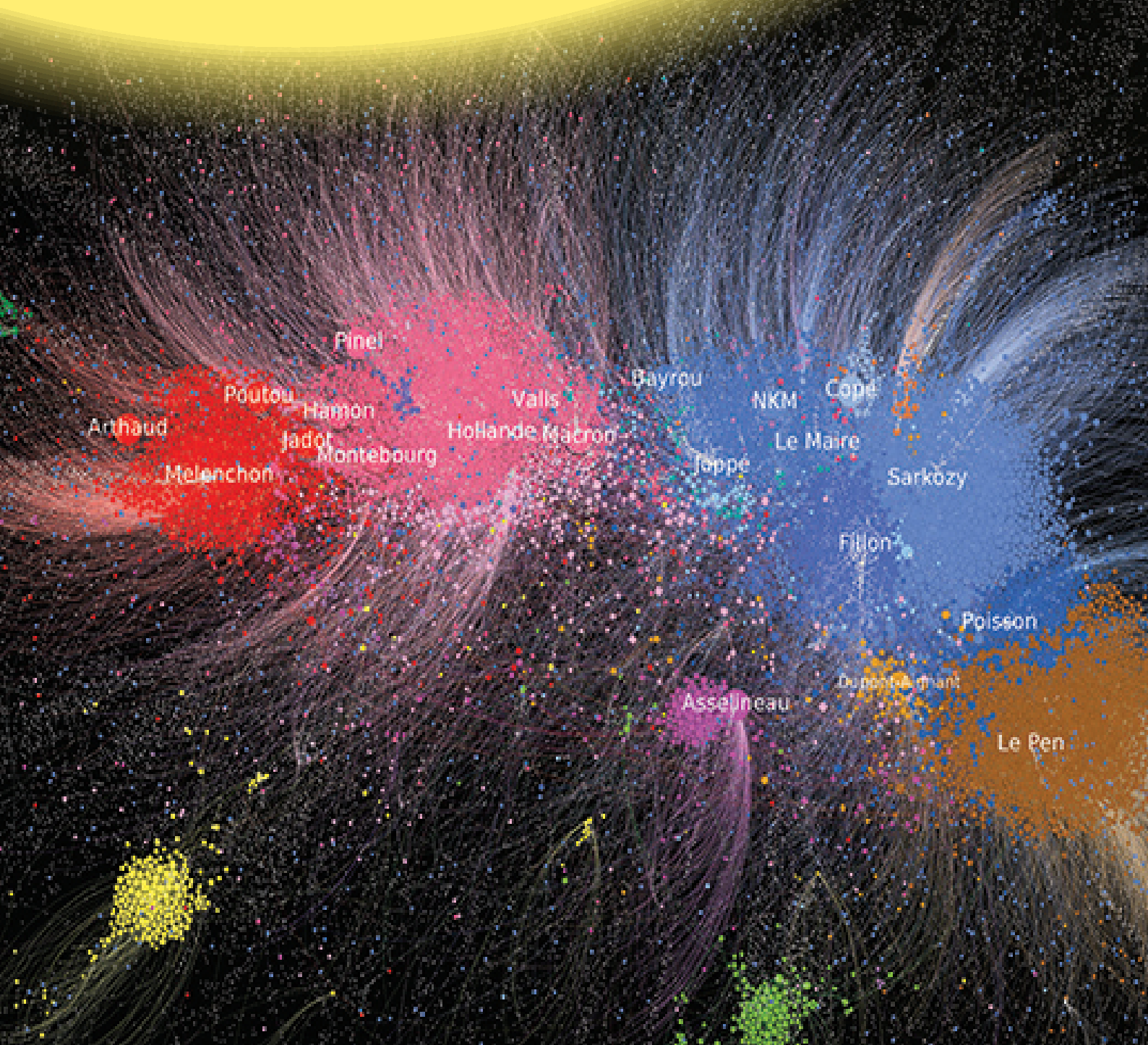
HUMANITIES & SOCIAL SCIENCES

Humanities and social sciences **and the protection of personal data** in the context of **open science**

A GUIDE FOR RESEARCH
VERSION 2
FEBRUARY 2021



Contents

« The political landscape of Twitter, before the first round of the 2017 presidential election » | CNRS Images
Photo credit: David Chavalarias/Noë Gaumont/Mazyhar Panahi/ISC-PIF/CAMS

Visualisation of the Politoscope modelling the political landscape of Twitter before the start of the 2017 French presidential campaign (tweets sent between August and December 2016). Each node represents a Twitter account (there are 65,000 in this figure) and the links materialise recurrent exchanges in the form of re-tweets between two accounts. The political communities are identified by colours. The nodes corresponding to the main presidential candidates' Twitter accounts are highlighted. Nearly 60 million political tweets were made during the presidential campaign and these have made it possible to better analyse the dynamics of the various political communities and how information circulates on Twitter. This study enables us to understand the mechanisms and challenges of social networks which have become a major communication tool for political parties during election campaigns.

Editorial

François-Joseph Ruggiu

Director of CNRS Humanities & social sciences

Nearly three years after the General Data Protection Regulation (GDPR) came into application a second version of the guide is being made available to Humanities and Social Sciences research communities. This is the result of a fruitful collaboration between INSHS teams, the CNRS Data Protection Office and several unit directors just like the first version. This cross-fertilisation of views and approaches made it possible to construct this guide in direct response to questions that may arise during the processing of personal data for research purposes.

This second version has been enriched by more tangible examples to illustrate developments in this field and with appendixes providing model documents and summaries to help make information more easily accessible for readers. The guide also benefits from the CNRS's experience gained over more than a year of applying the GDPR. The procedures described in the first version of the guide have been implemented, particularly the designation of unit directors as the persons responsible for the data processing carried out in CNRS laboratories. These directors can also rely on the support of the CNRS Data Protection Office which keeps a register of processing operations and advises researchers on how to implement the regulations.

The GDPR was the subject of a certain amount of concern in research communities when it was introduced because of its complexity and the constraints it looked like creating. In retrospect, it is now clear that it is in fact a well-balanced text. While it creates new obligations aimed at making controllers more responsible, it also provides for a special system for research activities which opens up a number of new avenues. Solutions can be found even for processing sensitive data as was demonstrated this year by the authorisations granted to CNRS research units by the CNIL (National Commission on Informatics and Liberty).

In 2020, the CNRS adopted an ambitious Research Data Plan which demonstrates the Institution's determination to fully respond to the challenge of data in the context of developments driven by digital technologies. Data need to be open when that is possible and protected when it is necessary which together create major organisational changes for research institutions. The pandemic has clearly demonstrated the essential need to share data widely among researchers while at the same time highlighting how important it is for the rights of individuals to be respected as set out in the GDPR.

We would like to thank the people involved in writing this guide particularly Gaëlle Bujan, the CNRS Data Protection Officer. A third version is already planned because the CNRS has committed to updating the guide regularly to keep pace with developments.

VERSION 2 OF THIS GUIDE WAS DEVELOPED BY :

Fabrice Boudjaaba, Deputy Scientific Director of CNRS Humanities & social sciences

Gaëlle Bujan, CNRS Data Protection Officer

Béatrice Collignon, Director of the Passages Research Unit

Christine Hadrossek, research data officer at the CNRS Open Research Data Department

Émilie Masson, legal supervisor, CNRS Data Protection Office

Lionel Maurel, Deputy Scientific Director of CNRS Humanities & social sciences

Clément Oliver, head of the legal affairs and partnerships department of CNRS Humanities & social sciences

Serge Pinto, deputy director of the Speech and Language Laboratory

Muriel Roger, Professor at the Sorbonne Economics Centre, Université Paris1 Panthéon Sorbonne; head of the «Méthodes et statistiques publiques» department of the PROGEDO Very Large Research Infrastructure (TGIR)

Paola Tubaro, Senior Researcher at the Interdisciplinary Laboratory of Digital Sciences (LISN)

Guide published in February 2021

Introduction

The European Regulation on the protection of personal data, which came into force in May 2018, raises questions among the scientific community, particularly in the humanities and social sciences, about the compatibility of research work with this regulation. It is very protective and provides every person free control over his or her personal data, under certain conditions. This is a fundamental right, the laws, statutes and regulations apply to everyone (see Article 8 of Chapter II «Freedoms» of the Charter of Fundamental Rights of the European Union)

Failure to comply with the laws and regulations on the protection of personal data is a criminal offence.

The scientific community in the humanities and social sciences only uses personal data for research purposes. Those who made this regulation have understood this and it takes into account the specificities of scientific activity, i.e. possible reuse of data for research purposes, processing of sensitive data (e.g. health data, data on trade union membership, ethnic origins) for research purposes by taking appropriate precautions, possible derogations under certain conditions from the obligation to inform individuals, etc (See article 89 of the GDPR).
The objectives of this guide are simple - to help humanities and social sciences researchers understand the legal mechanisms that impact their research and to provide them with the right habits and tools when they are led to process personal data.

It also aims to be a tool to support legitimate questions during the construction, implementation of a research program, the publication of results and the potential reuse of data. Data is essential material for scientific work. Personal data form part of this material, particularly for humanities and social sciences researchers. Such data need to be protected as much as they need to be used, reused or disseminated according to the organisational and operational principles of scientific research and in compliance with all regulations on the protection of privacy.

The final objective of the guide is to facilitate the definition of how to securely use personal data which belong to subjects of research who have agreed to contribute to helping knowledge advance. The measures taken and the transparency of the use of the data enhance the credibility of and trust in the work of researchers.

The various topics covered in this guide refer to the General Data Protection Regulation, its application to research in the humanities and social sciences and provide examples from situations encountered in laboratories.

The information and examples in this guide focus on research data.

The Guide is not intended to address issues of personal data protection and privacy related to the operation or administration of research in laboratories. For these types of data, the texts fall under European and French laws and regulations without there being any specific exceptions allowed for higher education and research. This is the case for data processing in the context of the organisation of scientific events, human resource management processes, professional travel management, etc.

As far as possible, this guide will be updated at least once a year by a committee of experts similar to the one that developed this initial version.

BACKGROUNG INFORMATION

In today's digital environment, the broad possibilities for the dissemination of information and data, including personal data, the daily use of social networks, and the vanishing boundaries between the public and private spheres make it crucial for the scientific community to maintain the trust of the individual citizens in the research activities and programmes being launched. Importantly, the quality of research and adhered-to research ethics combine to shape the simple principles that are part of the advancement of knowledge.

The research environment

Digital technology opens up new opportunities, access to mass and potentially reusable data, new techniques for storing, hosting and transferring information, which are all crucial resources for research whose reliability must be preserved.

Increasingly, progress in knowledge, scientific advances and innovation are partly based on data use/reuse and sharing. Open science, i.e. the unhindered dissemination of research publications and data, is now a «new paradigm» in which every researcher should work (Speech delivered by Frédérique Vidal, Minister of Higher Education, Research and Innovation, on 4 July 2018 to launch the French National Open Science Plan).

Open science relies on the opportunities to develop access to all publications and research data through digital technologies and it contributes to research efficiency, opens up opportunities to be part of international competition, and fosters citizens' trust through transparent research. This development impacts all scientific disciplines. The National Open Science Plan is in line with France's international commitments to transparent public action. It also meets the EU's Amsterdam Call for Action on Open Science to make research results accessible, without delay and without payment, to citizens, companies and research stakeholders.

The plan, with a budget of €5.4 million, is divided into three areas:

- Generalising access to open science: automatically publish in open access for any project financed by public funds; simplify the deposit of publications by researchers;
- Structuring and opening up research data: develop the open dissemination of publicly funded research data output; create the conditions and promote the openness of research data;
- Engaging in a sustainable European and international dynamic: develop skills in open science; encourage research operating organisations to adopt an open science policy; contribute to European structuring within the European Open Science Cloud.

The CNRS adopted its Roadmap for Open Science and finalised its Research Data Plan in November 2020. Various initiatives are underway which aim to enable data sharing and contribute to changing research practices under the guidance of the newly-created CNRS Open Research Data Department.

The increasing development of regulations and statutes in all areas of civil life also affects scientific work. Thus, in the fields of personal data protection, many other legislations apply to the processing of data: the law on digital trust, the public health code, the modified Jardé law (law on medical research), as amended, and its implementing decrees, the heritage code, the intellectual property code, etc.

The principles of research

Research quality, ethics and scientific integrity result from the practices and behaviours that contribute to civil trust and that of research stakeholders.

These principles must be complied with for all research, including that involving personal data.

Identifying the data adapted to the project, their relevance, volume, updating, stability over time, transparency as to how they are constructed are all part of science work reliability, and of the reproducibility of the results.

For any project, it is important to collect, use and analyse the data in connection with the research scope. An objective approach, respect for the data collected and shared research results are in line with the principles of ethics and scientific integrity which are to be followed.

Each higher education and research institution and many research funding institutions require these practices to be observed and make sure that they are complied with.

> The CNRS has set up an ethics committee and has entered into discussions on general ethical issues raised by the practice of research and related to the social and moral consequences of knowledge advances, the principles that should drive individual behaviour, and the making of science.

> In June 2014, the ANR adopted a policy on ethics and scientific integrity that sets out the fundamental principles

to be complied with by all research stakeholders and the rights and duties of those who evaluate and support scientific activity. To promote and make access to data easier and integrate its work in favour of open science into the National Plan, the ANR has required the implementation of a Data Management Plan (DMP) for all funded projects (see page 9 of the ANR action plan).

> The European Union in its H2020 programme requires research data management to be implemented. Research data must be «searchable, accessible, interoperable and reusable» Partners in EU-funded projects must build a data management plan.

See chapter 2, pages 19 et 20 for more information on data management plans.

Chapter 1 - The main definitions and their application to research in the humanities and social sciences

In this first part, the main concepts will be described and illustrated as much as much possible with examples in the different disciplines of the humanities and social sciences.
Personal data are involved daily in humanities and social sciences research projects, and protecting the information about people involved in scientific projects has to be a prominent concern.

1.1. PERSONAL DATA

Personal data means any information that makes it possible to identify a person directly or indirectly (Article 4 of the GPDR) :

- Directly identifying data: surname, forename, address, photo, voice, etc.
- Indirectly identifying data: a telephone number, or cross-referencing information such as the son of the research director, who lives on the island of Batz, etc.

Example : A research project is to develop a business travel plan in which the surnames and forenames of individuals are not collected (this information is not necessary) while data on the movements of individuals, their employers, their socio-professional categories and their place of residence allow the specific identification of these natural persons by cross-checking. This information is therefore personal data.

Note:

- Irreversibly anonymised data, whereby a person can no longer be re-identified, are not subject to the laws and regulations on the protection of personal data.

- Pseudonymised data are personal data that can no longer be directly attributed to the data subject. However, the use of additional information, such as a correspondence table, can be used to re-identify the person. In this case, the General Data Protection Regulation shall apply.

Among personal data, several are «sensitive» data under the Regulation: data disclosing alleged racial or ethnic origin, political opinions, philosophical or religious beliefs, trade union membership, sexual orientation, health data, biometric data allowing a person to be identified, genetic data.

Processing of such sensitive data is prohibited (article 9 of the GDPR) sunless explicitly listed in the Regulation (for example, with the consent of the person concerned, data manifestly made public by the data subject, substantial public interest, safety of human life). The use of this sensitive data is possible for public research purposes. In all cases, the reinforcement of data security and processing needs to be organised. More specifically, when the consent of individuals to process their sensitive data cannot be obtained, prior notice from the CNIL must be requested before any processing is carried out.

Other data are subject to specific requirements:

- The French Social Security number (aka NIR) is a directly identifying piece of data and its use is strictly regulated by law. This data may be used if the processing has an exclusively scientific purpose and provided that it has been encrypted prior to data processing.
- Data on offences or convictions can only be processed by the courts and a number of bodies specifically listed in the law. However, as part of an agreement with the Ministry of Justice, public research institutions and associated laboratories may sometimes be required, under certain strictly controlled conditions, to process these data, and in particular if, and only if, the purpose or result of the processing is not to re-identify a person.

1.2. STAKEHOLDERS AND ROLES

The EU regulation on the protection of personal data changed the concept of accountability for data processing.

First, the obligations of stakeholders, such as subcontractors, have been extended. Second, the controller must implement appropriate technical and organisational actions to ensure that the processing operation complies with the regulation.

For research projects, several parties are involved in achieving compliance.

- Researchers who conduct and lead any research project, whether funded or not, involving several partners or not, shall take the necessary steps to make sure that the project/processing complies with the regulation.

- Engineers and technians are involved in carrying out research projects, data collections, etc. They contribute to and carry out regulatory compliance procedures according to the roles they play in such projects.

- The role of PhD students: In the framework of research training, PhD students propose, construct and work on scientific projects. They are subject to the same regulatory requirements as researchers. They must comply with the GDPR when processing personal data. Any technical and organisational arrangements should be made in conjunction with their thesis director and need to respect the measures taken by the laboratory's supervisory institutions.

  For the humanities and social sciences, the unit director shall be the controller accountable for the processing of data related to a project. Doctoral students shall carry out the steps for compliance with the regulation under the supervision of their thesis director.

In the case of Cifre (Industrial agreements for Training through Research), responsibilities for data collection, processing or conservation should be jointly defined by the research unit and the company involved.
- the PhD student is paid by the company and signs an employment contract.
- an individual training through research agreement is signed between the company and the ANRT (National Research and Technology Association)
- a research collaboration agreement between the company and the supervisory Institution (on behalf of the unit) sets out the terms and conditions of the scientific collaboration project. A clause relating to the protection of personal data is to be included (role and responsibilities of each party) in this agreement. For units under CNRS supervision, the thesis director shall contact the Partnership and Technology Transfer Department at the CNRS Regional Office concerned.

In the case of *cotutelle* (jointly supervised) theses, the *cotutelle* agreement must define the role and responsibilities of each party in implementing personal data processing.

For units under CNRS supervision, the thesis director shall contact the Partnership and Technology Transfer Department at the CNRS Regional Office concerned.

. The controller responsible for processing is the person, public authority or body that determines the purpose and means of the processing operation. (Article 4 of the GDPR).

At the CNRS

For joint research units, the unit director is responsible for processing (controller). He/she must therefore make sure that the GDPR is complied with and appoint a Data Protection Officer. To this end, he/she relies on the scientific managers of the projects operated in the unit.

In most cases, when the unit director is a CNRS member of staff, he/she appoints the CNRS Data Protection Officer who shall assess the compliance of data processing work carried out by the whole of the unit.

Each controller is required to document the processing of personal data and maintain up-to-date records of processing operations, keeping track in particular of: :

- The purposes of the processing operation
- The categories of data subjects and related data
- The recipients of the data
- Information on the use of data, their storage and the rights of the data subjects
- The names and contact details of the controller and the Data Protection Officer

At the CNRS

The records of each controller (i.e. each unit) are maintained by the Data Protection Officer (DPO) on behalf of the controllers.

Formally, this task is carried out by the project's scientific manager on behalf of the DPO, who in turn provides counselling, and monitors and validates the registration of the data processing.

At their request and at least once a year, the DPO shall forward to the unit directors their unit's updated list of processing operations.

The procedures for registering processing operations were previously carried out and filed with the CNIL or the «Correspondents Informatique et Liberté». These are now to be carried out as a whole with the Unit's Data Protection Officer.

Remember, however, that the prior opinion of the CNIL is required before any processing operation that could create exceedingly significant risks for the data subjects, as highlighted by a privacy impact assessment (see page 14) and that the researcher cannot reduce without impacting his or her research.

Authorisation from the CNIL may also be required for health research (see the CNIL website and page 23).

In any case, it is advisable to contact the Data Protection Officer for a joint referral to the CNIL.

> The cocontroller of processing or joint controllers : Several institutions may define the aims and means of processing together. They must define their respective obligations in a transparent fashion by means of an agreement.

The persons concerned will be able to exercise their rights with regard to and against each of these

> The subcontrator is : « a natural or legal person, public authority, agency [or other body] which processes personal data on behalf of [on instructions from and under the authority of a] the controller » (Article 4 of the GDPR). The processor must provide appropriate safeguards to protect the security and confidentiality of the data, specified in particular in the binding contract between the controller and the processor. This contract shall also specify their respective commitments for data processing.

Examples of subsontractors :

Huma-Num TGIR data hosting service

A polling company when contracting a survey to such a firm. The processor is in charge of collecting the information from the data subjects while the researcher takes and uses the information thus obtained. In some cases the data may be pseudonymised by the processor.

> The Data Protection Officer (DPD)
Each public institution must have a Data Protection Officer to advise and make controllers aware of their obligations for the application of personal data protection regulations and rules (see article 37 to 38 of the GDPR). The Data Protection Officer advises on compliance with the regulations, cooperates with the supervisory authority, and ensures compliance with the regulation on the protection of natural persons.

Each unit director must appoint a Data Protection Officer. The appointed Data Protection Officer should be designated to the CNIL.

Note:

For unit directors who choose the DPO of the CNRS, the procedure is as follows: the unit director informs the DPO of his choice. The DPO contacts the CNIL for formalisation and designation

In all cases, a registration receipt is sent by the CNIL to the unit director with a copy to the designated DPO. This document must be kept by the unit and is integrated into the unit's corpus of documents relating to the protection of personal data.

All Higher Education and Research Institutions normally appoint a DPO whose contact details can be easily accessed within the institution.

For the CNRS, the DPO can be contacted using the following e-mail address: dpd.demande@cnrs.fr

> The Commission Nationale Informatique et Libertés (CNIL) is the supervisory and advisory authority in France responsible for monitoring, informing and supporting the application of the European regulation and French regulations on the protection of personal data (see article 51 of the GDPR) and the modified Data Protection Act of 6 January 1978.

1.3. THE TERRITORIAL SCOPE OF THE REGULATION

The European Regulation applies to data processing operations in the context of activities carried out by an establishment located on EU territories.
It also applies to processing operations carried out by a controller or processor established outside the European Union but involving individuals who are located in the territory of the European Union (see article 3).

Generally speaking, the GDPR applies in the following situations:

	The controller responsible for processing is located in a European Union country	The controller responsible for processing is located outside the European Union
People residing in a European Union Country	X	X
People residing in a country outside the European Union	X	NO

Scientific research is international by nature and the data protection regulations which need to be applied are the result of detailed analysis. It is advisable to contact your data protection officer for assistance.
It is important to maintain the principle of the objective protection of personal data whatever the applicable regulations while bearing in mind that levels of protection may differ from one country to another.

The CNIL has created and regularly updates a personal data protection world map.

Example: French researchers carried out a study on the genetic and linguistic diversity of the Cape Verdean population. European legislation is intended to apply because the controller is located on European territory, regardless of where the data collection takes place (Cape Verde in this example).

The transfer of data outside the European Union is possible provided that a sufficient and appropriate level of protection is provided. These transfers must be supervised using different legal mechanisms (see site CNIL website).

It is advisable to contact the unit's Data Protection Officer.

For an international research laboratory whose partners apply different laws and regulations, the applicable law must be subject to a thorough analysis.
A clause on the protection of personal data must be included in the relevant international cooperation contract. It is advisable to contact the Unit's Data Protection Officer.

1.4. DATA PROCESSING

A data processing operation is any operation involving personal data, whatever the process, the medium used, regardless of whether it is computerised. The data are used to meet objectives/purposes. The processing of data in the sense of «protection of personal data» goes beyond the analysis or exploitation of the data, it also covers the collection, analysis, reuse of data, archiving, etc. Article 4.2 of the GDPR.

Example:
Data hosting by Huma-Num and storage of health data.

The purpose of the processing is one of the essential principles of the Regulation. All data processing is carried out for a specific, explicit and legitimate purpose. The data may not be processed in a way that is incompatible with the defined purpose.
However, the data may subsequently be used for research purposes by providing safeguards to protect the privacy of the data subjects by the data collected (see Article 5 of the GDPR and article 89 of the GDPR).

For social science research, the research scope is often the purpose of data processing.

Example: a language science laboratory carries out a sociolinguistic study of the variation of the language used on Twitter
The purpose of this processing is not scientific research or the use of Twitter as a tool. It is the sociolinguistic study of the variations of language used on Twitter.

1.5. PRINCIPLES UNDERLYING DATA PROCESSING

Before starting one's research project and when it contains personal data, the person in charge of the scientific project shall undertake the analysis addressing:

- The lawfulness of the processing, which is the basis of the processing (a)
- The purpose of the processing operation (b)
- The relevance and proportionality of data (c))
- Data security and protection (d)
- Limited data storage (e)
- Transparency of information about the use of data (f)

In addition to the basis of the processing, it is necessary to ensure compliance with the principles of personal data protection in conducting the research.

(a) The leader of the project or the controller verifies whether the project is lawful, i.e. whether it complies with one of the following conditions: (Article 6 of the GDPR)

- The person has consented to the processing of his or her data
- The processing relies on a legal basis
- The processing is linked to the execution of a contract
- The processing is necessary to safeguard the vital interests of the data subject
- The processing is necessary for the performance of a task to be carried out in the public interest
- The processing is in accordance with a legitimate interest for the controller

In the humanities and social sciences, the basis most often involves consent, a task of public interest or a legitimate interest.

Examples:

- *Field surveys in metropolitan France involving personal data are often carried out on the basis of a consent given to the investigator.*
- *Sociological research with the collection of messages exchanged on Twitter can be based on a task of public interest.*
- *A joint analysis by a car manufacturer and a public research laboratory of the use of autonomous vehicles for people with reduced mobility can be justifiably based on legitimate interest.*

(b) The purpose of the processing corresponds to the objective pursued.

The purpose must match the missions of the institution or entity

Example:

Study on the evolution of territorial inequalities over the past 30 years in urban areas and the emergence of «urban traps» with the use of personal geolocation data using databases from the National Institute for Statistics and Economic Studies (INSEE).

(c) Relevance and proportionality of data

Data must be commensurate with purpose.

Example: In a research project on people's leisure activities, it may be relevant to collect certain complementary data such as religion. This may have an impact on the choice of leisure activities based on the days of practice of these leisure activities. The collection of this information is appropriate because it has an impact on research results.

(d) Data security and protection

The controller is required to take all measures to protect the data and prevent them from being diverted, reused for purposes not intended, to safeguard the integrity and confidentiality of the data.
Security mechanisms should be provided at all stages of the project, regardless of the nature of the data (see page 24).

(e) Limited data storage
The data may only be stored for a predefined and limited period. The length of the period of storage should be commensurate to the purpose of the processing. At the end of the processing operation, the data shall be either anonymised or stored for subsequent reuse for scientific research purposes only.
For research purposes, data can be archived according to specific provisions presented in chapter 2, page 26.

(f) Transparency of data processing
Information relating to the purpose of the processing operation, the name and contact details of the controller, the name and contact details of the Data Protection Officer and the storage periods shall be communicated in a transparent manner to the data subjects by the data controller.

See examples of information references on the CNIL website
See Appendix 2 for a sample information note developed by the Pacte joint research unit.

1.6. PRIVACY IMPACT ASSESSMENT

Its aim is to anticipate the risks of processing data for the privacy of the data subjects. This analysis is carried out by the controller (and by delegation, the scientific manager of the project) in conjunction with the Data Protection Officer and the Information Systems Security Officer.

It is mandatory where the processing is likely to generate high risks and in particular must be carried out if the processing includes at least two of the following criteria:

- Automated monitoring,
- Sensitive data,
- Large-scale processing,
- Data cross-referencing,
- Vulnerable persons (patients, elderly, children, etc.),
- Evaluation/scoring (including profiling),
- Automated decision making with a legal outcome,
- Innovative use or NICT use,
- Exclusion from a right or a contract.

The CNIL has published the list of processing operations for which a data protection impact assessment (DPIA) is mandatory and a list of operations for which it considers that a DPIA is not necessary. These lists are not exhaustive. Depending on the processing operations and the risks, the controller may decide to carry out a privacy impact assessment (PIA).

1.7. THE RIGHTS OF INDIVIDUALS

The GDPR has extended the rights of individuals.

> Precise information on the processing, purpose, use of the data, storage period must be provided to the data subjects. This information must be transparent and easily accessible (article 12 of the GDPR).It must be transmitted directly to the data subjects. Where the provision of such information is impossible or would require disproportionate effort, or where such information would be likely to render impossible the purposes of the processing or seriously impair their achievement, it is possible, by way of derogation, not to do so with the data subjects but to take appropriate measures to protect the rights and freedom of individuals, including by making the information publicly available (article 14.5 of the GDPR).

> Right of access to one's data (see article 15)

> Right to be informed of a data breach when there is a significant risk for the data subjects

> Right to objection (see article 21): a person may object, on legitimate grounds, to the use of his or her personal data unless the processing complies with a legal obligation. A derogation makes it possible to reject such a request where the processing is based on the performance of a task of public interest.

> Right of rectification (see article 16): a person may ask to modify his/her data.

> Right to erasure (see article 17): a person may request access to his/her data and request erasure. The request may not be granted, if exercising this right is likely to make impossible or seriously impair the achievement of the processing objectives

> Right to portability (see article 20): a person may ask to receive his/her data in a structured and machine-readable format and to transmit them to another controller. This right shall not apply to processing necessary for the performance of a task in the public interest or in the exercise of official authority vested in the controller.

> Right to a restricted use of one's data (see article 23)

Today, any individual can easily exercise his or her rights as soon as he or she knows the names and contact details of the controller and the Data Protection Officer, which are mandatory information that should be given to data subjects. The regulations provide for response times: as from the receipt of a request for access to the data, the data must be transmitted within one month.

At the CNRS, the exercise of rights is handled by the Data Protection Officer (DPO) and the controller

Case 1: When a request is made to the controller or the scientific manager
The request should be forwarded to the DPO who advises him/her. A reply is sent to the person and a copy forwarded to the DPO

Case 2: Request to the DPO
The controller or scientific leader is asked to draft the reply and send it with a copy forwarded to the DPO.
The request for the exercise of rights and the copy of the reply shall be entered in the records of the controller kept by the DPO.

One of the main changes in the GDPR covers the obligation for the controller (and the processor) to set out and organise measures to demonstrate compliance with the regulations at any time.

This accountability of stakeholders requires a thorough analysis of the data and their processing. During the development and implementation of research projects, this analysis is central and even imposed by the funders.

Chapter 2 - Research projects, data life cycle and the protection of personal data

This second part focuses on a tangible description of data analysis methods, the issues related to personal data protection legislation that may arise all along the path of data life during research projects, from data collection to dissemination, or potential reuse.

In the interest of compliance with the principles of ethical research, the reliability and quality of research and data require a structured approach prior to the establishment of any research project. This guides you through an early reflection about all stages of the project and helps you make appropriate decisions at the right time, thus facilitating the implementation of the project: for example, identification of the security measures to be taken, request for pre-project authorisations (filming, use of health data, sensitive data in certain cases, etc.), anticipation of final data storage issues at the end of the project and reflection on their availability at the end of the project.

Further reading: Huma-Num document on the main stages of a research project in the digital age

All projects now require prior reflection on research data. The aim is to identify data and describe the collection, storage and archiving methods, and thus to consider a data management plan.

Project grant applications are also facilitated or even automatically imply a prior clarification about the nature of the data the project is to use and the means. Therefore, a scientific project must formalise and explain all stages of data exploitation and processing. This is why research funding practices are gradually changing and funders increasingly tend to require DMPs.

Since 2007 and 2019 respectively, in an effort to achieve quality and open access to research, the European Union and the ANR have required a DMP to be drafted for every project funding application.

A Data Management Plan (DMP) : What is a DMP? Why a DMP?

A DMP is a structured document explaining how data are obtained and processed throughout their life cycle, from collection stage to archiving.
It must indicate:

- how research data are handled before, during and after the end of the project,
- the data that will be collected, processed and/or generated,
- - whether the data are shared, made accessible, and how they will be organised and stored (including after the end of a project).

A DMP:

- Ensures the quality of research
- Contributes to making data «findable, accessible, identifiable, reproducible», or FAIR (for H2020 projects)
- In our digital age, is a tool for reliability and knowledge for a potential reuse of Open Access data
- Meets the requirements of research funders like the European Union, ANR, etc. The associated costs are entitled to be included in project eligible expenses.

The elements relating to the protection of personal data are only a part of the information to be included in a DMP, even if compliance with the Regulation is to be observed during all stages of a DMP.

Several institutions provide approaches and templates to create data management plans:

DMP templates can be found at the Inist: see OPIDOR webside.

USR PROGEDO and its network of university data platforms , PUD, sur le territoire national, un accompagnement pour la réalisation des plans de gestion des données.

The INRAE provides a guide.

In 2020, gTIGRE, the inter-institutional working group for the Great East region of France published a self-training guide «s'autoformer aux données de la recherche : Guide à destination des professionnels de l'information et de la documentation» (Self-training for research data: A guide for information and document professionals). This lists a range of general training courses on data management and sharing, some of which are devoted to understanding the important legal and ethical issues.

Several professional networks in French Higher Education and Research are working on the development of data and data protection cultures. They can be contacted regarding the application of regulations and also for information on data management plans:

- The network of document specialists
- The network of engineers working for university data platforms
- The Inter-Network Data Working Group
- The database network

Also as an example, the CNRS Mission for Interdisciplinarity currently hosts and pilots 22 labelled business and technology networks on its platform. These cover the whole of the French territory and are transversal to the organisation. Thematic or institutional networks also exist such as the ISORE network, a network of INSHS document specialists and information professionals. Some of these are listed on the blog RH du CNRS.

These networks ensure the maintenance, exchange and development of skills. They can be called upon to provide support for the management of research data, writing data management plans and help with the application of regulations

The Inter-Network Data Working Group made up of representatives of several professional networks is particularly committed to defining best practices for research data management

The Inter-Network Data Working Group, made up of representatives of several professional networks is particularly committed to defining best practices for research data management

2.1. CREATING DATA (DATA COLLECTION) (COLLECTE)

2. 1. 1. Data categories

The regulation makes a distinction between the method used for collecting personal data and the method of producing research data in the humanities and social sciences.

Reminder: the GDPR makes a distinction between:

Data collected directly from the data subjects. These individuals must be informed precisely about the data collected, their use, the purpose of their processing, their storage period, and the procedures for exercising the relevant rights, the names of the controller and the Data Protection Officer.

Data indirectly collected are also subject to information on data processing as above. For this type of collection, it is also necessary to inform the data subjects about the categories of data collected and the source of the data (stating in particular whether they come from publicly available sources).
See articles 13 and 14 of the GDPR.

A - Data generated for research purposes directly by the researcher

Example:
Data from a sociological survey, an ethnological survey, a field survey, oral archives, questionnaires, forms, interviews, and data extracted from the web, etc.

If you contract to a service provider, a contract should set out the obligations and commitments of each party. The controller must communicate all information on the processing of data to the processor and comply with the principles of the data protection regulations. The obligations of each party shall be governed by a contract (Article 28 of the GDPR).

Note

The processor has crucial responsibilities for the protection of personal data. Thus, the processor must:

- Take into account the security, confidentiality and documentation aspects of the activities carried out on behalf of the controller
- Assist the controller in the implementation of his obligations (privacy impact assessment, data breach notification, security)
- Maintain a register of processing operations carried out on behalf of the data controllers
- Appoint a DPO under the same conditions as a controller.

In research projects involving contracted personal data processing, it is recommended to refer to the Partnership and Technology Transfer («Partenariats et Valorisation») and/or Financial («Département financier»)/Purchase department («Département Achats») of the supervisory authority institution that manages the project expenses in order to adapt the contract to the necessary protection of personal data.

Similarly, when the laboratory processes data as part of its services, it is recommended to contact the Partnership and Technology Transfer Department of the contract managing authority.

For surveys conducted by one or more students: provide a confidentiality commitment and ensure the security of the information systems used. Students will comply with the terms of the processing operation registered with the DPO of the unit in which the research is being carried out.

B - Data initially produced for non-research purposes and subsequently used for research purposes

Example:

Data from official statistics, surveys of polling agencies, government bodies, data from administrative files

In the case of data not directly collected from the data subjects but from a third party, which has collected the data lawfully, the GDPR provides that the pursuit of a research purpose is compatible with the original purpose of data collection. In this case, the regulations on the protection of personal data apply to the new processing operation.

In the case of data from official statistics, researchers can access different data files containing detailed and individual information on individuals or enterprises (household composition or income, business profits, location, etc.). There may be three types of files available, which may even come from the same resource. They differ in their accuracy and the risks involved in identifying individuals or companies.

. A distinction has to be made between:

- Confidential files giving specific information about the respondents that could be used to identify them. Access to this data is possible for researchers in a secure environment, on a project basis, after approval by the Committee on Statistical Secrecy. This is possible - against a fee - via the Secure Data Access Centre - CASD (https://www.casd.eu/)

- Production and Research files. These files are pseudonymised. The data made available to researchers are not directly personal but are considered to be indirectly identifiable (see definition on page 8). These files are produced specifically for research purposes. They are accessible to the scientificcommunity via Quetelet PROGEDO (http://quetelet.progedo.fr/) Standard files for which all the necessary processing has been carried out in order to make them anonymous (example: merging of professions and socio-professional categories with geographical areas of residence for households). They are accessible as open data.

The procedures for researchers' access to official public statistical data are set out in Appendix 3.

2. 1. 2. Types of personal data

Depending on their level of sensitivity, personal data may be subject to specific provisions in order to protect privacy.

For processing for research purposes, the data may:

1- have no specific level of sensitivity;

2- include so-called sensitive information (see page 11, chapter 1).

The processing of sensitive data is generally prohibited unless provided for by legislation: e.g. express consent or for research purposes. In this case, compliance with the regulations is of special importance (see page 11).

Example:

Comparative study of the eating habits of high school students residing in France, the United States, Senegal and Brazil with a follow-up of children aged 15 to 16 for 3 years comprising data on gender, socio-professional category of parents, religious beliefs, etc

Specific provisions must be implemented by the data controller for special categories of data relating to Social Security Numbers (NIRs) or criminal offences and convictions.
At the CNRS, the Data Protection Officer and also sometimes the INSHS (depending on the subject) must be involved once the planned processing of personal data has been defined.

3- target vulnerable groups of individuals: minors, elderly individuals, employees being subjects of surveys, prisoners, asylum seekers. There is no definition provided for the concept of a vulnerable population in the legislation. However, specific provisions must be complied with: e.g. obtain the individuals' consent and guarantee the security and confidentiality of information. For research involving minors, the information must be adapted and it is recommended to also inform parents accurately (or the person having parental authority).

Example:

A study aimed at specifying changes in the interactions between patients and caregivers. Patients in the situation being studied are considered to be vulnerable people and may not necessarily be able to express themselves freely.

4- cover research in the field of health
Health data are broadly defined as data relating to past, present or future physical or emotional health that gives an indication of the person's health status.

Example:

Monitoring the daily lives of men under 30 with autism with a collection of data on the condition and its course, and on the persons and their living habits.

Care should be taken with these two situations:

> The research involves health data (pregnancy, disability, chronic disease, etc.) but is not health research (it will not actually serve to improve health): the procedure is identical to that involving so-called sensitive data (see page 11).

> The research involves health data and its aim is to improve health. Specific provisions must be respected. These are set out in Appendix 4.

In all cases, for processing involving health data, it is advisable to conduct all steps in conjunction with the Unit's DPO whether the issue is linked to research in the health field or not.

In all cases, the processing operation will be registered in the processing records and documented with the privacy impact assessment, the compliance commitment and/or approval by the CNIL and any other relevant documents.

5 - Particular attention, including a data protection impact assessment, must be given to research involving data relating to criminal convictions and offences or related security measures.

2.1.3. The legitimate basis of the processing associated with data collection

As stated in Chapter 1, the regulation provides six bases, three of which are often used in research: consent, public interest, and legitimate interest.

In the context of research activities, processing should preferably be carried out on the basis of consent (compliance with the principle of informational self-determination), but may also be grounded on the performance of a task carried out in the public interest.

The choice always rests with the controller who analyses the nature of the data, the data subject population involved. This option's arguments should be detailed and the DPO should provide advice to determine the grounds for processing.

FOCUS on consent

Written or oral Consent may be given in different forms. In all cases, it is important to ensure the traceability of the collection of consent. It must include the information necessary for consent to be free, specific, informed and unambiguous.

You are advised to visit the CNIL website, on how to obtain the consent of individuals:
https://www.cnil.fr/fr/conformite-rgpd-comment-recueillir-le-consentement-des-personnes

Once given, one should be able to withdraw their consent as simply as when giving it. In the case of such an occurrence, the associated data can no longer be processed in the project.

The controller is required to keep proof that consent has been given to him/her to carry out a processing operation

See Appendix 1 : Sample consent form by the Pacte joint research unit.

2. 1. 4. Purpose

Article 5 of the GDPR allows for a certain degree of leeway regarding the final purposes of scientific research.
Many research projects are based on hypotheses that evolve and are refined through analysis, experiments, etc. Regarding the personal data required, this means that the purpose may be fairly broad at the start of the scientific project and be specified subsequently.

It is possible to achieve compliance with personal data protection regulations in such situations so that data processing can be carried out. The information given to the persons concerned must be adapted accordingly.
The purpose of the research will be refined according to how the project evolves. Technical and organisational provisions must also be updated and entered in the register of the data processing controller.

2. 1. 5. Proportional data

The question is whether the data collected are relevant and necessary for the processing operation.

Examples of questions that may arise when defining data for processing:

A research project requires the collection of the age of individuals, but this does not mean recording the day and month of birth. For the residence, is it necessary to collect the full address or only the city? For the occupational activity, is it necessary to know only the occupational category or the fact that the person works in a particular profession, in a particular company, etc.?

Similarly, personal data might be collected unintentionally.
For example, such data may result from an interview while not being related to a data processing purpose.

In this case, the researcher must decide whether the data should be saved for his initial investigation or erased, or reused subsequently (the same applies to so-called sensitive data).

2 . 2. DATA STORAGE

Preserving the security of data access, storage and hosting is crucial to protect personal data. The controller should use the tools provided by his supervisory institutions to comply with and/or enforce compliance with the institution's internal policies.

The basic rules (non-exhaustive list) for securing information systems, the use of digital tools, data exchanges and data storage are in line with the information systems security policy of the unit's and researchers' supervising institutions::

- Authentication should be provided for users of digital tools: digital certificates, passwords;
- Authorisation management: restricted access to sites and data should be granted only to persons approved by the controller or research project manager;
- Security of tools: computers and smartphones should be encrypted;
- Internal computer networks should be protected;
- Exchanges between organisations, units and researchers should be protected by a secured system;
- Secure tools should be used for videoconferences (for example at the CNRS: use of Skype Enterprise and Renater services).

Accessible tools
CNRS digital services
CNIL publication security of personal data, 2018 edition

FOCUS: Examples of bad practices:

Use of unsecured email messaging for information exchange
Private and professional email interactions
Use on-line survey tools hosted outside the EU (such as Google Forms). Save files on one's personal workstations when they are only accessible on secure workstations
Exchange files containing so-called sensitive personal data by email without message encryption

FOCUS: Retrieval advice by government authorities under the digital security programme:

Best practices are presented on the French government site «Cybermalveillance

Several entities offer solutions for data storage. Whatever the medium, it is important to anticipate the volume of data, the duration of storage and to estimate its cost (this can sometimes be taken into account in the eligible expenses of research project applications).

Huma-Num, is a Very Large Research Infrastructure (TGIR) whose mission is to ensure the preservation of scientific heritage in humanities and social sciences.
HumaNum offers various services that assist units at all stages of the digital data life cycle.

> storage: storing data in a secure personal space
> data processing: analysis, computing, coding, visualization, geographical information
> data dissemination for open access to data; shared web hosting, virtual machines, metadata dissemination
> data conservation, storage on the HumaNum servers hosted at the CNRS
> data indexing via the ISIDORE tools, data harvesting
> data presentation: the NAKALA service: data access and metadata presentation.

2. 3. DATA PROCESSING

In this section, data processing is to be understood as the stage of data analysis, consolidation and work on data that helps to provide insight on the issue investigated.

The data may match different statuses

1 — identifying data used on a small number of people and needed for qualitative analyses (common in ethnology, social or cultural geography, sociology).
It is important to remove the identifying character at the publication stage and/or at the end of the research project. Depending on the situation, anonymisation or pseudonymisation is required.

2 — anonymised data

This irreversibly removes the link with personal data.
If the identification of a person is not possible in any way, the GPDR does not apply. In qualitative surveys, anonymisation will generally not be possible, as there is a proven need to identify individuals.
When the research objectives requires the identity of the interviewees (personality, expert, etc.) to be cited, they should be informed that identifying data will be published and that they will be guaranteed access to the interview transcript.

For quantitative studies, anonymisation is often possible and is compatible with the research project. It is best to carry out anonymisation at the very moment of data collection.

3 — Pseudonymising data consists in separating directly identifying data (e.g. surname and forename) from other non-identifying data (for example, by assigning numbers to persons thus avoiding disclosing their surname, but by keeping a correspondence table to trace the identity of the person).

FOCUS on anonymisation

The anonymisation of data requires that the identification of persons becomes impossible, either directly or indirectly, an operation that follows a specific process.
Whatever the technique used, anonymisation must lead to compliance with three criteria:
- Total inability to single out an individual
- Total inability to link records relating to two individuals together
- Impossibility to deduce information about an individual

Examples of anonymisation techniques (Opinion of the G29) :

- *Adding noise: altering the accuracy of the information by adding randomness*
- *Switching: Mixing attribute values within the dataset*
- *Generalisation: Change the granularity of the values to form groups*
    - *k-anonymity: at least k people have the same profile*
    - *L-diversity: at least l values have the same attribute*

Big Data and Artificial Intelligence

The use of innovative data processing technologies such as Big Data or Artificial Intelligence (AI), requires special care, given the characteristics of these technologies.

Because it involves the processing - and often the cross-referencing - of large amounts of data, Big Data carries a higher risk of indirect identification of individuals, even when it is based on initially anonymous data.

The creation of a database from several sources requires special precautions. It must be ensured that the consent of the data subjects has been obtained or - in cases where processing is carried out on another basis - that the right to information of the data subject is respected.

The implementation of technologies such as Big Data or Artificial Intelligence will often require a privacy impact assessment (AIPD. see page 17). An impact study is necessary when faced with several criteria, including character profiling, data cross-referencing, automatic decisions or use of cutting-edge technologies.

With regard to Artificial Intelligence and Machine Learning, it should be remembered that profiling operations (in the sense of processing personal data for the purpose of analysing and predicting behaviour), as well as fully automated decisions that may result from them, are subject to special supervision under the GDPR.

In particular, persons may assert rights to transparency and the intervention of a human person, when such decisions have legal effect.

2.4. DATA ARCHIVING

In principle, the processing of personal data must have a fixed period of time, in relation to the fulfilment of the purpose for which they were collected. The GDPR states that this data storage period is set at the «strict minimum», at the end of which normally the data must be archived in accordance with the regulations on public archives.
Nevertheless, the GDPR also sets out that data may be kept for longer periods of time when they are processed «for archiving purposes in the public interest, for scientific or historical research, or statistical purposes» (see Article 5 of the GDPR and information on the CNIL website).

Data storage according to the regulations on archive provides for a three-stage cycle.

Phase 1: storage in the active database
It corresponds to the current use of data, i.e. the time span of the research project.

Phase 2: Intermediate archiving
Under certain conditions, personal data may be stored after the data have been processed but with restricted access. Personal data for research purposes are often archived in a so-called intermediate format, provided that the rights of individuals and the associated information are maintained.

At the CNRS, the intermediate archiving period is often two years after the last publication of the research results.

Phase 3: final archiving
Personal data that is not destroyed may be archived in accordance with the provisions of Book 2 of the Heritage Code (Code du Patrimoine). Final archiving cannot be carried out in the laboratory. It is carried out with the Departmental or National Archives in conjunction with the supervisory institutions of the laboratory. Certain French Higher Education and Research infrastructures like Huma-Num or the CINES can provide archiving solutions.
Once archived, research data containing personal information may be retrieved in accordance with the general rules governing the communication of archives laid down by the Heritage Code (deadlines for communication, possible exemptions for researchers, etc.).

Conservation periods must be set out and explicitly laid out when registering processing operations. This information must also be transparent to any subject data involved in a data procession operation. The conservation periods can be changed during processing operations.

2. 5. DATA SHARING IN PARTNERSHIP RESEARCH

Research activities are sometimes conducted in partnerships involving several entities, under different supervision, sometimes involving public and private partners.

In these cases, it is imperative to anticipate and lay out within a partnership agreement (such as a consortium agreement), the capacities the partners in the project will have, within the meaning of the GDPR, as controller, joint controller or subcontractor. These partnership agreements must make it possible to identify the roles and obligations of each party, particularly regarding data security.

The scientific leader of the project is the person best suited to become the controller of the processing operation under the supervision of his/her unit director.

When a research project is to be conducted by several specific partners, the scope of authorisations for data access and manipulation must be set up in anticipation when drafting partnership agreements.

In any case, for the research purpose of the project, funders will not have the status of data controllers, in particular insofar as they are not the recipients of raw data in most situations, but only of the finalised research output.

Data can be shared inside or outside the European Union.
Depending on the situation, an adapted framework needs to be set up. The transfer of data outside the European Union is possible as long as a sufficient and appropriate level of protection is ensured. These transfers must be regulated using different legal tools (see CNIL website).
It is also necessary to make sure these exchanges comply with regulations on the protection of the State's scientific and technical heritage

2. 6. DATA DISSEMINATION AND PUBLICATION

Several potential situations are to be taken into account:

- Dissemination of anonymised data (always possible when the data is truly irreversibly anonymised);
- Transmission of unanonymised data to other researchers;
- Publication of unanonymised data in research papers;
- Dissemination of unanonymised data with prior consent of the data subjects.

As for the transmission of non-anonymised data to other researchers, this is made possible with the authorisation of the controller pursuant to the decree of 1 August 2018 which provides (art. 100-1):

> *« The data resulting from these processing operations and stored by the controller or his processor may only be accessed or modified by authorised persons. These persons shall comply with the rules of ethics applicable to their sectors of activity»*
> *The authorisations granted to these persons by the controllers are strictly regulated, they must be in line with the specific purposes and safeguards provided for by the above-mentioned decree.*

Research is based on public personal data collected online or via social network services. The laws and regulations apply if the data are not anonymous.

- define the legal basis: generally, consent or performance of a task in the public interest
- state the purpose of the processing and the relevance of the data collected
- provide for the information of individuals (e.g. on laboratories' websites, in the press) and in particular about the procedures for exercising their rights
- use anonymisation of data as early as possible
- set up data storage periods according to the purposes of the processing and the project stage

Attention should be paid to the way data are collected, such as secure storage, sharing, etc

Further reading: see the CNIL's decision of 3 may 2018 authorising the University of Lorraine to implement automated processing of personal data for the purpose of research on the impact on privacy of publications of information openly accessible on social networks.

2. 7. REUSING DATA

The reuse of data makes it possible to share «personal data» resources with other researchers, particularly in the context of open science.

Since 2016, legislation (Digital Republic Act, Valter Act) have enshrined a general principle of openness and free reuse of public information (Open Data by default). In principle research data are such public information.
See: Guide d'ouverture des données de recherche du CoSo (a guide to research data openness by CoSO, or open science committee).

Nevertheless, the law links these obligations of openness to the protection of personal data, providing that when including personal data, public sector information may only be made public «after having been processed in such a way as to make it impossible to identify such persons» (anonymisation) or with the consent of data subjects.
Data may be reused for research purposes when they have been anonymised. They may also be reused under the persons' consent or if reuse had already been provided for in the initial processing

Except where the data have been anonymised, re-use does not exempt the researcher from procedures to update compliance with the GDPR: lawfulness of the processing, explicit, legitimate purpose, proportionality of the data, security of the data, information to individuals, etc.
See: Practical guide of the CNIL and CADA on the online publication and reuse of public data (Open data)

Example: Quetelet PROGEDO Diffusion and CASD provide access to data that can be reused for research projects.

The data can be reused for a commercial purpose, for example in research and technology transfer situations. In all cases, further processing will have to be set up and the personal data anonymised. Overall, except in research and technology transfer situations under the control of the laboratories' supervising entities, the purpose of the data processing operations must not be commercial. Precautions must be taken and counselling by the DPO and Technology Transfer Departments should be sought.

Appendix

APPENDIX 1 : SAMPLE CONSENT FORM

Proposed by the Pacte joint research unit, Social Sciences Laboratory (CNRS, Grenoble Alpes University, Institut d'Etudes Politiques de Grenoble)

CONSENT FORM FOR THE COLLECTION OF PERSONAL DATA

(To be completed on laboratory letterheaded paper, in duplicate to be signed by the respondent. The investigator will provide one copy to the respondent and the other to the project manager).

This form is intended to collect your consent for the collection of data/information about you, as part of the XXX project led by «specify team / laboratory».

By signing this consent form, you certify that:
- you have read and understood the information provided in the information sheet;
- you are satisfied with the answers given to your questions;
- you have been informed that you are free to withdraw your consent or withdraw from this research at any time, without prejudice

Information about the participant:
Surname:
Firstname:
Address:

To be filled by the participant: (to be adapted accordingly)
- I have read and understood the information provided in the information sheet and I freely agree to participate in this research:

[] YES [] NO

In the situation of an interview investigation:
- I agree that my comments may be recorded and used by the XXX project team:

[] YES [] NO
- I agree that my image and comments may be filmed and used by the XXX project team:

[] YES [] NO
- I accept that my image and my comments may be disseminated during scientific conferences, seminars or any form of promotion of the XXX project:

[] YES [] NO

In the situation of a questionnaire survey:
- I agree that my answers to the questions asked may be used by the XXX project team:

[] YES [] NO

Special cases:
- I agree to the use of an embedded system [or connected object] to collect data [geolocations, practices, etc.] and that these data [geolocations, practices] be used by the XXX project team:

[] YES [] NO
- I agree that «sensitive data» of the type (list the data concerned here) will be collected, stored and used by the XXX project team :

[] YES [] NO
- I agree that my personal data may be used for research projects with the same purposes as the XXX project:

[] YESI [] NO

1.2. Surname, forename - Date - Signature

A copy of this document is given to you, another copy is kept in our records.

APPENDIX 2 : SAMPLE INFORMATION NOTICE

Proposed by the Pacte joint research unit, Social Sciences Laboratory (CNRS, Grenoble Alpes University, Institut d'Etudes Politiques de Grenoble)

PERSONAL DATA COLLECTION INFORMATION SHEET
(to be given to the participant on a laboratory letter headed sheet).

(Language should be adapted to the target audience: for example, information sheets for children should be written in clear language with words used by children)

(Data controller)
The information collected [about you] (if direct collection) will be processed as part of the project XXX managed by «Surname, forename, title and institutional affiliation of the project manager, postal address and e-mail»

(In case of indirect data collection, e. g. web data, designate the data subjects)
The persons involved by the data processing operations will be:

(Purpose of the project)
The purpose of the processing operation is: «specify the main goal of the research and, where applicable, specify sub-objectives».
Specify what is expected of the person...

For example

For an interview investigation

We expect you to participate in an interview during which we will ask you questions about
(write a brief description of the purpose of the project). .... The interview will last «specify duration».
Specify the collection method:
Option 1: The interview will not be recorded.
Option 2: The information collected during this interview is recorded.
Option 3: The information collected during this interview is videotaped / photographed.
Option 4: In case of commented routes with satellite navigation system (SATNAV) data collection: The routes we will follow will be recorded by GPS/GSM sensor.

For a questionnaire survey

We expect you to participate in a questionnaire survey during which we will ask you questions about «write a reminder of the project purposes». The questionnaire will last «specify administering time span». In case of a longitudinal survey, specify the duration of participation and collection periods

Additional collections (e.g. logbook, SATNAV tracking, use of connected objects, etc.):
Adapt accordingly:
We would also like you to fill a logbook to learn about your practices of «specify types of practices» for a period of «specify time span».
We also want to use a SATNAV sensor (or other connected object) to understand your practices of «Specify type of practice» in «specify territory» for a duration of «specify time span». You may «turn off the GPS/deactivate the connected object» at any time

(Type of the data collected)
Only the data strictly necessary for the performance of our research will be collected and processed: List the types of personal data collected, for example:
Identification
Personal life (lifestyle, family situation)
Data on professional life (CV, education, training, distinctions, publications, etc.)
Economic and financial information (income, financial situation)
Internet connection (IP, logs, etc.)

Geo-location (GSM, SATNAV data, etc.)
Sensitive data (religious or philosophical beliefs, trade union or political affiliation, sexual orientation or life, offences or convictions, social security numbers, health, biometric or genetic data)
Data source (in case of indirect collection)
This information is collected from (source to be specified and indicate whether it comes from publicly available/ unavailable sources) between (specify collection period).

(Legal basis of the processing operation)

The legal basis of the processing operation is based on:
Adapt accordingly:

- the execution of a public research tasks => if projects financed with public funds only
- participants' consent => mandatory in case of sensitive data, joint consent of the child and holder of parental authority if the respondents are minors under 15 years of age.

(Voluntary participation)

Your participation in «specify project name» project is on an entirely free and voluntary basis.

(Withdrawal of consent)

You are free to withdraw or terminate your participation in this project at any time. This withdrawal will have no consequences.
[If you work with students, you can specify that the withdrawal will have no impact on the rest of their education].
Longitudinal data collection case: In the case of data collection over several periods, the withdrawal of consent will be effective from the date it was received by the controller

(Pseudonymisation/ confidentiality)

For an interview investigation
The project «specify the name of the project» makes the following commitments:

- Your identity will be obliterated using a random number in all writings produced on the basis of your comments (interview reports, observation notes, analysis notes exchanged between researchers, publications...).
- No other information will be kept that could reveal your identity: interview notes, interview reports, observation notes, analysis notes and publications will be completely anonymous

For a questionnaire survey

- Your identity will be obliterated using a random number for all types of information collected (list to be adapted according to the project: questionnaires, SATNAV data, logbook, etc.).
- Only the project manager holds the correspondence table that allows you to link your identity to the random number assigned in the different files (list to be adapted according to the project questionnaires, SATNAV data, logbook, etc.).

(Recipients of personal data)

The recipient or categories of recipients of this data are: «indicate who needs to access or receive it according to the stated purposes; specify the names of organisations, partners, entities, etc. ».

(Data transfers)

Option 1: All data will be kept in France.
Option 2: The data collected will be transferred / stored by one of the project partners in a European Union country, which is subject to the same privacy rules as France.
Option 3: The data collected will be transferred / stored by one of the project partners in a country outside the European Union. The transfer is based on standard contractual clauses of the European Commission or governed by approved specific contractual clauses, a code of conduct, certification, etc.

(Storage period)

Your personal data are kept in an active database until/for «specify date or duration».
Option 1 : After this date/period, your data will be permanently archived (if of significant scientific, statistical or historical interest)
Option 2 : After this date/period, they will be permanently archived anonymously (if there is no interest in keeping the personal data).

(Security)
In order to guarantee the confidentiality of your data and avoid their disclosure, the following measures have been put in place:

- The only entity(-ies) which will have authorised access to the data will be the following: «specify entity(-ies)».
- (If applicable) The external service provider «specify its role» is subject to contractual guarantees protecting your data.
- The following safety measures (hardware and software) have been secured: (e.g., fire protection, backup copies, antivirus software, regular change of minimum 8-character alphanumeric passwords, computer encryption)

(Dissemination)
Les résultats de cette recherche seront diffusés de façon anonyme dans des colloques professionnels et scientifiques, dans des rapports destinés aux autorités, dans des revues professionnelles et académiques et dans des médias destinés au grand public (liste à adapter selon le projet).

(Droits des personnes)
You are entitled to ask questions about this project at any time by contacting the project manager by e-mail: «specify address».
You have the right to access and require a copy of your personal data, object to the processing of your personal data, have them rectified or deleted. You also have a right to limit the processing of your data. You can exercise these rights by contacting: «indicate the contact details of the department or person responsible for the right of access - postal address and email».
You can also contact the Pact Laboratory Data Protection Officer at the following address:
CNRS - DPD — 2 rue Jean Zay — 54500 Vandoeuvre lès Nancy Cedex - dpd.demandes@cnrs.fr

After contacting us, if you believe that your Data Protection rights are not respected, you may file a claim online with the CNIL or by post. CNIL, 3 Place de Fontenoy, TSA 80715 - 75334 Paris Cedex 07 (https://www.cnil.fr/)

APPENDIX 3 :

FOCUS on the proceduree for researchers to gain access to public statistical data

Researchers are allowed to access individual household or business data collected through statistical surveys or from administrative files and then transmitted to the statistical office. These data are accessible in the context of research, scientific production and, in some cases, teaching.

The CSS , (Statistical Confidentiality Committee - https://www.comite-du-secret.fr/) is responsible for ensuring compliance with the rules on dissemination. It decides whether to approve requests for data release.

The access process depends on the nature of the data requested.

Confidential Data

Confidential data are statistical data, mostly from official statistics or tax data and also some administrative data. An online request must be made to access them via the Confidential Data Access Portal (CDAP).

The procedure for doing this is set out on the CCS website.

The administrative departments and offices which produce such data examine all applications and these are evaluated by the committee before access can be authorised.

Under the terms of Article 17 of Decree No. 2009-318 of 20 March 2009 on the National Council for Statistical Information, the Statistical Confidentiality Committee and the Committee for the Official Statistics Label, "the Statistical Confidentiality Committee shall deliver its opinion taking into account the nature and relevance of the work for which the application is made, the status of the person or body submitting the application and the guarantees it offers. It checks that the volume of information requested is not excessive in relation to the work justifying its communication and that this does not lead to excessive prejudice to the interests that the Act [No. 51-711 of 7 June 1951 as amended on the obligation, coordination and secrecy of statistics] was intended to protect."

All applications must therefore include a justification for the use of each source with regard to the research project's objectives.

Once the data producers' consent has been obtained, the CSS then decides whether to authorise access to the data and the material conditions for doing so.

FPR (Production and Research Files)

For some sources, data files are available which have less detail than confidential data sources as information is aggregated for a few potentially re-identifying variables. The procedure for accessing these data is covered by a simplified version of statistical confidentiality.

These data are accessible in the form of the so-called «production and research» files (FPR). FPR sources offer an intermediate level of information - between standard anonymised (and more aggregated) data and more detailed confidential data. A fairly extensive list of the available FPRs can be found as an appendix to the Opinion of the Statistical Confidentiality Committee dated December 14th 2018. It is an ongoing evolving document..

To access FPR sources for the first time, users must sign a confidentiality agreement before applying for authorisation from the Statistical Confidentiality Committee in the context of the Opinion dated December 14th 2018. To be authorised, a user must be a researcher, academic, PhD student or student working for a research or teaching unit listed in Appendix 2 of the CSS Opinion. However, it is also possible to apply even if the user's home institution is not yet listed in Appendix 2. In this case, the Quetelet PROGEDO Diffusion department in charge of dissemination to researchers shall submit a request to the Statistical Confidentiality Committee to add the institution to the list. The Committee then decides whether the proposed organization really is a research institution.

Once authorisation has been obtained, users can request access to all FPR sources listed in Appendix 1. Individual authorisations are personally assigned to a given user. They remain valid for all subsequent data requests for an unlimited period of time provided the user remains part of an authorised research or teaching institution.

Researchers should apply for access to these FPR data on the Quetelet Progedo diffusion website. The procedure is specified on the Adisp - PROGEDO website.

Teaching quantitative methods (source : Adisp website)

The dissemination of data in the context of teaching quantitative methods is only authorised to teachers and their students carrying out research into a given issue which will result in a dissertation (or other type of written report).

To speed up the process of student access to FPR data, teachers can anticipate requests for authorisation by providing the list of interested parties so their authorisation procedure can begin more rapidly (for example in June or July for a course scheduled to start in September or October).This preliminary request does not replace the applications that students will then have to make individually (even for group work) on the ordering portal, but it helps them to be processed faster.

Teachers are not authorised to redistribute any databases sent to them by Adisp.

In all other cases, particularly teaching on statistics or statistical analysis software, Adisp is not authorised to disseminate data. There are online databases - notably on the INSEE website - that can be used for this purpose (employment surveys, annual declarations of social data (DADS), civil registry, surveys of permanent living conditions (EPCV), etc.).

APPENDIX 4 :

FOCUS : Research in the field of health

Health data are broadly defined as relating to past, present or future physical or moral health, giving indications as to a person's current state of health.

Your research is not in the field of healthcare

However, it does involve the use of health data. For example, this may include projects in which it is important for researchers to consult certain health data (pregnancy, disability, chronic illness, etc.) or to know if participants have specific learning, sensory, motor, etc. disorders that could have an impact on their experiments. In this case, the procedure is the same as that for so-called sensitive data *(voir page 10).*

In short :

- Either consent can be obtained from the persons concerned (care should be taken to respect the correct formalities): an entry must be made in the register,

- Or the data have been made public by the persons concerned: an entry must be made in the register,

- Or a request for an authorisation from the CNIL is required.

Your research is in the field of healthcare

In this case, you must first decide whether you intend to carry out research involving the human person (RIPH) as defined by the Jardé law (law n° 2012-300 of March 5th 2012): research organised and carried out on human beings aimed at developing biological or medical knowledge (article L. 1121-1 of the amended public health code).

Your research is not an RIPH in the terms of the Jardé law

Mainly it is in a category of research involving the collection of pre-existing «retrospective» data (medical records, image banks, medico-administrative databases, etc.) and particularly concerns the evaluation of health practices and care. In all cases, it is advisable to carry out a privacy impact assessment (PIA) or data protection impact assessment (DPIA) with your unit's Data Protection Officer. This will enable a processing operation to be defined that respects privacy when the processing of personal data may lead to a high risk for the rights and freedom of the persons concerned.

If the research project does not comply with the CNIL Reference Methodology MR-004

«The reference methodology MR-004 provides a framework for processing personal data for the purpose of studies, evaluations or research not involving the human person. More specifically, this concerns studies that do not meet the definition of research involving the human person, particularly studies of the re-use of data. The research must be of public interest. The processing controller undertakes to only collect data that are strictly necessary and relevant to the research objectives» (Deliberation No. 2018-155 dated May 3rd 2018)»

A compliance commitment made to the CNIL is required. This commitment and the list of processing operations implemented in the framework of the reference methodology must be recorded in the register.

If the research project does not comply with the CNIL Reference Methodology MR-004

In this case, an opinion must be sought from CEREES (National Expert Committee for Health Research, Studies and Evaluations) via the INDS (National Institute for Health Data) and then an application for authorisation must be submitted to the CNIL. The processing worked will be recorded in the processing register.

Your research is an RIPH in the terms of the Jardé law

In all cases of this type of research, the person managing the study must request a favourable opinion from a Committee for the Protection of Individuals (CPP). A DPIA must be carried out by the unit's Data Protection Officer. Also, a distinction must be made between three categories of research involving the human person.

1. Interventional research (French acronym: RI

This research involves an intervention on the human person that is not justified by the person's usual healthcare (Article L.1121-1, paragraph 1). In addition to the opinion of the CPP, an authorisation from the ANSM (Agency for the Safety of Health Products) is required. It is compulsory to fully inform the person concerned. Their free informed consent to participate in the research must be obtained in writing. For example: double-blind research aimed at comparing the efficacy of a new drug treatment against a placebo carried out on a population group suffering from a particular pathology; research involving specific experimental tasks aimed at studying the pathophysiology of a specific disease for which patients must temporarily stop taking their usual medical treatment; the estimation of the effect of new physiotherapeutic treatment on patients with locomotion problems with longitudinal monitoring, etc.

2. Low-risk, low-burden interventional research (French acronym: RICRM)

This research is on the list set out in law by the Minister for Health (Article L.1121-1, paragraph 2). It is compulsory to fully inform the persons concerned and their free informed consent to participate in the research must be obtained. For example: research included in the list established by the order dated April 12th 2018 mentioned in 2° of Article L. 1121-1 of the Public Health Code which uses referenced methodologies (for example, magnetic resonance imaging, external mechanical, electrical or magnetic stimulation, psychotherapy and cognitive-behavioural therapy techniques, etc.), research concerning the evolution of a target pathology via regular and non-habitual examinations which are part of normal healthcare monitoring, etc.

In the case of interventional research and low-risk, low-burden interventional research,

Si la recherche est conforme à la Méthodologie de Référence MR-001

If the research project complies with the CNIL Reference Methodology MR-001:

«The reference methodology MR-001 provides a framework for processing work involving health data which are of public interest and are carried out in the context of research requiring the consent of the data subjects or of their legal representatives. The consent referred to in the reference methodology specifically relates to the participation of patients in the research and/or their agreement for an examination of genetic characteristics to be carried out rather than referring to the legal basis of the processing as set out in the GDPR. More specifically, it concerns interventional research, including research with minimal risks and constraints, clinical trials of medicinal products and research requiring the examination of genetic characteristics. Individual information to patients is mandatory. The processing controller undertakes to only collect data that are strictly necessary and relevant to the research objectives» (Deliberation No. 2018-153 of 3 May 2018)

A compliance commitment made to the CNIL is required. This commitment will be recorded in the register and the list of processing operations implemented in the framework of the reference methodology must also be recorded in the processing controller's register.

If the research does not comply with MR-001

In this case, an authorisation must be requested from the CNIL.

3. Non-interventional research (French acronym: RNI)

This type of research involves no risk or burden, with all acts and products being applied in the usual manner (Article L.1121-1, paragraph 3). This research is often observational. It is compulsory and sufficient to inform the persons concerned. For example: research included in the list established by the order dated April 12th 2018, mentioned in 3° of Article L. 1121-1 of the Public Health Code which uses referenced methodologies (e.g. audio, video and photographic recordings apart from any medical imaging; collection of electrophysiological data from implanted equipment or equipment being implanted for healthcare; anthropometric measurements which do not require invasive intervention; interviews, observations, tests and questionnaires that endanger the safety of the persons involved or lead to changes in their usual healthcare and for which the constraints and inconveniences to the subjects of the research are negligible, etc.), research into the evolution of a target pathology via regular examinations as part of normal healthcare monitoring, etc.

Dans le cas des RNI,

If the research project complies with the CNIL Reference Methodology MR-003:

«The reference methodology MR-003 provides a framework for processing operations involving health data and of public interest, carried out in the context of research involving the human person where subjects do not object to taking part in the research after being fully informed. More specifically, this concerns non-interventional research and cluster clinical trials of medicinal products. Individual information to patients is mandatory. The processing controller undertakes to only collect data that are strictly necessary and relevant to the research objectives (deliberation No. 2018-154 of May 3rd 2018)

A compliance commitment made to the CNIL is required. This commitment will be recorded in the register and the list of processing operations implemented in the framework of the reference methodology must also be recorded in the processing controller's register.

If the research does not comply with MR-003

In this case, an authorisation must be requested from the CNIL.

ANNEXE 5

Key issues to achieve compliance with legislation on the protection of personal data

Situation 1 :

The data used for your research will be (irreversibly non-identifying) anonymous data	The laws and regulations on personal data protection need not apply

Situation 2 :

The daata used for research are personal data : application of the GDPR with ajustments when the processing operations are for research purposes

		See Pages
Who is responsible?	• The data controller • The processor (subcontractor) • Partners, processing joint controllers	12, 13, 14
What kind of data?	• Non-sensitive data • Sensitive data • Social security number • Data on offences and convictions • Data on vulnerable populations	11, 22, 23
How re the cata collected, who are the recipients?	• Direct collection • Indirect collection	20, 22
What is the purpose of the processing?		15
Do the data correspond to the purpose and are they sufficient for the project?	Principles relating to data processing	15, 24
How long are the data to be stored?		27
How are people informed? What rights do they have?		17
What provisions are required to ensure data confidentiality and security?	• Data hosting • Storage • File/folder sharing	25, 27, 28

Is a privacy impact assessment necessary?	• The CNIL's open source PIA software	17
What steps should be taken? Who should they be carried out with?	Registration in the processing records • The Data Protection Officer • The CNIL	13 14
Are there any mechanisms, methodological aids, services available for the scientific community?	• Anonymisation and pseudonomysation techniques • Data hosting and archiving services	26
What data can I publish?		28
Can the data be reused?		29

ANNEXE 6

LISTE DES SIGLES

ANSM: Agence Nationale de Sécurité du médicament et des produits de santé (French Agency for the Safety of Health Products)

ANR: Agence Nationale de la Recherche (French National Research Agency)

ANRT: Association Nationale Recherche Technologie (National Research and Technology Association)

CADA: Commission d'Accès aux Documents Administratifs (Commission on the Access to Administrative Documents)

CASD Centre d'Accès Sécurisé aux Données (Centre for Secure Access to Data)

CDAP: Portail d'accès à des données confidentielles (Confidential Data Access Portal)

CEREES: Comité d'Expertises pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (National Expert Committee for Health Research, Studies and Evaluations)

CINES: Centre Informatique National de l'Enseignement Supérieur (National Computing Center for Higher Education)

CoSo: Comité pour la Science ouverte (Committee for Open Science)

CNIL: Commission Nationale de l'Informatique et des Libertés (National Commission on Informatics and Liberty)

CPP: Comité de Protection des Personnes (Ethical Research Committees)

CNRS: Centre National de la Recherche Scientifique (National Center for Scientific Research)

CSS: Comité du secret statistique (Statistical Confidentiality Committee)

DPD: Délégué à la Protection des Données (Data Protection Officer, DPO in this document)

DMP: Plan de gestion des données (Data Management Plan)

ESR: Enseignement Supérieur recherche (Higher Education and Research)

EU: European Union

FPR: Fichiers de « Production Recherche » (Production and Research Files)

G29: Article 29 Working Party which has now become the European Data Protection Board (EDPB)

GDPR: European General Data Protection Regulation

HSS: Humanities and Social Sciences

INDS: Institut National des Données de Santé (National Institute for Health Data)

INIST: Institut de l'Information Scientifique et Technique (Institute for Scientific and Technical Information)

INRAE: Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (National Research Institute for Agriculture, Food and Environment)

INSHS: Institut des Sciences Humaines et Sociales du CNRS (CNRS Institute for Humanities and Social Sciences)

MR: Research Methodology

NIR: National Membership Registry Number, commonly referred to as a Social Security number. This is a unique 13-digit identification number. The INSEE (France's National Institute of Statistics and Economic Studies) automatically assigns one to each person born in metropolitan France and in France's overseas departments (DOM). This number appears on the French health insurance card ("Carte Vitale") and is used to manage the holder's access to Social Security and medical services

OPIDOR: Optimisation du Partage et de l'Interopérabilité des Données de la Recherche (a support portal set up and hosted by the Inist-CNRS for research data sharing optimisation and interoperability)
Programme H2020 : Horizon 2020 research framework programme (EU)

PUD: Plateformes universitaires des Données (University Data Platforms)

TGIR: Très Grande Infrastructure de Recherche (Very Large Research Infrastructure)

TGIR Huma-Num : TGIR des Humanités Numériques (Very Large Research Infrastructure for Digital Humanities)

TGIR Progedo : TGIR Production et Gestion des Données en Sciences Sociales (Very Large Research Infrastructure for Social Science Data Production And Management)

HUMANITIES &
SOCIAL SCIENCES

3, rue Michel-Ange
75794 Paris Cedex 16
inshs.com@cnrs.fr | @cnrsshs.bsky.social