and the management of complex datasets (by lowering transaction costs with simple standard terms). It intends to facilitate greater voluntary and responsible data sharing, and improve trust and control over the data¹¹. It should spread the virality of open data to all data in an AI-era, as any input data under such terms used for AI training purposes will propagate its conditions to all aggregated and output data). Our proposal finally streamlines the circulation and reuse of data under certain conditions, take into account the collective dimension of data, and better guarantee the control of creators and subjects over their data.¹²

⁷ Cf. IGE 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, nor privacy rights, nor any other personality or similar rights (art. 2 lit. b).

⁸ One just has to think about connected objects, autonomous cars or smart city projects that bring a multitude of companies and individuals to collaborate and make data available.

⁹ About the benefits of *Creative Commons* for AI, cf. BRIGITTE VEZINA/SARAH PIERSON, Artificial Intelligence and copyright in the cultural heritage sector: views from Creative Commons, 2021, <<https://pro.europeana.eu/post/artificial-intelligence-and-copyright-in-the-cultural-heritage-sector-views-from-creative-commons>> (29.06.2021). About the benefits of open software for AI, cf. ALEXANDRA THEBEN/LAURA GUNDERSON/LAURA LÓPEZ FORÉS/GIANLUCA MISURACA/FRANCISCO LUPIÁÑEZ VILLANUEVA, Challenges and limits of an open source approach to Artificial Intelligence, study for the Special Committee on Artificial Intelligence in a Digital Age (AIDA), Luxembourg 2021, 16 ff.

¹⁰ Paul Keller, Alek Tarkowski, The Paradox of Open. <http://paradox.openfuture.eu/>

¹¹ Aligned with the GPAI IP Advisory Committee, Preliminary Report on Data and AI Model Licensing, November 2022, 12 ss.

¹² This collective dimension is often inspired by the theory of the commons, according to which tangible open access resources freely available to the community) are threatened by appropriation (enclosure), while common-pool resources-governed by communities can lead to better production and maintenance of resources. For references, BENHAMOU/TRAN (n. 4), 575. Beyond the collective management of data, there is also a movement towards the collective exercise of individual rights related to data, e.g. the right of access which, when exercised collectively, allows a better understanding of the functioning of the provider's algorithm, cf. RENÉ

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. However, particular attention will be given to EU law and in particular to the EU General Data Protection Regulation (GDPR) due to its extra-territorial scope and its global influence¹³.

II. Open licenses for all kinds of data: towards Open Data Commons Licences

1. Copyrighted data

1.1 Extension of open licenses to all copyrighted data, including AI training data

Open licenses promote the sharing of knowledge, initially software with Free, Libre and Open Source licenses, then non-software copyrighted works with Creative Commons licenses. With these standard licenses, the creator (licensor) allows any interested user (licensee) to access the licensed resource, to enrich it¹⁴ and then make it freely available again under certain conditions.¹⁵ The licensed resources are copyrighted works. The license has a scope determined by the license chosen by the author.¹⁶ It relies on copyright to apply it backwards and convert exclusivity to inclusivity.¹⁷ It promotes the circulation of the work and public access and reuse under certain conditions, unlike the public domain without legal protection and a risk of potential reappropriation. It facilitates collaboration and enrichment, aligned with the open source and free culture movements¹⁸, through an *ex-ante* permission to make derivatives.

The concretization of the viral character, is obtained by obliging to include the same terms in further distributions of the resource.¹⁹ Usage rights are subordinated to the obligation of integrating the license to the resource and its outputs. This mechanism allows the contractual conditions to be extended to any future copy or derivative of the work in circulation.²⁰ The contractual terms attached to each copy of the

MAHIEU/JEF AUSLOOS, Harnessing the collective potential of GDPR access rights: towards an ecology of transparency, in: Internet Policy Review, 1 ff and the many cited references.

¹³ See Ahmed Baladi, Can GDPR Hinder AI Made in Europe? (10 July 2019), Cybersecurity law Report, 1. In case of disputes around open licences, see BENHAMOU, RSDA 422 and references. The EU data act will not be dealt with, as it relates to B2B fair contractual terms, while this article focusses on B2C mainly, in particular with standard contractual terms that may be proposed by individuals,

¹⁴ Except in the case of non free licenses reserving derivative rights such as Creative Commons No Derivative and Creative Commons No Derivative Non Commercial.

¹⁵ For open source software, a widely used free license is the MIT License, but others exist (e.g. General Public License (GNU/GPL), Apache License, BSD License, CeCILL). With respect to literary and artistic creations, a widely used license is the Creative Commons License, but others exist (e.g. Free Art License). Despite variations in the level of openness, they all have the objective of encouraging the circulation of the object with or without the possibility of enriching it. About Open Source licenses, cf. GIANNI FRÖHLICH-BLEULER, Open Source Compliance, in: Jusletter November 12, 2012. About *Creative commons*, cf. MICHAEL JACCARD/EVA CELLINA, in: SJ 2017 II, 229 ff; MÉLANIE DULONG DE ROSNAY, Creative Commons Licenses Legal Pitfalls: Incompatibilities and solutions, 2009, 7; SÉVERINE DUSSOLIER, Le jeu du copyleft entre contrat et propriété, in: Cahiers de droit de l'entreprise, 49 ff.

¹⁶ Copyrighted works are defined as “*literary and artistic intellectual creations with individual character, irrespective of their value or purpose*” (art. 2 al. 1 CopA) and also cover software (art. 2 al. 2 CopA), photographic depictions of three-dimensional objects even if they do not have individual character (art. 2 al. 3bis CopA) and collections, databases or other original arrangements (art. 4 CopA).

¹⁷ Cf. DUSSOLIER (n. 14), 50, in reference to the reversal of “copyright” to “copyleft”; “all rights reserved” to “some rights reserved”.

¹⁸ EU Study, p. 16.

¹⁹ 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*”).

²⁰ E.g. The Creative Commons License requires that the terms of the license are included in the digital file of the work (cf. art. 4), the GPL requires that any copy of the code is subjected to the same license (section 5 “*the entire work under this License to anyone who comes into possession of a copy*”). Cf. DUSSOLIER (n. 14), 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, in the work itself, which makes it enforceable against any user like a property right.

work are imposed to every user, they become in fact almost enforceable *erga omnes*, as opposed to contractual provisions that are normally enforceable *inter partes*.²¹

In the AI-era, some argue that copyright does not extend to AI training data (and so neither do the exceptions), as they are used for their informational content only²². However, we could argue the other way round: most jurisdictions seem to consider that the use of AI training data triggers the reproduction right²³, so that their contractual terms would propagate to all output data (without prejudice to the exceptions), like the contaminating effect of open licenses that subject proprietary data to open licence terms²⁴. Under such premise, contractual terms, such as those of the proposed ODCL template, would become viral, propagate open and fair principles to all output data derived from training data, and redresses the bargaining and information asymmetry. Finally, a way to ensure the application of such licence is to indicate the terms under which the data can be used via technical tools (e.g robots telling AI whether or not to use certain data), a growing practice since the reservation of right of art. 4 of DSM Directive²⁵.

1.2 Contractual chain and the right to make derivatives

With open licenses, there is a multitude of users who may create a contractual chain between the primary author and the subsequent downstream users²⁶. A new license agreement is concluded between the author of the primary work and each and any downstream user, who cannot grant sublicenses.²⁷

A distinction is made between a “simple“ user and the creator of a derivative work (“derivative user“), whose status is acquired by a creative contribution to the primary work (and not by simple adherence to the contract to access or a technical modification of the work)²⁸. A distinction is made between being the creator of a derivative work, when a pre-existing work is modified in such a way that the modification results in a new work, and the status of co-authors when a pre-existing work is created jointly by several authors with a common purpose.²⁹ In principle, when a primary author puts his work under an open license so that it can be modified and a contribution is inserted (as it is or in a modified form), it is a derivative work, and not a joint authorship, because of the lack of close cooperation between the authors and lack of will to create a joint work based on a precise objective³⁰.

²¹ DUSOLLIER (n. 14), 51.

²² See Lazarova Ana et al., 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; Margoni Thomas / Kretschmer Martin, A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. In : GRUR International, 71(8), 2022, 685–701.

²³ See BENHAMOU, Big Data and the Law, RSDA 2020, 405, and references; Cherpillod, sic! 2023.

²⁴ See BENHAMOU/TRAN, 579, giving, as example of contaminating effect, article 5 of GNU GPL 3.0: all programs based on GNU GPL licence 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*” (article 2 GNU GPLv2).

²⁵ [...].

²⁶ See Andrés Guadamuz. Viral contracts or unenforceable documents? Contractual validity of copyleft licenses. EIPR 26 (8), 331-339, 2004.

²⁷ E.g. Creative Commons License, preamble and art. 1 lit. b. Cf. JACCARD/CELLINA (n. 14), 236.

²⁸ Cf. Creative Commons License, art.1 lit. a et 2 lit. a ch. 4 (“*Simply making technical modifications [...] never produces Adapted Material*”). BÉNABOU/FARCHY, 37.

²⁹ Cf. DENIS BARRELET/WILLY EGLOFF, in: BARRELET/EGLOFF/HEINZMANN/KÜNZI/MEIER/RIEDO (éd.), Commentaire de la loi fédérale sur le droit d'auteur et les droits voisins, 4th éd., Berne 2021, LDA 7 N 4 ; YVAN CHERPILLOD, in: DE WERRA/GILLIÉRON (éd.), Commentaire Romand – Propriété intellectuelle, Bâle 2013 (cité: “CR PI-AUTEUR”), LDA 7 N 3, stating that “*l'œuvre commune [est] le résultat d'une concertation entre les divers coauteurs, ce qui peut être le cas même si leurs apports sont séparables*“ and that the Swiss Copyright Act (CopA) does not (no longer) distinguish between joint authorship and collaborative works depending on whether the contributions can be disentangled, with a few exceptions (cf. art. 30 CopA) and in contrast to the US for example. For an analysis under American law, cf. SHUN-LING CHEN, Collaborative authorship: from folklore to the wikiblog, in: Journal of Law, Technology & Policy 2011, 132 ff, 140.

³⁰ CR LDA-CHERPILLOD (n. 24), LDA 7 N 3.

Lastly, when a derivative work is subject to many levels of derivation, a distinction must be made according to whether it remains a derivative work, or acquires a new status of a new primary work, which will be assessed on a case-by-case basis according to the degree of transformation of the original work.³¹ Under Creative Commons and Open Knowledge Foundation licenses, the Share Alike copyleft viral clause applies to derivative works.

The proposed model applies a Share Alike viral clause (requesting to keep the same license) to derivatives and also non derivatives. The right to make derivatives permits to reuse and modify the resource (e.g. extract, transform, adapt, aggregate, translate, update for scraping, as well as other kind of processing)³². The option of controlling the access to data (not only the right of making of derivatives) can be a useful restriction and is included in the model³³. Options controlling the right to access, not only the right to make derivatives can be deployed for public use, or granted within a closed consortium only, allowing to retain power, to address privacy, or to avoid that open data feeds closed platforms or surveillance models³⁴ for instance.

1.3 Liability of the users and of the authors

Open licenses can raise liability of the users, if they do not respect the license terms (eg. Omit the mention of the author when it is required) or liability of the author, if the resource causes harm to users (eg. Dysfunctional resource causing damage or resource infringing third party's copyright or personality rights).³⁵ Authors, users and third parties may then attempt to claim liability based on contractual or tort liability, regardless of eventual contractual disclaimers in the license.³⁶

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 accepts when they use the resource in a way that complies with the license terms.³⁷ There must be a wrongful conduct

³¹ BENABOU/FARCHY (n. 20), p. 37.

³² See the Montreal Data Licence towards standardization of data licences: <https://arxiv.org/pdf/1903.12262.pdf>.

³³ 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.

³⁴ Alek Tarkowski, Zuzanna Warso, AI_COMMONS. Filling the governance vacuum related to the use of information commons for AI training. Open Future report, 2023. <https://openfuture.eu/wp-content/uploads/2023/01/ai-commons-report.pdf>. Huang S and Siddarth D, 'Generative AI and the Digital Commons' [2023]. <https://arxiv.org/abs/2303.11074>

³⁵ As a concrete example, we recall the loss caused by the DAO (decentralized autonomous organization) platform to its users for an amount of approx. USD 50 million due to a vulnerability in its open source code (cf. <https://www.usine-digitale.fr/article/dao-perd-50-millions-de-dollars-lors-d-un-piratage.N397787>) (29.06.2021). As another example, one can think of an open source guidance software for autonomous cars that would cause the vehicle to lose control because of a design error. See also Mélanie Dulong de Rosnay, 2013, "Open Content Licenses Without Representation: Can You Give Away More Rights Than You Have?", *European Journal of Law and Technology* (EJLT), 4(3). ISSN: 2042-115X.

³⁶ With regard to the liability of owners towards users of Open Source software, cf. FRÖHLICH-BLEULER (n. 14), 5 ; with regard to the liability of users towards owners of works licensed under Creative commons, cf. JACCARD/CELLINA (n. 14), 236.

³⁷ EVA CELLINA, La commercialisation des données personnelles – Aspects de droit contractuel et de protection des données, Thèse Genève, Genève, Zurich, Bâle 2020, 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/TRAN (n. 4). We would also like to add 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 recognized by many courts, especially in France, Germany and the USA. For references, cf. FRÖHLICH-BLEULER (n. 14), 4. For a recent example, TGI de Paris, 3ème ch. 3ème section of June 21, 2019 recognizing the GNU-GPL license as a (membership) contract. In Switzerland, one should mention the AppGer Basel-City, 20 May 2016, ZK.2015.9, in: sic! 2016, 594 ("Panoramabild") about a violation of the terms of Wikimedia's free license for use of a Wikimedia photograph without permission or mention of the author, although to our knowledge there has been no ruling on the nature and validity of free licenses under Swiss law. In the US, US Court of Appeals, Federal Circuit, Jacobsen v. Katzer, 535 F.3d 1372.

or a 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*”.³⁸ Open licenses generally provide for an exclusion or limitation of liability clause³⁹, which is not valid in all jurisdictions, in particular in civil law jurisdictions for cases of fraud or gross negligence.⁴⁰ 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 resource 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.⁴¹

Our model proposes a liability clause corresponding to high European standards of consumer law since malpractice and infringement threatens the construction of safely reusable commons. This choice is at odds with free software and open licensing schemes, where the resource 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 data can be reused safely. This legal duty of care clause contributes to a virtuous 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 because of A the European Union Public License (EURL)⁴² which liability clause is included in the template, a “Chain of Authorship” could be developed to guarantee that both the licensor and licensee are in a position to grant the license and to exercise the rights (over 18 years old, or with parental approval, ownership or clearing of all the rights to the best of their knowledge the material does not contain an infringement to others’ rights). Similar to a “Developer Certificate of Origin”⁴³, a certificate of origin could also be issued with the material, and be kept intact with the license when redistributing the material or a derivative.

1.4 Limits of open licenses 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.⁴⁴

³⁸ With free software, the documentation sometimes contains an implicit guarantee that the software *fit-for-purpose*. In the field of cryptocurrencies, the documentation even states that the codes have been audited by an independent firm.

³⁹ 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.

⁴⁰ This explains the Creative commons license restriction (art. 5 lit. 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. DULONG DE ROSNAY (n. 14), 101 ff, recalling other legislation preventing such exclusions, such as consumer protection and contract law.

⁴¹ 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 Federal Act on Product Liability (PLA). For an analysis of liability under the PLA and the difficulties of applying it to artificial intelligence systems, cf. YANIV BENHAMOU/JUSTINE FERLAND, *Artificial Intelligence & Damages: Assessing Liability and Calculating the Damages*, in: D’AGOSTINO/GAON/PIOVESAN (éd.), *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*, Montréal 2021, 165 ff.

⁴² <https://joinup.ec.europa.eu/collection/eupl/eupl-text-eupl-12>.

⁴³ Linux Foundation Developer Certificate of Origin version 1.1 <https://developercertificate.org/>.

⁴⁴ JACCARD/CELLINA (n. 14), p. 236 ; JANE C. GINSBURG, *Author’s transfer and license contracts under U.S. copyright law*, in: DE WERRA (éd.), *Research handbook on intellectual property licensing*, Cheltenham, Northampton 2013, 3 ff.

This sort of waiver of the author's right to exercise his 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.⁴⁵

These difficulties led to the Creative Commons licenses 4.0 version to better suit some national legislations⁴⁶: 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”, minimize the effect of moral rights on the licensees' ability to use the work, and ensure that the license operates as intended internationally. Author also retain the right to be recognized as the copyright owner, thus ensuring that the attribution right is upheld.⁴⁷

However, the transfer of moral rights and of all rights in advance to all future works are considered to be a personality infringement.⁴⁸ On the contrary, authors should be able to reconsider their commitments at any time.

This implies offering an *opt-out* right to authors in order to comply with their moral and personality rights, in particular so that they can object to an alteration contrary to their personality, if the use of the work is denaturing.⁴⁹ The exercise of this prerogative to revoke proves however difficult to put into practice. The return or withdrawal of the publicly available or commercialized copies seems unrealistic, if not impossible, in the digital context⁵⁰. As a solution, the exercise of the right of revocation is only valid for the future (*ex nunc*) and does not oblige users to return or withdraw the copies that circulated.⁵¹

Our model supports the assertion of authorship and producer attribution. It also 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 challenging⁵². Alternatively, as detailed in the following section, waiving these can fulfill requirements related to the withdrawal or erasure of personal data.

⁴⁵ 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*” (Civ. 1^{ère} 28 mai 1991 Grands Arrêts de la Propriété Intellectuelle D. 2004, comm. 11), cf. BÉNABOU/FARCHY (n. 20), 26.

⁴⁶ <https://creativecommons.org/faq/#how-are-publicity-privacy-and-personality-rights-affected-when-i-apply-a-cc-license>. Prior to that, there were national versions in order to comply with moral rights. E.g. French CC 3.0 version stated that licensee had to respect licensor moral rights to the extent licensor could not renounce to them (see 4.d clause). 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 versions.

⁴⁷ Cf. License CC BY 4.0, art. 2 lit. b ch. 1.

⁴⁸ On the prohibition of excessive commitments, cf. SYLVAIN MARCHAND, in: PICHONNAZ/FOËX (éd.), Commentaire Romand –Code civil, Bâle 2010 (cité: “CR CC I-AUTEUR”), CC 27 N 14, and the cited references, including the criteria for determining whether the commitment is excessive, including its intensity, duration, purpose, or indefiniteness ; BENHAMOU/TRAN (n. 4 citing as an example the fact that an author's prior contractual cession of all his copyright in all his future works is considered excessive. Concerning the non-transferability of moral rights, cf. CR PI-DE WERRA (n. 24), LDA 16 N 19, recalling that the existence or the extent of this non-transferability is however controversial, some considering moral rights as non-transferable, others proposing to distinguish the hard core of the moral right corresponding to the protection of the right of the personality (which is non-transferable) from the outer sphere of the moral right (which is transferable).

⁴⁹ Cf. BÉNABOU/FARCHY (n. 20), p. 27.

⁵⁰ Cf. BÉNABOU/FARCHY (n. 20), 28, indicating that these modalities (transaction costs and compensation) are likely to neutralize 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.

⁵¹ Another solution is to consider that the right to revoke one's commitment contradicts the authorization 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.

⁵² In line with Open Definition 2.2.1. <https://opendefinition.org/od/2.1/en/>

2. Personal data

2.1 Extension of open licenses to personal data

Personal data are collected and processed or used everywhere in a data-driven society (eg. to feed AI systems, to use social media, or with the use of captcha mentioned above)⁵³. They are often transferred directly or by third parties (e.g. Data Brokers) through personal data licenses, either by consenting to the temporary use of the right, or by waiving the right to claim protection during the contractual period.⁵⁴ The question then arises whether it is possible to apply open licensing mechanisms to personal data, thus allowing a better circulation on one hand, and a better control of individuals on the other hand, if data subjects may determine the conditions of access and use of their data. Standardized licenses already exist in the area of copyright and of non-personal data with model contracts for the exchange of technical data between companies.⁵⁵ However, if there are personal data licenses initiatives, we are not aware of any finalised standardised licenses applicable to all types of personal data⁵⁶, thus, we explore this avenue, drawing inspiration from existing open copyright licenses and research towards personal data open licences⁵⁷, data commons or collective data governance⁵⁸, and personal data commons⁵⁹.

⁵³ Personal data (data) is all information relating to an identified or identifiable person in the sense of art.3 lit. a de la Federal Act of 19 June 1992 on Data Protection (FADP ; RS 253.1). Cf. also art.5 lit. a of the Federal Act of 25 September 2020 on Data Protection (nFADP).

⁵⁴ JULIEN ROUVINEZ, La licence des droits de la personnalité, Thèse Lausanne, Lausanne 2011, N 170. For an in-depth analysis of contracts for the commercialization of personal data, cf. CELLINA (n. 29), 158-160 For an in-depth analysis of contracts for the commercialization of personal data, see CELLINA (n. 29), 158-160, which distinguishes between a proper license when the data controller collects the data himself and an improper license when the data controller processes the data in the absence of a proper license or in cases where the data comes from an indirect collection.

⁵⁵ P.ex. IGE, 6, and templates for transfer, subscription or exchange contracts, available at <<https://www.ige.ch/fr/propriete-intellectuelle/pi-et-societe/traitement-et-securite-des-donnees>> (29.06.2021). Cf. also the Microsoft model contracts, MICROSOFT, Removing Barriers to Data Innovation, <<https://news.microsoft.com/datainnovation>> (29.06.2021), and the “Montreal Data License“, cf. BENJAMIN MISHA/PAUL GAGNON/NEGAR ROSTAMZADEH/CHRIS PA/YOSHUA BENGIO/ALEX SHEE, Towards Standardization of Data Licenses: The Montreal Data License, 2019.

⁵⁶ With the exception of the 2012 Privacyleft proposal. Thomas Saint-Aubin, Design your privacy, pour une licence de partage des données personnelles, InternetActu, 2012, <http://www.internetactu.net/2012/06/22/design-your-privacy-pour-une-licence-de-partage-des-donnees-personnelles/>. However, the licenses drafts are not available for comment anymore.

⁵⁷ However, mention should be made of licensing standardization efforts in health data between patients and data controller (in a cooperative) <<https://www.saluscoop.org/licencia>> (29.06.2021) and between providers (licensees and sub-licensees) <<https://sphn.ch/services/dtua/>> (29.06.2021), as well as in urban data, (cf. “Données métropolitaines pour le Grand Lyon <<https://download.data.grandlyon.com/files/grandlyon/LicenceEngagee.pdf>> (29.06.2021), and consumer data (cf. <<https://customercommons.org/>>) (29.06.2021). Mention should also be made of standardization initiatives with regard to information to be provided to individuals (“privacy-icons”) such as <<https://privacy-icons.ch/fr/contact/>> (29.06.2021) in Switzerland and <https://cyber.harvard.edu/i4bi/Privacy_Icons> (29.06.2021) abroad in order to remedy the lack of information of the data subjects, specifying that such initiatives derive directly from the law (cf. art. 12.1 GDPR, which requires information to be provided in a concise, transparent, intelligible and easily accessible manner, in clear and simple terms, and art. 12.7 GDPR indicating that information may be provided in a standardized manner in the form of icons). More generally, cf. ZOHAR EFRONI/JAKOB METZGER/LENA MISCHAU/MARIE SCHIRMBECK, Privacy Icons: A Risk-Based Approach to Visualisation of Data Processing, in: European Data Protection Law Review 2019, 352 ff; PAUL JURCYS/CHRIS DONEWALD/JURE GLOBOCNIK/MARKUS LAMPINEN, My data, my terms: a proposal for personal data use licenses, in: Harvard Journal of Law & Technology 2020, 13.

⁵⁸ Zygmontowski, J.O. Data Governance in a Trilemma: A Qualitative Analysis of Rights, Values, and Goals in Building Data Commons. *DISO* 2, 30 (2023). <https://doi.org/10.1007/s44206-023-00058-y>; Kariotis, T., Ball, M. P., Greshake Tzovaras, B., Dennis, S., Sahama, T., Johnston, C., ... Borda, A. (2020). Emerging health data platforms: From individual control to collective data governance. *Data & Policy*, 2, e13. <http://doi.org/10.1017/dap.2020.14>

⁵⁹ See Wong, J., Henderson, T., & Ball, K. (2020). Data protection for the common good: developing a framework for a data protection-focused data commons. In *Data for Policy Conference 2020* Data for Policy. <https://doi.org/10.5281/zenodo.3965670>. More practically, see all the development of a data commons infrastructure by the DECODE project (<https://tools.decodeproject.eu/>) including the Digital Data Commons Privacy Pledge of the DECODE EU project in Bassi, E., Ciurcina, M., De Martin, J. C., & Fenoglio, S. (2019). Licensing of digital commons including personal data – update. DECODE European project deliverable 1.9. <https://decodeproject.eu/publications/licensing-digital-commons-including-personal-data-update> and Andrea D'Intino, Denis Roio, Smart Contracts for Data Commons integrated with GDPR compliant legal rules and tested in pilots, DECODE European project

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

We claim that open licenses designed for copyright could apply *mutatis mutandis* to personal data: in both cases, the resource of the license is an absolute subjective right (copyright and personal data protection).⁶⁰ The terms are attached to the resource which thus allows the resource to be tracked and contractual rights to be extended to any user⁶¹, creating a contractual relationship between the data subject and any downstream user (who would become a data controller). 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.⁶² 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, all the more with reidentification and sensitive health or research data. So, the analogy doesn't 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 personal data open license. This could be operated by combining this licence with technical tools (e.g. robots.txt on a web site)⁶³ and a collective governance structure such as cooperative-like data trusts (III)⁶⁴.

2.2 Limits related to consent validity

There remains the question of consent validity, given that the open licences relate to the use of the resource by an undefined number of users and uses (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)⁶⁵ and there must be a certain control over the data. Consequently, it cannot be excluded that courts invalid open license applied to personal data, based on a strict application of GDPR consent requirements.

As a counter-argument, 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 disposal of data means that data also can be made freely

deliverable D4.12, 2019. See also as early as 2014 on personal data commons by Valérie Peugeot, *Données personnelles : sortir des injonctions contradictoires*, 2014, Vecam blog, <https://vecam.org/2002-2014/article1289.html> and the reframing of personal data from individual rights to collective negotiation by Antonio Casilli, *Quatre thèses sur la surveillance numérique de masse et la négociation de la vie privée*, in Jacky Richard & Laurent Cytermann (dir.), *Etude annuelle du Conseil d'Etat “Le numérique et les droits fondamentaux”*, Paris, La Documentation française, 2014, p. 423-434, as well as several blog posts by Lionel Maurel (such as <https://scinfolex.com/2014/09/01/une-gouvernance-en-communs-des-donnees-personnelles-est-elle-possible/> and <https://scinfolex.com/2016/01/15/eriger-le-reseau-des-donnees-personnelles-en-bien-commun/>)

⁶⁰ BENHAMOU/TRAN (n. 4), 579.

⁶¹ Cf. above I.1.1.

⁶² Cf. above I.1.2. Cf. also CELLINA/JACCARD (n. 14), 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*”).

⁶³ See II.1.1.

⁶⁴ Below III.

⁶⁵ PHILIPP FISCHER/NICOLAS TSCHUMY, *Commentaire Romand LPD*, art. 6 N 91; PHILIPPE MEIER, *Protection des données*, N 883.

available⁶⁶. Strict privacy laws, such as the GDPR, focus on the data subject as the origin point of the licence rather than the controller or the processor which should 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 work⁶⁷. Transposing to personal data, one could argue that open license relates to specific data, not to all data relating to a data subject.

2.3 Limits related to the right to self-determination and the right to erasure

There also remains the question of irrevocability of personality rights, given that the personal data subject has the right to withdraw consent at any time. More generally, the non-transferability of informational self-determination right covers continuous control rights, like the right to access, rectify, erase, integrity and withdraw consent at any time.⁶⁸

This irrevocability is known under copyright and has been addressed by Creative Commons licenses, which effect is in some jurisdictions limited by the non-transferability of moral rights and the prohibition of excessive commitments.⁶⁹ Just as copyright moral rights allows the author in some jurisdictions to object to any use that denatures the work, or to revoke their consent, in the context of personal data, we foresee that the data subject can withdraw consent and request the deletion of their data (*opt-out*). However, unlike copyright Creative Commons licenses, the personal data withdrawal right shall apply retroactively (*i.e.* to use and copies done in the past) (*ex tunc*), and not only for the future (*ex nunc*)⁷⁰.

Therefore, our proposal includes a right of withdrawal, revocation, rectification, to be forgotten (right to erasure), which must be complied on request of data subjects: the license grant has to be reversible in order to allow the right to erasure, moral rights and personality rights.

This requirement is impossible to enforce *ex tunc* (for the past, including the derivatives realised in the past) and practically difficult to enforce *ex nunc* (for the future), and contradicts the logic of irrevocable open licenses⁷¹, in which licensors waive moral rights⁷², and which do not terminate even if the licensor

⁶⁶ 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 GDPR 49 al. 1 let. a; LPD 17 al. 1 let. a et 30 al. 3; Message LPD, FF 2017 6688.

⁶⁷ CR PI-de Werra LDA 16 N 15 ; Oger ZH, sic! 2006, p. 851.

⁶⁸ CELLINA (n. 29), 257 ; BENHAMOU/TRAN (n. 4), 579, recalling that this irrevocability was relativized in a ruling of the Federal Court (ATF 136 III 401) which admitted the irrevocable transfer of the image rights of a model on her photos, on the grounds that these were "*biens de la personnalité qui ne font pas partie du noyau intangible de l'essence humaine (p. ex. le nom, la voix ou l'image vu l'importance prise ces dernières décennies de leur commercialisation)*". 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 para. 3 nFADP; Art. 12 para. 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.

⁶⁹ Cf. above II.1.4.

⁷⁰ Another solution is to consider that the right to revoke one's commitment contradicts the authorization given to third parties to use, modify and distribute the data, and that the exercise of the right of revocation should thus constitute a breach of the data licence agreement. As a solution, the exercise of this right of revocation could be conditional upon prior compensation of all those who would suffer harm as a result of the revocation, in a similar way to abusive use of moral rights under French law, except for data intimately linked to the person, which are non-transferable.

⁷¹ Creative Commons FAQs : <https://creativecommons.org/faq/#what-happens-if-the-author-decides-to-revoke-the-cc-license-to-material-i-am-using>: "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."

⁷² 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."

stops the distribution⁷³. However, solutions must exist for non-transferrable and irrevocable moral rights and self-determination right. For personal data, solutions must cover the past (*ex tunc*), as the right to be forgotten can be used retroactively. Practically, it is expected that if usages takes place within the boundaries of the conditions reflecting shared values, corresponding to the data subject self-determination, and monitored by the collective data trust, the right of withdrawal should be less exercised.

Our model includes a reservation of personality rights (following the reservation of moral rights in CC 3.0) and to consider an *opt-out* right if the data subject wants to withdraw consent as allowed by the GDPR. Therefore, a scope of anonymity is provided: full anonymity means strict confidentiality with fully anonymized data; limited anonymity means semi-private with pseudonymity, or aggregation for research or statistical purposes⁷⁴, clear data means ordinary clear data which can freely flow, not confidentiality.

The challenge is practical, since data may be shared with a multitude of actors⁷⁵. A technical solution is to incorporate a metadata to trace data uses. Indicate the terms under which the data can be scrapped via robots.txt is a growing practice in line with art. 4 of DSM Directive rights reservation. If such solution is neither available nor practical, it is worth remembering that the right to be erasure has limits. Where the controller has made the personal data public, the controller must only inform the 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 into the output and, from an input data standpoint, it seems impossible to make an AI model ‘forget’ the training data⁷⁶. As a policy solution to implement self determination, data trusts could enforce their members rights and requests to erasure, which justifies coupling open licenses with collective structures as we propose below⁷⁷.

2.4 The case of mixed datasets

In an AI and data-driven society, the multiplication of contractual relations corresponding to the number of uses and modifications creates a risk of incompatibility between users expectations and licences, which is even more complex when data subject to different legal regimes are inextricably linked (so-called mixed datasets). In the absence of a clear compatibility system, one question is to know which legal regime is applicable, or prevails another.

In the case of cumulative application of copyright and data protection to mixed datasets, one solution is to sort out the different resource and apply different contractual terms to each resource of the license separately. If the resources of the license are so intertwined as to make the sorting impossible, one solution is to apply the protection of personal data to the whole resource, by a phenomenon of

⁷³ 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.”

⁷⁴ For different techniques of anonymisation, see for instance Thiago Guimarães Moraes and others, Open data on the COVID-19 pandemic: anonymisation as a technical solution for transparency, privacy, and data protection, *International Data Privacy Law*, Volume 11, Issue 1, February 2021, Pages 32–47, <https://doi.org/10.1093/idpl/ipaa025>

⁷⁵ This practical challenge explains why, in medical research, medical institutions often continue to use data even after revocation of consent, based on the ease of *opting out*: in the event of revocation of consent, the data must be anonymized, unless it is clear from the beginning of the research project that anonymization is not possible, cf. art. 10 of the Ordinance of 20 September 2013 on Human Research with the Exception of Clinical Trials (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, 9.

⁷⁶ Tiffany Li, Algorithmic Shadow Harms: The Problem With Data Deletion; <https://fortune.com/2023/08/30/researchers-impossible-remove-private-user-data-delete-trained-ai-models/>, indicating that machines do not “forget” due to their design, once an AI has learned something, unlearning proves complicated.

⁷⁷ Cf. below III.

absorption of the protection and non-transferable nature of personal data.⁷⁸ In other words, the licensing conditions of the personal data protection would prevail over other regimes, given the fact that they are more sensitive and more closely related to the personality. In the case of a revocation request by the licensor (*opt-out*), one solution is to recognise the users' interest in continuing to use the data, in order to respect the user's obligation to share their data with an unspecified number of users, except in the case of data that is closely linked to the person, in which case the *opt-out* right must be guaranteed and absolute: in other words, a real independent right of the subsequent users to continue using the data, or at least in preventing the licensor from terminating the license agreement.⁷⁹ In order to take into account the fact that moral rights, personality rights, and self-determination cannot be waived in Europe, a solution has already been a clause to ask the licensor to not exercise them.⁸⁰ Another solution is to think about ways to implement moral rights and privacy. Implementation of such a mechanism is still not practical, but could in theory be achieved *ex tunc* through a combination of collective management and watermarking, or through personal data stores allowing decentralized data processing⁸¹.

Consequently, in the case of mixed datasets, it is unclear which legal regimes and/or contractual relations apply. A way to address this would be to rely on licenses that settle the compatibility among them.

3. Technical data

3.1 Extension of open licenses to technical data

Another important question is to analyse the application of open licenses to technical data (eg. machine-generated data, or commercial data) that are often contained in databases and used for marketing or training purposes⁸². Distinction must be made between such technical data subject to an exclusive right and those with no exclusive right.

3.1 Technical data subjected to an exclusive right

Technical data may be protected in many countries through competition law, while they may be protected in the EU by a *sui generis* database right⁸³. In practice, the database supplier could prohibit the downloading of data or datasets contained therein, even if the underlying data are not protected by an exclusive right. The EU *sui generis* database protection requires that substantial investments have

⁷⁸ This “default“ data protection solution is advocated by the European Regulator, see European Commission, Guidelines for the Regulation on a framework for the free flow of non-personal data in the European Union, 2019, § 2.2 ; YANIV BENHAMOU, Big Data and the Law: a holistic analysis based on a three-step approach – Mapping property-like rights, their exceptions and licensing practices, in: RSDA 2020, 393 ff, 414, comparing this situation with the “contaminating” effect of the GNU GPL license, which makes the whole software subject to the free license when a piece of code under the Free Software license is inserted into proprietary software.

⁷⁹ DE WERRA (n. 4), 194, citant BGH, GRUR 2009, 946, “Reifen Progressiv“ et TF 4A_598/2012 du 19.03.2013, c. 5.5, in: sic! 2013, 603 ff ; BENHAMOU/TRAN (n. 4), 590.

⁸⁰ This solution is already favored with Creative Commons licenses (art. 2 lit. b ch. 1) : “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“.

⁸¹ Heleen Janssen and others, Decentralized data processing: personal data stores and the GDPR, *International Data Privacy Law*, Volume 10, Issue 4, November 2020, Pages 356–384, <https://doi.org/10.1093/idpl/ipaa016>

⁸² In EU Law, such data are generally defined as non-personal data as opposed to personal data and includes such technical data (eg. aggregated datasets used for big data analytics or weather conditions generated by sensors installed on wind turbines or data on maintenance needs for industrial machines) as well as anonymized 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.

⁸³ 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

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”)⁸⁴.

In EU Law (or other jurisdictions providing an exclusive database right), licenses over technical data or datasets (“database licenses“) are common and allow data suppliers to shape an ecosystem around their data aligned with their values. For B2B, there are efforts to standardise licenses of non-personal data with model contracts for the exchange of technical data between companies.⁸⁵

For B2C, few standardised licenses can be mentioned too (eg. for urban and map data).⁸⁶ In the open ecosystem, both the Open Data Commons and the Creative Commons 4.0 licenses include the EU *sui generis* database right (the definition of “Work“ include databases), allowing the optional conditions (attribution, non-commercial use, no derivative work and share alike) to also apply to the *sui generis* database right and not only to the copyrightable elements⁸⁷.

3.2 Technical data subjected to no exclusive protection

In the absence of an exclusive right, licensing technical data can only be envisioned by contractual mechanisms. Distinction must be made between data provided by a data supplier 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 “B2B“ respectively “B2C“ licenses).

In B2B, 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 we have⁸⁸. If there is not exclusive right to the data, then how can the data supplier or user claims to own it, and if not, how can they license it? Regardless this legal uncertainty, some courts have recognized 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 exploitation⁸⁹.

In B2C, such licensing mechanisms may be also envisioned. Consequently, standardized licenses applicable to databases are beginning to emerge⁹⁰. This raises the delicate issue of validity of such contracts with respect to third parties that have no contractual relationship with the data supplier (eg. e-commerce websites or social media that provide access to data to an undefined number of users, often

⁸⁴ BENHAMOU, Big Data, 398.

⁸⁵ See IGE, 6, and templates for transfer, subscription or exchange contracts, available at <<https://www.ige.ch/fr/propriete-intellectuelle/pi-et-societe/traitement-et-securite-des-donnees>> (29.06.2021).

⁸⁶ 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 model contracts, MICROSOFT, Removing Barriers to Data Innovation, <<https://news.microsoft.com/datainnovation/>> (29.06.2021), and the “Montreal Data License“. Cf. BENJAMIN MISHA/PAUL GAGNON/NEGAR ROSTAMZADEH/CHRIS PA/YOSHUA BENGIO/ALEX SHEE, Towards Standardization of Data Licenses: The Montreal Data License, 2019.

⁸⁷ CC licenses prior to the 4.0 version excluded databases from the scope of the licence. 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 to decide under which licensing terms their database would be licensed. To fill this gap, the UK-based Open Knowledge Foundation issued the Open Data Licence that includes database into the licence followed by Creative Commons with its CC 4.0 in 2012 to correct a “market failure“ and “to align with expectations of users.“ <https://blog.okfn.org/2012/08/15/cc-license-version-4-0-helping-meet-the-needs-of-open-data-publishers-and-users/>.

⁸⁸ In civil law jurisdictions, this principle arises from the principle of “*nemo plus juris ad alium transferre potest quam ipse habet*“, see Gilliéron, 443.

⁸⁹ Gilliéron, 443, quoting BGH, NJW 2012, p. 3512, spec. p. 3514.

⁹⁰ 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 analyze and use for AI training“ and used notably by IBM (cf. <<https://developer.ibm.com/blogs/new-collaborative-data-license-agreement-makes-sharing-data-sets-easier/>>) (04.09.2021).

via general terms of use). This implies to analyze whether and how the conditions of the open license can be imposed on any third party user. In EU law, such licenses seem valid, as it seems possible to restrict the use of databases through contractual mechanisms⁹¹. The CJEU 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 scraping⁹².

However, abroad where there is no exclusive database protection, such licenses may be challenged, or even invalid. In US-law for instance, while many companies continue using different techniques to bind website's visitors (e.g. browsewrap⁹³, sign-in-wrap⁹⁴, clickwrap⁹⁵, or scroll wrap agreements), some techniques such as browsewrap agreements are often unenforceable, in particular against consumer due to insufficient notice. On the other hand, the browsewrap agreements is enforceable when the website contains an explicit textual notice that continued use will act as a manifestation of the user's intent to be bound, courts have been more amenable to enforcing. Consequently, website owners can still rely on the enforceability of browsewrap agreements, provided that the presentation of those agreements complies with the case-law, in particular on web design.

4. Restricting openness to protect the commons and privacy

4.1 In general

The terms proposed at the end of this paper are inspired by open licenses, such as Creative Commons licenses. They contain elements such as (i) definitions and (ii) scope of license (optional and compulsory elements). Among the optional elements, data may be stored, copied and shared under the same license, may or may not be enhanced, correlated to create derivative data, data must remain free or can be monetized, data must be fully or partially anonymized, and, crucially, terms may limit who may access the data and reserve to categories of downstream users or uses.⁹⁶ The terms also contain (iii) a liability clause: the licensor must ensure to have obtained all necessary rights to license the resource. They also contain (iv) a term and termination clause: if terms are not respected by the licensee, the licence grant is interrupted.

Privacy and commons-based purposes may be achieved through setting restrictions applicable to the nature of the user or to the kind of use, even if these proposed restrictions break the non-discrimination openness principle at the core of free software and open licenses⁹⁷. Instead, an Open Definition revision could include the respect of the rights contained in a Privacy Pledge (below 4.2)⁹⁸ and the openness could be achieved within a closed consortium or among certain users based on their status or the type of usage, as long as the purpose is commons-based (below 4.3)⁹⁹. These restrictions enforce certain values and translates a certain vision of data altruism. They would require to be governed and monitored by a collective data trust.

⁹¹ Cf. C-30/14, Ryanair vs PR Aviation, [2015] ECLI:EU: C:2015:10.

⁹² The enforceability of such a clause contained in the online terms of use is then a matter of national law.

⁹³ Link on the screen with notice that using the site indicates acceptance of terms.

⁹⁴ Notify the user of the existence of the contract and advise to click on the button to proceed to the next screen.

⁹⁵ Also known as click-accept, or clickthrough.

⁹⁶ For suggested terms to be included in a license for personal data, cf. JURCYS/DONEWALD/ GLOBOCNIK,/LAMPINEN (n. 45), 11 ff ; for non-personal data, cf. Technical Data Sharing Model Contracts of August 31, 2020 proposed by id is lawyers at IGE ; for urban data including personal and non-personal data, cf. License committed to "Données métropolitaines pour le Grand Lyon <<https://download.data.grandlyon.com/files/grandlyon/LicenceEngagee.pdf>> (29.06.2021) of which art. 9 deals with personal data and contains an anonymization obligation and a prohibition on cross-referencing data allowing the reconstruction of personal data.

⁹⁷ <https://opendefinition.org/>, Robin Romback and others, 'CreativeML Open RAIL-M' (22 August 2022), indicating that the Openrail licence already proposes use-based restrictions banning "social cost, harmful use, restricted scenarios".

⁹⁸ <https://privacy-pledge.com/>; <https://zenroom.org/privacy/>

⁹⁹ Eleonora Bassi, Marco Ciurcina, Juan Carlos De Martin, Selina Fenoglietto, Antonio Santangelo, D1.6 Licensing of digital commons including personal data – update, DECODE (DEcentralised Citizens Owned Data Ecosystem) European project, 2019.

4.2 Restrictions related to Privacy (Privacy Pledge)

The Privacy Pledge is 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 material in respect of personal data, such as high technical standards of cybersecurity and human rights. Encryption protocols and privacy-by-design benchmarks can serve as standards to guarantee a high level of human rights. In case of breach of the Privacy Pledge, the licence terminates. This is intended to set a viral effect for privacy and a compliant downstream behaviour. The breaching party may be further held liable for breaching the contract, which should be enforced by the data trust. To facilitate recognition, benchmarks could be used and also certified compatible. Minimum privacy criterias are expressed in the Privacy Pledge which both licensor and licensee promise to respect. It could be used as a compulsory standard, similar to other definitions in the FLOSS licensing world¹⁰⁰. These texts can be upgraded and subsequent versions go through certification process by licenses issuer organisms, and serve as benchmark to certify other licenses as compatible. Other 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.

4.3 Restrictions related to authorized users and uses to protect the commons

Restrictions to support commons-based purposes are based on the last generation of open licenses, rethinking communing and giving it more value than full openness (e.g. Peer Production licence, the Copyfair, Copyfarleft, Coopyright policy¹⁰¹). All conceptualises models that restrict openness in order to protect the commons. They develop a sort of Non Commercial No Derivatives license outside of certain boundaries¹⁰². The commercialization of the ressource remains free for those who contribute to the commons, while a fee must be paid by others. This is similar to traditional knowledge or other community governance benefit sharing mechanisms¹⁰³. It also echoes the Open Source AI principle of the Open RAIL-M licence¹⁰⁴ governing the use of Stable Diffusion which contains use-restrictions for ethical concerns¹⁰⁵.

In order to ensure interoperability with a plurality of political economy models between commercial and hard core commons, an option of our proposed licenses also accommodates compatibility with commercial reuse, for those who are fine with authorising some data monetisation. In order to respect EU privacy high standards and to support collective goals such as the contribution to the commons, the licenses can be adapted in a more granular way, for instance to accept data processing only for selected accredited public interest purposes (e.g. better transport or participatory pollution measurement in one's city) and under strict conditions (e.g. excluding marketing targeting or military use).

Such granularity is operated by optional terms related to authorized users and uses: rights could be reserved, or granted only to accredited users (e.g. non-profit education, open access research, climate change or public health research or any other high-level public interest or common purpose defined by the consortium, the data trust, or the local government development priority). Rights could be reserved, or granted only for accredited usages (e.g. Access to Knowledge, UN sustainability goals, ethical or responsible use, competition-friendly use, clinical data use, vulnerable persons-friendly use,

¹⁰⁰ The Open Definition : <https://opendefinition.org/od/2.1/en/>, the [Open Source Definition](#), the [Free Software Definition](#) and the [Definition of Free Cultural Works](#).

¹⁰¹ Peer Production License. http://wiki.p2pfoundation.net/Peer_Production_License; Copyfair license. http://wiki.p2pfoundation.net/CopyFair_License and <http://wiki.p2pfoundation.net/Copyfair>; Copyfarleft license. <http://wiki.p2pfoundation.net/Copyfarleft>; Coopyright. <https://coopdescommuns.org/fr/le-coopyright-pour-la-coop-des-communs/>.

¹⁰² E.g. For instance, the Peer Production licence, an adaptation version of the Creative Commons Non Commercial option: only other commoners, cooperatives or non-profit can share and reuse, and commercial benefits must remain in the collective.

¹⁰³ Singh, Parminder Jeet and Gurumurthy, Anita, Economic Governance of Data: Balancing individualist-property approaches with a community rights framework (January 1, 2021). <https://ssrn.com/abstract=3873141>

¹⁰⁴ <https://huggingface.co/spaces/CompVis/stable-diffusion-license>

¹⁰⁵ <https://huggingface.co/spaces/bigscience/license>

environmental-friendly use, no surveillance use). Clauses selecting compatible users and usages could be merged, but since they correspond to different ontological categories, they are kept separate for the sake of technical compatibility with existing Rights Expression Languages¹⁰⁶, 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 revenues sharing for the data trust to benefit and support the commons, the community, the infrastructure, fund data maintenance, or data care operation, through a fee collected and managed through collective management, like radio broadcasting since 1970's or file-sharing legalisation proposals in the 2000s¹⁰⁷. The question of offering an individual a remuneration in exchange of digital labour or the use of their personal data is highly controversial, because introducing data ownership could lead to an unbalanced power relation between users and platforms. This proposal is not meant to dispossess users from fundamental rights on 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 value-based.

These optional terms are meant to foster the construction of sustainable data commons which are not necessarily fully open access. These optional terms will require a certification, called Shared Values Network and Stamped Usage, and based on characteristics of the reuser or of the reuse. Such a process would be similar to a license compatibility assessment. Certification is a process which started in the free and open software licensing world, in order to manage licenses proliferation, and avoid both formal incompatibility between licenses and putting a legal research burden on users.

4.4 Compatibility among open licences

Licensing frameworks have been created separately without resolving the question of their compatibility. Attempts have been made to assert compatibility between different licenses, but the fact that they sometimes follow radically different logics and stem from fragmented or even conflicting legal traditions makes the exercise arduous.¹⁰⁸

Legal compatibility is needed when merging data are licensed under different licences. Among the options available to ensure interoperability, there is the insertion of a downstream compatibility clause declaring that the resource (and its derivatives) must be licensed under the same licence but also under licenses deemed compatible.¹⁰⁹ It may occur after an evaluation process which checked whether the letter and the spirit of the will be protected along the line by the other license recognised compatible.

This can be achieved through three solutions. First, unilateral compatibility only requires a declaration by the source. It has been performed by Creative Commons BY SA to GNU-GPL, by Open Government License to Creative Commons¹¹⁰.

¹⁰⁶ Coyle, K. (2004, February). Rights expression languages. In *Report, Library of Congress, February*. <https://www.loc.gov/standards/relreport.pdf>

¹⁰⁷ Aigrain, P. (2012). *Sharing: Culture and the economy in the Internet age*. Amsterdam: Amsterdam University Press.

¹⁰⁸ BÉNABOU/FARCHY (n. 20), 35.

¹⁰⁹ DULONG DE ROSNAY (n. 14), 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."

¹¹⁰ 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/>. 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".

Second, reciprocal or bidirectional compatibility requires evaluation and bilateral negotiations, sometimes versioning. It has been achieved between Free Art Licence 1.3 and CC BY-SA 4.0¹¹¹. Licensees may reuse works under both licenses and the product can be further licensed under any of these two licences.

Third, in the same vein the Creative Commons compatibility recognition process refers to respecting the Free Cultural Work definition as "minimum compatibility criteria", there is the solution of using a standard as a proxy to help assess compatibility among the proposed Data Commons Licences and other licenses. This reinforces the proposed need to include a revamped Open Definition and a Digital Data Commons Privacy Pledge, which would include openness, commons principles, and personal data protection standard as minimum criterias for open data commons, based on an Europeanist extra territorial effect, reflecting the imperative nature of data protection.

Our proposal invites to think about compatibility before releasing it for public use, in order to avoid silos. The ideal solution would be reciprocity, if not achievable, unilateral compatibility, and in the absence of declared compatible licence by the licensing framework issuer, the third, long-term option is a process led by data trusts (III).

III. Collective management through *Data Trusts*

1. Introduction and attempt to define

Open licensing can be combined with collective data governance to improve its effectiveness and enforcement, including licensor opt-out, necessary for personal data and moral rights. Data trusts are now widely discussed as a new model of collective data governance, allowing to strengthen the control of individuals, the protection of their data, and the balance between corporate and public power.¹¹²

Indeed, bottom-up data trusts, conceptualized as “independent intermediary between data subjects and data collectors”¹¹³ or “data rights intermediaries”¹¹⁴ have the potential to allow trustees to transfer their rights according to specific terms. However, to be effective, models need to be further developed and adapted to the different types of data and expectations.¹¹⁵

The *trust* is a legal means of managing the rights relating to an ressource for the benefit of another person. It can be defined as a unilateral act of a settlor, who transfers the ownership of assets (*trust fund*) to another person (*trustee*).¹¹⁶ The *data trust* can be defined as a “*legal structure providing independent data management services*”.¹¹⁷ Like the *trust*, the *data trust* begins with an “asset” (e.g. data) that a settlor (e.g. organisation or individual) transfers to a trust (e.g. organisation or individual charged with

¹¹¹ "The [Free Art license 1.3](https://creativecommons.org/licenses/by-sa/4.0/) was declared a “BY-SA-Compatible License” for version 4.0 on 21 October 2014." <https://creativecommons.org/share-your-work/licensing-considerations/compatible-licenses>

¹¹² Cf. ELEMENT AI/NESTA, Fiducies de Données. Un nouvel outil pour la gouvernance des données, Livre blanc, 2019, 9, 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, 1% of privacy policies are 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.

¹¹³ Sylvie Delacroix, Neil D Lawrence, Bottom-up data Trusts: disturbing the ‘one size fits all’ approach to data governance, *International Data Privacy Law*, Volume 9, Issue 4, November 2019, pp. 236–252.

¹¹⁴ Alexandra Giannopoulou, Jef Ausloos, Sylvie Delacroix, Heleen Janssen, Intermediating data rights exercises: the role of legal mandates, *International Data Privacy Law*, Volume 12, Issue 4, November 2022, pp. 316–331

¹¹⁵ Expectations include raising public awareness of the benefits of data trusts and taking into account varying objectives and levels of governance, reflecting their values and privacy preferences. Cf. ELEMENT AI/NESTA (n. 70), 15.

¹¹⁶ Cf. GUILLAUME GRISEL, *Le trust en Suisse*, Zurich 2020, 3.

¹¹⁷ Open Data Institute, Data trusts: lessons from three pilots, Report published on April 15, 2019 <<https://theodi.org/article/odi-data-trusts-report/>> (29.06.2021).

managing the data in the overriding interest of the beneficiaries) for the benefit of a beneficiary (e.g. person or group of persons), the purpose and terms of which are set forth in the trust instrument.¹¹⁸ According to the Open Data Institute, a data trust should have the following characteristics: a legal structure, a clear purpose for managing the data, defined rights and obligations over the managed data, a decision-making process, a description of how the benefits will be distributed, and sustainable funding.¹¹⁹ According to Element AI (an AI product provider) and Nesta (a research foundation), data trusts must also be inclusive and flexible, advocating for the settlor to also be a beneficiary (i.e. the settlor-beneficiary) and for ease of transfer (i.e. the right to transfer and share data on behalf of beneficiaries with other trusts or entities).¹²⁰ They can take different legal structures, depending on the context and the objectives pursued.

2. Success factors of a data trust

Several collaborative governance projects identified success factors, including expectations of the founders and external users of the trust, as well as criteria for the viability of a trust, including the legal structure and decision-making process.

In the area of *smart cities*, several data governance projects aim to improve the urban environment (e.g., reduce pollution and/or improve transportation services) by collecting “urban data“ data related to the environment and individual behavior. For example, Sidewalk Labs, a subsidiary of Google, won a bid from the City of Toronto in 2017 to transform the Quayside neighborhood into the beginning of a *smart city*.¹²¹ This project was met with criticism from the public, who feared the privatisation of public space, the ubiquity of sensors in the city, and the resulting exploitation of data. In response to this criticism, Sidewalk Labs explored various governance models, including data pooling in the form of a simple partnership (“*société simple*“) and cooperatives, and ultimately proposed a “Civic Data Trust“, an independent entity (in a public-private partnership). In the face of persistent criticism, the city ultimately opted for a government agency to authorize and control data collection and access to information without granting privileged access to Sidewalk Labs.¹²² This project confirms the need to build a governance model that accomodates the plurality of interests (those of authorities, businesses, citizens and the public) and builds public trust with a bottom-up approach and clear governance.¹²³

Another example, European civic participation platform Decidim, is based on a social contract and compliant with GDPR and right of erasure, to the extent data can be removed. This data sovereignty tool “provides decentralized, privacy-enhancing, rights preserving tools to give back data sovereignty to people and enable citizens' digital rights”¹²⁴.

The Open Data Institute studied data trust models in three different contexts (urban data, health data, online platforms)¹²⁵ and identified the following drivers of adherence: the importance of privacy, which can vary according to personal preferences and sectors involved ; the importance of social equalities,

¹¹⁸ ELEMENT AI/NESTA (n. 70), 17 ; GRISEL (n. 72), quoting art. 2 of the Convention concluded on 1 July 1985 on the law applicable to trusts and on their recognition (CLH Trust ; RS 0.221.371) (“*trust refers to the legal relationships created by a person, the settlor - by inter vivos deed or mortis causa - whereby property has been placed under the control of a trustee for the benefit of a beneficiary or for a specific purpose*“).

¹¹⁹ ODI (n. 73).

¹²⁰ Cf. ELEMENT AI/NESTA (n. 70), 22, defining the approach as a “bottom-up approach“ to involve constituents in the trust.

¹²¹ Cf. <<https://www.toronto.ca/city-government/planning-development/waterfront/initiatives/current-projects/quayside/>> (29.06.2021).

¹²² LAURA FERNANDEZ RODRIGUEZ/ROMAIN MAZON, Quelle sera l’issue de la partie d’échecs qui se joue entre Toronto et Google?, 2019, <<https://www.lagazettedescommunes.com/646652/quelle-sera-lissue-de-la-partie-dechecs-qui-se-joue-entre-toronto-et-google/>> (29.06.2021). It is interesting to note that the project continues in the form of smart zone management, cf. <<https://www.coord.com>> (29.06.2021).

¹²³ Cf. ELEMENT AI/NESTA (n. 70), 22.

¹²⁴ <https://tools.decodeproject.eu/>

¹²⁵ Cf. ODI (n. 73).

which can be enhanced through inclusion and representativeness on bodies; the importance of transparency, which can be fostered by subjecting the trust to access to information laws or permission regimes ; the importance of the values conveyed by the trust and the collection modes (e.g. intrusive or non-intrusive, such as collection by simply entering a site).¹²⁶

Success factors may also emerge from Ostrom classical eight institutional governance design principles, conceptualised to explain success and sustainability factors of commons¹²⁷. Beyond the factors of object and community delineation, which can be implemented through the adoption of an appropriate legal structure and dispute resolution mechanisms¹²⁸, this implies the recognition of a plurality of interests, a bottom-up approach with adequate representation of licensors, increased transparency through direct or analogous application of transparency laws, and a decision-making process that takes into account the trust values and degree of openness (e.g. social models that guarantee the circulation of public data and avoid its reappropriation)¹²⁹. Ostrom model is also about implementing monitoring of the use of data, and having external rules that allow users to recognize and respect the object, which implies binding rules and contracts that allow taking into account the interests of licensors, such as data confidentiality.

3. Legal structures inspired by copyright collective management

The choice of the legal structure is part of the success factors. The trust in the sense of trust law is not necessarily the most appropriate form, especially because of its lack of flexibility. However, the underlying concept - the settlor transfers his data to a service provider over which he retains control - can be realized through different legal forms. It is also interesting to combine both a corporate structure and a contractual architecture to impose internal requirements on the trust and external rules on users through contractual mechanisms such as those discussed above, but the more complex the environment (e.g. changing membership and evolving goals), the more appropriate a corporate structure will be.¹³⁰

In the EU, data trusts could develop rapidly with the EU Data Governance Act (DGA), which aims to facilitate data flows within the EU while strengthening the EU's digital sovereignty, in particular through new trusted intermediaries. Data sharing services are encouraged (art. 10 ff) with the emergence of “*data intermediation services providers*” subject to a notification regime (art. 11), specific sharing conditions (art. 12) and a supervisory authority (art. 13). “*Data altruism organizations*”, subject to a registration regime, would encourage donations of “*data in the general interest*” voluntarily provided by individuals or companies for the public good, such as scientific research or the improvement of public services (art. 16 ff.)¹³¹. The altruism organization will have to meet certain registration requirements, in particular to operate in a legal form that pursues purposes of general interest, without profit motive and legally independent from any other entity pursuing other activities (art. 18). These two new actors can thus be qualified as data trusts, insofar as they are “*legal structures providing independent data management services*”, specifying that they can be established in the EU or offer their services from abroad, provided that they comply with the conditions and appoint a legal representative in the EU (art. 10 § 3 and 17 § 3). The choice of their legal form remains free¹³².

¹²⁶ Cf. ODI (n. 73).

¹²⁷ Ostrom E, *Governing the Commons: The Evolution of Institutions for Collective Action* (The Political economy of institutions and decisions), Cambridge University Press 1990.

¹²⁸ Cf. ODI (n. 73).

¹²⁹ Cf. ODI (n. 73). On openness contextual meaning, see also Chuang, Tyng-Ruey, and Rebecca C. Fan, and Ming-Syuan Ho, and Kalpana Tyagi. 2022. "Openness". *Internet Policy Review* 11 (1). DOI: 10.14763/2022.1.1643. <https://policyreview.info/glossary/openness>

¹³⁰ Chris REED/IRENE NG, Data trusts as an AI governance mechanism, 2019, 18, available at <<https://ssrn.com/abstract=3334527>>.

¹³¹ However, the criticism of data altruism in a joint opinion of the European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS), cf. EDPB-EDPS, Joint opinion on the on the Proposal for a Regulation of the European Parliament and of the Council on European Data governance (Data Governance Act), 2021.

¹³² NB: the Digital Decade European Digital Infrastructure Consortium could also provide the legal form to govern the infrastructure for personal data trusts as it is meant to host multi-country projects https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

In other jurisdictions, where there are no comparable legislative proposals yet, it is however interesting to think about the legal structures that could play the role of data trusts and achieve the success factors. For instance in Switzerland, legal structures range from simple partnership (“*société simple*“) and trust agreement (“*contrat de fiducie*“) to corporate structures, including associations and cooperatives.¹³³ All are likely to achieve the goals of data pooling and sharing. The choice may vary depending on the context, but the main disadvantage of partnership and trust agreement is that they are purely contractual in nature, with no *erga omnes* effect.¹³⁴ Therefore, it seems advisable to focus on the association and the cooperative with *erga omnes* effect. In France, the instrument of (common law) public trust exists as *fiducie* and is an appropriate vehicle for public domain estate. It could be considered for data trusts too.

It is also possible to draw inspiration from the model of copyright management organizations (CMOs) because of their similarity to data trusts and to the managed resource (copyrighted works). CMOs are private societies created by rights holders, whose objective is to collectively administer copyright, namely to assert their prerogatives to a multitude of users, by collecting the royalties due according to the copyright, and by monitoring usages¹³⁵. The legal structure depends on the jurisdiction but they mostly take the form of corporate structures (e.g. cooperatives or associations)¹³⁶. These structures make it possible to comply with the legal requirements, in particular the non-profit purposes, the equal treatment of all members, the accessibility to all copyright holders (with the principle of an open door to all new members) and the appropriate participation in the decisions of the society (with the principle of an equal vote of all members, i.e. regardless of their economic weight).¹³⁷ Externally, the assignment of copyright makes the transfer enforceable against all, and internally (*inter partes*), the trust deed (*pactum fiduciae*) limits the effects of the assignment to the agreement of the parties.¹³⁸ Membership in the collective administration and the transfer of copyright are generally effected by the conclusion of a management (or membership) contract, which is supplemented by the general management conditions.

Consequently, data trusts could take the form of most CMOs, a non-profit association or cooperative that guarantees equal treatment of all members, accessibility to all rightholders, and appropriate participation in the society's decisions.¹³⁹ It would transfer, on a fiduciary basis, the rights of the owners it represents, all of whom would become members of the society and take part in its activities. In addition to management, societies could issue price recommendations to help individuals or companies determine the amount of remuneration they wish to charge for the use of their data.¹⁴⁰ Collection of royalties, for commercial use and/or to support the maintenance or the development of commons under strict conditions, could be enforced by such governance body. It can be managed either individually or collectively by a data, i.e. by or for an individual rightholder (individual rights approach, exercised individually), and by or for a collective rightholders (collective approach). Both can mandate a CMO or data trust, a local or global platform, a thematic or universal intermediary.

¹³³ GRISEL (n. 72), 14.

¹³⁴ GRISEL (n. 72), 14.

¹³⁵ In Switzerland, see CR PI-FEHLBAUM (n. 24), LDA 40 N 3 ; CR PI-SALVADÉ (n. 24), LDA Intro. art. 40-60 N 1 ; JEAN CAVALLI, La gestion collective des droits d'auteur exploités sur Internet, in: Internet 2005, 181 ff, 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.

¹³⁶ For instance in Switzerland, the CMOs take the form of cooperatives (Pro Litteris, SSA, SUISA, Suissimage) or associations (Swissperform). Cf. <<http://www.swisscopyright.ch/fr/recettes-et-repartition/repartition/reglements.html>> (29.06.2021).

¹³⁷ Cf. art. 885 CO respectively art. 67 al. 1 CC, specifying that the association may deviate from the principle of equality of votes by giving more weight to certain votes than to others.

¹³⁸ In trust law, cf. GRISEL (n. 72), 3.

¹³⁹ It is specified that the non-profit condition refers to the corporate structure, not to the members who may want to monetize their data, and that it is not part of the identified success factors, but of the concept of altruistic organization and trust law, so it could be removed depending on the *Data Trust* models chosen.

¹⁴⁰ YANIV BENHAMOU, Dommages-intérêts sous l'angle de la méthode des redevances suite à la violation de droit de propriété intellectuelle en droit suisse et comparé, Thèse Genève, Genève 2012, 251.

Finally, it is possible to draw inspiration from the copyright model of extended collective licenses (ECL) that could develop for massive use of copyrighted data (e.g. AI training data). ECL allows an authorised CMO to extend an existing collective license to non-member rights holders of the same sector, except those who have explicitly opted out¹⁴¹. Because of the similarity between the ECL model and data of all kinds (e.g. personal data used for AI training purposes), data trust could rely on similar principles and justifications as those for ECL.

4. Towards legal and political compatibility definition

Data trusts 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 licences drafters organisations, or later by data trusts based on declaratory shared values¹⁴². There remains a risk of demultiplications of data trusts and licensing proliferation, a caveat of the FLOSS world. Legal compatibility requires cooperation with other trusts, which may follow different political economy values¹⁴³.

Members would delegate the power to manage their data to data trusts and CMOs, and exercise their self-determination through their memberships. Members would select their data trust 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 ontologies¹⁴⁴. Values can also 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. Data trust implements a commons-based governance, collects 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.

IV. Conclusion

This article suggests exploratory paths to develop open licensing to all kind of data (including personal and non personal data and the case of mixed datasets covered by multiple intricate rights, avoiding users to have 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 cooperatives, commons-based pilot projects and copyright management societies.

In the annex, we propose a template for open data common licences (ODCL). This template has the triple view of (1) implementing the findings of this article to apply open licensing to all kind of data, including personal and non-personal data, (2) implementing both privacy and commons-based values, and (3) relying on the development of collective management organisations (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 licences”. This proposal is a first step towards the

¹⁴¹ For references relating to ECL, see BENHAMOU YANIV/FERLAND JUSTINE, Digitization of GLAM Collections and Copyright: Policy Paper. GRUR international 2022, 403.

¹⁴² 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.

¹⁴³ Again in line with Ostrom #8 principle, which recommends, as a success factor for commons, coordination with other groups.

¹⁴⁴ See <https://www.w3.org/TR/odrl-model/> and Abelson, H., Adida, B., Linksvayer, M., & Yergler, N. 2012. 10. CC REL: The Creative Commons Rights Expression Language. In Dulong de Rosnay M., De Martin J.C. (ed), *The Digital Public Domain : Foundations for an Open Culture*. Open Book Publishers. <http://books.openedition.org/obp/557>

design of a revamped open licensing framework which would go beyond existing open licences¹⁴⁵ and be compliant with copyright, privacy and other laws¹⁴⁶. Regarding the applicable law, we use EU regulations as a high standard baseline because privacy and self-determination are essential. However, we do not impose any applicable law in order to make the licenses as internationally valid as possible, but the license respects and is compatible with the strictest standards of data protection and moral rights.

While we are aware that not all practical questions are solved, such as its enforceability or the technical implementation of the withdrawal rights, the template offers a first panorama of the rights and values for a data licensing scheme to be applicable to the widest range possible of data, and compatible with other existing and future schemes, between open access and semi-closed commons.

To facilitate reading and future compatibility recognition, the template drew inspiration from the Creative Commons skeleton. A collaboration with public, private and commons actors and licensing experts is planned to further refine and develop this scheme with the Open Knowledge Foundation.

V. Annex: Template for Open Data Commons Licences (ODCL)

1. Table of elements




The combination of the various compulsory and optional elements to be selected by licensors may look as follows, illustrated by pictograms using legal design techniques and inspired by the bricks of Creative Commons licences. It does not follow a principle of no proliferation, but it allows to implement values by preserving privacy and supporting the construction of personal and non personal data commons.



Clause 1. Definitions




Clause 2. Scope of licence. Optional elements

Condition 1: Attribution BY  or No Attribution (~~BY~~)

Condition 2: Derivative  or No Derivative 

Condition 3: Confidentiality -> Anonymity , Limited Anonymity , or Clear (free flow, visible) 


Condition 4: Scope of use/users -> Open data (public use) , or Controlled Use 

Condition 5: Consortium only (closed doors) , Compatible user (shared value network) , or Compatible use (stamped use diploma) 

Condition 6: Finance -> For free (0€) , or For a fee (piggy bank) 

Clause 3. Values. Compulsory elements

Compulsory element 1. Privacy pledge (compulsory in all licenses) 

Compulsory element 2. Right to erasure (compulsory in all licenses) 

¹⁴⁵ Melanie Dulong de Rosnay, Chapter 9. Openness and Licensing, in Mathieu O’Neil, Christian Pentzold, Sophie Toupin (eds), *The Handbook of Peer Production*, Wiley Handbooks in Communication and Media Series, pp. 109-122. Mélanie Dulong de Rosnay, Alternative policies for alternative Internets, *Journal of Peer Production*, Issue 9 on Alternative Internets, September 2016, 10 p. <http://peerproduction.net/issues/issue-9-alternative-internets/experimental-format/alternative-policies-for-alternative-internets/>

¹⁴⁶ As envisioned in the ODECO EU project towards a sustainable Open Data ECOSystem Marie Skłodowska-Curie grant agreement No 955569.

Compulsory element 3. Share Alike (compulsory in all licenses) 

Clause 4. Other standards clauses

2. Clauses to be combined

Clause 1. Definitions

Consortium: Closed group authorized to exercise rights.

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

License elements: The options with icons corresponding to the licence optional conditions, clustered to minimise the cost of entry for licenses users, and illustrated by pictograms or icons. The scope of rights (2.) and the conditions (3.) can be selected by individuals or communities according to their needs.

Licensed material: includes works, databases, personal data, mixed data, technical data, any sort of data.

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

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

Compatible license: Licenses recognised compatible include futures versions of the Open Data Commons Licences (revisions) and other licenses recognized compatible by the license issuer or the Content Management Organisation and listed below¹⁴⁷.

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.

Clause 2. Scope of license. *Optional elements.*

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

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

Condition 1: Request attribution, or not

- Authorship and producer attribution (option BY) 
- No attribution needed (option no BY) (~~BY~~)

Condition 2. Authorise the making of derivatives or not

- Right to make derivatives (option Derivative) 

This option grants the right of adaptation, modification or data processing, aggregation, reuse, any treatment should be considered as data processing¹⁴⁸, derived version, compilation, modification.

¹⁴⁷ Based on the CC clause: "means a license listed at creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License."




¹⁴⁸ It will be assumed that processing requires prior access. On the debate, see Arye Schreiber, Mere access to personal data: is it processing?, *International Data Privacy Law*, Volume 10, Issue 3, August 2020, Pages 269–277, <https://doi.org/10.1093/idpl/ipaa005>

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


- No derivative (option No Derivative) 

This option reserves data integrity or for selected users/uses (see condition 4).


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

- Full anonymity¹⁴⁹ 
- Limited Anonymity 
- Clear (free flow, visible) 

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

Condition 5. Determine the scope of restricted users/uses

- Rights can be exercised for free only within the consortium (Option closed doors) 
- Rights can be exercised for free only by compatible trusts based on shared values (certified compatibility) (option compatible user / shared values network) 
- Rights can be exercised for free only for agreed purposes or by special categories of reusers certified by data trust CMO (option compatible use / stamped usage) 

Condition 6: Financial conditions:




- Rights can be exercised within the consortium or by compatible use or user for free (option royalty-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 (option shared revenues) 

¹⁴⁹ From Jurcys, op cit, p.9.

Clause 3. Values. *Compulsory elements.*

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.

These conditions are compulsory and part of the grant of all licenses baseline (in the same way BY is a compulsory element in all Creative Commons licenses)

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

5. Other standard clauses

Term and Termination:

Adapted from Creative Commons text: "This Public License applies for the term of the ~~Copyright and Similar~~ Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically."

Notice and license persistence:

The licensee must keep a link to this license, or a compatible licence when sharing it even in an unmodified version. Copies and derivatives must carry a link to this licence 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 licence, mediation, arbitration or alternative dispute resolution boards. The Creative Commons 3.0 International Governmental Organizations licenses have a mediation and arbitration clause.¹⁵⁰

Most open licenses usually do not mention any applicable law, and that the practice of legal porting or national legal adaptations has been abandoned by Creative Commons, since it was creating internal incompatibilities. Some open licenses (such as the European Union Public Licence¹⁵¹) contain an applicable law clause, which did 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 license¹⁵².

* * *

¹⁵⁰ See clause 8. h. at the very end of <https://creativecommons.org/licenses/by/3.0/igo/legalcode>

¹⁵¹ See EU Public Licence (EURL) article 15 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017D0863>

¹⁵² See EURL article 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."