

Occitanie Data's Ethical Charter for trustworthy development of the data economy

Why this Charter?

The development of the data economy holds great opportunities. However, in order to prevent an increasing number of violations of privacy and of society's interest, an ethical framework that promotes citizens' confidence in the new technologies must be established. That is precisely the purpose our Charter.

We started from the observation that while there is a large number of ethical frameworks applicable to artificial intelligence (for instance the EU's "Guidelines for trustworthy AI" ¹ and the "Montreal Declaration for responsible AI" ²), there is currently no comparable initiative applicable to data science, which is a broader issue. Indeed, although these two fields are linked, there are some ethical issues raised by data science that remain independent of those raised by artificial intelligence. Therefore, we considered the elaboration of an ethical framework specific to data science particularly necessary et relevant.

This Charter is the result of a collective work. Presented publicly in its first version elaborated by experts, a process of citizen co-construction will lead to a second version. In parallel to this process, everyone can already contribute to the improvement of the Charter through a feedback form, available on Occitanie Data's website ³. Moreover, the Charter is a dynamic instrument that will be regularly updated in line with scientific, technological, social and environmental advances.

Our association, Occitanie Data, pursues 3 objectives :

- developing a trustworthy data economy to offer innovative services in an aim of general interest ;
- define an ethical and sovereign framework for data science and artificial intelligence ;
- support and organise the digital transition in a compatible way with the ecological transition.

More precisely, Occitanie Data aims to facilitate the sharing and mutualisation of data between a multitude of actors - from the public sector, the private sector, civil society or research institutes - in order to give them the opportunity to develop new and relevant uses. Because of this atypical and original composition, the nature of shared data is of a wide variety : economic data, meteorological data, geo-

¹ High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019.

² Montreal Declaration for the Responsible Development of AI, 2018.

³ <https://occitaniedata.fr/consultez-la-1ere-version-de-la-charte-ethique-du-big-data-elaboree-par-occitanie-data/>

graphical data, mobility data, energy data, environmental data, infrastructure data, health data, etc. Thus, mixing non-personal and personal data.

Concerning the specific case of personal data, the framework of Regulation n°2016/679 relating to the protection of individuals with regard to the processing of personal data and the free circulation of such data, of 27 April 2016 (hereinafter "GDPR") and, in France, the framework set by Law n°2018-493 of 20 June 2018, naturally apply to all Signatories.

As a complement to this legal framework, our Charter sets out the key principles applicable to data sharing and related activities. Initially intended for Occitanie Data's members and partners, its general character will allow an appropriation by other actors at the French, European or even international level. This Charter, to which the parties voluntarily adhere, has no legally-binding value : it defines core ethical principles to guide the organizations that sign it (hereafter referred to as "Signatories") towards a greater consideration of these themes in their digital activities, and in particular those carried out in the context of Occitanie Data.

All the principles will then be translated into a set of detailed rules, enabling their optimal application by the Signatories. In other words, these rules will provide a more precise framework for the data processing that the Signatories plan to operate. Finally, this set of principles and rules will be complemented by recommendations for their implementation in the Signatories' policies, methodologies and technologies.

Finally, this triptych "Charter - rules - implementation" reflects the conditions that must be fulfilled to create a genuine area of trust, within which the Signatories can share their data without fear.

Nature of this Charter

The Charter constitutes a socle of ethical principles applicable to the activities of production, collection, storage and processing of data operated through a consortium as Occitanie Data, as well as to the uses of the results obtained from such processing. These principles are in accordance with the specific legal provisions applicable to the Signatories' fields of activity and may not be interpreted in a way to reduce their scope or undermine their application.

The integration of any interested organisation into such a consortium will be conditioned by the signature of this Charter. Furthermore, each consortium claiming to be guided by this Charter will have to enact precise rules concerning the application of the Charter's principles, and make them adopted by its members.

The Charter embodies 4 core values :

- **ETHICS**, to ensure that shared-data is used in a way that respects our planet as well as human being,
- **TRUST**, to enable the full potential of data share to be realized,
- **RESPONSIBLE INNOVATION**, to encourage the creation of innovative services in the interest of citizens,

- **DEVELOPMENT OF AN EQUITABLE DATA ECONOMY**, allowing an equitable distribution of value between each party having substantially contributed to the achievement of a shared-data project.

The principles set out in the Charter are not ranked in order of importance : they form of a coherent whole that should be interpreted in a comprehensive and constructive manner. If this Charter seeks to address the ethical issues raised by data sharing, it also seeks to address the ethical issues raised by the analyse of such shared-data by AI systems. For this reason, and in order to prevent the potential misuse of these tools, certain principles of the Charter are directly related to their use to process shared-data (*cf. principles 3.3 and 4.2*). Concerning the fields of AI that are not covered by the Charter, Signatories are invited to refer to the principles defined in the EU's "Guidelines for trustworthy AI", and those of the « Montreal's Declaration for Responsible AI".

Furthermore, the interpretation of this Charter must be guided by the fact that any project is part of a humanized system, composed of competent agents, in which technology is only a support for innovation. This human factor is intended to be as pluridisciplinary as possible, i.e. representative of all the skills mobilized by the exploitation of data (data analysis and management, human and social sciences, philosophy, etc.).

0. Preamble

We, Signatories of this Charter

Deeply committed to the strict respect of the rights and fundamental freedoms of the citizen as set out in the Declaration of Human Rights of 1789, the Preamble of the Constitution of 1946, the Charter of the Environment of 2004, the European Convention on Human Rights, of 1950 and the Charter of Fundamental Rights of the European Union of 2000,

Determined to explore all the possibilities offered by the legal framework to carry out innovative projects in a trustworthy framework, like the re-use of personal data for archival purposes in the public interest, for scientific or historical research or for statistical purposes⁴,

Aware that the current framework does not provide specific guarantees for data sharing actors, and that a thriving data economy can only prosper in an area of trust in which ethical principles are respected,

Inspired by the innovative method used to establish the « Montreal's Declaration for Responsible AI", consisting in a citizens co-construction process that aims to take into consideration the individuals point of view regarding to the conditions that must be fulfilled to allow the ethical development of data economy,

Convinced,

⁴ Article 89 of the GDPR.

- that data sharing is a powerful lever in order to achieve social progresses and to ensure a better quality of life, particularly in terms of health, ecology and culture,
- that the uses of shared-data have to serve the interests of humanity, for instance by developing useful innovative services, by satisfying its well-being aspirations and by addressing major global challenges such as climate change or epidemics,
- that ensuring a high level of data protection is an essential prerequisite for the use of shared-data,

Conscious,

- that citizens express serious concerns about the capacity of both public authorities and economic and social operators to manage data in an responsible way,
- that the processing and the sharing of data raises fundamental ethical questions and entails high social risks,
- that, to our knowledge, there is no ethical charter or declaration on data science with a similar scope to those that exist for AI,

Animated by the ambition to remove all unjustified barriers to data sharing, in particular the lack of a trustworthy framework and of an appropriate business model's,

Resolved,

- to provide a legal and ethical framework that facilitates data sharing, pursuing the aim of developing a data economy serving everyone's sake,
- to set the foundations for a social pact of trust between citizens and data science,

Proclaim the principles set out below as necessary for the growth of an ethical data economy :

1. Data Science and Society

1.1. Beneficence

The principle of beneficence and its corollary, the principle of do no harm, require Signatories to include objectives of collective welfare and of sustainability in the conduct of their activities. That means that the activities of the Signatories must be oriented towards improving the quality of daily lives of both current and future generations. Thus, Signatories should use the data in such a way to contribute to the achievement of the Sustainable Development Goals adopted by the 193 member states of the United Nations.

1.2. Sustainable innovation

The principle of beneficence leads Signatories to develop sustainable innovations. Thus, any technological, social or organizational disruption project carried out by Signatories thanks to the shared-data

is implemented using processes that respect both people and the environment. With this in mind, Signatories pay particular attention to their ecological footprint, with a view to reducing it. To this end, they respect the principle of digital sobriety (which implies processing only the data strictly necessary and preferring the most sober algorithms to achieve the project's purpose), prefer environmentally friendly data centers to store the shared-data and are cautious about the rebound-effect of their innovations.

1.3. Solidarity and inclusivity

Signatories shall ensure that their projects based on shared-data do not create or exacerbate social inequalities. In accordance with the principle of non-discrimination, they shall also ensure that such projects do not create, neither by purpose or by effect, a discrimination against an individual or a small group of individuals.

To this end, they shall be attentive to biases that may affect shared-data (*cf. principle 3.1*) and, where appropriate, to biases that may affect algorithmic systems used to process such data. At the same time, Signatories will progressively develop strategies to overcome this issue in a sustainable manner, in particular by promoting multidisciplinary in their project teams.

Signatories are also vigilant on the « digital divide » issue : they seek to reduce as much as possible the disparities regarding people's digital equipment and digital literacy. Also, they ensure that the most significant political and social digital devices are widely accessible.

2. Data Science and the Individual

2.1. Respect for individual autonomy

When they process personal data, Signatories consider the respect of individual autonomy as a central element. In this regard, when personal data is processed on the basis of the consent of the data subject concerned, Signatories shall implement best practices to provide an information that truly enables these persons to exercise their free will, in a clear, specific and unambiguous manner. They pay particular attention to the modalities of collecting consent from vulnerable persons, in particular minors, elderly persons and dependent persons.

2.2. Protection of personal data and privacy

Signatories shall comply with the applicable rules on privacy and personal data protection⁵. In addition, concerning personal data process for archival purposes in the public interest, for scientific or historical research or for statistical purposes⁶, all Charter's principles set out the appropriate guarantees for the respect of the rights of individuals. Also, the results of such data processing shall not provide any information enabling to retrieve data related to an individual or a small group of individuals.

⁵ For instance in France, the GDPR and the "Data Protection Law" n°2018-493 of 20 June 2018 (non-exhaustive list). While the national rules apply only to Signatories established in France, the GDPR applies to all controllers (or processors) who use personal data concerning persons who are on European Union's territory, as long as they offer services to these persons or monitor their behaviour within the Union (*cf.* Article 3-2 of the GDPR).

⁶ Article 89 of the GDPR.

Signatories shall pay particular attention to ensuring that the privacy of individuals is protected during the all "life cycle" of data, in particular through the careful application of the following principles : data minimization, data protection by design and data protection by default.

Signatories are aware that data sharing (even the sharing of anonymized data) considerably increases the risk of re-identification of individuals at the end of their processing. In this regard, they undertake to apply, on a case-by-case basis, the anonymization techniques they consider most appropriate to optimize the protection of their privacy.

3. Data quality and information system security

3.1. Data quality

Data quality is essential because it directly affects the quality of the processing results. Therefore, Signatories will take all measures they deem necessary to optimize shared-data quality ; for instance, by ensuring that the data provided are relevant for the intended use and are representative of the project target. Therefore, data quality requires Signatories to investigate and eliminate, as much as possible, bias affecting the data sets they intend to share. In addition, they endeavour to share data in interoperable formats.

3.2. Data centers security

In order to protect shared data from physical or virtual attacks that could compromise their availability, integrity and confidentiality, Signatories give preference to data centers that meet the highest security standards. Their vigilance is increased when they process sensitive personal data⁷ or data related to national public security.

3.3. Algorithms robustness

Algorithm robustness depends on the reliability⁸ and reproducibility⁹ of its results. Thus, Signatories processing shared-data using an algorithmic system should choose the most robust. In this matter, when the uncertainty of processing results cannot be avoided, they take care to make this clear. As far as statistically possible, they specify the margin of error to be taken into consideration.

Concerning algorithmic systems based on machine learning, their robustness also relies on the quality of data sets used to training them (*cf. principle 3.1*). Signatories using these tools to process shared-data are particularly vigilant on this point.

⁷ This includes data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership, genetic data, biometric data, health data or data concerning sex life or sexual orientation.

⁸ Reliability assumes that from the same data set entered, the algorithm calculates correct predictions for different situations.

⁹ Reproducibility assumes that from the same batch of data entered, the algorithm calculates identical predictions for identical situations.

4. Transparency

4.1. Clear and accessible information

In order to build trust, Signatories shall provide clear and accessible information on the method of data collection, where it is stored, the methods used to analyze them and the purposes of such analysis. These informations are provided within the limits of industrial, business, defence and professional secrecy.

More broadly, the Signatories shall endeavour to provide to citizens, by the means they deem appropriate, clear and accessible information about both advances and risks that may be entailed by data science.

4.2. Algorithms explainability

Signatories using algorithms to process shared-data shall provide an intelligible explanation of the result obtained, in particular by pointing out the principles underlying and the criteria determining the « choices » made by these algorithms.

Signatories using classical algorithmic systems to analyze shared-data shall take care to detail the way algorithms are implemented (i.e. detail the rules pre-set by a human operator and then embedded into the system). Signatories using algorithmic systems based on machine learning, which are currently difficult to explain, should at least explain the general logic of their functionalities. These explanations include for instance the data input, the purpose of the analysis of these data, and the data output.

Due to the opacity that affects algorithmic systems based on machine learning, and especially those based on deep learning, Signatories intending to use these tools will first have to demonstrate the significant advantage they provide in comparison with a more explainable method of data analysis.

4.3. Projects' auditability

Signatories recognize the importance of facilitating the monitoring concerning the compliance of their data sharing activities with the legal framework that applies to them and, as far as possible, with the rules they implement in order to apply this Charter.

In this regard, each step of a project they carry out thanks to shared-data is documented in adequate terms, these documents being intended to provide information or to serve as a basis for control :

- each project participant shall keep a description of its own data ;
- each participant shall ensure traceability of its data by mechanisms that make possible to list and detail all the transformations it carries out on them ;
- if an algorithm system is used to process the shared-data, at least one participant shall keep a document that describes its functionalities ;

- each participant shall document the various impact and risk assessments carried out prior to project implementation (*cf. principle 5.2*) and, concerning projects with a high societal impact, shall make these documents publicly available.

5. A trustworthy framework for data governance

5.1. Extension of the Charter by specific rules

This Charter contains a set of principles reflecting the guidelines to be followed by Signatories in defining the infrastructure in which they wish to share their data. Thus, these principles guide the conduct to be adopted by Signatories when carrying out a project using shared-data. With a view to making them sustainable, these general principles are interpreted taking account of scientific, technological, social and environmental advances.

Each consortium claiming to be guided by this Charter must translate these principles into a set of rules in order to provide a concrete framework for shared-data processing. These rules, which are more precise than the principles, will respect the legal and regulatory specificities of the country and sector of activity involved.

Then, these rules will be implemented organizationally and technologically to ensure that shared-data processing complies with the Charter's principles. The rules will be publicly available and submitted to independent control procedures to guarantee Charter's effectiveness.

5.2. Collective learning

Signatories are up to date on data sharing best practices. They produce an annual report Charter's implementation, indicating in particular any difficulties encountered in implementing one or more principles. These reports help to identify common issues and thus, open up possibilities for collaboration in order to resolve them.

5.3. Risks assessment

Signatories shall endeavour to assess direct and indirect risks that may arise from their shared-data projects regarding privacy of individuals, society and the environment.

On the basis of the results of risk assessments, Signatories seek to maximize the beneficial effects and minimize the adverse effects of their shared-data projects, at both individual and collective scales.

5.4. Citizen inclusion

Signatories provide citizens tools to develop their digital literacy. For this purpose, they shall set up mechanisms allowing them to be informed, or even to express their point of view, in particular concerning shared-data projects that are part of a public policy ; for instance by setting up specific, appropriate and effective means of communication, such as citizen consultations or citizen co-construction processes.

5.5. Integrity

Any person involved in a shared-data project scrupulously respect the deontological rules that applies to him or her, and acts in a spirit of intellectual integrity and cooperation. This covers both the aims of the work and the methods used, the management of human resources, including supervision of students, dissemination of knowledge and scientific communication. Above all, Signatories shall refrain from using shared-data for purpose of falsification, plagiarism or unlawful retention of data.

6. Reciprocity

6.1. Acknowledgement

When a shared-data project involves the collaboration of several Signatories, each essential contribution to the achievement of this project is explicitly acknowledged and made public. These essential contributions include for instance the production and supply of data, the supply of algorithms or the supply of research work for the project (non-exhaustive list).

6.2. Equitable distribution of value creation

Signatories recognize that value creation (economic, social or environmental), should not be captured by one or more dominant players. They shall develop models, and in particular economic models, that allow a fair return to each Signatory who provided an essential contribution to a shared-data project.

7. Ethics and emergencies

Data science can produce useful knowledge and tools in times of health, environmental or safety crisis. In this context, Charter's principles should continue to apply : they constitute a frame of reference for the use of data in a democratic and accountable environment. Nevertheless, it is possible that emergency leads to a temporary adjustment of these principles. Under any circumstances, the following principles should continue to be fully applied :

- proportionality principle : the least intrusive method of processing data should always be preferred to achieve the desired purpose ;
- if personal data are processed without the consent of the individuals, it stays necessary to provide them an accessible information about data processing characteristics (in particular the length of time of data retention and the purposes of the processing) ;
- if "anonymized" data, which are not covered by the GDPR, are processed in this context, particular attention should be paid to the application of adequate anonymization techniques to prevent any re-identification of individuals after the end of the crisis period ;
- in any case, population monitoring solutions should be based on a prior assessment of the data processing impact on the fundamental rights of individuals (including social impact, in order

to avoid discrimination) and, if possible, be pre-tested on a small scale before being deployed on a large scale ;

- risk management decisions and their underlying logic should be documented ;
- concrete procedures allowing the return to "normal" data processing regimes must be foreseen, paying particular attention to databases containing health-related data and databases created for the purpose of tracking, tracing and profiling individuals.

8. Regular review of the Charter

This Charter is a dynamic tool : it is not intended to be immutable, but to be regularly updated to keep pace with scientific, technological, social and environmental advances and also with evolutions of the legal framework applicable to digital activities.
