



White Paper

**Open Science
in a
Digital Republic**

March 2016



dépasser les frontières

Scientific and Technical Information Department



White Paper

Open Science

in a

Digital Republic

Scientific and Technical Information Department

CNRS

Assistance, Expertise, Consulting: Cabinet ALAIN BENSOUSSAN

CONTENTS

Preface	6
Summary and proposals	12
I/ Summary	12
II/ Findings	16
III/ Main recommendations	17
Commented plan	19
1. OVERVIEW: “SCIENCE IN TRANSITION”	22
1.1 Snapshot of the uses of Science	22
1.1.1 Science at the heart of the digital transition.....	22
1.1.2 New uses by researchers.....	24
1.1.3 Pressing demand from laboratories, institutes, agencies	35
1.2 The legal vacuums	39
1.2.1 Lack of a legal framework for science	39
1.2.2 Law on platforms: developments in progress	39
1.2.3 A right to TDM: an absence with serious consequences.....	40
1.2.4 The need for reformed rights in scientific digital publishing	44
1.3 The risks of misappropriation	52
1.3.1 Appropriation by economic uses	52
1.3.2 Appropriation by the uses of scientific publishing	52
1.3.3 Appropriation by the contracts.....	53
1.4 Massive validation of these findings	57
1.4.1 Summary of the national consultation	57
1.4.2 The opinion of the French Digital Council.....	61
1.4.3 The impact assessment for the Bill	63
1.4.4 The text adopted by the National Assembly	65
2. THE FUTURE: OPEN DIGITAL SCIENCE	67
2.1 Personal testimonies recorded for the White Paper: converging principles for an approach to Open Science	68
2.1.1 A shared value: science, a "common good" of humanity.....	68

White Paper - Open Science in a Digital Republic

2.1.2	Science, driving the economy	70
2.1.3	Priority for Open Access and the sharing of scientific data	71
2.1.4	The numerical assessment of peer-reviewed results and publication metrics.....	79
2.1.5	Publication and embargo periods.....	82
2.1.6	Analysis and exploration of corpora of data	85
2.1.7	Digital Intellectual Property and recognition of authorship.....	88
2.1.8	The limits of exploitation and Open Science	89
2.1.9	Towards an ethical charter for digital scientific and technical information (STI)	89
2.1.10	Toward a radical change of paradigm?	90
	In sum	92
2.2	Open Science around the world	93
2.2.1	The European Union clearly in favour of Open Science	93
2.2.2	Abroad: creating a legal framework for Open Science	95
2.2.3	International bodies: the tendency is in favour of Open Science	97
2.3	The key concepts of Open Science.....	99
2.3.1	A right to Openness: an unstoppable international movement toward openness and data sharing.....	99
2.3.2	The French Research Code: the basis of digital STI legislation	103
2.3.3	A need for consistency with rights governing public data.....	104
2.3.4	An indispensable legal principle: literary and artistic copyright	105
2.3.5	Exceptions to be protected: the public interest and legal secrets	107
2.3.6	The protection of privacy and personal data.....	108
2.3.7	Exploitation: a legitimate interest to be preserved	108
2.4	The gap between current practice and the law	110
2.5	The amendment of Article 17 of the Digital Republic Bill.....	114
2.5.1	Guiding principles.....	114
2.5.2	Proposal for amendments to the text.....	116
2.6	Proposals to be firmed up, paths for the future	118
2.6.1	Positive right to text and data mining	118
2.6.2	Defining a legal framework to underpin Open Science: asserting certain values....	121
2.6.3	Reference guidelines for the different practices regarding digital STI.....	124
2.6.4	An ISO standard, or an AFNOR standard.....	125

2.6.5	Definition of model contracts for the transfer of copyright: how to protect researchers	127
2.6.6	An ethics charter for digital science	127
2.6.7	An Agency for Open Science	128
2.6.8	An international convention for universal Open Science	128
2.7	Summary of proposals for Open Science	129
2.7.1	Strategic opinions	129
2.7.2	Proposals	133
	Acknowledgements	134
	Glossary	135
	Annexes	140
1.	Presentation of the White Paper	140
2.	Contribution of the CNRS Scientific Board	161
3.	Recommendation of the CNRS Scientific Board	165
4.	Opinion of the Ethics Committee of 7 May 2015 "The ethical issues of scientific data sharing"	167
5.	Interview guide for the hearings – the contribution of research to the themes of the Digital Republic Bill	182
6.	Minutes from the hearings	184
a.	UPMC, Jean Chambaz and Paul Indelicato, 9 June 2016.....	184
b.	The CNRS Scientific Board: Bruno Chaudret, Claire Lemerrier, François Bonnardel, François Tronche, 30 June 2015.....	190
c.	University of Strasbourg, Paul-Antoine Hervieux and Françoise Curtit, 10 July 2015	203
d.	French Digital Council: Benoit Thieulin and Yann Bonnet, 1 September 2015.....	211
e.	ISTEX Committee, Laurent Schmitt, Jean-Marie Pierrel and Grégory Colcanap, 24 September 2015.....	218
f.	ABES hearing: Jérôme Kalfon, 5 October 2015	223
g.	CCSD, Claude Kirchner, 15 October 2015.....	229
7.	Table summarising foreign legislation: Open access and TDM	236
	Credits	239

Preface

The national consultation on the Digital Republic Bill, launched by the Prime Minister, prompted an unprecedented debate on the preparation of a major government text. Among the many original results from this consultation, it is worth noting the importance attached by the debaters (researchers, publishers, scientific institutions or research groupings) to the theme specifically covered by this White Paper:

How should scientific information in the digital age be characterised, exploited and shared?

Admittedly, France is not the first country to ask itself the question: an abundance of solutions have been offered, both inside and outside the European arena. This White Paper intends to innovate by providing an overall study, based on possible solutions developed in several major fields of scientific practice. This work takes place alongside the recent efforts prompted by the Academy of Science and its Permanent Secretary, Jean-François Bach, to conduct an in-depth review of the foundations of access to scientific evaluation and publication. The aim of this wide-ranging study is to inform the choices which, in the new digital practices adopted in science as in other fields, must inevitably take the international dimension into account, all the more so in view of the impact of and issues raised by information systems that now stretch right round the planet, thanks to the Internet.

The challenges of Open Science: international convergence is necessary

The issue of Open Science, specifically, is neither a distant vision nor a slogan: Open Science, as currently accepted by Carlos Moedas, European Commissioner for Research, Science and Innovation, who in 2015 made it one of the European Commission's three priorities, is an approach that has to simultaneously reconcile new digital uses, new ways to exploit these uses (intellectual property in particular), and new procedures and rights guaranteeing access to and sharing of scientific results. On these themes, North America especially is actively pursuing the construction of new bases that make the field of scientific research particularly competitive at an international level.

Each country is therefore currently seeking its own solutions, which is only logical, even though we should all ultimately be seeking convergence: scientific publishing practices, whether digital or "paper", depend on national publishing ecosystems, historically well established and constructed, with their own legal and contractual structures, and their acquired rights. These ecosystems are themselves dependent on the influence of

international scientific publishing in each of the contexts: the major science publishers¹ each take a slightly different approach to their national scientific publication systems, through a wide range of contractual practices.

In the Digital Republic Bill, our country will therefore seek to define the positioning called for by its history and its scientific publishing sector, in light of the needs of its research communities.

This White Paper will confine itself to addressing only the need expressed by the research communities. It should be emphasised that these parties expressed their views on several major converging themes during the consultation on the Bill². Axelle Lemaire, Secretary of State in charge of the Bill, has acknowledged this³.

A public research approach for a national dialogue

By deliberately adopting this single point of view confined to the needs of "end producers" and "direct users" of the scientific information, this White Paper is obviously open to criticisms about the deliberate restriction of its horizon of concern: indeed, would it not have been logical to involve all the stakeholders in the production and sharing of scientific information more broadly? While they are aware of the limitations imposed by this choice, the authors of this White Paper have also weighed the benefits in the current context: the debate with publishers and all providers of digital scientific information services has clearly still failed even to approach a consensus at any of the levels – whether national, European or international – at which it is currently being conducted, while the actual options considered by this debate continue to evolve.

By adopting a scope of analysis confined solely to internal research uses, the authors of this White Paper believe they can offer sufficient "added value" to contribute to a debate that exceeds the scope of the Bill itself. As a reminder, this White Paper was initiated before the debate on the Digital Republic Bill was launched, and many of its themes exceed the scope of the legislative debate, with a view to providing benchmarks in an overall framework of reflection on the needs for digital information in science. This approach is widely inspired by the debate initiated at and by INIST, alongside its director, Raymond Bérard: in the digital age, this approach updates for the 21st century one of the key ideas set out by Diderot and d'Alembert starting from the Prospectus of 1750, the

¹ A recent study by Livres-Hebdo (26 June 2015) shows that of the world's 12 leading publishers in 2014, the top four, which all have a turnover in excess of 3.5 billion euros, are professional publishing groups, largely rooted in science and its promotion.

² A detailed summary of all contributions is accessible from: http://www.cnrs.fr/dist/z-outils/documents/Projet%20de%20loi%20-%20analyseCNRS_DIST.pdf

³ Axelle Lemaire, summary of the consultation on the Bill.

famous work programme for the *Encyclopédie*, which aimed, literally, to find the representations and the means to "make science work".

In this regard, two key observations serve as the guiding thread to the two phases of this White Paper.

Firstly, the review of the current situation on uses suggests a **pressing need to catch up**, where today, the digital uses of science currently lag behind the major emerging and/or established practices in the leading countries of science. A second key idea concerns the direction of the changes under way: **moving towards a "right of shared resources and protected uses"** might broadly summarise the proposals and recommendations of the White Paper. We can clearly see here a digital perspective of the proposals made in the report "*Pour une société apprenante*" ["For a society that learns"]⁴ presented to the President of the Republic in September 2015.

The authors of this collegiate approach also wish to stress that they attempted to share and debate with all the actors contributing their services and know-how to the work of science: the needs and constraints of researchers have been addressed with a view to leaving open the question of the economic and legal models by which these needs could be met in the future. In fact, rather than entering the arena of a debate whose scope is clearly far from settled, whether about current concerns regarding models of convergence between the needs and uses of scientific information, or the means to satisfy them, the authors of this White Paper hope that their contribution will feed usefully into the indispensable debate with all the present and future suppliers of science.

In taking this option, the authors of this White Paper affirm their commitment to a dialogue at national, European and international level, in particular with all the representatives of scientific publishing and providers of services to science. Only this dialogue can give meaning to these proposals and thoughts.

It should also be recognised that the aim of this analysis and deliberation is not to appropriate and reflect the opinion of the national scientific community as a whole, but to provide an initial constructed contribution to a national debate that is just getting started, and in which individual positions are not yet being adequately identified or substantiated.

⁴ Report presented to François Hollande, President of the Republic, in the presence of Najat Vallaud-Belkacem, Minister of National Education, Higher Education and Research, and Thierry Mandon, Secretary of State for Higher Education and Research, by Sophie Béjean, Chairwoman of the StraNES Committee and Bertrand Monthubert, General Rapporteur.

The lack of studies has been emphasised on several sides, and is summarised by the Opinion on the Bill by the Council of State⁵. On these aspects, CNRS's surveys of the Joint Research Units (UMRs) constitute isolated examples, and must be further developed, however thorough these have attempted to be⁶.

A White Paper: testimony, deliberations, expertise

Lastly, the authors of this White Paper take collective responsibility for the choices regarding the construction and presentation of the study that has brought them together, which is based on both third-party testimony and expertise. A detailed presentation of this approach is offered in Annex 1 of the White Paper.

The testimony is firstly that created by the ISTEEX project, the first national digital "Investment for the Future" programme for scientific documentation, whose instigators together examined the new uses of digital resources dedicated to scientific literature.

ISTEX offers a dramatic validation of the approach proposed by the Digital Scientific Library (BSN): the Ministry of National Education, Higher Education and Research (MENESR), the ABES, the Couperin Consortium, Lorraine University, and CNRS (DIST and INIST) are all stakeholders of this major project. This White Paper originated in the context of the ISTEEX Seminar, chaired in 2014 by Jean-Pierre Finance, representative of the Conference of University Presidents (CPU) and initiator of ISTEEX, and aimed to lay the foundations for a legal debate on the fundamental reorientations required for science in the age of digital practices.

This White Paper also came about thanks to the deliberations and testimonies of two major university members of the League of European Research Universities (LERU): the Pierre and Marie Curie University (UPMC) and the University of Strasbourg, represented by their respective Presidents and Vice-Presidents: Jean Chambaz and Paul Indelicato, for the first, and Alain Beretz and Paul-Antoine Hervieux for the second.

The national Open Access system, the CCSD (Centre for Direct Scientific Communication), which is paving the way for new forms of knowledge sharing, expressed its view via its current president, Claude Kirchner.

⁵ By the publishers themselves through the French Publishers' Association (SNE), and more recently by the Opinion on the Bill by the Council of State

⁶ <http://www.cnrs.fr/dist/z-outils/documents/Enqu%C3%AAte%20DU%20-%20DIST%20mars%202015.pdf>

White Paper - Open Science in a Digital Republic

CNRS, which as early as 2013 promoted a national strategy based on the pooling of scientific information⁷ needs and resources, expressed its view here via the Chairman of the Scientific Board, Bruno Chaudret, after nearly a year of internal debate by the Scientific Board. Lastly, Benoit Thieulin, President of the French Digital Council, made a valuable contribution to this overall process of reflection, by establishing some coherence between the themes of science and all those issues for which he is responsible in the organisation of a national transition to digital uses across our country.

The expertise provided by the international consulting firm Cabinet Alain Bensoussan has given an indispensable legal underpinning to this work, through its analysis of the texts and options offered by the emerging law of digital technologies applied to science, at national, European and international level. The Cabinet Alain Bensoussan has brought this project to life by organising all the hearings and the construction of this White Paper, in phase with the expectations of all the actors involved in this process, and by bringing openness and a spirit of innovation.

Lastly, several key witnesses wanted to share their thoughts on two major areas central to key issues for scientific information in the digital age: Bruno David, President of the French Natural History Museum addressed the diversity of uses of scientific information and new ways of sharing it, and lastly Daniel Egret, former President of the Astronomical Observatory of Paris, described the new uses of scientific publication metrics.

CNRS orchestrated this approach through especially productive joint contributions, by its Scientific and Technical Information Department (DIST) and Office of Legal Affairs (DAJ).

The many testimonies offered by researchers and research managers, combined with the international legal expertise presented here, have enabled this White Paper to compile an outline of current information on the rights and uses of scientific digital publication: this responds to the need expressed by all actors, both professional and non-professional, for objective information on the changes under way introduced by digital technology.

May this collection of testimonies, thoughts and expertise contribute to the common debate and enable France to make the necessary choices for it to take its place in the great 21st century project of digital scientific creation, which is now under way.

The signatories:

⁷ <http://www.cnrs.fr/dist/strategie-ist.htm>

For the members of the Executive Committee of the ISTEX Investments for the Future project:

Grégory Colcanap, Coordinator of the Couperin Consortium

Renaud Fabre, Director of the CNRS DIST

Jérôme Kalfon, Director of the ABES

Jean-Marie Pierrel, Professor at Lorraine University

Laurent Schmitt, Head of the Projects and Innovation Department, INIST CNRS

For the key witnesses:

Alain Beretz, President of the University of Strasbourg

Jean Chambaz, President of the UPMC

Bruno Chaudret, President of the Scientific Board of CNRS

Bruno David, President of the French Natural History Museum

Daniel Egret, Astronomer (Paris Science et Lettres), Former President of the Observatory of Paris

Claude Kirchner, Adviser to the President of INRIA, Senior Researcher

Benoît Thieulin, President of the French Digital Council

Summary and proposals

I/ Summary

"OPEN SCIENCE", NEW RIGHTS FOR DIGITAL USES

Main orientations:

- **Create:** Create a right to Open Science guaranteeing free access and free reuse of data from public research
- **Balance:** Redefine the economic balance of the digital science ecosystem
- **Secure:** Adopt Article 18 bis (new) of the the Digital Republic Bill creating an exception to copyright and the right of database producer in favour of text and data mining for data from public research (research articles and data) in order to secure automated data processing practices and reduce the risk of misappropriation
- **Compete:** Enable French public research to acquire legal and technical resources that are at least equivalent to those of its European and American counterparts, and in phase with the international Open Science movement
- **Protect:** Protect legitimate interests: exploitation, secrecy, patents, copyright, privacy and personal data.

What is Open Science?

Open Science is a new horizontal approach to access to scientific work and objectives, and to sharing of scientific results, as well as a new way of DOING science, by opening up its processes, codes and methods.

The Open Science project offers a *renaissance* of global "encyclopaedic" views, through such themes as the decompartmentalisation and large-scale sharing of knowledge: in the digital age, this concept stresses the "leveraging" of knowledge such as occurs, for

instance, through the in-depth exploration of digital databases containing scientific journals⁸.

Open Science is therefore a **change in perspective** that can be compared with other earlier major stages, such as the advent of the telescope or the microscope.

Open Science thus seeks to take into account the changes brought by the major "open" international platforms: they provide access to new arrangements for research actors (digital innovation, civil society), facilitated by new sharing approaches (digital laws), and leading to novel types of results and regulations (data and analysis platforms, scientific social networks, new forms of collaboration, etc.).

Open Science, a field that is far deeper than open access limited solely to publication, refers to all the different ways and means of enhancing scientific work offered by digital technologies.

An inevitable international movement

Open Science is part of an international movement towards greater openness. Many countries have already legislated in favour of open access, and of text and data mining. International and European governing bodies are advocating this step forward.

In June 2015, the European Commissioner for Research, Science and Innovation, Carlos Moedas, defined three priorities of action: "Open Innovation, **Open Science**, Openness to the World". Open Science is defined by Brussels as encompassing all the "transitions" that accompany digital changes in science⁹. In 2015, the OECD developed a similar approach with emphasis on the possible global breakthrough that Open Science¹⁰ could represent, conditional on concerted action.

In France: elements for a national Open Science strategy

The CNRS strategy "A better sharing of knowledge"¹¹ revealed the need to catch up with regard to digital practices of scientific publication on the different platforms.

⁸ These fertile explorations may be either synchronic, as on PubMed Central, <http://www.ncbi.nlm.nih.gov/pmc/> or diachronic, as in the example quoted by the NSF, http://www.nsf.gov/news/news_summ.jsp?cntn_id=135258

⁹ Open Science "describes the on-going transitions in the way research is performed, researchers collaborate, knowledge is shared, and science is organized. It is enabled by digital technologies".

¹⁰ OECD 2015: "*Open Science is more than open access to publications or data; it includes many aspects and stages of research processes.*"

¹¹ <http://www.cnrs.fr/dist/strategie-ist.htm>

These themes were also addressed by the Scientific Board of CNRS in its unanimous recommendation, as well as by the Ethics Committee.

Many voices have come out in favour of Open Science. The Government itself has taken an ambitious stance, particularly in its "Digital Strategy" of 18 June 2015, as well as in the explanatory statement for the Digital Republic Bill. The national consultation on the draft bill set science apart as a priority theme of the national consensus on digital technologies and as a theme where the views of researchers and institutions coincide, on the basis of simple principles:

- science is a common good of humanity;
- legitimate interests of protecting secrecy and exploitation should be preserved;
- text and data mining is a natural right of digital observation necessary to researchers in their scientific process;
- clauses on exclusive transfer of copyright laid down in publishing contracts should be declared null and void;
- it should be possible to freely exploit knowledge industrially or commercially in a consolidated ethical framework.

This White Paper proposes to step back and reflect on the uses of research results, in particular by means of key witnesses who wished to express themselves and whose opinions generally converged. These contemporary digital uses are presented in the context of the existing legal framework and the resulting conflicts are discussed.

The absence of legal antagonism

The analysis of the legal framework revealed that current French and European laws are no hindrance to the introduction of positive rights. The rules of law applicable to digital practices need to be updated in order to secure access to and use of scientific and technical information via online platforms.

In its current Articles 17 and 18bis (new) as adopted at first reading by the French National Assembly on 26 January 2016, the Digital Republic Bill proposes significant advances. The French text may go further in affirming the common values of the world of research and consolidating the right to conduct text and data mining.

The need for new rights for science

Several findings underlie the presentation of this White Paper:

1. **The need to update the rights on the uses of digital science in France:** these uses should be fully in line with all those that have been adopted, in particular by

our larger neighbours and by Europe. This is not the case today and this updating is necessary.

2. **The diversity of scientific publishing ecosystems and the digital practices that accompany them:** our country is specific in this regard and adaptations must be found, in France as elsewhere, to enable scientific publishers to work alongside all the scientific communities.
3. **The dominant direction is the development of digital Open Science:** this direction is developing in all the major countries today, offering terms for open access and open process for practices and according to models that are still far from stable.
4. The lines of work for new rights on uses are the subject of **recommendations** detailed in this White Paper, and in particular the amendment of Article 17 of the Digital Republic Bill in its draft version resulting from the adoption of the text at first reading by the French National Assembly (26 January 2016).

May the national representation and our society in general take full advantage of this prospective and expertise study on the digital uses of science: this work is now available to all those with a stake in the future sharing of knowledge that France intends to define.

II/ Findings

Listed below are the findings that emerged from the hearings with the representatives of the research communities and researchers.

FINDINGS

The multiplication of platforms and the weakness of their contractual framework has generated a need for new governance.

The multiplicity of STI objects calls for a clarification of the law and a balance between access to scientific knowledge and preservation of the potential for STI exploitation in all its components.

The absence of a legal status for data exploration is a source of legal insecurity for which the law must provide.

The absence of a legal status for data exploration and the unsuitability of the right of database creators to the dynamic processing of knowledge are sources of legal insecurity for which the law must provide.

Publishing contracts signed by researchers for articles they wrote in the framework of publicly funded research largely provide for exclusive transfers of rights in favour of the publisher. These constitute one-sided standard form contracts.

To carry out their work, researchers need open and free access to all scientific data in digital form, consisting of:

- scientific results, including the results published by a scientific publisher;
- research data in the sense of the data used to establish these results.

Researchers have expressed the need to share scientific data.

The practice of depositing articles in archives or on platforms in specific fields should be generalised.

Researchers have expressed the need for:

- a "one-stop-shop" for scientific knowledge;
- legal regulation of the platforms.

Researchers need access to the latest state of knowledge. If an embargo period can be defined as part of a compromise with the publishers, it must not exceed the maximum time limits provided for in the Recommendation of the European Commission (C(2012)

4890) (6 and 12 months depending on the category of discipline) and the time limits observed in other countries, as otherwise French research runs the risk of marginalisation and discrimination. The principle of a distinction between the exact sciences and the human and social sciences has been challenged.

The provisions ensuing from Act No. 2015-1779 cannot be used to adapt the public provision of data produced by the education and research establishments and institutions.

These provisions are not in line with the needs of researchers and the uses of scientific communities, and do not take into account the nature of the data (data from ongoing research, know-how, a restricted regime area, etc.).

Scientific texts and publications can be protected by copyright if they are original in their form of expression. Researchers own the copyright over their scientific articles and texts.

Open Science must preserve secrets as well as public safety.

The French Research Code already contains in its principles the foundations of a digital law for Open Science.

III/ Main recommendations

Listed below are a series of proposals to be discussed with all parties, with a view to optimising the uses of digital scientific information.

- 1 Adoption of Article 17 of the Digital Republic Bill (adopted text no. 663)**
- 2 Adoption of Article 18 bis (new) of the the Digital Republic Bill creating an exception to copyright and the right of database producer in favour of text and data mining**
- 3 Participation in the creation of a European process of "Open Science" (a priority of the Commissioner for Research, Science and Innovation for 2016)**
- 4 Creation of reference guidelines on the use of digital STI**
- 5 Definition of a set of ISO standards on the uses of STI**
- 6 Definition of model contracts for the transfer of copyright**

- 7 Drafting of an ethical charter for digital science**
- 8 Creation of an Agency for the development of Open Science**
- 9 Creation of an international convention for Open Science**

Commented plan

1. Like many other types of data, scientific data are central to the changes in a system whose equilibrium has been upset by the advent of new digital tools and the affirmation of the values of openness, sharing and collaboration. These developments inevitably lead to questions, fears and concerns about shifts in power, and call for the legal provisions to be clarified and practices to be secured.

2. Faced with the observation of "a profound change in the production and dissemination of science (characterised by rapid growth in the number of scientific publications and, at the same time, the ever-increasing cost of access to these resources in the framework of financial resources that, at best, remain constant)", the stated objective of the "Government's Digital Strategy" as presented by the Prime Minister on 18 June 2015 is to "Foster Open Science by the free dissemination of research publications and data."

3. This White Paper proposes a two-part argument.

4. **The first part provides an overview of the current science situation** in the digital environment, by noting:

- the practices of researchers and their teams. These were mainly identified from a survey on STI uses and needs in research units, conducted with CNRS Unit Directors; this survey was carried out by the CNRS Scientific and Technical Information Department in mid-2014 among 1250 units publishing articles;
- the inhibiting legal embargo periods and the need for rights to be reformed;
- the risks of misappropriation.

5. Understood as a genuine working tool of scientific communities, the scientific data or information made available on digital platforms is subjected to use, exchange, manipulation and multiple processing operations that radically alter the traditional concepts of material and intellectual property rights.

6. When compared with foreign legislation, the existing legal framework prevents researchers from:

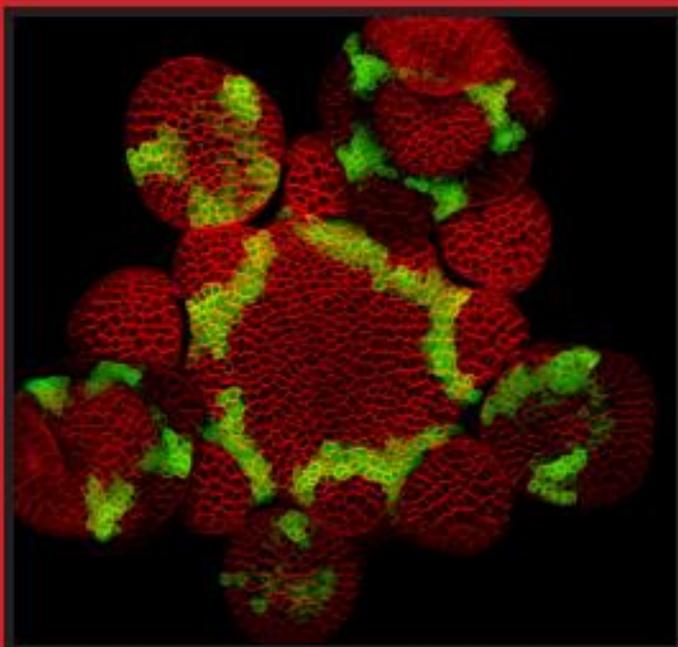
- gaining free access to scientific results;
- fully using the analysis or processing features offered by digital platforms.

7. These findings were also overwhelmingly validated in the framework of the national consultation on the Digital Republic Bill.

8. The second part formulates legal proposals in favour of Open Science, that the public authorities are invited to take up. These proposals result from:

- the values and the needs expressed by the key witnesses for this White Paper. A consensus approach is preferred, which contributes to the emergence and the sharing of values common to public research;
- an analysis of the positions adopted by EU and international bodies, as well as foreign legislation.

9. The Digital Republic Bill could embody France's ambition to promote the values of Open Science. This framework is an opportunity to include French public research in the global Open movement, while preserving the interests of all science stakeholders, whether they are the laboratories or public bodies funding the research project, the scientific publishers, the private partners, or the researchers themselves. Strategies for the development of a new law for digital scientific and technical information are proposed, with a view to rebalancing the digital science ecosystem.



1. L'Etat des lieux :
« La science en transition »

1. **OVERVIEW: “SCIENCE IN TRANSITION”**

10. Science is in transition towards a new system of rights and major changes in its practices.

11. The new digital STI practices followed by researchers and laboratories (1.1) are out of phase with the existing legal framework (1.2), which is lacking in some areas and has shortcomings in others.

12. Science is also grappling “with the new risk of appropriation of its data, mainly by scientific publishers demanding licence transfers for the datasets integrated with or associated with the research articles they publish”¹² (1.3). The explanatory statement for the Digital Republic Bill affirms these findings, which were confirmed in particular by the national consultation, by the French Digital Council, as well as by the impact assessment for the Bill (1.4).

1.1 **Snapshot of the uses of Science**

13. Its practices place science at the heart of the digital transition.

14. This section offers a snapshot of the practices in force in France and abroad concerning:

- the use of STI:
 - o by researchers according to the scientific community to which they belong, in particular the use of STI as a research tool;
 - o by laboratories, institutes and agencies dealing in STI, with regard to organisation of access to STI;
- scientific publishing contracts with regard to the conditions of publication, in particular in public/private partnerships.

1.1.1 **Science at the heart of the digital transition**

15. The digital transition is marked by an explosion in the quantity of data that are available and accessible at any time and any place in the world. This age of big data is characterised by:

- the absence of borders and the globalisation of information;
- a spirit of cooperation and sharing;

¹² Digital Republic Bill, Explanatory Statement, page 6

- the automation of certain activities, especially professional ones;
- the creation of value.

1.1.1.1 The absence of borders

16. The Internet is the flagship tool of the digital transition. This global network is notable for the absence of physical borders and entry visas, being instead characterised by cross-border flows and exchanges through the world of information.

17. Thanks to the IP protocol and the interconnection between the various operators, the Internet constitutes a seamless, end-to-end network.

18. Despite flaws and inequalities such as the digital divide or equipment with differential flow rates according to a country's public policies, this global network provides everyone with access to globalised information.

1.1.1.2 The right to cooperation and sharing

19. The digital transition is marked by the passage from paper to digital, in other words going from a single physical property to a multiplication of that property.

20. The Web 2.0 generation is characterised in particular by the ease of placing content online and the ability of users to interact; it is based on users themselves being able to generate, disseminate and consult multimedia content directly.

21. The principle of the collectivisation of content has fundamentally transformed uses. Sharing and a spirit of cooperation characterise the Web model. Its archetypes include YouTube, Wikipedia and social networks.

1.1.1.3 The right to automation

22. Numerous technological applications have resulted in authorities, research organisations and businesses holding large quantities of data and metadata.

23. Semantic and lexicographical analysis tools need to be developed in order to process these data. Dematerialised or automated work processes enable the creation of content by the user: user-generated content.

1.1.1.4 The right to creation of value

24. The digital transition is generating a new so-called digital economy. It is based on online trade as well as on non-commercial exchanges such as the sharing of knowledge and user-generated content (usage data, wikis, discussion forums, blogs, tweets, etc.).

25. The OECD's report of 12 April 2007¹³ had already highlighted the very significant economic impact of user-generated content, and the creation of the resulting value and innovation.

1.1.2 New uses by researchers

26. The CNRS survey and the hearings revealed the practices of researchers and institutes regarding the use of digital STI.

Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015

“Uses are changing and the paradigm for research data (BSN10 - research data¹⁴) is undergoing a transformation. Researchers and academics are beginning to realise the value of their data. Once we start talking about value, we need to start thinking about rights. We are now in a world where private businesses predominate and are increasingly interested in data with a view to commercialising its uses.”

27. The following findings emerged regarding the way researchers use STI as part of their research activity:

- STI is a working tool that can be used, shared and exploited freely as part of research work;
- databases in the form of open archives are used in a heterogeneous manner by the different communities;
- scientific communities lack sufficient knowledge of copyright and the *sui generis* right for databases to effectively manage the use of STI.

1.1.2.1 Digital STI: a tool for exploration and analysis

28. STI is primarily seen by researchers as a working instrument to which they must have access, which they can share, exchange, reuse or reprocess for the needs of their research subject. As a raw material, STI is part of the scientific process.

29. Scientific communities distinguish two main categories of data in STI:

¹³ OECD, Working Party on the Information Economy “Participative web: user-created content” 12-4-2007 <http://www.oecd.org/sti/38393115.pdf>

¹⁴ <http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche/>

- research data;
- publications.

30. Other categories of data are specific to certain communities:

- material: especially in biology;
- third-party data: in particular in the human and social sciences (HSS).

31. **Access to STI.** The contribution of the CNRS Scientific Board helped identify the main practices of researchers in the area of access to STI: access to publications, access to research data, and for some communities, access to the research material and to the third-party data used in the framework of their research.

Community	Access to STI
Human and social sciences	<p>Access to publications: “In the human and social sciences, with regard to recent scientific publications, while more and more French-language journals offer free access immediately or after a few years (mainly through HumaNum, BSN, OpenEdition), English-language journals are often confined to rather expensive platforms.”</p> <p>Access to numerical data or the data produced by research: “Platforms for exchanging numerical data (...) have also been established (Quetelet Network, DIMESHS, etc.): they provide better circulation of data, compliance with the necessary constraints, such as anonymisation, and the documentation (“metadata”) without which the figures would be unusable.”</p> <p>Access to third-party data: “The problem here is that some of the data used by scientists in HSS are not produced by them (this may concern a song, a company's annual report or the architecture of a monument): other natural or legal entities have rights over them.”</p>
Science of the Universe	<p>Access to research data: “The data are freely accessible in astronomy for the entire community following expiration of a proprietary period.”</p>

Community	Access to STI
Biology	<p>Access to publications: “Digital publishing is widespread. Academic institutions have developed platforms to help researchers find articles, and access to abstracts is free. The most important of these (PubMed) is managed by the US National Institutes of Health (NIH).”</p> <p>PubMed also provides access to an unformatted version of any article published commercially describing work funded by the NIH.</p> <p>Access to research data: “Many publishers, including Nature, also make publication of an article conditional on the depositing of mass data associated with a publication on a platform that is accessible to all, free of charge.”</p> <p>Access to material: “It should be noted that this requirement goes beyond digital data and also concerns material produced within the framework of a publication.”</p>
Physics	<p>Access to research data: “Free access to raw data is not yet very widespread.”</p> <p>“Many digital libraries have been formed and made freely accessible by groups of researchers.”</p>
Chemistry	<p>Access to publications: “The rule is still that papers appear in paid journals published by learned societies (American Chemical Society, Royal Society) or commercial companies (Wiley, Elsevier, etc.), with the timid development of 'gold' type open access, paid for by the authors.”</p>

Community	Access to STI
	<p>“Freely accessible databases are developing, especially the Cambridge Structural Database, which contains all the published molecular structures.”</p>
<p>Computing and Mathematics</p>	<p>Access to publications: “Databases relating to publications are very important for both individual and community work. A unique feature of this discipline is the importance of easy access to both recent and “old” (several years, decades or even centuries old) publications. Long-term access to these publications is therefore crucial for research.”</p>

32. Access to these categories of data is obtained by means of several technical tools:

- publisher portals;
- digitised archives;
- voluntary submission platforms;
- paper (question PAP 1-6 from the Shared Action Plan in the CNRS survey¹⁵, notes “the historical attachment of mathematics and the HSS to paper documentation resulting from the needs for/uses of old documents”).

33. **Processing of STI.** The scientific communities have different approaches to the sharing of data as well as to the techniques of text and data mining.

Community	Processing of STI
<p>Human and social sciences</p>	<p>“Data sharing and 'text and data mining' techniques are thus unevenly spread according to the types of data, mainly due to legal obstacles [third-party data], and a lack of human resources for the production and maintenance of quality metadata.”</p>

¹⁵ <http://www.cnrs.fr/dist/z-outils/documents/Enqu%C3%AAtes%20DU%20-%20DIST%20mars%202015.pdf>

Community	Processing of STI
Science of the Universe	“The formats, descriptions and modes of access to archive data, metadata and the applications likely to be used to process them should be harmonised and standardised, in order to achieve interoperability.”
Biology	“While text mining techniques are not a priority for most fields of biology as a discovery tool (but rather in terms of documentary collection), 'data mining' itself is playing an increasingly important role”.
Computing and Mathematics	“Furthermore, mathematics and computing play an important role in the analysis, management and exploitation of masses of data (questions around Big Data). It is certainly very important for the data to be accessible, but when they become more and more massive, it must also be possible to exploit them effectively.”

1.1.2.2 Sharing of knowledge: depositing of scientific results in open digital archives

34. **Disparate practices.** The CNRS survey and the contribution of its Scientific Board revealed different uses of open archives according to the scientific communities.

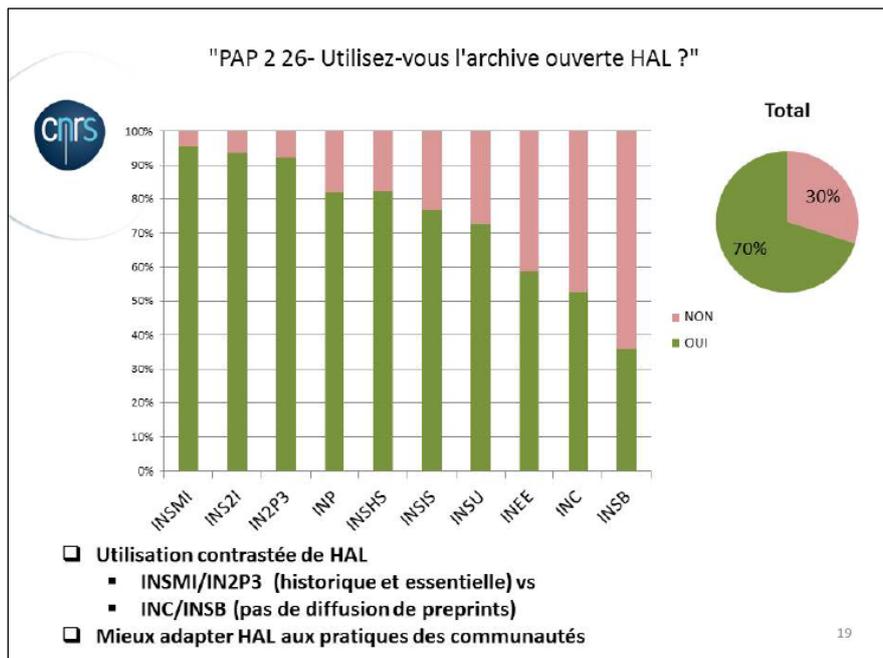
35. The HAL (Hyper Articles OnLine) multidisciplinary open archive is the one used most widely by the national community. Its purpose is the depositing and dissemination of scientific documents in any field of research; they have not necessarily been validated by an editorial board or programme committee. These documents come from researchers in French or foreign, public or private teaching and research institutions.

White Paper - Open Science in a Digital Republic

36. Question PAP 2-26 from the Shared Action Plan shows that different scientific communities use HAL in different ways. Researchers at the National Institute for Mathematical Sciences (INSMI) and those at the National Institute of Nuclear and Particle Physics (IN2P3) use HAL extensively; these institutes use it for historical purposes; while researchers from the Institute of Chemistry (INC) or the Institute of Biological Sciences (INSB) use it very little, and use other archives rarely or not at all.

37. Sixty-nine per cent of those communities not using HAL as an open archive do not use other archives either (PAP 2-29). Many of the publications are therefore merely archived by the publishers and are not freely available to the scientific communities after an embargo period. Access to publications is only possible via the publisher's platform, by subscription or the one-off purchase of the article.

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 2 26 – Utilisez-vous l’archive ouverte HAL?

PAP 2-26: Do you use the HAL open archive?

NON / OUI

NO / YES

Utilisation contrasté de HAL

Mixed use of HAL

- INSMI/IN2P3 (historique et essentielle) vs
- INC/INSB (pas de diffusion de preprints)

- INSMI/IN2P3 (historical and essential) versus
- INC/INSB (no dissemination of preprints)

Mieux adapter HAL aux pratiques des communautés

Better adapt HAL to the practices of the communities

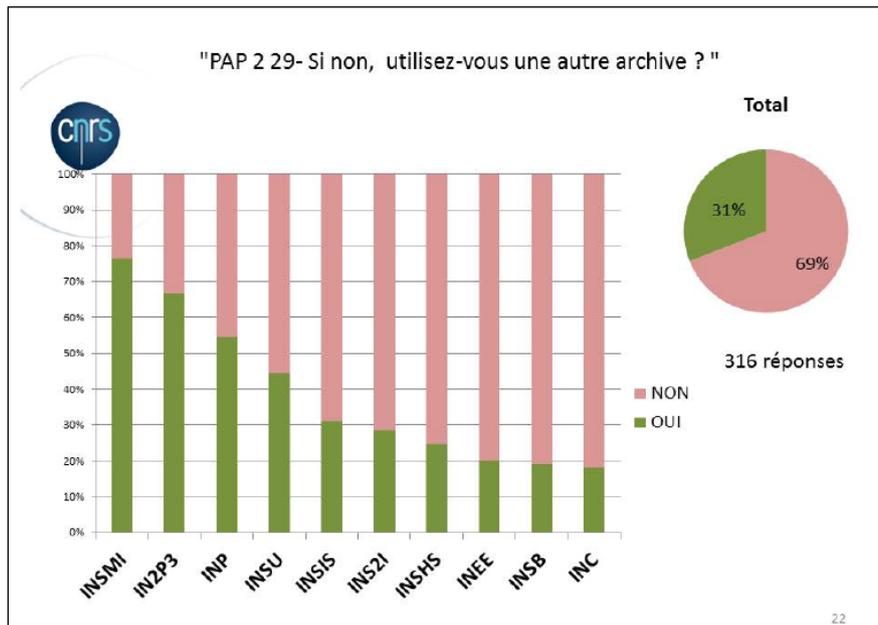


DIST - Direction de l'information scientifique et technique



ISTEX

White Paper - Open Science in a Digital Republic



PAP 2 29 – Si non, utilisez-vous une autre archive?

PAP 2-29: If not, do you use another archive?

316 réponses

316 responses

NON / OUI

NO / YES

38. The contributions of the Scientific Board provided some clarification on the practice of depositing in open archives:

Community	Submission practice
Human and social sciences	The “parallel depositing of articles in open archives is not greatly developed.”
Science of the Universe	Very widespread practice
Physics	Pre-publication server (preprint): “Digital technologies also play an essential role in the dissemination of results, with the almost systematic use of pre-publication servers. Articles are deposited on these servers at the same time as they are sent to a peer-reviewed scientific journal: it enables readers to take early notice, prior to publication.”

Community	Submission practice
Chemistry	"There is no pre-publication archive like ArXiv."
Computing and Mathematics	Archiving platform: "Publication archiving platforms such as HAL or ArXiv thus respond in part to this problem and should be supported, along with metadata platforms (MathSciNet , Zentralblatt , etc.)."

39. **Data deposited.** The data deposited by researchers in open archives differ from one community to another:

- preprint publications¹⁶ (article, book, chapter of a book);
- postprint publications¹⁷ (article, book, chapter of a book);
- thesis/dissertation (PhD, Master, etc.);
- lessons;
- images, videos, sounds, maps;
- bibliographic records only;
- bibliographic records accompanied by the publication;
- research data;
- metadata associated with these data.

40. The CNRS survey revealed that of the communities that deposit data in HAL:

- 63% deposit bibliographic records (PAP 2-27);
- 56% deposit the full text (PAP 2-28);

41. Sixty-nine per cent of the communities that do not deposit data in HAL do not use other archives either.

1.1.2.3 Rights on uses to be created from scratch

42. Publication, deposit in open archives, text and data mining, or operations relating to other data or text processing or mining techniques are carried out by researchers with an almost complete lack of awareness of intellectual property rights.

¹⁶ **Version submitted** ("Preprint"): version submitted by the author(s) to the journal before any process of revision (peer reviewing by publishers and referees).

¹⁷ **Accepted version** ("Postprint"): version after the author(s) have taken into account the remarks of the evaluators and the article is accepted by the editorial board.

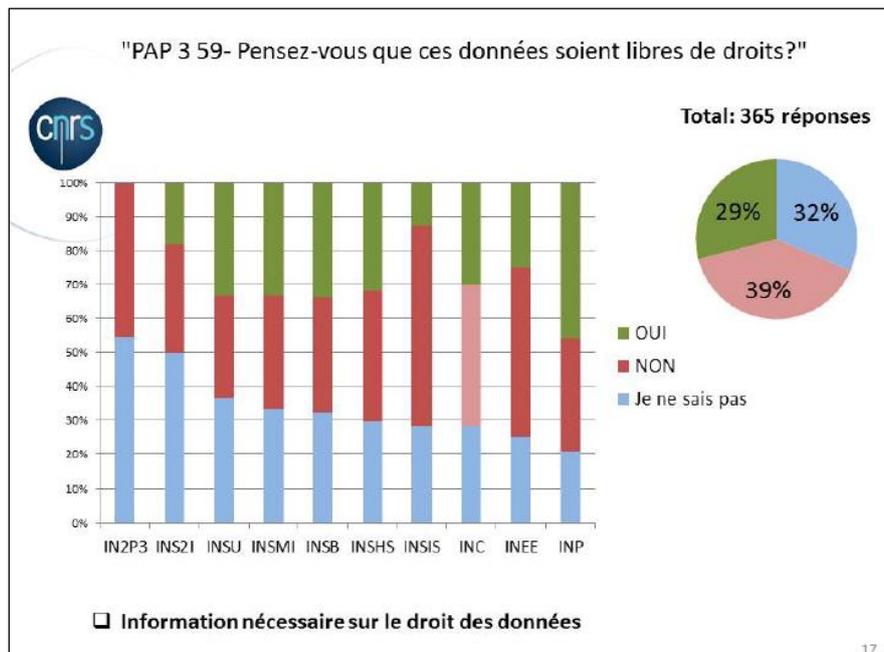
Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015

“In academic communities there is an almost across-the-board lack of awareness of copyright and its implementing measures, regardless of the type of document or data.”

43. In the framework of the CNRS survey, two questions were used to assess the researchers' level of knowledge about their rights.

44. To the question “Do you think that these data [the raw data] are copyright-free?”, the answers are divided between “Yes” at 29%, “No” at 39% and “I don't know” at 32%.

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 3 59- Pensez-vous que ces données soient libres de droits?

Total: 365 réponses

OUI / NON / Je ne sais pas

Information nécessaire sur le droit des données

PAP 3-59: Do you think that these data are copyright-free?

Total: 365 responses

YES / NO / I don't know

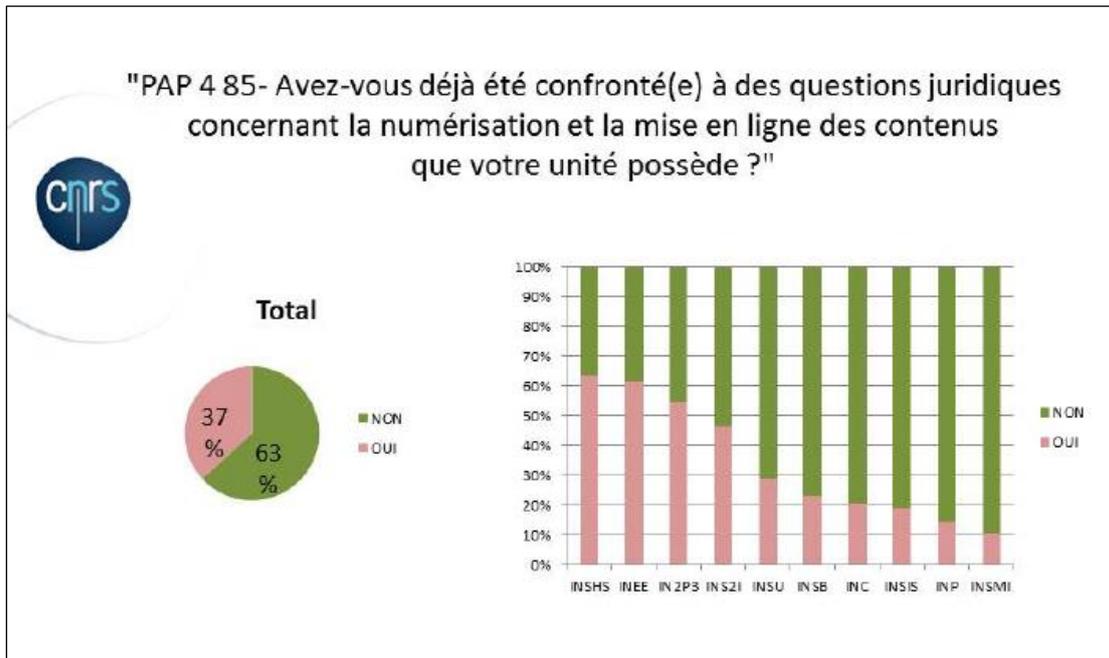
Information is needed on data rights

45. The breakdown of responses to this question in three equivalent groups shows that the researchers:

White Paper - Open Science in a Digital Republic

- cannot distinguish between what is copyright-free and what is protected by an intellectual property right;
- do not know their rights over the raw data and by extension the protected data (the publications).

46. The second question revealing their ignorance is the following:



PAP 4 85- Avez-vous déjà été confronté(e) à des questions juridiques concernant la numérisation et la mise en ligne des contenus que votre unité possède ?	PAP 4-85: Have you already been faced with legal issues about the digitisation and posting online of content owned by your unit?
NON / OUI	NO / YES

47. It can be seen that 63% of respondents have never been faced with legal issues concerning the digitisation and posting online of content. To be “faced with legal issues”, there must first be an awareness of the existence of problems raised by digitisation and posting online with regard to copyright.

48. This lack of awareness of rights concerning research data reveals that illegal practices can occur through ignorance, and indicates the need for clarification or even affirmation of the legal position with regard to these practices and the needs of researchers. Training initiatives will be necessary to support these desirable legislative developments.

1.1.3 Pressing demand from laboratories, institutes, agencies

49. The multiplication of platforms and the multiplicity of STI objects have generated a growing need for governance and definition of a legal framework for science in the digital age.



DIST - Direction de l'information scientifique et technique



ISTEX

White Paper - Open Science in a Digital Republic

1.1.3.1 Multiplication of platforms: a need for governance

50. **Multiplication of platforms.** A multiplication in the number of STI platforms has been observed in recent years:

- institutional platforms;
- thematic platforms;
- submission platforms;
- bibliographic platforms;
- archive platforms, etc.

51. This multiplication has resulted in a dilution of the information and a risk that it could lose its value. It also prevents cross-referenced information searches as well as multidisciplinary searches. Some platforms offer value-added services while others offer simple consultation of the texts, or simple access to the bibliographic data. This multiplication has a particularly high cost, whether regarding the quality of the resulting services or in human or financial terms.

52. **Weakness of the legal framework of the platforms.** These various platforms have more or less restrictive general terms and conditions of use with regard to the use of STI, which are not always in phase with intellectual property rights and the publishing contracts entered into between researchers and publishers.

53. The Digital Republic Bill proposes to introduce a definition of the concept of “platform” and to associate it with a duty to act in good faith. These provisions are introduced in the French Consumer Code but could be extended, as “good practices”, to science platforms.

+ The multiplication of the platforms and the weakness of their contractual framework has generated a need for new governance.

1.1.3.2 Multiplicity of STI objects (data, analyses, articles): a need for consistency for their exploitation

54. Exploiting the fruits of public research is an essential concern, and the filing of patents is at the centre of the exploitation process.

55. According to the 2014 SIR (Scimago Institutions Rankings) classification, CNRS is the world's leading filer of patents¹⁸ from scientific publications from institutions. Most of these patents are jointly owned with universities. This achievement is the result of cutting-edge research but also an incentivising exploitation policy. French research is generally active and produces many patentable innovations.

56. Public research has established a policy to incentivise the research units from which the inventions and their inventors come. Staff members thus receive a lump sum bonus for patents, whose amount is set by joint Order of the Ministers for the Budget, Public Service and Research. A coefficient reflecting their contribution to the invention is assigned to each staff member involved¹⁹.

57. Patents are not however the only way to exploit results, and may prove to be unsuitable for some innovations, for which the filing of a patent is impossible or inappropriate.

58. STI covers multiple elements: data, analyses, results, processed results, articles, search queries, user-generated content, etc., all of which are potential sources of exploitation that should be taken advantage of by French research.

59. Open access to scientific platforms should not compromise the exploitation of STI in any of its different components.

60. The search for a balance between exploitation and open access requires clarification of the rights and the legal regime applicable to the STI objects, and a dividing line to be drawn between the common good and the protected innovation, between freedom of access and private reservation.

+ The multiplicity of STI objects calls for a clarification of the law and a balance to be struck between access to scientific knowledge and preservation of the potential for STI exploitation in all its components.

¹⁸ <http://www.gfii.fr/fr/document/le-cnrs-conforte-sa-premiere-place-mondiale-en-nombre-de-publications-scientifiques>

¹⁹ http://www.cnrs.fr/dire/termes_cles/interessement.htm

1.2 The legal vacuums

61. The emergence of digital technologies in STI practices is creating a discrepancy between law and practice. While the French Research Code organises public research and defines its objectives, it does not at any time affirm the common values of science.

62. STI is managed by multiple platforms with a non-existent legal model that offer data processing tools of questionable legality.

63. In addition, the contractual practices are not aligned with the practices of the research communities, to the detriment of science.

1.2.1 Lack of a legal framework for science

64. **No law for science.** There is no legal provision, no text that reflects the values of the scientific communities and affirms the best interests of science.

65. This legal vacuum is increasingly felt in the framework of the digital transition and the development of value-added services for scientific data.

1.2.2 Law on platforms: developments in progress

66. The concept of “platform” has no legal status or regime. This legal vacuum entails a certain legal insecurity that was already highlighted by the French Digital Council in its Opinion of 13 June 2014, as well as by the Council of State in its 2014 Report “*Le numérique et les droits fondamentaux*” [Digital technology and fundamental rights].

67. Article 22 of the Digital Republic Bill provides for the introduction of a definition of the concept of “platform”:

- “activities consisting in classifying or referencing content, goods, or services offered or placed online by third parties, or putting several parties in contact with each other, by electronic means, with a view to selling goods, providing services, including non-paying ones, or exchanging or sharing goods or services.”

68. A duty to act in good faith is also imposed on the platform operator:

- “any operator of an online platform is required to give the consumer fair, clear and transparent information on the general conditions of use of the intermediation service it offers and on the terms for referencing, classifying and dereferencing the content, goods or services to which this service provides access.”

69. Although the text of the Bill seems to govern relationships with consumers, Open Science platforms will need to apply these principles and provide researcher-users with fair information with respect to their conditions of use.

1.2.3 A right to TDM: an absence with serious consequences

1.2.3.1 Challenges of TDM

70. **Major challenge.** TDM is a major issue for science, research and innovation in that it enables scientists to identify new research subjects, produce new knowledge and address economic, social and societal issues²⁰.

71. It also creates opportunities in terms of the exploitation of this new knowledge with all this entails for innovation, growth and employment.

72. The scientific and economic issues are especially important in that TDM is practised around the world and is governed by different standards in different countries, including within Europe. The United States and the United Kingdom have affirmed the right of researchers to conduct TDM; Germany has introduced a right of secondary exploitation of scientific publications.

73. A draft revised text of the InfoSoc Directive²¹ (Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society) was expected for the end of 2015 but had not yet been circulated on the date this White Paper was published. The process for accepting a draft Directive is long: at least two to three years are needed before this Directive is accepted by the Member States, and an additional two years for it to be transposed into French law.

²⁰ In 2001, the Massachusetts Institute of Technology presented data exploration as one of the ten emerging technologies that would “change the world in the 21st Century”. (*Data mining et statistique décisionnelle – l'intelligence des données* ["Data mining and Statistics for Decision Making"] Stéphane Tuffery Editions Technip 2012)

²¹ <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000266350>

74. French research cannot afford to suffer discrimination with regard to its European neighbours and allow an unbridgeable gap to open with multiple harmful consequences: delayed development of digital research techniques, delay in emerging research subjects, loss of partnership contracts at European level, fall in the number of patents filed, risk of privatisation of data mining techniques, etc.

75. **Draft Bill V.1.** By inserting provisions in the first version of the draft Digital Republic Bill allowing researchers to carry out data mining operations, the government seemed to have understood what is at stake regarding TDM for research, and especially public research. However, these provisions were removed and are no longer included in the Bill as sent to the Council of State.

76. The removal of these provisions from the Bill was raised in many of the contributions by the scientific community as part of the public consultation on the Bill.

77. The Government offers two arguments to justify this removal:

- The proposed text permitted an exception to copyright, however the text of the InfoSoc Directive regulating copyright between the members of the European Union prevents Member States from creating exceptions not provided for by the texts;
- The InfoSoc Directive will undoubtedly be revised, and the reports submitted in this regard are virtually unanimous in advocating the introduction of this TDM exception to the copyright provisions.

78. **Bill, adopted text (TA) No 663.** As part of the parliamentary debate on the Digital Republic Bill, Deputies of different political persuasions supported the introduction of an amendment creating an exception for text and data mining. An exception to copyright and the right of the database creator was introduced in the text of the Bill (Article 18bis (new) of the adopted text No 663).

79. Fleur Pellerin, Minister of Culture and Communication and Thierry Mandon, Secretary of State for Higher Education and Research, together entrusted Charles Huot, President of the French Professional Group for BtoB Information and Knowledge (GFII), with a “mission of consultation and proposal to facilitate the development of the use in France of text and data mining technologies”.

1.2.3.2 Text and data mining with regard to copyright

80. Copyright includes a monopoly on reproduction, including the adaptation of works. The publisher owning the property rights for the scientific literature it publishes can

therefore prohibit third parties, as well as authors, from making any full or partial reproduction as well as any translation, adaptation or transformation, arrangement or reproduction by any technique or process (Article L.122-4 of the Intellectual Property Code).

81. TDM does not have its own legal status and this lack is a source of legal insecurity. Indeed, data exploration services reflect multiple technical operations, including:

- operations to analyse or process knowledge alone: these are acts undertaken freely;
- technical operations involving the full-text reproduction of data that can be protected by copyright, and their modification in forms such as sections, extracts, mergers, compilations, etc.: some of these acts may concern author monopolies and consequently require the prior authorisation of the copyright owner.

82. The absence of a legal status for TDM and the lively doctrinal debate on the incompatibility of these exploration techniques with the copyright provisions are a source of legal uncertainty and call for rapid legislative clarification.

+ The absence of a legal status for data exploration is a source of legal insecurity for which the law must provide.

1.2.3.3 Text and data mining with regard to the right of the database creator

83. **Principles of the *sui generis* right.** Although, in principle, data cannot be individually protected (except in the event that the data is protected by a private right, intellectual property right, right of personal data, right to privacy), the aggregation of a significant amount of data can, where appropriate, be protected under the *sui generis* right of the database creator.

84. Databases are defined in the French Intellectual Property Code (CPI) as a “collection of independent works, data or other materials, arranged in a systematic or methodical way, and capable of being individually accessed by electronic or any other means”²².

85. The legal framework for the protection of data is defined by the provisions of the Directive of 11 March 1996 on databases²³ (transposed in France by the Act on the legal

²² [CPI Art. L. 112-3.](#)

²³ <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31996L0009> on the legal protection of databases.

protection of databases²⁴) that create a “*sui generis*” right in favour of the database creator.

86. The database creator is defined as the person who takes the initiative and the risk of the investment. The creator may prohibit:

- the extraction of all or a substantial part of the content of the database;
- the reuse of all or a qualitatively or quantitatively substantial part of the content of the database;
- and/or the repeated and systematic extraction or reuse of qualitatively or quantitatively non-substantial parts of the content of the database when these operations clearly exceed the normal conditions of use of the base²⁵.

87. **Research database.** Digital STI is accessible from the databases of scientific publishers that have, little by little, mainly replaced their paper editions by online access to their journal via their platform. As such, the publisher is the database creator and can therefore prohibit any qualitatively or quantitatively substantial extraction of its base.

88. Digital STI is also available from institutional databases, overlay journals and Open Access databases. The creators of each of these databases are also holders of the *sui generis* right and may prohibit any qualitatively or quantitatively substantial extraction from them.

89. ***Sui generis* right vs TDM.** To perform TDM on corpora of data, the following operations are necessary:

- extraction from databases covered by the exclusive right of the database creator;
- technical operations not covered by the regime of the database creator.

90. Some argue for a revision of the right of the database creator, with the original text corresponding “to a static vision of data processing that is now giving way to dynamic processing”.

“On the basis of a review of the legislation, the existing case law and the issues facing the actors in the data production, processing and analysis sector, proposed changes can

²⁴ [Act 98-536 of 1-7-1998](#) on the transposition into the Intellectual Property Code of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

²⁵ [CPI Art. L. 342-3.](#)

be made to adapt the right of database creators to its new technological and commercial environment, transforming it into a right of data and database producers and operators”²⁶.

+ The absence of a legal status for data exploration and the unsuitability of the right of database creators to the dynamic processing of knowledge are sources of legal insecurity for which the law must provide.

1.2.4 The need for reformed rights in scientific digital publishing

91. Scientific publishing practices are fundamentally different from those of literary publishing:

- the purpose of the publication is different: unlike a literary work, in scientific publication, the informational content prevails over the form of expression, which may be incidental;
- the content of the publication is different: a literary creation is specific to its author while scientists, whose “raw material” is science, mostly exploit the work of their predecessors;
- the work provided by the publisher is different: the literary publisher assists their authors in the drafting of the book, encourages them (including financially by the provision of credit), involves them in sales, works on formatting and presentation, organises publication, distribution and promotion, follows the authors in these promotion activities, etc. Conversely, the scientific publisher only receives completed articles (after they have undergone peer review), does not involve the authors in the selling of journals or subscriptions, works on layout and online distribution;

92. Digital scientific publishing is leading to practices being transformed and necessarily imposes a revision of the contractual rules of the game:

- in the relationship between publisher and researcher;
- in subscription contracts;
- in public/private partnership contracts and the organisation of publication of articles resulting from the partnership.

²⁶ “La nécessaire évolution du droit du producteur de base de données pour permettre son adaptation à l'émergence du Big data” [“The necessary change in the rights of database creators to enable their adaptation to the emergence of Big Data”] by Nicolas COURTIER for “[La propriété intellectuelle & la transformation numérique de l'économie](#)” Inpi, 10-9-2015

1.2.4.1 Publisher/researcher contract: contract of transfer of copyright

93. Articles by researchers are published by publishers in the framework of a publishing contract. In addition, the publication is especially important to the researcher, who is evaluated mainly on the basis of this indicator.

94. Publishers are responsible for several major tasks:

- dating of the article;
- ensuring that the article is reread and validated by an editorial board consisting of specialist researchers (peer-reviewing), usually for no payment;
- possible page layout of the article;
- possible correction of the language;
- dissemination of the article by its own channels;
- registering the article in a database or assigning a DOI²⁷;
- archiving of the article;
- management of copyright;
- publicity for the journal and the article.

95. The publishing contract between a researcher and a publisher most often takes the form of an adhesion contract. It provides for a transfer of the researcher's copyright to the publisher, most often **on an exclusive basis and free of charge, for exploitation worldwide and for the entire legal term of copyright**. Many of the testimonies mentioned the practice of publishers of getting researchers to sign a "Copyright transfer form". This contract is "written in a way that only a specialist in 'copyright' law can understand".

Researchers often sign these even without reading them because they often lack support from their establishment in offering a reasoned opinion and the means to defend their interests as creators. Lastly, the speed of publication is often an important point in the context of international competition, and researchers are rarely given the time to implement a procedure to validate the appropriate contract²⁸.

96. **Hybrid model.** Authors making their work freely available in hybrid electronic scientific journals (free access and subscription access) must generally pay a fee (Article Processing Charges) to cover the journal's costs.

²⁷ Digital Object Identifier - <https://www.doi.org/>

²⁸ Contribution of Marie Farge, Senior researcher, CNRS

97. The CNRS Ethics Committee, in an Opinion “on the relationships between researchers and scientific publishing houses” of 31 January 2011²⁹, describes this situation as follows:

- “The transfer of copyright for an article accepted by the editorial board of a journal, which may be based in one country or another, on the recommendation of one or more reviewers, is most often requested by the publisher free of charge. If an author refuses to sign the copyright transfer form, their article, despite having been accepted by the editorial board, will generally not be published. If on the other hand they sign this form, in principle they surrender the right to disseminate their article themselves and to use the figures and tables of data it contains, since the publishing house has become the owner, and in most cases has not even paid the author for the loss of these rights.”

98. Indeed, entering into a publishing contract with exclusive transfer prevents researchers, in particular, from:

- placing their article online on the institutional platform of their employer organisation, which funded the research that led to the article;
- sharing the article with other researchers interested in the work;
- disseminating the article on the researcher's website;
- reusing the graphics and media in other publications or oral presentations;
- depositing the article in an open archive.

99. Some publishers, aware of the importance to research of making articles available and the trend towards Open Science, authorise articles to be placed online on an open archive after an embargo period has been respected (postprint). The HELOISE site <https://heloise.ccsd.cnrs.fr/> is an information service on publishers' policies with respect to the filing of articles. This service only concerns articles deposited on the websites of the scientists themselves and of scientific institutions.

1.2.4.2 Publisher/institute contract: licence contract

100. In order to gain access to scientific journals and books, the institutes enter into subscription contracts with publishers, which make access available to the publisher's journals and online services.

101. This subscription relating to access to the publisher's current collections and/or archives, is:

²⁹ http://www.cnrs.fr/comets/IMG/pdf/03-avis_relations-chercheurs-maisons-edition-2.pdf

- either entered into directly between the institute and the publisher;
- or entered into via a group order (as part of a “national licence”).

102. The licence agreements contain different provisions depending on the publishers but generally provide for a common set of rights in favour of the institutes³⁰:

- a right to access, consult or display the collections;
- a right to print or make an electronic copy for the subscriber's own use;
- sometimes a right to practice TDM via the API (Application Programming Interface) of the publisher, who then retrieves the data on usage; the dissemination of the result of the TDM may include extracts from the full text limited to a certain number of words or a percentage of the text, under the Creative Commons CC BY-NC licence, and with a link to the full text of the article on the publisher's site.

103. Users are not permitted to extract, modify, translate, or create any derived work of any kind from the data made available by the publisher as part of the subscription.

1.2.4.3 Publication in industrial exploitation contracts

104. The provision of research data must be organised and must take into account the nature of the data. Data from a restricted regime area (ZRR)³¹, from ongoing research, data related to know-how or secrets, or to industrial property titles must not be made available systematically.

105. Moreover, in the framework of research collaboration contracts and public/private partnership contracts, the terms of publication of research results are covered by specific provisions, especially when the contract leads to the filing of a patent or is subject to a duty of confidentiality.

106. **Patent clause.** The contract may stipulate that if the research results are patentable, the partners shall file a patent.

107. The contract must then stipulate the terms for filing the patent (single or joint ownership), and the terms for exploiting the patent (exclusive operating licence,

³⁰ Analysis of contracts:

- Elsevier - ABES of 31/01/2014, subscription to the Freedom Collection
- Elsevier - ABES for ISTEX 2013-20 contract
- Elsevier - CNRS 2010-09 contract

³¹ Regime that emerged from the Decree of 3 July 2012 relating to the protection of the nation's scientific and technical potential and the Inter-ministerial Circular on establishment of the scheme for protection of the nation's scientific and technical potential of 7 November 2012

operating licence by business sector or by geographic area, conditions for sharing of royalties, transfer).

108. An invention is patentable if it constitutes a novelty with regard to the state of the art. The latter is understood to mean anything that has been made accessible to the public before the patent filing date, by a written or oral description, a usage or any other means, including the content of French, European or international patent applications, provided that they designate France, and were filed earlier, and were not yet published on the date the patent application in question was filed³².

109. Thus, if the invention has been made public in any part of the world, not only by a prior patent but also by a publication, a public exhibition (at a trade fair for example) or even a simple oral disclosure, it is no longer new.

110. However, if the disclosure of the invention resulted from wrongdoing³³, it does not affect the novelty of the invention if it took place in the six months preceding the filing of the patent application.

³² <http://www.entreprises.gouv.fr/propriete-intellectuelle/inventions>

³³ [Article L. 611-13 of the CPI](#)

111. Disclosure may be regarded as improper if it occurs:

- without the agreement of the inventor (theft of the invention, industrial espionage);
- in violation of a secret (trade secrets, for example);
- in violation of a contractual duty of confidentiality.

112. When a contract provides for the filing of a patent at the conclusion of the cooperation, the researcher(s) responsible for research and development are prohibited from publication as this risks destroying the novelty of the innovation and preventing any patent from being filed.

113. **Confidentiality clause.** A growing number of research contracts provide for clauses governing the confidentiality of the "own knowledge" of each of the parties and the conditions of publication of common results derived from the research.

114. The following partial clauses can for example be included in research contracts entered into by CNRS with industrial companies:

- Example 1 is drawn from a standard research collaboration contract between CNRS and an industrial partner;
- Examples 2 and 3 are derived from framework contracts between CNRS and industrial partners.

	Example 1	Example 2	Example 3
Key definitions	Confidential information Own knowledge New knowledge	Confidential information Results Common results	Confidential information (list) Business data from X Results Exploitable results
Confidentiality	Duty of confidentiality concerning the Confidential Information during the term of the contract and in the 5 years following the termination or expiry of the contract	Duty of confidentiality concerning the Confidential Information during the term of the contract and in the 5 years following the termination or expiry of the contract	Are considered as Confidential Information: <ul style="list-style-type: none"> - another Party's Own Knowledge and the Business data from X for a period of confidentiality of 10 years after contract termination, for whatever reason - the Exploitable Results
Publication	<p>Written agreement of the other party for any publication about the new Knowledge during the term of the current contract and in the 6 months following its expiry.</p> <p>Any draft publication or communication shall be subject to the consent of the other party, which may delete or modify certain statements whose disclosure could be detrimental to the industrial and commercial exploitation of the new knowledge under favourable conditions. Such deletions or amendments must not affect the scientific value of the publication.</p> <p>If the information contained in the publication or communication must be protected as Industrial Property, one of the parties may withhold publication or communication by a maximum period of 18 months from the date of the corresponding request.</p>	<p>Publications or communications relating to the derived Common results shall make reference to the cooperation of the Parties.</p> <p>Any draft publication or communication relating to the Common results must, during the term of the specific agreement and in the twenty-four (24) months following its expiration or termination, receive the prior and written agreement of the other Party.</p> <p>They may decide:</p> <ul style="list-style-type: none"> - to accept without reservation the draft publication or communication; or - to request that the confidential information belonging to it be removed from the draft; or - to request that the Common results that have been the subject of a secret technical file be removed from the 	<p>A prior request in writing, by registered letter with acknowledgement of receipt, must be made by one Party to the other Party for any draft publication or communication, regardless of the form or media, relating to the research programme and the Results, for the period of confidentiality of the Confidential Information.</p> <p>The other Party may request the deletion or modification of certain elements of the publication whose disclosure it believes could be detrimental to it, or detrimental to the industrial or commercial use of the Results of the Research Programme or to the protection of an intellectual property title.</p> <p>In particular, a Party may request that the publication or communication be delayed, for a maximum period of eighteen (18) months, if certain information is to be protected under an</p>

	Example 1	Example 2	Example 3
		draft; or <ul style="list-style-type: none"> - to request changes, especially if some of the information contained in the draft communication could be detrimental to the industrial and commercial exploitation of the Common results; or - to request that the communication be deferred, if genuine and serious reasons seem to warrant it, in particular if an application is to be made for protection of the information contained in the draft communication as Industrial Property. 	intellectual property right.
Exception	Free for use in researcher activity reports and thesis defences	Free for use in researcher activity reports and thesis defences	Free for use in researcher activity reports and thesis defences. Free for use in communications and filing of patent application on own results.

115. This growing contractual practice, increasingly involving negotiations, whose aim is to regulate publication by patent and/or confidentiality clauses, must be taken into account for the provision of research data, in order to preserve the balance necessary for the exploitation of innovations.

1.3 The risks of misappropriation

116. The appropriation of scientific data and results can be legitimate if it responds to legitimate interests of exploitation, preservation of secrets or respect of privacy. On the other hand, when it responds to private interests, to the detriment of Science, even more so when it concerns results from public research, it becomes misappropriation.

117. This misuse can occur in:

- the choice of the economic model of scientific publishing;
- the abuse of intellectual property rights;
- contractual practices.

1.3.1 Appropriation by economic uses

118. The hybrid Gold model, which some consider as transitional, allows free access to scientific publications subject to the payment of “article processing charges” and thus represents an initial source of appropriation.

119. The scientific publishing market is divided into two groups between which a gap is widening irreversibly: the “majors” (Elsevier, Springer, Nature), which are growing faster than the market by capturing the publishing activities of learned societies, have been able to grow their catalogues far more quickly than smaller academic and commercial publishers.

120. In addition, the scientific publishing sector is a place of imperfect competition where price competition is inexistent since the journals are not mutually substitutable. This explains the very high margin rates of the major publishers and the unilateral and uncompetitive setting of the amounts of the subscriptions and “article processing charges”³⁴.

1.3.2 Appropriation by the uses of scientific publishing

121. The right to intellectual property applied to scientific data in fact leads to privatisation of knowledge to the publishers' benefit, which has the effect of:

- erecting legal and financial barriers to access to scientific and technical information;
- slowing down scientific research and progress;

³⁴ Information extracted from the study by the DIST-CNRS “*L’Edition de sciences à l’heure numérique : dynamiques en cours*” [“Publishing of Science in the digital age: dynamics in progress”] (2015)

- prompting concentrations of scientific information and research themes, financed at least 50% by public funds, in the hands of private publishers.

122. Publishers exploit the publishing of scientific articles and make significant profits, while:

- the costs of digital publishing are lower than the costs of the paper edition;
- neither the authors, the editorial board, nor the peer-reviewers are paid by the publishers; often the authors even have to pay “article processing charges” and the organisations have to pay their subscriptions.

123. Publishers' profits have never been higher than they are today. The global market for scientific research publishing is estimated to be worth 12.8 billion euros. Digital services account for 60% of the revenue, on average. Among the major publishers, which invested in digital technologies and the platform strategy early on, this ratio is 75%.

124. The conditions under which researchers can re-use their own articles are highly restrictive or even non-existent, as in the vast majority of cases the authors have transferred their rights on an exclusive basis by means of true adhesion contracts.

1.3.3 Appropriation by the contracts

1.3.3.1 The scientific publishing contract

125. **Nominate contract.** Authors of scientific texts can, by a publishing contract, transfer “under specified conditions to a person referred to as the publisher the right to manufacture or have manufactured a number of copies of the work, it being for the latter to ensure publication and dissemination thereof”³⁵.

126. This is a nominate contract governed by Articles L.132-1 to L.132-17 of the Intellectual Property Code (CPI), which places major obligations on the publisher, including:

- a duty to publish: In the absence of publication, the contract will be terminated (Article L. 132-17 CPI);
- continuous and sustained exploitation: Article L.132-12 of the CPI imposes on the publisher the requirement to ensure the permanent availability of the work, and therefore to make automatic reprints, and to make or have made popular editions;

³⁵ [CPI Art L132-1](#)

- accountability: Article 132-13 specifies that “the publisher shall be required to render accounts”.

127. Publishing implies a transfer of rights from authors to their publishers, and this transfer of rights, with the assigned objective, is the element that characterises publishing contracts compared to other contracts for the transfer of intellectual property rights.

128. Article L.132-8 of the Intellectual Property Code stipulates that “the author shall guarantee the publisher the undisturbed and, unless otherwise agreed, exclusive exercise of the right assigned.”

129. **Remuneration.** The transfer of rights in favour of the publisher is exchanged for proportional remuneration. In principle, therefore, the law provides for the remuneration of authors in proportion “to revenue” or “to the products of exploitation”. In other words, in order to protect authors and allow them to participate in the success of their work, Article L. 132-5 of the Code provides for remuneration of authors proportional to the proceeds of exploitation of their work, which includes not only publishing, but also the right of representation, translation or adaptation of their work.

130. By the rule of proportionate interest, the legislation is intended to protect authors against any transfer of exploitation rights agreed by them for an amount that is very small in relation to the profits made by the publisher. By way of derogation, remuneration in the form of a lump sum may be provided for in the cases listed exhaustively in Articles L. 131-4 and L. 132-6 of the CPI.

131. Authors may however contractually waive remuneration that is proportional to the exploitation of their work. Indeed, Article L. 122-7 of the CPI relating to the transfer of the right of reproduction of a work stipulates that such a transfer may be without payment. For this, a clause on transfer of the right of reproduction free of charge must be formally stipulated in the publishing contract.

132. Violation of the provisions relating to remuneration is sanctioned by relative nullity whose action shall lapse after five years, a period that runs from the date of signature of the contract³⁶.

133. **Standard contract.** By transferring to the publisher their exclusive property rights over the article, author-researchers can no longer exploit their articles, share them or self-archive them, even if this is free of charge and for the benefit of the scientific

³⁶ Dalloz Action “*Droit d’auteur*” [“Copyright”] Chapter 112 - Transmission, transfer and contracts relating to copyright - André R. Bertrand - 2010

community. In the majority of cases, the signed contract is an adhesion contract providing for an exclusive transfer of rights and is not accompanied by any remuneration. These contracts could be revised in light of the provisions of the Intellectual Property Code.

134. It would be interesting to return to a regime providing more protection to authors, authorising author-researchers to share their articles without any limits, including contractual ones, in the name of:

- the best interests of science;
- the financing by public funds of the research that led to the writing of the article.

135. This system enabling author-researchers to freely share their publications should not penalise the world of scientific publishing; arrangements can be found in particular by defining embargo periods.

 Publishing contracts signed by researchers for articles they wrote in the framework of publicly-funded research largely provide for exclusive transfers of rights in favour of the publisher.

These constitute true adhesion contracts.

Publishing contracts that do not state formally that the rights are transferred free of charge shall be subject to nullity.

1.3.3.2 The contract for the subscription to the publisher's platform

136. **Terms and conditions of use or subscription.** The terms and conditions of use as well as the subscription contracts with publishers lay down the conditions under which the articles and services accessible from the platforms can be used. These contracts generally provide for limited conditions of use of the articles:

- access to the full text, printing or downloading for the subscriber's sole use;
- inability to conduct searches of articles outside the platform.

137. For example, the general terms and conditions of use of Elsevier Masson, available on its website, stipulate in the article on Intellectual property:

- "6.1 Copyright and other intellectual property rights to all Elsevier proposals, publications and other Products and or Services shall remain with Elsevier unless agreed otherwise in writing. The rights granted by Elsevier are restricted to use

solely by the Client and may not be assigned, transferred or sub-licensed without the prior written permission of Elsevier. The rights granted by Elsevier are non-exclusive and for the purpose expressly agreed upon. Any other use shall require the prior written permission of Elsevier. The Client shall not acquire any intellectual property rights in the Products.

- 6.2 No part of the Elsevier proposals, publications or Products may be stored in any automated data file and/or reproduced, whether electronically, mechanically, by photocopying, recording or in any other manner or form, without the specific prior written permission of Elsevier.”

138. **User-generated content.** Aware of the need to propose information processing services to their subscribers (cross-referencing, semantic and lexicographical analysis, automatic synthesis, translation, etc.), publishers have developed value-added services that are available from their platforms.

139. For example, Elsevier offers a TDM licence contract via its API:

- limiting the use of the API to non-commercial purposes;
- limiting the exploration of results from using the API on datasets, the “user-generated content”, to:
 - the placing online of bibliographic data accompanied by a DOI (Digital Object Identifier) link leading to the full-text article;
 - the addition of a proprietary notice;
- prohibiting:
 - the use of an extract of more than 200 characters from the full text;
 - the modification, translation, or creation of derivative work based on the datasets;
 - the reproduction, retaining or redistribution of the datasets;
 - the extraction or use of the datasets for any commercial activity;
 - the use of robots or other automated programs, or algorithms for searching;
 - the use of the output generated by the API to enhance institutional repositories in a way that would compete with the final peer-review journal article.

140. The ownership of the user-generated content is not clearly stated by the publisher, but by licensing rights over the “TDM output”, the publisher assumes that it is the holder.

141. This appropriation by private publishers of publicly-funded science calls for legislation in favour of researchers' rights:

- on free access to the data and results of research;
- on free exploration of the data and results of research;

142. In order to respond to the needs of researchers in light of the science macro-environment, Part 2 offers an approach and legal solutions in favour of Open Science.

1.4 Massive validation of these findings

1.4.1 Summary of the national consultation

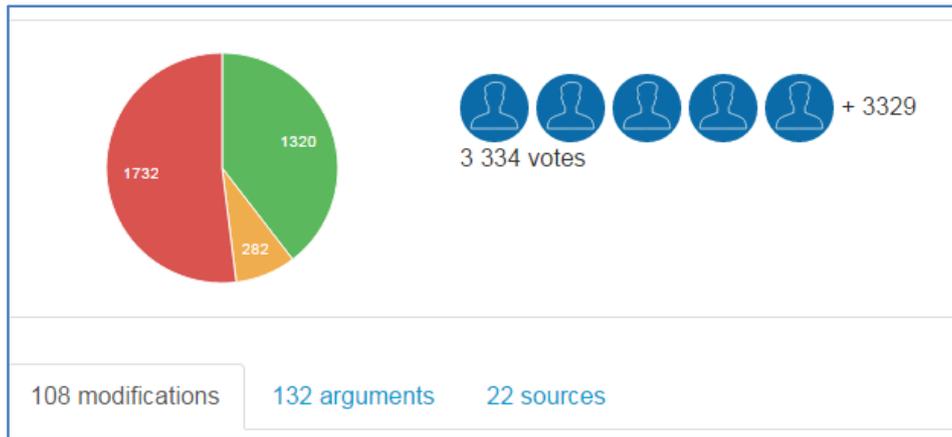
143. For the first time in legislative history, the Digital Republic Bill was placed online, to enable it to be publicly discussed and to receive contributions from citizens, between Saturday 26 September and Sunday 18 October 2015.

144. The initiative greatly mobilised Internet users: when the consultation closed, the counters showed:



Gouvernement, 26 Septembre 2015	Government, 26 September 2015
votes	votes
contributions	contributions
participants	participants

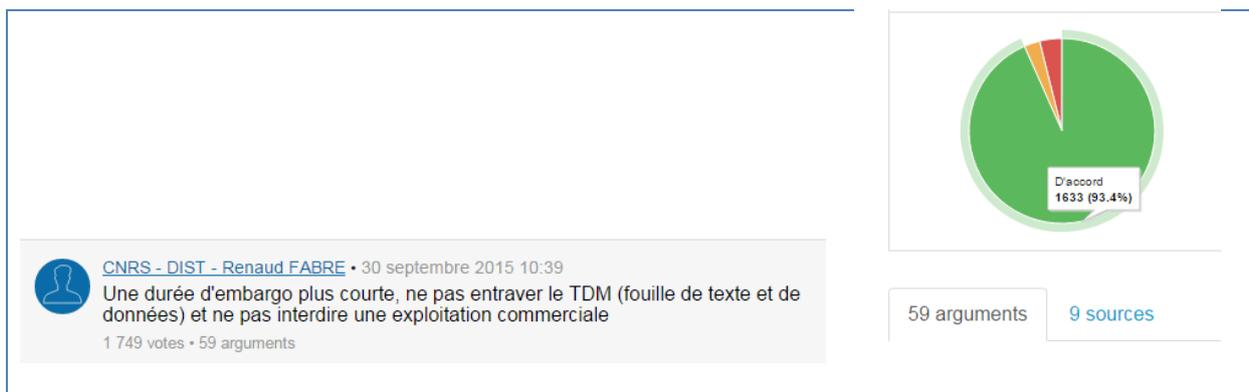
145. The result of the consultation on the article “Free access to scientific publications from public research” is as follows:



3334 votes	3334 votes
108 modifications	108 changes
132 arguments	132 arguments
22 sources	22 sources

146. When the public consultation on the Digital Republic Bill ended on 18 October, the former Article 9 of the draft bill devoted to “Free access to scientific publications from public research” had generated the most reactions among Internet users with 3334 votes and 108 proposals for changes, ahead of the articles dedicated to open data, or to the free reuse of data from industrial and commercial public services.

147. The proposal that received the most votes was the one by the CNRS Scientific and Technical Information Department entitled “A shorter embargo period, no hindrance to TDM (text and data mining) and no prohibition of commercial exploitation”³⁷ with 1633 votes “for” out of 1749.



³⁷ <https://www.republique-numerique.fr/consultations/projet-de-loi-numerique/consultation/consultation/opinions/section-2-travaux-de-recherche-et-de-statistique/article-9-acces-aux-travaux-de-la-recherche-finanee-par-des-fonds-publics>

CNRS - DIST - Renaud FABRE 30 Septembre 2015 10:39 Une durée d'embargo plus courte, ne pas entraver le TDM (fouille de texte et de données) et ne pas interdire une exploitation commerciale 1749 votes 59 arguments	CNRS - DIST - Renaud FABRE 30 September 2015 10:39 A shorter embargo period, no hindrance to TDM (text and data mining) and no prohibition of commercial exploitation 1749 votes 59 arguments
D'accord 1633 (93,4%) 59 arguments 9 sources	Agree 1633 (93.4%) 59 arguments 9 sources

148. Roberto Di Cosmo, Professor in Computer Science at Paris-Diderot University, posted the modification receiving the second largest number of votes, with 1511 votes for his proposal to “Protect copyright on scientific articles, to enable free access to scientific research”, filed 10 days before the end of the consultation.

149. Institutions and research organisations have also taken formal positions. These include INRIA, INRA, the Couperin Consortium with the ADBU, SNESUP, UPMC, INP Toulouse, the Association of French Archivists (AAF), Cairn.info, etc. Their proposals were predominantly the following:

- enable free access to the results of scientific research;
- reduce embargo periods to 6 and 12 months, or even remove them in some cases, and for others make no distinction between the scientific fields;
- enable articles to be deposited in open archives;
- authorise text and data mining operations or data searches;
- guarantee the possibility of exploiting the results generated by knowledge processing.

150. The French Publishers' Association (SNE) and the French Specialised Periodical Publishers' Federation (FNPS) were favourable to the embargo periods of 12 and 24 months stipulated by the text; the French Professional Group for BtoB Information and Knowledge (GFII) proposed that they be decided on the basis of impact assessments to be implemented.

151. On 18 October 2015 a discussion forum (Gouv'camp) was organised for the closure of the national contribution platform, to bring the different bill contributors together in working groups. A report presenting a consensual position was placed online by Alain Bensoussan, co-rapporteur with Grégory Colcanap of the “Article 9” group (which has become Article 17 in the latest version of the Bill). This report insists on the fact that:

- “scientific texts must become common assets”;
- the question of TDM must be addressed;
- embargo periods before publications can be made freely accessible must be reduced to a maximum of six and twelve months;
- depositing articles “in a long-term public open archive” must be encouraged;
- the commercial exploitation of the contents of a scientific article must remain open.

152. The needs expressed by the researchers and key witnesses interviewed in the context of this White Paper cover the same crucial points mentioned above by the contributors to the platform.

153. **Official summary.** An official summary of the public consultation prefaced by Ms Axelle Lemaire lists the contributions that led to the amendment of the draft bill, including the contribution of the DIST. The Government's comment explains the scope of the amendments made in light of the proposal:

An almost general consensus emerged from the consultation regarding a clear demand to strengthen the rights of researchers to disseminate their work freely, when the work has been financed by public funds. Seeking a new balance between the positions of the different stakeholders in the digital age and the knowledge society, the Government has developed the measure in the following way:

- The “embargo” periods, at the end of which authors of publications financed by public funds may, at the latest, make their texts freely available, have been reduced by half.
- If articles are made available by the online publisher free of charge, authors will be able to exercise their right immediately.
- The text also now states that the research data associated with these texts can be reused immediately, and that their circulation may not be impeded at the time the texts are published.

Among the requests made that were not followed up at this stage, it is important to mention the creation of an exception to copyright for the analysis of texts and data for research purposes (“text and data mining”, TDM), which is clearly supported by the scientific community. European law does not currently make it possible to create new exceptions, and the Government hopes that this issue can be addressed in the framework of the

European work in progress.

154. Significant changes have emerged from this national contribution; some important points remain, however, that have not been taken into account in this Article 17 (concept of publication in its author's version/in its publisher's version) or have been postponed to a future text. Proposals for amendments to the text of the Bill are included in Part 2 of this White Paper (2.5 "The amendment to Article 17 of the Digital Republic Bill").

1.4.2 The opinion of the French Digital Council

155. On 30 November 2015, the French Digital Council (CNNum) issued an opinion on the Digital Republic Bill based on the outcome of the public consultation.

156. The White Paper takes up here the opinion of the CNNum on the provisions relating to Article 17 of the Bill and shares it entirely.

On the free access to scientific publications and data from public research (Open Access): supplement the provisions of the Bill

The CNNum welcomes the limitation of exclusive transfer periods for scientific publications from public research to 6 months for the sciences, technology and medicine, and 12 months for the human and social sciences, by the recognition of a secondary right of exploitation for researchers.

In its report “*Ambition Numérique*” [“Digital ambition”], the Council recommended supplementing this provision with a requirement to make these publications accessible free of charge on an institutional website, in an open journal or on an open archive site. It could be applied to research organisations. The United States have taken this route by stipulating an embargo period equal to 1 year. The United Kingdom has decided to create incentives for free access by taking open scientific publications into account in the evaluation – and funding – of research.

Lastly, the CNNum welcomes the inclusion of data from public research activities made public legally under a regime of “commons”, within the meaning of Article 714 of the French Civil Code (*choses communes*).

157. The French Digital Council also notes certain shortcomings in the Bill, in particular with regard to the circulation of data and knowledge and more particularly on text and data mining.

158. The White Paper is also in line with the opinion of the French Digital Council on this point and recommends an exception to copyright.

Authorise text and data mining

Text and data mining refers to a series of computer processing operations that consist in extracting knowledge according to a criterion of novelty or similarity in texts or databases. For example, it enables searches to be conducted for “weak signals” that are difficult to grasp from cursory reading, identification or analysis of reports of failed experiments.

It has been regarded as highly promising for scientific discovery and the development of new knowledge. It should enable research to take advantage of progress in the analysis of Big Data, which is destined to become a major factor of international competitiveness. Ireland, the United Kingdom and also the United States and Japan allow it today.

Considering that:

- the automated searching of texts and data, as an information reading and extraction activity, is a practice that is not fundamentally different from the manual reading of information, which has always been carried out by research;

- copyright, which protects the form of expression and not the ideas, today makes it possible to read and reuse information or data included in a text for which a right of access has been obtained,

the CNNum considers that there is no legitimate reason to restrict this right in the framework of automated processing.

The major publishers that own the majority of scientific publications can today, by means of contracts, prohibit researchers from searching texts and data – in particular temporary copies, which are technically necessary in order to do this searching – even when the researchers have legal access to all of the scientific publications included in the databases searched. This ban is based mainly on the *sui generis* right concerning databases. This practice therefore requires the creation of an exception to copyright, on the basis of a reinterpretation of the exception for research, identical to the interpretation by the United Kingdom.

Taking into account the limits and constraints imposed by the contractual solutions, the CNNum therefore recommends **establishing a real exception to copyright authorising text and data mining.**

1.4.3 The impact assessment for the Bill

159. While the Council of State in its Opinion of 9 December 2015 regretted the lack of impact assessment, in particular on Article 17 of the Bill³⁸, on the same day the

³⁸ Council of State 3-12-2015 Opinion on a Digital Republic Bill No. 390741 page 5

“With regard to the provision free of charge on the Internet of the results of publicly-funded research, provided for by Article 14, the Council of State noted that the impact of such a measure on future contracts between publishers and authors is determined by its public policy nature, which can only apply on French territory, whereas dissemination on the Internet has a global effect. It considered that this inconsistency was an obstacle to the adoption of this measure. In addition, the Council of State was unable to retain this provision, nor the one that qualifies the data from publicly-funded research as

Government proposed this impact assessment whose terms reinforce the present analysis.

160. The assessment indeed insists that the objectives of such new legislation:

- “Involve opening the possibility of free access dissemination of publicly-funded scientific work, upon expiry of what is known as an “embargo” period that preserves the exclusive rights of publishers. At the same time it involves legally securing existing practices in the scientific community that are well tolerated by publishers. The creation of this new right for the authors of the work requires the intervention of the legislator, in order for it to be imposed on all forthcoming publishing contracts. This provision creates new rights for the authors of publications and promotes a new balance in the relationship between researchers and their publishers.”
- “The measure also aims to promote and protect the free reuse of research data, from the time they are made public.”

161. The impact of the law has been analysed at various levels:

- the public authorities: it “promotes better regulation of the costs of scientific and technical information, today widely borne by the public authorities”;
- economic and social: “the sharing of research data contributes to economic and social development”;
- on research: “open access to publications and the free reuse of research data promote the sharing of knowledge and discoveries, earlier and recent, within the scientific community. It encourages cooperation and interdisciplinarity, limits the duplication of research efforts, and contributes to the overall improvement in the quality of work. It also paves the way for greater account to be taken of the expectations of civil society, promoting responsible research and innovation. Lastly, it will benefit companies looking to innovate, in particular small and medium-sized enterprises that do not have the capacity to invest in research and development.”
- on the scientific publishing sector: the impact is very low:
 - o foreign experience has shown “a limited decline in access via publishers’ websites”;

“commons”, within the meaning of Article 714 of the French Civil Code, and that allows the free reuse of these data once published, on the grounds that there was no real assessment available of the positive or negative impacts, either legal or economic, that can be expected.”

- the risk of loss of subscribers is very low;
- in French scientific publishing, the impact should be put in perspective since most of the turnover today comes from subsidies provided by the research establishments or laboratories. However, given the French publishing landscape, the Government has decided to take into account the concerns expressed by many publishers and journal directors in the human and social sciences by defining a plan for the transition to free access for HSS journals;
- on world scientific publishing: it “is today characterised by a high concentration, oligopolistic in nature, around a few international groups” and “scientific and technical information constitutes an exceptionally profitable activity”.

1.4.4 The text adopted by the National Assembly

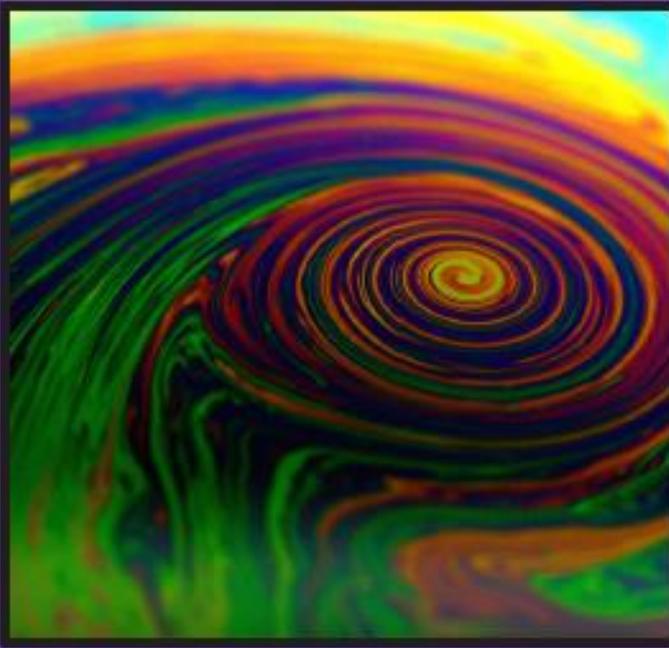
162. The Deputies have grasped the challenges for public research, not only by validating the principle of free access to scientific data but also by introducing a legal framework for text and data mining practices.

163. A report on the impact of the principle of free access to scientific data on the scientific publishing market and on the circulation of ideas and scientific data will be submitted to Parliament by the Government “no later than two years after the promulgation” of the Digital Republic Act.

164. A new Article 18bis has been introduced and creates an exception to copyright and an exception to the right of the database creator in favour of text and data mining³⁹.

165. A table in the Annex offers a comparison of the different versions of the text of the Bill (Annex 1) from the public consultation to the adoption of the Bill by the National Assembly.

³⁹ Digital Republic Bill, text adopted by the National Assembly No. 663 on 26-1-2016
<http://www.assemblee-nationale.fr/14/ta/ta0663.asp>



2. L'Avenir : « La science numérique ouverte »

2. THE FUTURE: OPEN DIGITAL SCIENCE

166. This section presents a set of proposals (2.5 and 2.6) aimed at changing the law on scientific and technical information to take into account the needs and practices of the scientific community, secure the uses of public research, and rebalance the protection of the interests of those concerned, all in the higher interest of Science.

167. These proposals are made on the basis of:

- topics investigated by the key figures interviewed for this White Paper and the expression of the common and universal values of Science that emerged from hearings with these figures (2.1);
- analysis of the rules emerging around the world (2.2);
- a study of the gap between the law as it stands and the practices and needs (2.4).

168. France must comply with the values and rules emerging in supranational and foreign bodies and legislation, otherwise French science could become marginalised.

169. The necessary emergence of this new legal framework for Open Science is based on existing key concepts and legitimate interests that must be preserved (2.3).

170. The adoption of Article 17 of the Digital Republic Bill resulting from a consensual position that emerged during the GouvCamp initiative and the confirmation of Article 18 bis (new) in favour of text and data mining would enable the world of public research to secure its practices. A proposal for the creation of a genuine positive and comprehensive law for Open Science is formulated.

171. In order to ensure that the legal provisions can be applied flexibly, particularly in the light of the very different practices implemented by different scientific communities, it has been proposed that reference guidelines on use should be drawn up, together with a standard.

172. Furthermore, as agreements for the transfer of rights between authors and publishers are one-sided standard contracts, a model contract for the transfer of copyright providing better protection for the legitimate interests of researchers and the scientific community could be made compulsory.

173. Open Science also requires the definition of common ethical rules whose values could be extended internationally and whose application could be guaranteed by an Open Science Agency. Finally, the influence of French science around the world gives

France the legitimacy necessary to propose an international convention for universal Open Science.

174. Training initiatives will be necessary to support the legislative changes desired.

2.1 Personal testimonies recorded for the White Paper: converging principles for an approach to Open Science

2.1.1 A shared value: science, a "common good" of humanity

175. Historically, science has always been considered as a common good; the scientific method itself implies the collective accumulation of knowledge (work in partnership, exchange of information, peer review, etc.). The growing place of information technologies in science to facilitate research, sharing and collaboration has reactivated the notion of "common good" associated with science.

176. **A shared international position.** "Scientific articles have a unique role as products for the common good." This position was affirmed in a press release by the Conference of University Presidents, the *Conférence des Grandes Ecoles*, the Conference of Directors of French schools of engineering and the Couperin Consortium.

177. This universal dimension of scientific knowledge had already been upheld in the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" of 2003. Open Access is defined as "a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community"⁴⁰.

178. The opinion piece entitled "*Favorisons la libre diffusion de la culture et des savoirs*" (Let's all agree to facilitate the dissemination of culture and knowledge) published in *Le Monde* on 10 September 2015⁴¹, signed by nearly 1820 people, reaffirmed that "common assets – or commons – have always benefited from the exchanging and sharing practices on which scientific production and cultural creation depend".

⁴⁰ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 22 October 2003, <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>

⁴¹ Article "*Favorisons la libre diffusion de la culture et des savoirs*" http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs_4751847_3232.html

179. In a motion approved on October 2015, the Conference of University Presidents states "that knowledge is a common good of humanity and that scientific data should be regarded as information of general interest"⁴².

180. Open Science must be given the status of a "universal principle" to allow access to the commons that scientific data really are, for the good of humanity and scientific progress.

The recommendation of the CNRS Scientific Board

"Science is a common good of humanity which cannot be misappropriated by commercial interests."

181. Indeed, Open Access has a very real impact on progress in research and even in some cases on the protection of public health:

- the team combating the Ebola virus in Liberia was unable to access certain articles because of their high cost, although this would have enabled them to identify the virus earlier and thus choose suitable measures for prevention and treatment more rapidly. In this case, the private retention of knowledge resulted in a number of deaths; free and immediate access to knowledge is a vital necessity in such cases;
- The project to sequence the DNA of the entire human genome was possible thanks to large-scale collaboration between researchers from all over the world, and to immediate public dissemination of the research results. The Internet also acts as a catalyst that reduces the time necessary for a study of this magnitude. Open, free and immediate access to scientific results was an indispensable condition for this major scientific achievement.

182. Moreover, this change to the existing economic and technological system is necessary in order to:

- prevent the use of digital platforms centralising research results and data from being governed by commercial law alone;
- avoid the waste of public money that occurs when research that has already been carried out is duplicated by other institutions;
- avoid extra costs for research institutes, laboratories and universities.

⁴² <http://www.cpu.fr/actualite/les-donnees-de-la-science-un-bien-commun/>

183. Everywhere in the world, positions in favour of science as a common good are being expressed:

- in Quebec, the association "Science and the Common Good" was set up in July 2011 to defend and promote a vision of the sciences as useful for the common good;
- the Open Science Federation, which campaigns for Open Science;
- at the international level, the Creative Commons organisation launched a project named Science Commons in 2005 that proposed to spread the principles of openness and sharing within the scientific community by establishing the notion of science as a common good and extending the use of Creative Commons licences to include scientific and technical research⁴³.

2.1.2 Science, driving the economy

184. Open Science facilitates research and innovation by allowing the sharing of knowledge, the identification of new research subjects, the production of new knowledge and responses to economic, social and societal issues. It also creates opportunities in terms of the exploitation of new knowledge, with all that this entails for innovation, growth and employment.

185. Its role in driving innovation has often been recognised:

- by UNESCO: open access "promotes global knowledge flow for the benefit of scientific discovery, innovation and socio-economic development"⁴⁴;
- for the OECD, "Open science has the potential to enhance the efficiency and quality of research by reducing the costs of data collection, by facilitating the exploitation of *dormant* or inaccessible data at low cost and by increasing the opportunities for collaboration in research as well as in innovation." Open Science also helps reduce the divide affecting access to science and strengthen capacity in developing countries;
- in its digital strategy document published on 18 June 2015, the French Government stated: "The free movement of scientific knowledge and its free exploitation contributes to innovation, encourages collaboration, improves the quality of publications, avoids the duplication of effort, allows the exploitation of

⁴³ <http://creativecommons.org/science>

⁴⁴ <http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-access-to-scientific-information/>

the results of previous research and promotes the participation of citizens and civil society."

186. In its contribution, the CNRS Scientific Board emphasises the advantages for researchers of having scientific data and publications in digital format, for easier sharing and searching:

Contribution of the CNRS Scientific Board

"The digitisation of data used by scientists and of their publications enables automated processing, fast transfer, the harmonisation of methods of access and descriptions; all these advantages help bring vast, rich and diverse resources within the reach of researchers, in much shorter timeframes."

187. The CNRS Ethics Committee (COMETS)⁴⁵ has also stressed that:

- "Facilitating access to and the reuse of these data has thus become a crucial issue for sharing and circulating research results more rapidly."

188. During the hearings held in preparation for the drafting of this White Paper, the following needs were consistently expressed by the people interviewed:

- open access to scientific data;
- the need for peer-reviewing and for new assessment indicators;
- publishing and the publisher's embargo period;
- ease of searching through data;
- recognition of origin and visibility;
- respect for legitimate interest (patents, confidentiality);
- implementing an ethical charter for scientific and technical information (STI).

189. Proposals for a complete change of paradigms have also emerged.

2.1.3 Priority for Open Access and the sharing of scientific data

190. Everyone agrees that it is necessary to have free and massive access to scientific data, in the greater interest of research and its ability to address human, social and economic issues.

⁴⁵ Self-referral by the COMETS, "The ethical issues of scientific data sharing", by the Data Sharing Group, 12-12-2014

191. This notion of scientific data includes not only research data but also the results of research, whether published by a scientific publisher or not at all.

INRIA hearing: Claude Kirchner, 15 October 2015

"All scientific data must remain under the control of scientists."

The recommendation of the CNRS Scientific Board

"Any hindrance to open access to the results of scientific activity (publications, research data, metadata, value-added services) would compromise the development of science."

192. There is a consensus within the scientific communities on the need for open data, the opening up of research data, and open access to scientific publications protected by copyright.

193. In addition, researchers mostly access data via platforms for which the technical and legal model is not secure.

194. **Scope of the data.** In order to determine their needs in terms of free access, it is first necessary to define the scope of the data required by researchers as part of the scientific approach and when conducting their research.

195. The Digital Republic Bill appears inadequate in this respect.

196. The original draft version of the Digital Republic Bill of July 2015 employed the concept of "scientific contributions". But there is no statutory definition of "contribution", so it was imprecise and subject to different interpretations.

197. The Digital Republic Bill preferred the notion of "Scientific Texts," a term taken from Article L.112-2 of the French Intellectual Property Code to qualify creative work capable of being protected by copyright: "books, pamphlets and other literary, artistic and scientific works."

198. The text only concerns "Scientific Texts", where such works can be protected by copyright. This interpretation is confirmed:

- by reasserting the principle of copyright protection in scientific texts;

- by a provision limiting the right of researchers to make their texts available in the "latest version of the manuscript accepted by the publisher, excluding the formatting, which is the publisher's contribution".

199. The Bill limits the possibility of open access to scientific work in "postprints".

INRIA hearing: Claude Kirchner, 15 October 2015

"The 'author's accepted version'⁴⁶ of a scientific article, entirely created by this author up to its transmission to the publisher for publication, must remain free of any restrictions and it should be possible to post it on line in whatever form may be chosen by the author (or their institution), in particular in an open archive.

Any embargo period could only concern the "publisher's version" in its final form, in order to retain its commercial potential. Such restrictions are only acceptable if the "author's version" can be freely distributed, and the duration of the embargo should then be set in compliance with international practices.

200. **The needs.** Researchers have expressed the need to be able to access all the scientific data and results of any research activity financed at least 50% by public funds.

201. This need was reaffirmed by the Scientific Board in its recommendation attached to the White Paper.

202. The communication of the European Commission of 17 July 2012⁴⁷ also states that "many of the publicly-funded research results that exist in the form of data are not made widely available for others to verify or build upon, and this makes research investment highly inefficient."

203. It must be possible to put this "Lost Science" financed by public funds, and which has undeniable economic value, to good use and for it to be exploited by public research.

204. In addition, the French Research Code defines the following among other missions of public research (Article L.112-1 of the Research Code):

- "sharing and disseminating scientific knowledge";

⁴⁶ A reminder of the definitions of the different versions, "author's initial version" (or preprint), "author's accepted version" (or postprint) and "publisher's version" is available in the glossary.

⁴⁷ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "A Reinforced European Research Area Partnership for Excellence and Growth" C(2012) 401 of 17-7-2012

- "open access to scientific data".

+ To carry out their work, researchers need open and free access to all scientific data in digital form:

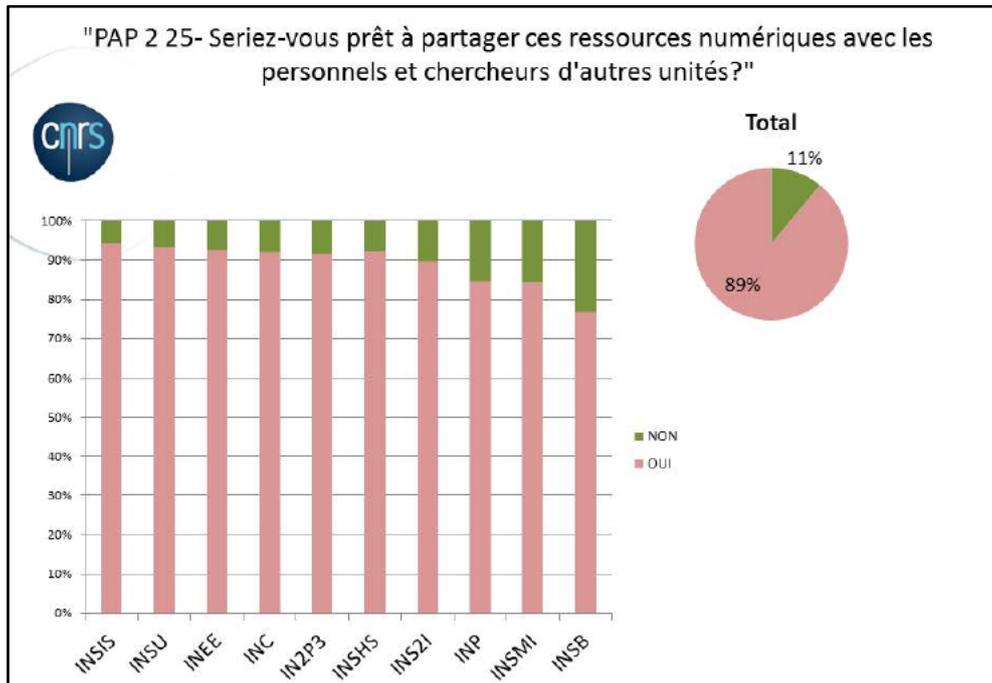
- scientific results, including the results published by a scientific publisher;
- research data in the sense of the data used to establish these results.



205. **Sharing data.** The sharing of knowledge is the vital and historical basis of the scientific approach, and indispensable for research. The digital transition has disrupted the practice by giving access to a growing and comprehensive mass of data, instantaneously and anywhere in the world.

206. More than 89% of researchers are ready to share digital resources with the personnel and scientists of other units.

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 2 25- Seriez-vous prêt à partager ces ressources numériques avec les personnels et chercheurs d'autres unités ?	PAP 2-25: Would you be prepared to share these digital resources with the personnel and researchers of other units?
NON / OUI	NO / YES

207. The sharing of scientific data extends the scope of knowledge.



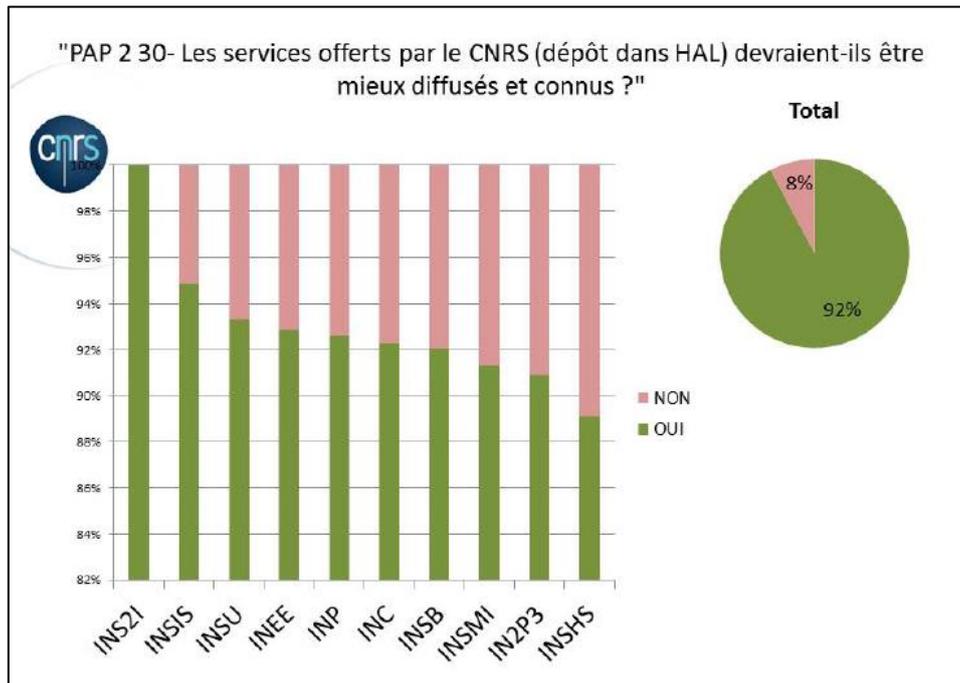
Researchers have expressed the need to share scientific data.

208. **Multipurpose platforms.** Scientific data can be accessed and shared from "Innovative and user-friendly tools"⁴⁸, simple to use, and enabling broad access to knowledge.

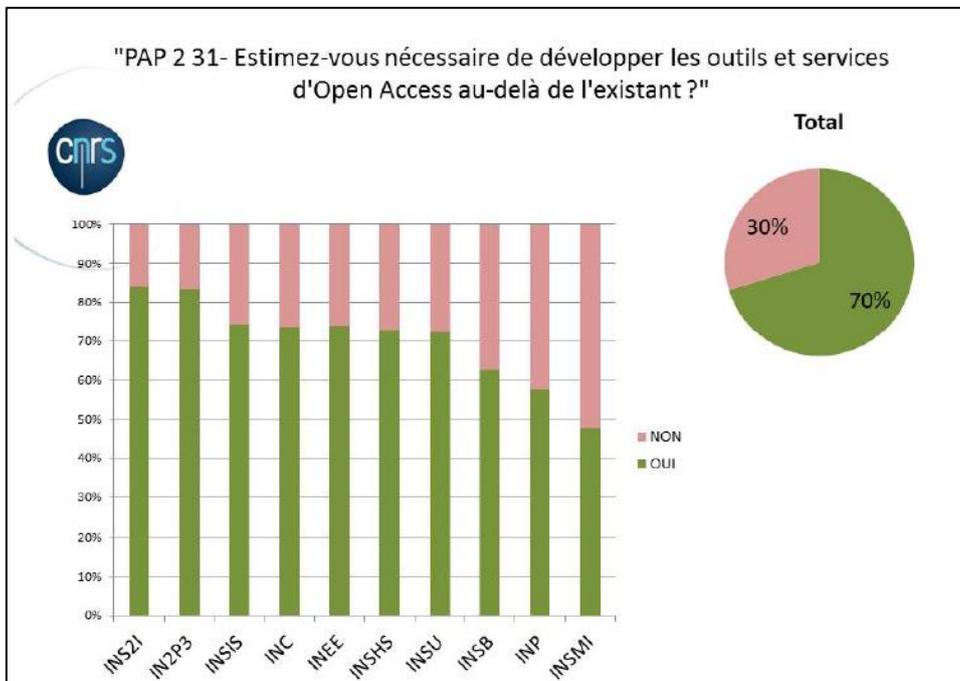
209. Online platforms such as HAL or ArXiv have been developed. The scientific communities would like to have better information on the services of submission platforms such as HAL, and 70% consider that other Open Access tools and services need to be developed.

⁴⁸ Results of the survey on the uses and needs of STI - CNRS - March 2015, page 5

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 2 30- Les services offerts par le CNRS (dépôt dans HAL) devraient-ils être mieux diffusés et connus ?	PAP 2-30: Should the services offered by the CNRS (deposition on HAL) be better publicised and understood?
NON / OUI	NO / YES

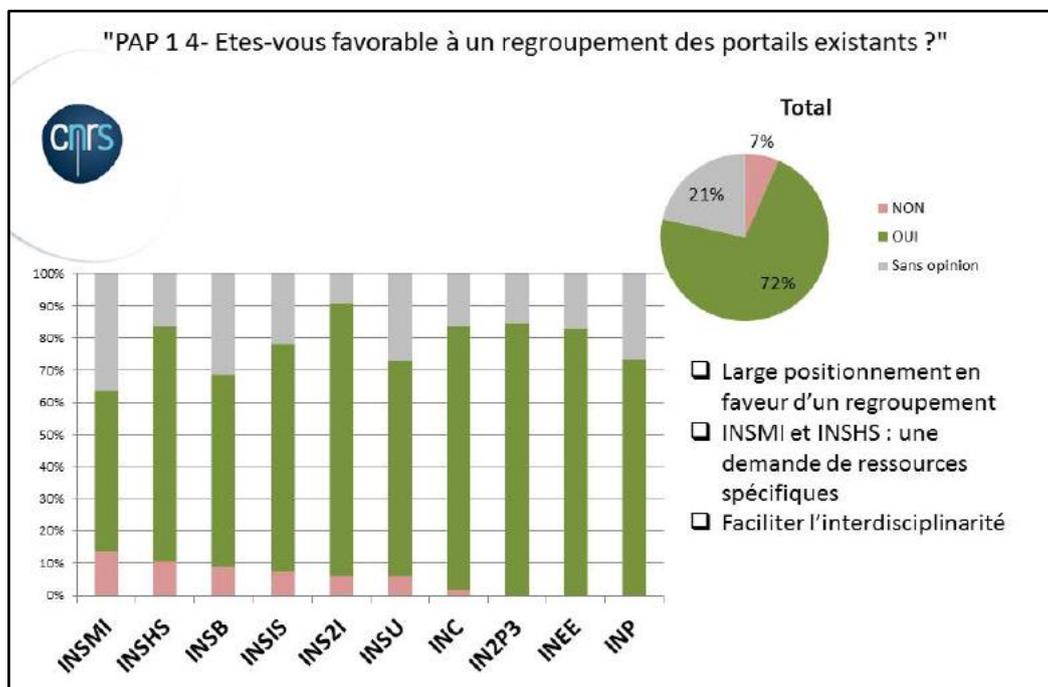


PAP 2 30- Estimez-vous nécessaire de développer les outils et services d'Open Access	PAP 2-30: Do you consider it necessary to develop more advanced Open Access tools and
--	---

au-delà de l'existant ?	services?
NON / OUI	NO / YES

210. A 72% majority of researchers are in favour of consolidating the existing portals. Gathering data on a single portal helps limit the loss of knowledge and enables scientists to work from documents produced by different disciplines, thus facilitating trans-disciplinary research from the corpora of different publishers.

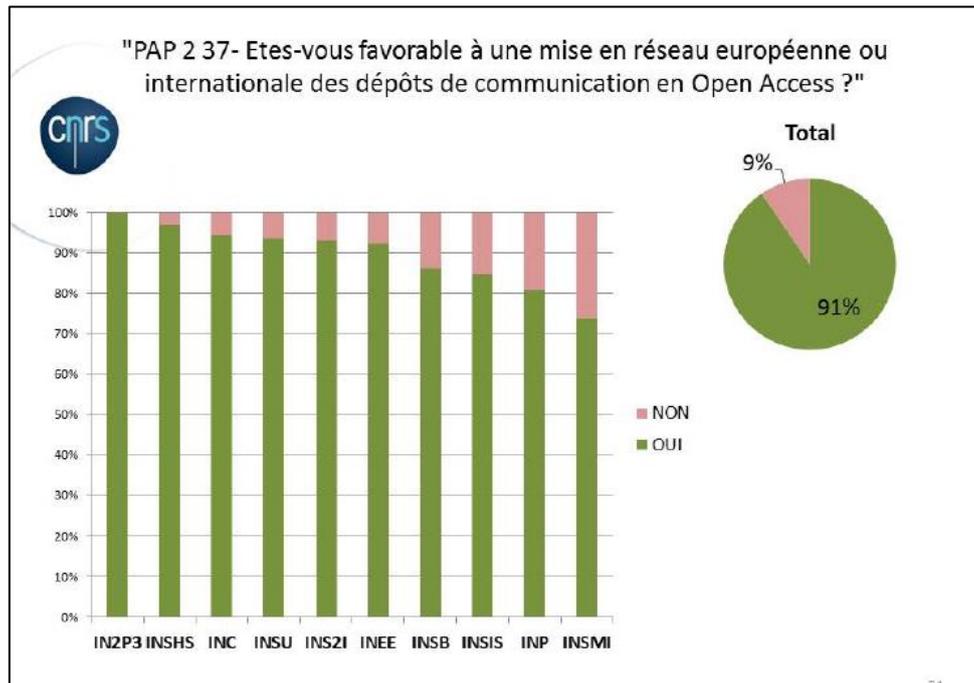
Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 2 30- Estimez-vous nécessaire de développer les outils et services d'Open Access au-delà de l'existant ?	PAP 1-4: Are you in favour of merging existing portals?
NON / OUI	NO / YES
Sans opinion	No opinion
Large positionnement en faveur d'un regroupement	Broad position in favour of mergers
INSMI et INSHS : une demande de ressources spécifiques	INSMI and INSHS: a demand for specific resources
Faciliter l'interdisciplinarité	Facilitate interdisciplinarity

211. In the same way, 91% of researchers are favourable to a Europe-wide or international network of repositories for Open Access communication.

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 2 37- Etes-vous favorable à une mise en réseau européenne ou internationale des dépôts de communication en Open Access ?	PAP 2-37: Are you in favour of creating a European or international network of repositories of papers in Open Access?
NON / OUI	NO / YES

212. A majority of researchers expressed a technical need for digital tools such as online platforms, enabling access to and the sharing of data and results, at least at the national level.

+ The practice of depositing articles in archives or on platforms in specific fields should be generalised.

213. Concerns about the rules governing the use of platforms were also expressed:

- the need for a definition of what a platform actually is;
- definition of the scope of the rights of researchers using these platforms;
- format and interoperability of the data;
- relevance of metadata.

+ Researchers have expressed the need for:

- a "one-stop-shop" for scientific knowledge;
- legal regulation of the platforms.

2.1.4 The numerical assessment of peer-reviewed results and publication metrics

214. The notion of assessment refers to two distinct procedures:

- peer review, i.e. the reading and evaluation of a researcher's work, usually by two or three experts, before publication in a scientific journal;
- the procedure for assessing a researcher by a research unit for the purpose of internal promotion or during a recruitment procedure.

215. **Peer review.** Researchers are strongly in favour of this system for the assessment of scientific work before publication in scientific journals. Peer reviewing is generally performed by researchers working in the same field as that of the proposed article. The peers are responsible for judging the scientific quality of the article, and the methodological validity of the demonstration described. Their opinion decides whether the article will then be accepted or rejected, with the final decision lying with the editorial board.

216. However, the way this assessment is organised has been criticised, for example in the article "*Peer review*": *déontologie et fraudes chez les chercheurs scientifiques* (Peer Review: ethics and fraud among scientific researchers)⁴⁹, published on 2 February 2014:

- a preliminary shortlist is usually drawn up by the editorial board of articles to be submitted for peer review;
- there is often only a single peer reviewer, who performs the evaluation on a voluntary basis on behalf of the publisher;
- this peer is often overwhelmed with many requests for assessment and must make a selection of the articles to assess;
- the result is a significant loss of articles and scientific knowledge;
- the risks related to the spread of "article processing charges" for the publication of an author's article, in particular in terms of quality.

217. **Assessment.** Researchers from CNRS must submit an activity report and a comprehensive list of their scientific output in view of their assessment by their section(s) and/or an interdisciplinary commission. This is a statutory obligation stipulated by Articles

⁴⁹ <http://www.contrepoints.org/2014/02/02/155325-peer-review-deontologie-et-fraude-chez-les-chercheurs-scientifiques>

10, 29 and 49 as amended of Decree no.83-1260 of 30 December 1983 laying down the statutory provisions common to all officials of public institutions in science and technology.

218. Publication is a criterion for the recruitment and promotion of researchers, and the financing of research projects.

219. However, this quantitative criterion of the number of publications has been criticised, and the CNRS Ethics Committee has issued the following recommendation: "Qualitative assessment by peers must remain the rule,"⁵⁰ bearing the crucial implication for assessment that publications must actually be read.

220. Posting data online and sharing them should be taken into account as part of the assessment of researchers. On this point, the CNRS Ethics Committee recommends that:

- "The contribution to the work of data sharing must be recognised in assessments and decisions concerning the promotion of researchers. To facilitate this recognition, the COMETS recommends that appropriate indicators be created and that a section on these activities be added in the activity report and the annual activity sheet concerning researchers."⁵¹



New criteria for the evaluation of researchers will need to be introduced, and in any event publication in Open Science will need to be taken into account.

221. **Need for metrics.** A need for changes in publication metrics has also been expressed:

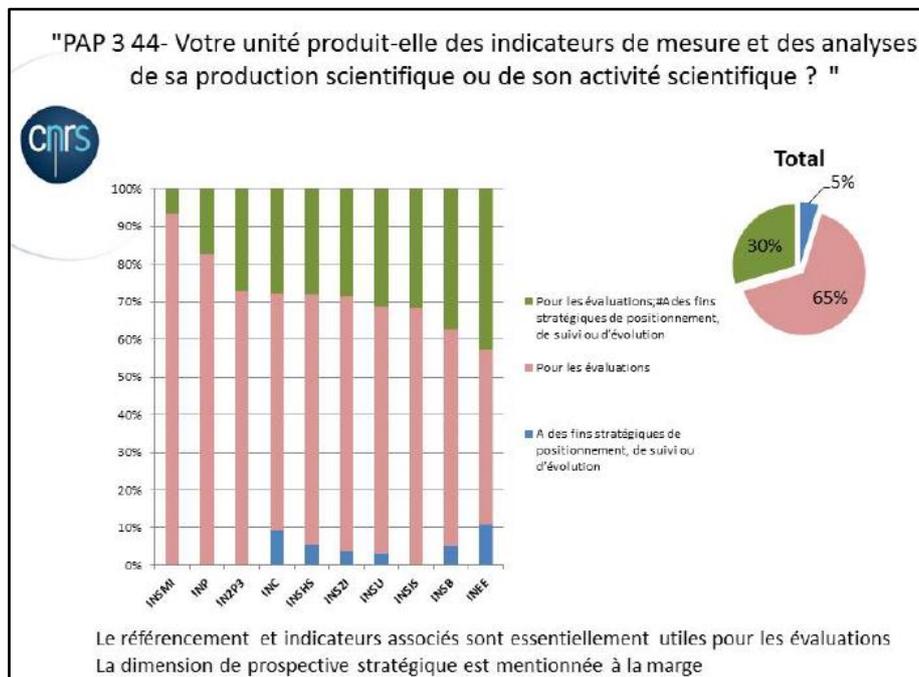
- Most sections express the "demand for consolidation and a sharing of new data management practices by publishers, together with tools for analysing results, and for innovative publication metrics."⁵²

⁵⁰ Opinion of CNRS's COMETS, "Ethical problems for the evolving occupations of public research", 12 February 2014

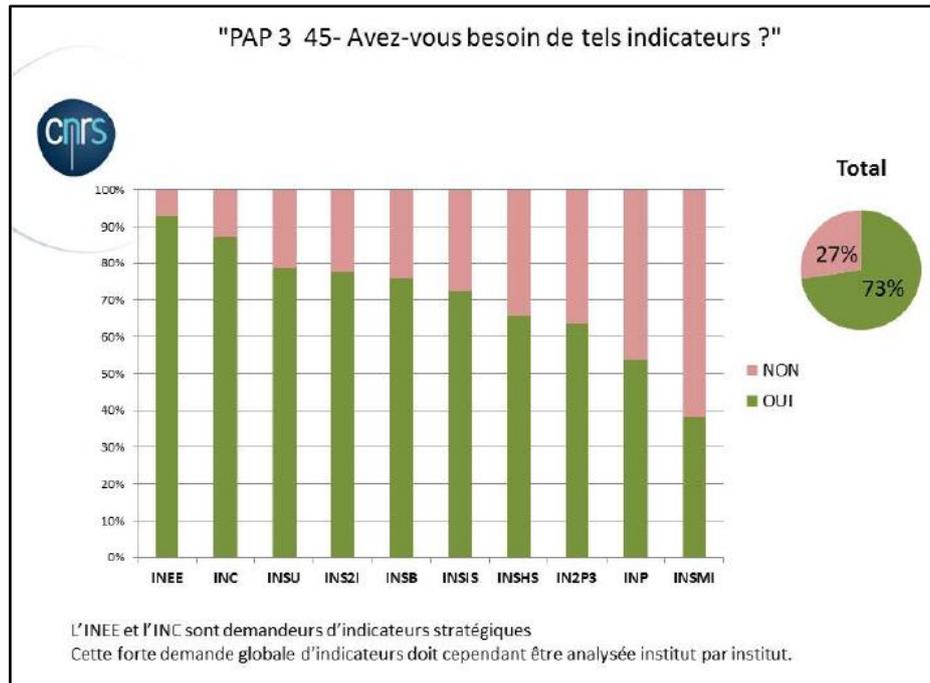
⁵¹ Self-referral by the COMETS, "The ethical issues of scientific data sharing", by the Data Sharing Group, 12-12-2014

⁵² Results of the survey on the uses and needs of STI - CNRS - March 2015, page 5

Results of the survey on the uses and needs of STI - CNRS - March 2015



PAP 3 44- Votre unité produit-elle des indicateurs de mesure et des analyses de sa production scientifique ou de son activité scientifique?	PAP 3-44: Does your unit produce measurement indicators and analyses of its scientific output and scientific activity?
Pour les évaluations à des fins stratégiques de positionnement, de suivi ou d'évolution	For assessments and for strategic purposes concerning positioning, monitoring and evolution
Pour les évaluations	For assessments
A des fins stratégiques de positionnement, de suivi ou d'évolution	For strategic purposes concerning positioning, monitoring and evolution
Le référencement et indicateurs associés sont essentiellement utiles pour les évaluations	Referencing and the corresponding indicators are mainly useful for assessments
La dimension de prospective stratégique est mentionnée à la marge	The strategic outlook aspect is only mentioned marginally



PAP 3 45- Avez-vous besoin de tels indicateurs ?	PAP 3-45: Do you need such indicators?
NON / OUI	NO / YES
L'INEE et l'INC sont demandeurs d'indicateurs stratégiques	There is a demand from INEE and INC for strategic indicators
Cette forte demande globale d'indicateurs doit cependant être analysée institut par institut	This strong overall demand for indicators must however be analysed institute by institute

222. It is therefore necessary to “create platforms equipped with tools to calculate publication, data and analysis metrics and develop networking approaches to link platforms and e-infrastructures”⁵³.

2.1.5 Publication and embargo periods

223. Open Science is not incompatible with the publication by a publisher of scientific articles – on the contrary, these two modes of dissemination are complementary, with publishers often providing services concerning the data.

224. **Publication.** The publisher, in addition to the peer review, “curates” the article in terms of layout, insertion in a review, posting online and dissemination. The publisher’s work is therefore complementary to the work of posting the article online and the processing of the transdisciplinary and multi-corpus data.

⁵³ Results of the survey on the uses and needs of STI - CNRS - March 2015, page 39

225. Furthermore, in countries and disciplines where Open Access has become an accepted part of the practices of publishers and researchers, there has been no decline in turnover of scientific publishers, who on the contrary gain visibility by providing their articles to scientific communities.

226. **Embargo periods.** Research work involves reading articles, gathering knowledge, and comparing old and recent data. To do this, scientists need instant access to articles published by scientific publishers.

227. During the hearings as well as in the recommendations of the Scientific Board and the Ethics Committee, researchers unanimously stated that, because science is a common good of humanity, no embargo period should apply⁵⁴.

228. Moreover, the principle of the distinction made in the bill between "science, technology and medicine" and "human and social sciences" (HSS) for the embargo period has been challenged. A longer embargo period for HSS does not fit with the needs of researchers, as expressed particularly by Maya Bacache-Beauvallet, Françoise Benhamou and Marc Bourreau in a report by the French Institute of Political Studies No. 11 of July 2015.

229. This report, entitled "*Les revues de sciences humaines et sociales en France : libre accès et audience*" (human and social science journals in France: free access and readership) concludes on Page 5 that the results of the study "therefore point to the imposition of a relatively short embargo period, compared to the periods discussed in the public debate for HSS studies".

Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015

"The key factor in science is the immediacy of research. Each data item has a certain lifecycle, but the fresher the better. For science to advance, the embargo period should not be an obstacle to the dissemination of results. However, depending on the scientific community concerned and for various reasons (e.g. competition, assessment process under way, etc.) an embargo period may be introduced."

230. However, in order to ensure the transition to Open Science and preserve the economic interests of publishers – especially French publishers – the scientific communities agree in granting publishers an embargo period.

⁵⁴ Recommendation of the Scientific Board of 25-9-2015: "Scientists must be able to make these data and results available for no fee, in digital form, a priori without any embargo period imposed by publishers."

231. This period of exclusivity must be "long enough to enable digital journals to survive financially, and short enough to significantly broaden the readership able to access the article in its open-access version⁵⁵."

ABES hearing: Jérôme Kalfon, 5 October 2015

"We quit wishful thinking and return to reality."

232. The French Digital Council also stressed in its contribution the need for a "short embargo period to allow the publisher a degree of commercial activity".

233. The embargo periods accepted by researchers are the maximum deadlines provided for by the Recommendation of the European Commission (C(2012) 4890):

- "there should be open access to publications arising from publicly funded research as soon as possible, preferably immediately and in any case no later than six months after the date of publication, and twelve months for social sciences and humanities."

234. In any case, the embargo periods cannot be longer than those stipulated by the national legislation of our European neighbours:

- in Germany: embargo period of 12 months with no distinction between disciplines;
- in Spain: filing in an institutional archive as early as possible, without exceeding 12 months, with no distinction between disciplines.

235. The French National Assembly's "Working Group on rights and freedoms in the digital age" presented a report to Claude Bartolone, President of the National Assembly, on 8 October 2015 with recommendations in favour of Open Access, including:

- "Make publicly-funded scientific publications freely accessible" after "a period of exclusivity of 6 to 12 months".

236. In its revised version following that contribution, the Digital Republic Bill took these arguments into account and returned to reasonable embargo periods of 6 and 12 months.

⁵⁵ [Report](#) by the French National Assembly's "Working Group on rights and freedoms in the digital age" submitted on 8 October 2015, Page 241



Researchers need access to the latest state of knowledge. If an embargo period can be defined in the framework of a compromise with the publishers, it must not exceed the maximum time limits provided for in the Recommendation of the European Commission (C(2012) 4890) and the deadlines observed in other countries, as otherwise French research runs the risk of marginalisation and discrimination.

The principle of a distinction between the exact sciences and the human and social sciences has been challenged.

2.1.6 Analysis and exploration of corpora of data

237. **Historic right.** Researchers have always analysed and explored data, as an essential part of the scientific approach. Only the tools for observation have changed, with digital tools making it possible to scan a greater volume of information in less time.

238. **Right of observation.** Researchers are agreed that the freedom to explore data is a natural right of observation that must not be restricted. TDM enables the observation of scientific objects in the same way as a microscope does.

Hearing of Jean-Marie Pierrel and Gregory Colcanap, 24 September 2015

"TDM is a right of observation of scientific objects, indispensable for science."

"IT is just a particular kind of tool for observing data."

Contribution of the French Digital Council

"TDM is not in itself a new activity.

It just means reading and extracting information and meaning from documents. It is not really so different to gathering information manually, which has been the way research has proceeded since the birth of science."

239. In this framework, researchers need free and open access to digital tools for processing data:

- "A demand for the consolidation and sharing of new management practices for data and published material, together with tools for analysing results, and innovative tools for calculating publication metrics"⁵⁶;
- "Make tools and services available to facilitate the exploitation of research data"⁵⁷.

Contribution of the CNRS Scientific Board

"This requirement to make data available extends to added-value services (massive processing such as Big Data, data mining, relationship with metadata, interoperability) which must also be public and open-access to avoid any misappropriation."

Hearing at Pierre et Marie Curie University: Jean Chambaz and Paul Indelicato, 9 June 2015

"Scientific data are constantly evolving, and although data science will never replace the scientific method, data and even more so the reuse of data are at the heart of this new approach."

240. **Risks.** While this right of observation seems theoretically to be an acquired right, it is currently being questioned:

- by scientific publishers who sell access;

⁵⁶ Results of the survey on the uses and needs of STI - CNRS - March 2015, Page 5

⁵⁷ Results of the survey on the uses and needs of STI - CNRS - March 2015, Page 35

- by the many questions and differences of opinion expressed on whether text and data mining practices are compatible with copyright and the rights of whoever produces the database.

241. Risk of appropriation. The production of value by the use of these data-processing techniques, particularly today with TDM, must not be pre-empted by commercial publishers. However this commercialisation is already happening through:

- the way researchers are obliged to use the publisher's own application programming interface (API) to process that publisher's data;
- the way that, under the General Terms and Conditions of Use of the API or the subscription contracts, the publisher reserves the right to distribute all the results from the use of TDM techniques, the "TDM output" or "user-generated content", to third parties.

Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015

"Publishers seek to protect themselves by controlling the possibility of TDM via Application Programming Interfaces (APIs), under contractual clauses (Creative Commons licences, restricting the number of 'searchable' words)."

242. This appropriation prevents researchers from conducting transdisciplinary searches, from the corpora of several publishers, and deprives science of fundamental knowledge and of scientific and added value created in user-generated content.

Hearing at Pierre et Marie Curie University: Jean Chambaz and Paul Indelicato, 9 June 2015

" It is essential that scientists should be able to carry out TDM on scientific data. If this possibility depends on publishers, researchers will be subject to strict controls with no apparent limits."

243. What is at stake. The right to search and process data is a major issue for science, research and innovation in that it enables scientists to identify new research subjects, produce new knowledge and address economic, social and societal issues. Text and

data mining (TDM) is one of the technical applications of this right to explore data, on which much attention is focused today⁵⁸.

Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015

"The relationship with publishers is central to the concerns of researchers, particularly as regards TDM. TDM will be the ultimate research tool in the years to come."

244. Exploration rights thus open up enormous opportunities in terms of the exploitation of new knowledge and all that this implies regarding innovation, growth and employment, and there is no reason why economic forces in the sector should not also benefit. "Data-mining and similar services play a considerable role in the scientific exploitation of open access data and texts"⁵⁹.

245. The scientific and economic issues are especially sensitive in that TDM is practised around the world and is governed by different standards in different countries, including within Europe (see 2.2 "Open Science around the world"). If rules authorising TDM practices are not adopted, there is a significant risk that a two-tier research system could arise within the European Union, which would threaten certain research partnerships between France and England, for example.

+ The new Act must assert the right of observation of scientific data by the use of digital search techniques as a universal principle, if French research is not to be penalised.

2.1.7 Digital Intellectual Property and recognition of authorship

246. Open Science does not mean that the author need renounce all moral rights to ownership. Researchers want to retain their right to authorship, which is particularly important as the number of citations is part of the assessment criteria for researchers.

247. "Good practice" for researchers using STI must include citing the name of the author and the article for quotations or in the bibliographies of study reports. Ethical rules could back up these good practices.

⁵⁸ In 2001, the Massachusetts Institute of Technology presented data exploration as one of the ten emerging technologies that would "change the world in the 21st Century". (*Data mining et statistique décisionnelle – l'intelligence des données* [Data Mining and Statistics for Decision Making]. Stéphane Tuffery, Editions Technip 2012)

⁵⁹ Recommendation of the CNRS Scientific Board

+ Authorship is an intangible aspect of an author's moral rights. It must be strictly respected and strengthened by ethical rules.

2.1.8 The limits of exploitation and Open Science

248. Open Science must not hamper the economic aspects of research.

249. The provision of scientific data on Open Science platforms must not jeopardise:

- the exploitation of data, in particular through patents;
- respect for secrecy and specific provisions, such as restricted access areas;
- respect for contractual rules of confidentiality.

250. The way research data are made available must also be organised to take into account the different practices of different scientific communities.

+ Open Science must protect legitimate interests, including those related to the exploitation of innovations and to the protection of secrets, and adhere to the practices of the different scientific communities.

2.1.9 Towards an ethical charter for digital scientific and technical information (STI)

251. The researchers expressed a need for regulation at different levels:

- at the legal level, in order to make these practices secure;
- at the ethical level, by drawing up a "Charter of STI Ethics" laying down "Ethical principles designed to transcend instrumental considerations and affirm the goals of public research in a global context of Open Science"⁶⁰.

252. The Ethics Committee has also argued that "confronted by this dynamic movement of data encouraged by their supervisory authorities and by their community, researchers must:

- be aware of their individual, deontological⁶¹ and ethical responsibilities, with respect to the community to which they belong;
- abide by the international undertakings of the institutions to which they belong;

⁶⁰ Results of the survey on the uses and needs of STI - CNRS - March 2015, Page 59

⁶¹ "Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences", *Committee on Responsibilities of Authorship in the Biological Sciences*, National Research Council. National Academy of Sciences.

- participate in the definition of ethical principles specific to their discipline in the field of data sharing and of Big Data in general⁶²."

2.1.10 Toward a radical change of paradigm?

253. More radical proposals have also emerged. Considering that the existing economic model is changing before our eyes, some have proposed moving toward a thorough structural change.

254. **Model 1.** The first proposal is to consider scientific publication as open in principle, devoid of economic rights, immediately accessible, with the author alone retaining authorship rights. Under this model, scientific publishers provide a "labelling" service (peer review), disseminate knowledge and develop other services, and are paid for this.

Hearing: Jérôme Kalfon, 5 October 2015

"Publishers must be paid for the work they perform, on a 'jobbing' basis."

"Scientific knowledge is a special field, a factor of collective enrichment and development, and it must be possible to access it freely with no economic rights."

255. **Model 2.** The second proposal is based on two fundamental principles:

- all scientific data must remain under the control of scientists;
- the services around these data are open to competition.

256. According to this model, the concept of data is very wide and covers any article (published or not), web page, communication, blog, video, photo, research data including data measured by sensors, either automated or human numerical simulation, lab book, source code, query, etc.

257. These data do not give rise to any property rights and must be freely accessible and freely reusable within the limits of scientific ethics.

258. The services developed around data and especially search techniques can be developed by commercial publishers and open to competition, because the data generated by processing will be open access and will enable scientists to verify the results of the studies concerned on the basis of these data.

⁶² Self-referral by the COMETS "The ethical issues of scientific data sharing", by the Data Sharing Group, 12-12-2014

259. However, the requests made to a publisher today are biased by the search algorithm. For example, it is currently impossible to query *h*-index⁶³ on Google Scholar. In addition to the principle of Internet neutrality there is a need to assert a right to transparency concerning data.

INRIA hearing: Claude Kirchner, 15 October 2015

"A right to transparency must be affirmed and this right exists only if the researcher is able to check the data; this verification is an essential component of the scientific method."

"This right to transparency in queries extends to all information covering all the questions raised by researchers (queries, discussions, etc.)."

⁶³ *h*-index aims to quantify the scientific productivity and impact of a scientist on the basis of the number of citations of his/her publications. Source: Wikipedia.

In sum

260. The following diagram summarises the needs of researchers for the use of STI as an analytical tool regarding existing technical, contractual, and legal constraints.



2.2 Open Science around the world

261. The principle of making scientific data "open" has been declared several times by European bodies as well as by international forums.

262. Many countries, and in particular those where scientific research is seen as helping to drive the economy, have adopted legislation in favour of Open Science and/or TDM.

263. The march toward Open Science seems to have an historic inevitability.

2.2.1 The European Union clearly in favour of Open Science

264. **European Commission Recommendation, 2012.** In a Communication issued on 17 July 2012 (C(2012)4890) on access to and preservation of scientific information, the European Commission recommends that States:

- "define clear policies for the dissemination of and open access to scientific publications arising from publicly funded research" and ensure that there is "open access to publications arising from publicly-funded research as soon as possible, preferably immediately and in any case no later than six months after the date of publication, and twelve months for social sciences and humanities".
- "define clear policies for the dissemination of and open access to research data arising from publicly funded research" and ensure that "research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures".

265. In addition, stressing the fundamental importance of works published in the way researchers are assessed, the Recommendation suggests revising the system for academic advancement:

- by rewarding "researchers who participate in a culture of sharing the results of their research, in particular by ensuring open access to their publications";
- by "developing, encouraging and using new, alternative models of career assessment".

266. **Horizon 2020.** In the framework of the Horizon 2020 programme (an EU financial Instrument for the development of an Innovation Union, which provides funding for research and innovation for the period 2014-2020), the European Commission has made free access to scientific publications a general principle.

267. The "Guidelines on Open Access to Scientific Publications and Research Data" were drafted in the framework of this programme and first published on 16 December

2013⁶⁴. They provide that from 2014, all scientific publications arising from projects financed or co-financed in the framework of the Horizon 2020 programme will need to be made available in open access:

- either immediately by the publisher, who will publish them online (approach known as "Gold Open Access"); the costs of publication incurred will be reimbursed by the European Commission;
- or by the researchers, 6 months at the latest after publication (12 months for the human and social sciences), via open access archives (approach known as "Green Open Access"). Publications and scientific data from publicly-funded research will be available to a wider public more quickly, which will enable researchers and businesses to exploit them more easily.

268. **Reda Report.** The European Parliament adopted on 9 July 2015 the report on the implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, whose rapporteur was Julia Reda. In the framework of the revision of the Information Society Directive (InfoSoc Directive) this report:

- suggests "that the Commission examine and propose solutions for automated analytical techniques for text and data ("text and data mining") for all purposes, provided that permission to read the work has been obtained (Point 48 of the final report);
- "calls for an exception for research and education purposes, which should cover not only educational establishments but also accredited educational or research activities", including informal education (Point 51 of the final report).

269. **Revision of copyright.** The European Commission also announced on 9 December 2015 its "first steps to broaden access to online content" and outlined "its vision to modernise EU copyright rules". In this context "The Commission intends to work on key EU exceptions to copyright" and, in particular, "will revise EU rules to make it easier for researchers to use 'text and data mining' technologies to analyse large sets of data."

⁶⁴ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2.2.2 Abroad: creating a legal framework for Open Science

270. There have been many legal and other initiatives around the world in favour of opening up access to scientific data and text and data mining operations on scientific data.

271. **United States.** The USA was one of the first countries to introduce legal provisions (Consolidated Appropriations Act, 2008) on making research work funded by the National Institutes of Health (NIH) publicly available. This Act provides that all articles published in journals as the result of work funded by the NIH must be deposited in the NIH's own online open archive, the "National Library of Medicine's PubMed Central". Contracts with publishers must allow it explicitly. There is even a list of publishers who deposit articles systematically, thus requiring no action by the researcher. This Act specifies that articles must be so deposited at the latest 12 months after the actual date of publication in a journal.

272. In February 2013, the "Fair Access to Science and Technology Research Act" (FASTR) was submitted to Congress.

273. In May 2013, the "networking and information technology development program" added a supplementary item to the Federal Budget for 2014. This supplementary item presents several research and development programmes in the field of technological development and in particular the sharing of knowledge in the public sector.

274. There have been initiatives such as that taken by the University of Berkeley, which set up a fund dedicated to financing articles for free access.

275. American judges also explicitly recognised that TDM operations can be covered by the "fair use" exception, in the case of Authors Guild v. Google (14 November 2013), in the framework of the implementation of a vast programme to digitise books and build up a universally accessible digital library.

276. Finally, the very recent Google Books decision of 16 October 2015 has broadened the scope of TDM for American researchers. In 2004, Google had launched a vast project to digitise books; in 2005, a group of authors and publishers challenged the search engine for breach of copyright. The US Court of Appeal recognised Google's right to benefit from the "fair use" exception, considering that the service offered by Google to users was likely to provide a benefit to society in terms of access to knowledge, and arguing that the exclusive rights of the authors should be set aside. This Decision therefore grants Google, its competitors and also public institutions the right to digitise

data and provide data-mining services. With this ruling, the United States gives its researchers a significant advantage by enabling them to digitise very large legally accessible datasets, to pool these corpora and to develop search and algorithmic data-processing systems⁶⁵.

277. United Kingdom. The United Kingdom is a leader in the development of Open Access. The Finch Group, an independent working group set up by the Department for Business Innovation and Skills, published a report in June 2012 on how the results of research could be made more accessible.

278. In September 2013, the British Parliament published a report on Open Access and seemed to adopt a dual solution for "green" and "gold" access to scientific work.

279. Lastly, researchers whose work is funded in whole or in part by the British non-governmental *Wellcome Trust* funding agency, which has clearly declared a position in favour of free access, must deposit an electronic copy of any article accepted for publication in a peer-reviewed journal in PubMed Central and in PubMed Central UK. Articles must be deposited as quickly as possible, and no later than 6 months after the date of publication.

280. In 2014, the United Kingdom also introduced an exception for data exploration (the right to make a copy exclusively for the purpose of TDM operations for non-commercial research without requiring agreement from or financial compensation for the copyright holders)⁶⁶ on the basis of "fair dealing"⁶⁷.

281. US-UK. Moreover, the United States and the United Kingdom have together launched an initiative entitled "UK-US Global Innovation Initiative" which aims to facilitate academic collaboration between the two countries as well as with emerging countries over the next five years.

282. Germany. Germany has created a platform for filing scientific contributions. The federal agency subsidises the purchase of journals through grouped orders, imposing the

⁶⁵ Article entitled "*Comment l'affaire Google Books se termine en victoire pour le Text mining*" [How the Google Books affair ended with a victory for text mining] 21-10-2015 <http://scinfolex.com/2015/10/21/comment-laffaire-google-books-se-terme-en-victoire-pour-le-text-mining/>

⁶⁶ Copyright, Designs and Patents Act, Article 29A "Copies for text and data analysis for non-commercial research", October 2014

⁶⁷ Copyright, Designs and Patents Act 1988, Art 29 "Fair dealing with a literary, dramatic, musical or artistic work for the purposes of research or private study does not infringe any copyright in the work or, in the case of a published edition, in the typographical arrangement."

condition that publishers accept that articles be free to access after a period of 12 months from the date of publication. To date, acquisition remains limited to certain niche subscription packages, which cost less than the main “big deal” subscription packages.

283. In addition, a bill⁶⁸ was proposed in February 2013 and adopted in November 2013 amending the German Copyright Act (*Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen of Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes*). This text modifies copyright law in the field of science publishing by introducing “a right of secondary exploitation” (*Zweitverwertungsrecht*).

284. The adopted text has been translated as follows:

- “The author of a scientific article, arising from research funded at least 50% by public sources and published in a journal appearing at least twice a year, has the right, even if they have transferred an exclusive exploitation right to the publisher, to make the article publicly accessible in the accepted version of the manuscript, after a period of twelve months following its first publication, to the exclusion of any commercial purpose. The source of the first publication must be indicated. Any waiver agreement to the detriment of the author is null and void.”⁶⁹

285. This amendment provides a legal framework for a right of secondary exploitation of scientific texts arising from teaching or research that is at least 50% publicly funded, even if an exclusive exploitation right has been transferred to a publisher. It lays down the practical conditions for exercising this right of secondary exploitation in such a way as to take into account both the interests of the publishers (12-month embargo period, publication in scientific periodicals, availability for non-commercial purposes) and those of the authors (inalienable rights of secondary exploitation, strengthening the position of the author by a second publication and distribution of results).

2.2.3 International bodies: the tendency is in favour of Open Science

286. Finally, positions in favour of open access to scientific data have been adopted in international forums such as UNESCO, the G8 and the OECD.

⁶⁸ *Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen des Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes*

⁶⁹ French translation of the adopted text: <http://openaccess.inist.fr/?Point-sur-le-Libre-Acces-en>

287. **UNESCO**. In the framework of the World Conference on Science of 1 July 1999 organised by UNESCO, entitled "Science for the Twenty-First Century, a new Commitment", two documents were endorsed:

- "The Declaration on Science and the Use of Scientific Knowledge": this text asserts the need to share data and scientific knowledge and to promote and facilitate cooperation, "essential for undertaking scientific work and for translating the results of scientific research into tangible benefits for society"⁷⁰. The text states that "Parliaments and governments should be invited to provide a legal, institutional and economic basis for enhancing scientific and technological capacity in the public and private sectors and facilitate their interaction"⁷¹;
- A "Science Agenda - Framework for Action": this document defines the guiding principles "for dealing with the problems, challenges and opportunities confronting scientific research". "Sharing scientific information and knowledge" is one of these principles. The document calls on the different participants in research to collaborate at the international level⁷² and to facilitate "the publication and wider dissemination of the results of scientific research (...) through training, the exchange of information and the development of bibliographic services and information systems better serving the needs of scientific communities around the world"⁷³. The text also suggests that research institutions should encourage the use of new information technologies, "through the development of electronic publishing and the establishment of virtual research and teaching environments or digital libraries"⁷⁴. Lastly, the text exhorts governments to ensure that "relevant infrastructure and other costs are adequately covered in research budgets" and that appropriate legal frameworks are set up⁷⁵.

288. **G8 London 2013**. In a joint declaration on 12 June 2013, the G8 Science Ministers stated that international scientific collaboration is a new global challenge requiring the modification and improvement of research infrastructure in order to make published, peer-reviewed scientific data globally accessible. The G8 Ministers propose a framework for:

⁷⁰ "Declaration on Science and the Use of Scientific Knowledge", UNESCO, Point 38

⁷¹ Ibid, Point 37

⁷² "Science Agenda - Framework for Action", UNESCO, Point 17

⁷³ Ibid, UNESCO, Point 19

⁷⁴ Ibid, UNESCO, Point 20

⁷⁵ Ibid, UNESCO, Point 21

- a global research infrastructure;
- open scientific research data;
- expanding access to scientific research results.

289. **OECD Report.** The OECD published a report entitled "*Enquiries into Intellectual Property's Economic Impact*" in August 2015. Chapter 7 "*Legal Aspects of Open Access to Publicly Funded Research*" provides an overview of different national legal regimes on access to, and the distribution and use of the results of publicly-funded research in the framework of open access.

290. The report also draws attention to two other questions:

- the problem of open access in the context of partnerships between the public and private sectors;
- the regimes governing text and data mining.

291. These positions in favour of Open Access taken by bodies outside France reflect the natural evolution of scientific publishing toward the free and broad provision of scientific data. France has also repeatedly affirmed its commitment to Open Access.



Open Science is an international movement marked by commitments from and strong positions taken by European, supranational and other foreign institutions.

The French research community must declare its participation in this movement, or risk being discriminated against and left marginalised and uncompetitive.

2.3 The key concepts of Open Science

292. The new right to Open Science is one of a body of existing concepts and legitimate interests that must be preserved or indeed strengthened.

2.3.1 A right to Openness: an unstoppable international movement toward openness and data sharing

293. Open data, open format, open source, open access, open process are some of the areas in which a common philosophy is advocating the sharing and the free reuse of knowledge.

2.3.1.1 Open data

294. The movement to open up access to data, otherwise known as "Open Data", first emerged in 1957-58 in the United States with the creation of the World Data Center System. The movement acquired legal status in 1966 with the passing of the *Freedom of Information Act*⁷⁶. Then, in 2007, an amendment to this Act by the *Open Government Act* made the concepts of the transparency, governance and opening up of public data central to the work of the American government.

295. This movement was echoed in Europe, particularly in the public sector. In the United Kingdom, a project similar to the one conducted in the United States was officially launched in January 2010⁷⁷. In France, the release of public sector data has attracted considerable attention within the civil service since 2009. The Etalab Mission was created in 2011 under the authority of the Prime Minister. Its brief is to oversee application of the policy in favour of openness and sharing of public data.

296. In this framework, Etalab administers the single interdepartmental webportal at data.gouv.fr intended to collect and to make freely available all public information concerning the State, its public institutions and, if they wish, the various local authorities and entities under public or private law, responsible for a public service mission.

297. Many towns are also developing open data platforms making public information available to their citizens, in particular in the areas of culture, citizenship, transport, town planning, the environment, public finances, the administration, services and parking.

2.3.1.2 Open format

298. The concept of an "open standard" or "open format" is defined by Article 4 of Act no. 2004-575 of 21 June 2004 as "Any communication, network or exchange protocol and any interoperable data format whose technical specifications are public and with no restriction of access or of implementation".

299. The Act no. 2015-1779 "relative to the freedom and conditions of the reuse of public sector information" as well as the "Digital Republic Bill" reaffirm the need to make data available in an "open and easily reusable standard".

⁷⁶ <http://www.foia.gov/>

⁷⁷ <http://data.gov.uk/project>

2.3.1.3 Open source

300. Open licences enable data, databases, digital creations and software to be made available to third parties under licences granting varying degrees of freedom.

301. The most widely-used licences, especially as regards open data, are the following:

- The Etalab licence;
- The Open Database licence (ODbl);
- Creative Commons licences;
- The Public Domain Dedication and Licence (PDDL);

302. In the field of software, the most widely-used licences used are the *CEA CNRS INRIA Logiciel Libre* (CeCILL) licence, the GNU-GPL licence, the MIT licence, etc.

2.3.1.4 Open access and Open process

303. **Definition.** "The Open Access movement is a position taken by the international scientific community, requiring the results of scientific research be made available openly and freely"⁷⁸.

304. The Open process implies the right to the free observation of data by the use of digital processing, analysis or exploration tools, such as text and data mining (TDM).

305. **International commitment.** This position in favour of open access first saw the light of day nearly 15 years ago:

- in "The open letter of the Public Library of Science" of 2001 encouraging publishers to establish an online public library to ensure free access to research documents published in their journals;
- in the framework of the "Budapest Open Access Initiative" (14 February 2002)⁷⁹. This global campaign advocated free access to all new peer-reviewed research⁸⁰.

306. Today, this movement is building momentum and can particularly be seen in:

- publications:
 - o Opinion piece: "*Favorisons la libre diffusion de la culture et des savoirs*" (Let's all agree to facilitate the dissemination of culture and knowledge) and the associated petition, published in *Le Monde* on 10/09/2015;

⁷⁸ http://corist-shs.cnrs.fr/gold_open_access

⁷⁹ <http://openaccess.inist.fr/?Initiative-de-Budapest-pour-l>

⁸⁰ <http://www.budapestopenaccessinitiative.org/>

- French Digital Council: Opinion and Report on Digital Ambition, "For a French and European policy to address the digital transition" (submitted to the French Prime Minister on 18 June 2015);
 - SavoirCom1: <http://www.savoirscom1.info/2014/10/savoirscom1-soutient-le-projet-dune-charte-universelle-de-lopen-science>;
 - INIST and the website <http://openaccess.inist.fr/>
 - Science Commons <http://scoms.hypotheses.org/458>;
- in conferences or symposia:
- CNRS symposium on "The dynamics of scientific publishing, the information industry, and documentation. An Agenda for 2015 for open publicly-funded science" held on 4 and 5 November 2014;
 - CERN/CNRS (DIST) Workshop on Multi-core platforms for science;
 - Symposium on "Innovation and governance of STI in the ESR", 18 and 19 March 2014;
 - GFII (French professional think-tank for BtoB Information), Open Access group on 15 October 2014;
 - Congress of the ADBU (Association of Directors of University Libraries) on 2 and 4 September 2014: "The law is under challenge by developments in STI and the needs of science";
 - The Paris Book Fair, "Facts & Knowledge" area, 21 March 2015, Talk on "Towards Open Science: What are the impacts on scientific publishing?";
- by the creation of archiving platforms, including:
- HAL: an open archive that has recorded 9 million unique visitors in 2014;
 - HAL-SHS: the social sciences version of HAL;
 - ArXiv;
- by the creation of Open Access archiving platforms, including:
- Openedition.org and revues.org
 - Persée;
- by self-help practices:
- <http://rue89.nouvelobs.com/2015/09/08/hashtag-clandestin-partager-science-inaccessible-261102>: hashtag created ([#IcanhazPDF](#)) helping scientists access scientific articles.

2.3.2 The French Research Code: the basis of digital STI legislation

307. The Research Code lays down an institutional framework for the organisations that participate in scientific research in France but does not define the principles or values of the scientific community.

308. Article L.112-1 of the Research Code defines the goals of public research, among which are:

- "sharing and disseminating scientific knowledge, with priority for open access formats";
- and "organising Open Access to scientific data".

Article L112-1 [En savoir plus sur cet article...](#)
Modifié par [LOI n°2013-660 du 22 juillet 2013 - art. 16](#)

La recherche publique a pour objectifs :

- a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;
- b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;
- c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;
- c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;
- d) La formation à la recherche et par la recherche ;
- e) L'organisation de l'accès libre aux données scientifiques.

Article L112-1

Modified by LAW No.2013-660 of 22 July 2013 – Art 16

The goals of public research are:

- a) The development and advancement of research in all fields of knowledge;
- b) The exploitation of the results of research for the benefit of society, based on innovation and technology transfer;
- c) The sharing and dissemination of scientific knowledge with priority open access formats;
- c bis) The development of an expertise and support capability for associations and foundations, recognised as being for public benefit, and for public policies addressing societal issues, social, economical and sustainability needs;
- d) Training for research and by research;
- e) The organisation of open access to scientific data.

309. In affirming as objectives of public research "the sharing and dissemination of knowledge by giving priority to Open Access formats" and "organising Open Access to scientific data", it provides a basis for the legal protection for Open Science in the digital age.

2.3.3 A need for consistency with rights governing public data

310. **General principles.** The Order no. 2005-650 of 6 June 2005 sets up as a general principle the right of the public to reuse the data held by public legal or physical entities for any purpose whatsoever, and especially for commercial and private purposes⁸¹.

311. The public entities subject to this Act are listed in Article 1 of the "CADA" Act of 1978⁸²: the State, local authorities, the other entities under public law and entities under private law, responsible for a public service mission.

312. Article 11 of the CADA Act provides that "the conditions under which information can be reused are laid down, where appropriate, by the authorities referred to in A and B of this article when they are contained in documents produced or received by:

- teaching and research establishment and institutions;
- cultural institutions, agencies or services."

313. This exception meant that:

- there should be no obstacle to the policy for making research data available, because it favoured the principle of communication and reuse, but under the conditions determined by the institutions themselves;
- the data could be made available under conditions that suited the pace of work of researchers, the practices of each community and the nature of the data;
- reuse was not necessarily a right that could be claimed by any third party, without the nature of the data being taken into account (data from restricted regime areas, ongoing research, data related to know-how or industrial copyright, etc.). As a result, the absence of an automatic right of reuse places limits on the right to communication of data media.

314. **Evolution and the Valter Act.** This exclusion of the principle of free reuse of public information produced by establishments and institutions of education and research was quite simply deleted from the CADA Act by Act no. 2015-1779 of 28 December 2015 on the free access to and the terms of reuse of public sector information ("Valter" Act).

⁸¹ Article 10 of Act No. 78-753 of 17 July 1978 on the freedom of access to administrative documents and the reuse of public information as amended by Order no. 2005-650 of 6 June 2005.

⁸² CADA: Commission d'Accès aux Documents Administrative. Act No. 78-753 of 17 July 1978 on the freedom of access to administrative documents and the reuse of public information.

315. The wording of the Act leaves no room for the principle of making public data available or for reasonable protection of the interests of researchers and intangible public assets.

316. The legal principle adopted also seems to disregard the discussions on the draft Digital Republic Bill and the needs of the scientific communities concerning the regulated circulation of published material and data as expressed in the framework of the public consultation.



The provisions of Act No. 2015-1779 do not allow establishments and institutions of education and research any discretion as regards making publicly available any data they produce. These provisions are not consistent with the needs of researchers and the practices of scientific communities, and do not take into account the nature of the data (data from current research, know-how, restricted regime areas, etc.).

2.3.4 An indispensable legal principle: literary and artistic copyright

317. **Origin.** Historically, the legal recognition of the right of copyright proposed by Beaumarchais was proclaimed by the Constituent Assembly on 13 January 1791 (Law ratified on 19 January 1791 by Louis XVI). This was the first law enacted in the world to protect authors and their rights: it gave authors the exclusive right of authorising the reproduction of their works throughout their lives, and also granted it to their heirs for a period of five years. At the end of this period, the work fell into the public domain.

318. This right of copyright was a reaction to:

- printers-booksellers-publishers who automatically acquired full property rights to the texts they purchased from their authors, often at a ridiculously low price, and went on to exploit them without further consideration for those same authors;
- performers who kept the rewards that should rightfully have gone to the authors.

319. Copyright has thus historically protected the author of a work and not its legal owner.

320. **Copyright and the protection of scientific texts:** scientific literature, one of the components of STI (in the form of articles, books, etc.) is eligible for protection by copyright under Article L. 112-2 of the Intellectual Property Code (CPI) which provides a non-exhaustive list of the works considered intellectual creations and includes:

- 1° books, pamphlets and other literary, artistic and scientific works.

321. Knowledge, know-how and purely technical or descriptive written works do not fall within the scope of copyright, however. They belong to the public domain, are freely reusable by the public without prior authorisation and therefore cannot be the subject of exclusive protection; only the original form in which they were expressed and published may be subject to copyright.

322. Only a work that formalises an idea or item of knowledge can be protected by copyright⁸³. In this respect, Article L. 111-2 of the CPI lays down that:

- "The work is deemed to have been created, whether or not it has been made public, by its very production, even in unfinished form, by the imagination of the author."

323. Only original creations are eligible for protection by copyright.



Scientific texts and publications are subject to protection by copyright if they are original in their form of expression.

However, the knowledge and scientific information contained in these texts are by principle freely accessible.

324. **The researcher: right-holder.** Under the terms of Article L. 111.1 of the CPI, it is the author, understood as a physical person who created the item in question, who owns the intellectual property rights to the work. The same article specifies that neither the employment contract, nor the contract to supply the work can set aside this principle.

325. The rule is invariable regardless of the public or private status of the contracting authority. However, the rights of copyright of public sector employees can be adjusted in the interest of public service. While recognising that authors who are public sector employees should have copyright on their work, the Act modifies this in the interest of the public service provided by the body employing the researchers:

- on the one hand, by restricting the scope of the moral rights of their employees:
 - o the right to disclosure is limited;
 - o researchers who are public sector employees may not object to the modification of their work when this is decided in the interest of the public service;

⁸³ Court of Cassation (French supreme court of appeal), Civil Chamber 1, 17-10-2000, RG No. 97-20820: "The protection of an idea as an intellectual creation supposes that the work springs, even if in unfinished form, from the imagination of the author."

- researchers who are public sector employees may not exercise their right to retract unless so authorised by their management;
- on the other hand, by granting them certain prerogatives.

326. This situation, resulting from the Act, also in some cases provides remuneration for researchers who are public sector employees.

327. However, certain categories of public sector employees are not subject to this specific regime. "Employees who are the authors of works whose disclosure is not subject, by virtue of their status or the rules which govern their positions, to any prior permission from their management" are subject to the general principle of ownership. Faculty, researchers and, more generally, as expressed during the parliamentary debates "employees who by virtue of their positions have wide intellectual autonomy, or independence of judgement, even in the context of hierarchical prerogatives" are subject to the general regime applicable to any author and enjoy full rights of copyright.



Researchers hold the copyright of their articles and scientific texts.

2.3.5 Exceptions to be protected: the public interest and legal secrets

328. The availability of research data and results must be limited by considerations of public interest, such as:

- national security, public safety or the safety of persons;
- secrets protected by law.

329. This limit was already stipulated in the Act No. 78-753 of 17 July 1978 on various measures to improve relations between the administration and the public and various provisions of an administrative, social and fiscal nature (known as the CADA Act). Article 6 provides for a list of restrictions to the principle of communication of administrative documents when such consultation or communication would prejudice:

- "the confidentiality of the proceedings of the Government and of the responsible authorities attached to the executive;
- the confidentiality of national defence;
- France's foreign policy dealings;
- national security, public safety or the safety of persons;

- the currency and public credit;
- the conduct of proceedings before the courts or of activities preliminary to such proceedings, subject to authorisation by the competent authority;
- inquiries by the competent services into fiscal and customs offences;
- or, with the exception of Article L.124-4 of the Environment Code, secrets protected by legislation (...)."



Open Science must preserve secrets, as well as public safety.

2.3.6 The protection of privacy and personal data

330. The free provision of data and results as an aspect of Open Science must not compromise the protection of privacy and personal data, under the same terms as the limits imposed by the French Data Protection Act.

2.3.7 Exploitation: a legitimate interest to be preserved

331. The availability of scientific data must also be limited by the possibility of exploiting results.

332. **Legal framework.** Article L.112-1 of the Research Code defines the goals of public research, among which are:

"(b) **The exploitation of the results** of research in the service of society, which is based on innovation and technology transfer."

333. Chapter III of the Research Code lays down the terms and conditions for the exploitation of the results of research by research institutions and organisations (Articles L.533-1 to L.533-3). The Research Code thus encourages "public sector employees and public institutions entrusted with a research mission, authors of patentable inventions (in the framework of research funded by grants from the State and local authorities or from national funding agencies) to declare their inventions so that their employer may exploit the patented invention".

334. **Goal.** The exploitation of research aims to increase the value of the results of research and development.

335. The National Council for the Evaluation of Higher Education (CNE, France) defines exploitation as the means for "making the results, knowledge and skills generated by research usable or marketable"⁸⁴.

336. Consequently, the exploitation of research involves:

- bringing the world of research in contact with the socio-economic world;
- giving value to the results of research;
- returning to society the results of the research it has helped to finance⁸⁵.

337. Act no. 99-587 of 12 July 1999 on innovation and research promotes the transfer of technologies from public research to the economic sector and the creation of innovative enterprises. The following are the major routes for exploitation:

- bringing the world of research into contact with the socio-economic world;
- the provision of equipment;
- providing expertise or consulting services;
- the protection of results and the transfer of intellectual property rights to a partner via licensing or transfer contracts;
- the creation of enterprises as well as the mobility of researchers toward firms.

338. **Means.** There are three types of exploitation:

- patents;
- confidentiality;
- secrecy.

339. **Patents.** To be patentable, inventions must be new, involve an inventive step and be applicable in industry⁸⁶. To be patentable, the invention must also constitute a novelty with regard to the state of the art. An invention is not new if it already exists in its entirety in the technical state of the art, either because there is a prior example, or because the inventor has already made his or her invention public before filing a patent application.

340. If the invention has been made public in any part of the world, whether by a publication, a public exhibition (at a trade fair for example) or even a simple oral disclosure, it is no longer new (except in the event of wrongful disclosure).

⁸⁴ <http://www.senat.fr/rap/r05-341/r05-3411.html>

⁸⁵ <https://www.univ-lille3.fr/recherche/valorisation/valorisation/>

⁸⁶ Art L.611-10.1.CPI

341. **Confidentiality.** Research contracts or public/private partnerships in the field of research provide for the confidentiality of research results produced under the contract. This "negotiated" confidentiality is often limited in time, and the conditions of publication for the results of research are contractually framed.

342. Any effort to make scientific data and results available must take into account the confidentiality obligations governing the scientific results of any research activity that is at least 50% publicly funded. There are two possible types of legal regime:

- one under which confidentiality clauses on the results of research that is at least 50% publicly funded cannot apply, as a result of the public nature of the provisions;
- one with the possibility of waiving the principle of open access and the reuse of results of research that is at least 50% publicly funded, where confidentiality clauses are contractually imposed, with a time limit.

343. **Secrecy.** There are two types of secrecy:

- legal secrets that constitute a limitation on the principle of Open Science as developed in § 2.3.5 "Exceptions to be protected: the public interest and legal secrets";
- contractual secrets, in other words confidentiality as described above.

344. The protection of the intangible assets of public research fits within the framework of this goal of exploitation. Open Science must also fit within this framework and be compatible with this goal of exploitation.



The Research Code already contains among its principles:

- the foundations of digital rights to underpin Open Science;
- the necessary balance between Open Science and exploitation.

2.4 The gap between current practice and the law

345. The following table analyses the gaps between:

- the existing legal framework, the gaps and shortcomings identified particularly by the key witnesses;
- the practices of researchers.

346. For each practice or need identified, the gap separating it from the existing legal situation is rated on a scale of 1 to 5. The following scale is used:

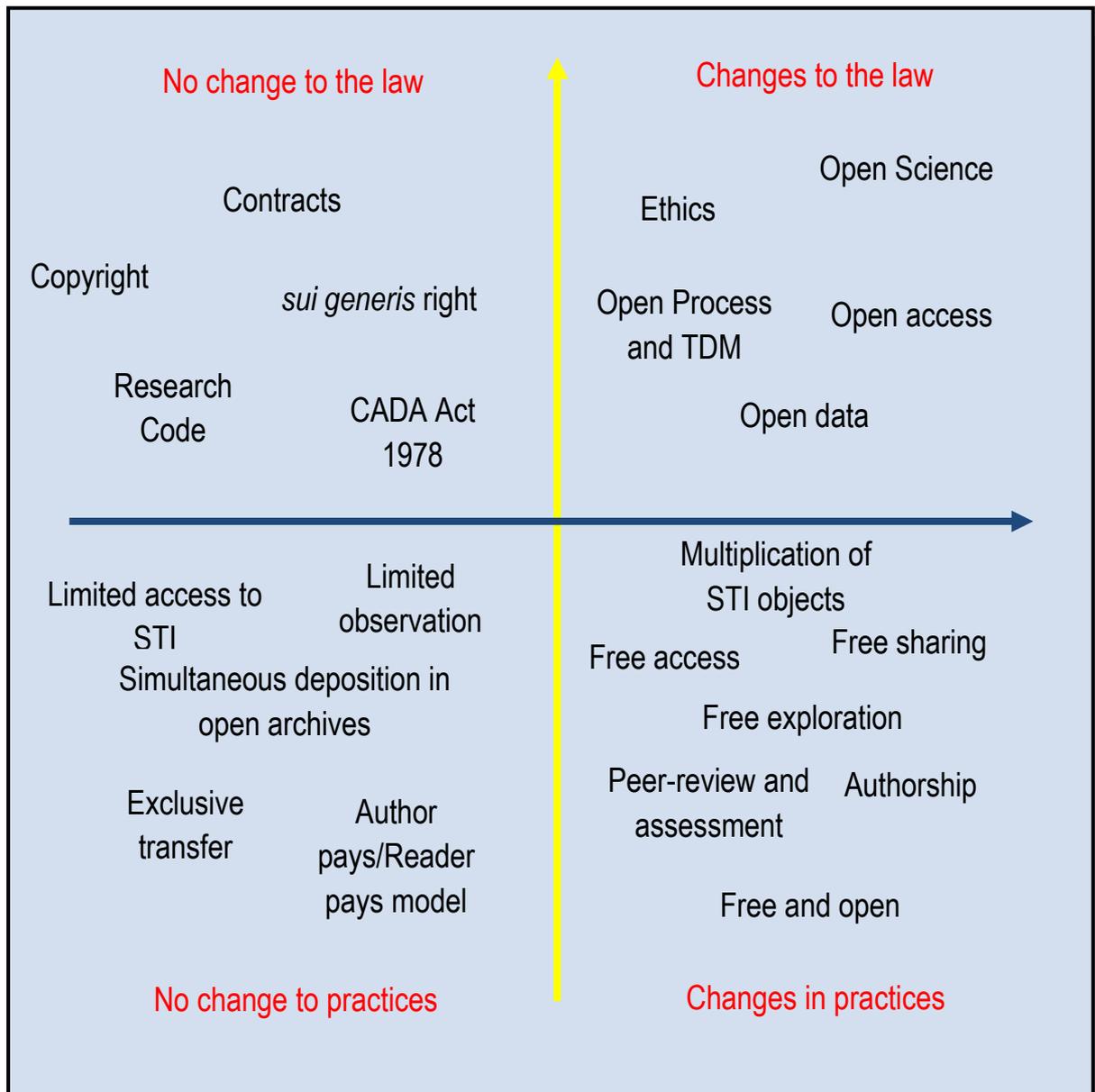
- 1/5: no difference;
- 2/5: slight difference;
- 3/5: some difference;
- 4/5: considerable difference;
- 5/5: total incompatibility.

347. A brief comment justifying this gap in light of the developments presented in this White Paper has been added in the right-hand column.

Digital practice	French legal framework	Difference	Comments
Open access and free exploration of data	Research Code Act of 1978 as amended by the Valter Act 2015-1779	3/5	If there is no change to the law: incompatibility
Open access to published scientific texts	Protection by copyright. Publishing contract and exclusivity clause	5/5	Total incompatibility, especially in publishing contracts with exclusive transfer
Free exploration of published scientific texts	No legal framework No consensus on incompatibility with copyright and the <i>sui generis</i> right General conditions of use of publishers' platforms/subscription contracts	5/5	Legal uncertainty Private exploitation by contract and by publishers' own APIs
Deposition of data in open and permanent archives	Act of 1978 as amended by the Valter Act 2015-1779	3/5	Deposition already organised in certain communities
Deposition of published material in open and permanent archives	Protection by copyright. Publishing contract and exclusivity clause	3.5/5	Total incompatibility, especially in publishing contracts with exclusive transfer In practice, some publishers allow this after an embargo period
Peer review	No legal framework	Not applicable	No legal framework Ethical rules to be defined

Digital practice	French legal framework	Difference	Comments
Assessment of researchers, taking "Open" publications into account	Decree no. 83-1260 of 30 December 1983 laying down the statutory provisions common to employees of public institutions in science and technology.	4/5	Obligation of assessment prescribed in the Decree Changes to assessment criteria necessary Ethical rules to be defined
Recognition of authorship	Copyright	2/5	Application of the author's moral rights Ethical rules to be defined
Exploitation	Research Code	2/5	Existing legal provisions Issue to be taken into account in digital Open Science (multiplication of STI objects)
Ethics of STI	No legal framework	Not applicable	Ethical rules to be defined, in particular in the field of peer review/assessment of researchers/recognition of authorship

348. The legal square graph below formalises these discrepancies:



2.5 The amendment of Article 17 of the Digital Republic Bill

349. Article 17 of the Government's final version of the Digital Republic Bill took into account many of the guiding principles that emerged during the national consultation and the GouvCamp discussion forum. Improvements taking into account the ideas formulated by key witnesses are proposed.

2.5.1 Guiding principles

350. The proposed text is based on the eight following guiding principles. These guiding principles were presented by Alain Bensoussan and Grégory Colcanap, co-rapporteurs in the framework of the consensual report by the GouvCamp's Working Group on the Open Access article (GouvCamp workshop on the draft Digital Republic Bill 16-10-2015) and were the subject of a report submitted to Axelle Lemaire, Secretary of State responsible for Digital Affairs, on Friday, 16 October 2015.

1. Scientific texts must become common assets. Indeed, scientific knowledge and results are "knowledge commons" intended for universal use in the interest of humanity. Scientific text may not be a means to prohibit or restrict access to scientific knowledge.

2. The text concerning "data mining" must be reintroduced without fail. It is both an economic issue (for the sake of innovation and to drive research) and important for competitive positioning (acquire legal provisions comparable to those of other countries such as the United Kingdom). There is strong demand for French research to have regulations at least as favourable as those governing English research (TDM) in order to be competitive. Data mining is a "telescope" granting a right to digital observation with total freedom.

3. Scientific data that are more than 50% publicly funded must become knowledge commons. The goal is for the basic research data to be deposited simultaneously with the corresponding articles. Making such data available would make it easier to reproduce the research while also fostering innovation in civil society.

4. The use of the content of a scientific article must include the possibility of commercial exploitation. Scientific text as such may not be commercially exploited without the authorisation of the copyright holders. On the other hand, the content of a scientific text and the scientific and technical information it contains is potentially a source of innovations with considerable commercial benefits. The

sharing of scientific progress, and therefore of the texts that describe it, is one of the fundamental missions of research organisations and universities. To prohibit the commercial exploitation of the contents of a scientific article by its authors and their employers would be contrary to the fundamental missions of schools, universities and teaching and research institutes, and would seriously handicap innovation in France. Limits placed by the author or a third party on the commercial exploitation of open access online publications can apply to the article itself, but not to the content or findings of the article.

5. Exclusive transfer clauses must be declared null and void. A new balance for the various interests at stake should be found by taking the risks of contractual asymmetry into account.

6. The final accepted version of a manuscript must be available immediately, or within a period of 6 to 12 months. The proposed embargo period would be a handicap for French research and its dissemination to other countries; nor does it concur with the European recommendations, which would create inconsistencies in the case of European contracts.

7. Deposition of published material must be in open and permanent archives. It appears essential to mention the preservation of the right to file published material in open archives. The role of these infrastructures is to collect and preserve scientific production, and make it freely available in accordance with international standards. Failing to mention them would run the risk of denying them fair recognition as a strategic tool, with production being posted online haphazardly and with researchers refusing to deposit their work in open archives because of a preference for other digital forms.

8. The act should apply to contracts according to the rules of application of law over time.

351. Some of these principles were taken into account in the drafting of Article 17 as approved by the Council of Ministers, and in particular:

- the clarification of the provision relating to commercial exploitation;
- certain transfer clauses to be declared null and void;
- the maximum embargo periods to be reduced to 6 and 12 months;
- the act should apply to contracts according to the rules of application of law over time.

2.5.2 Proposal for amendments to the text

352. Published scientific results. The additional amendments proposed to Article 17 in the version adopted by the National Assembly appear below in green.

In Chapter III of Title III of Book V of the Research Code, Article L. 533-4 has been inserted, worded as follows:

Article 17

At the end of Chapter III of Title III of Book V of the Research Code, an Article L. 533-4 shall be inserted as follows:

"Art. L. 533-4. - I. - When a scientific ~~text~~ **result**, arising from a research activity financed at least 50% by grants allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, is published ~~in a periodical appearing at least once a year~~, its author, even after having granted exclusive rights to a publisher, has the right to make available free of charge in an open format, in digital form, **in particular in an open, public and permanent archive**, subject to the rights of any co-authors, all successive versions of the manuscript until the final version accepted for publication, as soon as the publisher itself makes the latter available free of charge in digital form, and, failing this, on expiry of a period running from the date of first publication. This period is six months for a publication in the field of the sciences, technology and medicine, and twelve months in that of the human and social sciences. A shorter period may be provided for certain disciplines, by order of the Minister for Research.

The version made available in application of the first subparagraph may not be exploited in the framework of a commercial publishing activity.

"II. – Once the data from a research activity financed at least 50% by grants allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, are no longer protected by specific rights, or special regulations, and they have been made public by the researcher, the research establishment or organisation, they can be freely reused.

"III. - The publisher of a scientific text mentioned in I shall not limit the reuse of research data made public in the framework of its publication.

"IV. - The provisions of this Article are public policy and any clause to the contrary is deemed to be unwritten."

353. Notion of scientific text. The notion of scientific text used by the text of the Bill does not correspond to a typology of data as used by researchers. It would be preferable to use the vocabulary used in the practice of law ("published scientific result").

354. The notion of scientific texts refers to Article L.112-2 of the Intellectual Property Code. Scientific text is considered as an intellectual creation, protected by copyright as originally expressed. This approach does not take account of the scientific or informational value of scientific text, nor of the scientific data that it contains.

355. **Elimination of the notion of common assets.** The reference to the notion of the common good and to Article 714 of the Civil Code has been deleted from the version of the bill adopted by the National Assembly; the text prefers to assert a principle of free reuse. This rewording has been thus:

- "The Government has taken account of the opinion of the Council of State, which considers that the effects of a reference to Article 714 of the Civil Code would be uncertain, due to the lack of a sufficient jurisprudence. The wording chosen, resulting from discussions with the Council of State, has the same objective, namely the free reuse of data."

356. **TDM.** The provisions of the Digital Republic Bill in its version of July 2015 incorporated in the Intellectual Property Code an exception to copyright and the right of the database creator in favour of text and data mining. This text was first deleted and then reintroduced by amendments before the National Assembly. Proposals for amending and strengthening this text paving the way for TDM are given below.

2.6 Proposals to be firmed up, paths for the future

2.6.1 Positive right to text and data mining

2.6.1.1 Guiding principles

357. TDM is a crucial scientific, economic and human challenge for French research.

358. French researchers must not be discriminated against relative to their foreign counterparts, or the result will be a two-tier research community, which would threaten partnerships with foreign research institutions.

359. It is necessary to establish a legal framework for TDM, which can be accomplished in two ways:

- the "exception to copyright" path, which is the one chosen by the French parliament;
- the "positive right" path, which could be an alternative model to the exception.

2.6.1.2 Introducing an exception to copyright and to the rights of database creators

360. In the framework of the revision of the InfoSoc Directive 2001/29/, the European Union is aware of the need to regulate the practice of TDM. This will undoubtedly lead to a proposal to insert in the revised InfoSoc Directive an exception to copyright and to the rights of database creators to facilitate text and data mining.

361. Several reports have been written, all of which are in favour of introducing regulations covering TDM:

- The Sirinelli Report for the CSPLA "*Rapport de la mission sur la révision de la directive 2001/29/CE sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information*" (Report of the mission on the revision of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the Information Society) of December 2014 states that "calls for the creation of new copyright exceptions, in particular for so-called "text and data mining" (TDM) activities"⁸⁷;

⁸⁷ CSPLA report, page 8

- The study by Wolf & Partners in March 2014, entitled "Study on the legal framework of text and data mining (TDM)"⁸⁸ for the European Commission;
- A group of experts of the European Commission also published in April 2014 a report entitled "Standardisation in the area of innovation and technological development, notably in the field of text and data mining"⁸⁹;
- The Reda Report: this report, adopted by the European Parliament on 9 July 2015, "stresses the need to properly assess the enablement of automated analytical techniques for text and data (e.g. 'text and data mining' or 'content mining') for research purposes";
- The European Commission press release of 9 December 2015 presenting the measures to improve access to online content and the Commission's vision of an overhauled copyright.

362. The official summary of the public consultation specifies that "European law does not currently make it possible to create new exceptions, and the Government hopes that this issue [TDM] can be addressed in the framework of the European work in progress."

363. Indeed, a revised draft directive is expected for early 2016. However, at least two to three years will be required before this Directive is accepted and an additional two years for it to be transposed into French law. French Research cannot afford such a delay and the risk of reducing its partnerships with foreign universities or research units, slowing the pace of research for scientists, and relegating French research irremediably to the practices of another age.

364. **Bill – adopted text no.663.** As part of the parliamentary debate on the Digital Republic Bill, members of the French National Assembly of different political persuasions supported the introduction of an amendment favouring text and data mining. The following text was adopted:

Article 18bis (new)

The Intellectual Property Code is modified as follows:

1° After the second subparagraph of 9° of Article L. 122-5, a 10° shall be inserted as follows:

⁸⁸ http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf

⁸⁹ http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf

"10° Digital copies or reproductions made from a lawful source, in view of the exploration of texts and data for public research needs, excluding any commercial purpose. A decree lays down the conditions under which the exploration of texts and data is implemented, as well as the terms for storage and communication of the files produced on conclusion of the research activities for which they were produced; these files constitute the research data;"

2° After 4° of Article L. 342-3, a 5° shall be inserted as follows:

"5° Digital copies or reproductions of the base made by a person with lawful access, in view of text and data mining in a research framework, excluding any commercial purpose. The storage and communication of technical copies resulting from processing, on conclusion of the research activities for which they were produced, are carried out by organisations appointed by decree. Other copies or reproductions are destroyed."

365. The parliamentarians chose to establish a legal framework for text and data mining via an exception to copyright and to the right of database creators. This choice can only be welcomed, and the authors of the White Paper encourage the definitive adoption of this article. The creation of a positive law, sectorised to Research disciplines, could be a potential alternative to the exception.

2.6.1.3 The creation of a positive right in the Research Code a potential alternative

366. Taking into account the fundamental importance of TDM for research and of the calendar constraints mentioned in §363, it would be recommended to consider a potential alternative with the introduction of a positive right in the Research Code to carry out processing and mining operations on scientific data, research data and scientific publications

367. **Higher interests of research.** This proposal does not constitute an exception to copyright even though it seems constitutionally fragile and therefore does not contradict the InfoSoc Directive's ban on creating exceptions other than those expressly provided for by the Directive.

368. It does not deal with text in general, but focuses exclusively on the data and results of public scientific research, which include scientific publications, among other works.

369. This positive right introduced in the higher interests of research is fully compatible with the right to information and the right of access to knowledge. This general interest "is no longer managed exclusively via exceptions to copyright; it is increasingly to be found in the form of external limitations on copyright, thus in addition to the exceptions"⁹⁰.

370. **The pre-existing right of observation.** The introduction of a right to explore digital data is the affirmation of a general principle of observation and the switch from practices dating back to the age of paper and the analysis of text using tangible tools (highlighters, comments, etc.), to the practice of digital analysis using automated tools. The objective is the same but the tools have changed and now allow large masses of data to be processed simultaneously.

371. The confirmation of this pre-existing positive right would remove the uncertainty as to the rights of researchers to explore and analyse legally accessible scientific data, and would reduce the risk of the private retention of research data by publishers.

372. It has been proposed that this right to TDM should be created and added to the Research Code in the following terms:

In Chapter III of Title III of Book V of the Research Code, Article L. 533-5 has been inserted, worded as follows:

Art. L. 533-5:

- "Technical copies of research data and scientific texts under the conditions mentioned in Article L. 533-4 of the Research Code can be made freely and without charge for purposes of observation, processing and digital exploration, for the needs of public research and with due respect for the moral right of the author."

2.6.2 Defining a legal framework to underpin Open Science: asserting certain values

2.6.2.1 Guiding principles

373. The Research Code lays down an institutional framework for the organisations that participate in scientific research in France, but there is no text defining the principles or values of the scientific community.

⁹⁰ Article on "Copyright and general interest" (*Droit d'auteur et intérêt général*) by Yves Gaubiac, in *Propriétés Intellectuelles*, July 2010 - No.36

374. A right of Science, established by consensus among scientists to cover public research, would incorporate the values of the scientific communities such as:

- knowledge sharing;
- open access to scientific data;
- open processing of scientific data;
- the issues surrounding exploitation.

375. A text establishing the principles of an Open Science would enable France to be a pioneer in this field.

2.6.2.2 Proposal for a text

376. It is proposed that the following provisions should be added to the Research Code:

The following shall be added to the

Research Code:

BOOK I: THE GENERAL ORGANISATION OF RESEARCH AND TECHNOLOGICAL DEVELOPMENT

TITLE I: GUIDELINES FOR RESEARCH AND TECHNOLOGICAL DEVELOPMENT

Chapter II: The goals and institutional resources of public research;

Articles L.112-6 to L.112-11 as follows:

Article L.112-6

1° Without prejudice to the provisions of Article L. 112-1, the scientific data generated by public research available to the scientific community and freely reused for the needs of public research under the conditions laid down by this Chapter, are publicly-owned assets falling under the regime of knowledge commons.

2° Scientific data are considered to have been generated by public research when they are the result of research activity that was at least 50% publicly funded or when they result from work undertaken entirely by research units set up by establishments participating in public service research.

3° Scientific data is taken to include all the results of research and also all the research data used to establish these results.

Article L.112-7

Unless the requirements of exploitation or technology transfer or the general interest prevent their being made public, scientific data are made available as stipulated by Article L. 122-6 by deposition on digital platforms open to the scientific community. These platforms include features providing online access, referencing, sharing and processing of data, without prejudicing copyright or the rights of database creators.

Public institutions responsible for a research mission may set up digital platforms for this purpose.

Article L.112-8

Research institutions and organisations that have made data public in any form whatsoever have the right to reserve exploitation of the data for themselves, including for research purposes, under conditions and for a period of time defined by the said institutions and organisations. For this reason, it is necessary to take usages and practices in the scientific fields concerned into account. In any case, such a period may not exceed five years.

Article L.112-9

When scientific data are published by a private publisher, they must also be deposited, in parallel, in appropriate form, on an open digital platform as referred to in Article L. 112-7.

These scientific data are made accessible and reusable immediately in their author's version and no later than at the end of a period of six months for the physical sciences and twelve months for the human and social sciences, from the first publication in their publisher's version.

Article L.112-10

Scientific data must be reused with due respect for the right of authorship of the academics, researchers and, more generally, any person having directly contributed to obtaining them.

Reuse includes the reproduction, modification and processing of data in whatever form by all computer tools and all automated techniques for the exploration, analysis, indexing, aggregation, classification and processing of data such as text and data mining.

Article L.112-11

Contracts concerning the publication of scientific data from public research may not have the purpose or effect of preventing, wholly or partially, the application of the provisions of this Chapter, in particular those in Articles L. 112-9 and L. 112-10.

2.6.3 Reference guidelines for the different practices regarding digital STI

377. The reason for creating a directory of the different practices of scientific communities is to provide guidelines that are "soft" (flexible) and adaptable. This directory would list:

- practices common to all scientific communities;
- practices specific to each scientific community;
- structural definitions of science.

378. These guidelines could take one of two forms:

- a directory created by a body representative of the scientific community; this could be the role of the future Agency for Open Science;
- general reference guidelines for practices, published by Decree.

2.6.3.1 "Soft law": good practice guidelines

379. **Soft law.** The Council of State recommends that the public authorities adopt a "soft law" approach and apply it in line with the policy of simplifying standards and providing high-quality regulations (codes of good conduct, recommendations of good practices).

380. "Soft law" is defined by the Council of State as satisfying three cumulative criteria:

- the purpose must be to change or to guide the behaviour of those it is aimed at by securing their support as far as possible;
- it should not itself create rights or obligations for those it is aimed at;

- it should, by its content and the way it is drafted, be similar to a legal requirement in its form and structure⁹¹.

381. **Custom.** In addition to being described as soft law, these good practice guidelines could be a compilation of the "customs" applied by the scientific communities, which would thus be incorporated into positive law.

382. "Custom" is defined in the legal dictionary "*Vocabulaire juridique*" by Gérard Cornu as the "standard of objective law based on a popular tradition which becomes constant practice, of a legally binding nature; a real rule of law but of non-state origin, that the Community has endorsed by habit in the conviction of its compulsory nature."

383. **Agency for Open Science.** The guidelines would need to be drafted, maintained, modified and updated by a body representative of the scientific community, which could be the future Agency for Open Science.

384. They could also be posted online for comments, so that proposals for changes could be made by the scientific communities, and the guidelines could become a real tool to help the scientific community, for the benefit of Science.

385. The French ISTE platform could host a prototype for these good practice guidelines.

2.6.3.2 Guidelines by decree: a general directory of practices

386. This directory of practices could take the form of "General good practice guidelines" which, like France's "*Référentiel général de sécurité*" (General Safety Guidelines), would be issued as a Decree and would thus become a binding framework while remaining adaptable and suited to the challenges and needs of public research.

387. General guidelines of this nature also provide a reference document for the private sector.

2.6.4 An ISO standard, or an AFNOR standard

2.6.4.1 Reference guidelines embodying the state of the art and current practices

388. A "standard" is a set of specifications describing an object, an entity or a procedure. The result is a principle that can be used as a rule and a technical reference.

⁹¹ <http://www.conseil-etat.fr/Decisions-Avis-Publications/Etudes-Publications/Rapports-Etudes/Etude-annuelle-2013-Le-droit-souple>

389. A standard is not mandatory (as a general rule), but is complied with as a matter of choice.

390. Nevertheless, even if the application of standards is not mandatory, the courts systematically refer to such standards (when they exist) as describing the state of the art or current professional practice.

391. As a result, the courts will often rule that failure to comply with a standard (whether optional or mandatory) can be proof of a fault, providing grounds for liability on the part of the professional.

392. Thus, although not mandatory, standards have considerable importance. In particular, the fact that they are recognised as representative of current best practice and usage can provide an indication as to whether the regulations are being complied with.

2.6.4.2 Practical arrangements

393. Standards are established by agencies, the best known of which are:

- at the international level: ISO (International Organization for Standardization);
- at the European level: the European Committee for Standardisation (CEN)
- in France: AFNOR (*Association Française de Normalisation*)

394. A request for standardisation should be sent to the chosen certification agency. The certification agency will assess whether it is appropriate – or feasible – to initiate work in the area concerned. Depending on the result of this feasibility study, the agency may decide either to entrust the work to an existing standardisation board, or to create a new area of activity for standardisation, or extend an existing area of activity.

395. An initial working document will be drafted by a working group or by a project manager designated for this purpose. A public inquiry is then launched. If the consultation carried out with a view to obtaining final agreement on the document is positive, the text can be finalised. The agency then pronounces its approval of the standard.

396. The standard may contain:

- the scope;
- structural definitions of the field;
- the scientific approach and the associated practices;
- the conditions under which the results are to be published;
- the depositing of articles in an open archive (format for deposition, metadata, etc.);

- respect for authorship, etc.

2.6.5 Definition of model contracts for the transfer of copyright: how to protect researchers

397. In order to guarantee the rights of researchers regarding their published material and to take into account the risks of contractual asymmetry, a model contract could be promulgated by decree for transferring copyright for use in public research.

398. This contract would lay down the rules governing the relationship between the parties and protect researchers in their relationship with publishers. It would in particular ensure that there was no exclusive transfer, and guarantee the rights of researchers to:

- authorize the filing and the reproduction in an open archive of the publication in the author's version immediately, and in the publisher's version after expiry of an embargo period;
- allow the immediate exploration of the content of the article using digital data-processing tools;
- prevent all forms of private retention or reservation of property concerning the content of the article.

399. This contract could be promulgated by decree and thus have a regulatory value which could be imposed on publishers for any scientific publication resulting from public research.

2.6.6 An ethics charter for digital science

400. Le CNRS Ethics Committee, supported by the CNRS Scientific Board, is in favour of the introduction of an ethics charter for digital science. This charter would define the values associated with accessing and sharing scientific data, as well as good practice for researchers, such as:

- depositing scientific data on Open Science platforms;
- ensuring that authorship is clearly mentioned.

401. An ethics committee would guarantee compliance with this Charter, in particular by ensuring:

- that its content is disseminated and understood;

- that researchers are aware of the importance of ethics: "Researchers and all personnel involved in research must be trained to understand the ethical dimensions of data management, in particular in respect of privacy, intellectual property, and the quality and integrity of data. They must be informed as to the current status and evolution of the legal rules concerning responsible sharing of data used"⁹²;
- issue opinions with recommendations to clarify the good practice guidelines laid down in the Charter.

2.6.7 An Agency for Open Science

402. An Agency for Open Science could be created.

403. Among the roles of such an Agency would be:

- monitoring compliance with of the ethical rules defined in the Ethics Charter;
- ensuring that the principle of open access is observed;
- an advisory role;
- tracking technical developments and changing practices as well as managing the good practice guidelines;
- proposing changes to the existing legal framework in the light of the evolution of practices and needs.

404. The Agency could also be responsible for writing a report on "the impact of the principle of free access to scientific data on the scientific publishing market and on the circulation of ideas and scientific data" under the terms of Article 17ter (new) of the Bill adopted by the National Assembly.

2.6.8 An international convention for universal Open Science

405. France could very well propose an international convention for universal Open Science.

406. The concept of "international convention" is used in international law to describe formal declarations of principles that initially do not have binding force. These

⁹² Opinion issued by COMETS, "The ethical issues of scientific data sharing", 7-6-2015

conventions generally have to be ratified by the signatory states to become binding and thus become real international treaties.

407. Considering the positions already taken by UNESCO in favour of Open Science, this Convention could be organised under its auspices.

2.7 Summary of proposals for Open Science

2.7.1 Strategic opinions

Opinion No.1 - *Le numérique et la science : faire circuler les connaissances* (Science in the digital age: getting knowledge moving) Bruno David, President of the French Natural History Museum

Multiple practices among the different scientific communities

The [results of the survey](#) on the uses and needs of STI among research units conducted with the Directors of the CNRS institutes and the Directors of publishing units revealed a continuum of practices between the disciplines, over a wide range of uses.

The developments in scientific and technical information will therefore have to be defined on the basis of what they have in common, without setting the different communities against one another, and while outlining guidelines that incorporate the diversity of practices and uses.

Scientific publishing in the digital age

In addition to the routine management of the publications that pass through their laboratories, research teams expect tools with which to analyse the content to allow them to work at the forefront of scientific endeavour. First by analysing external publications, to keep abreast of current progress, but also by analysing the research produced in their own laboratories to assist with assessment and guidance.

If we compare the digital repositories of scientific publications to traditional libraries, where it took days to summarise content, the quasi-automatic and quasi-instantaneous access to content and the possibility of extracting signals make such repositories formidable strategic tools.

The exploration and sharing of research data

Access to large knowledge bases and the possibility of text and data mining (TDM) is of major importance in the digital age. TDM is like using a telescope, because of the quantity of data it brings within reach, but also like a microscope, because of the detailed examination it allows.

The organisation and condensation of data related to scientific articles are essential to allow researchers to examine content as broadly as possible, that is to say by having access to both the signals from the content of publications and to all the data on which the results are based.

This is where there is greatest room for progress: data are currently accessible on a case-by-case basis, so the digital transformation would be a great step forward and would allow global access and the possibility of leading on to new questions and investigations.

Research data are a formidable commonly-owned asset. It must be possible to share them without constraint, between the scientific communities.

These changes are before us, and research must be able to grasp the opportunities they present.

Bruno David, President of the French Natural History Museum

Opinion 2: New practices for sharing knowledge: metrics for scientific publications – Daniel Egret, astronomer, Adviser to the President of Paris Sciences et Lettres University (PSL) on bibliometrics

Sharing knowledge better by digital means

In only a few decades, the eruption of digital technologies has profoundly changed the practices and the challenges of the sharing of scientific information. Every laboratory or research unit must now have a website or a digital portal opening a window onto the unit's scientific activities. This portal often goes no further than to provide a general overview, pointers and contacts, but there is a considerable thirst for information among the scientific community – and among society in general, particularly flagrant regarding those fields of science that attract most public interest – and there is likely to be growing impatience for something meatier. It has become imperative to better inform our research colleagues and our fellow citizens about our research activity and the results we produce (and above all our scientific publications, our instrumental achievements, our innovations, etc.). It is also imperative that we organise wider access to the datasets produced by this activity. Digital technologies offer a wide range of technical solutions, but of course the human cost and the necessary mobilisation of resources and skills remain the most difficult obstacles to overcome.

Characterising scientific production by metrics

More specifically, in the field of scientific publications, it is necessary not only to provide even better ways of sharing knowledge and research data, but also to establish ways of making the scientific production of laboratories and institutions easier to identify, view and access, so that it can be analysed, compared and assessed.

This is crucial for French research because it is indispensable for extending the reach and impact of the research produced in our laboratories. Providing users, regardless of their function, with the professional tools and services they need to assess the impact of publications is one way of helping them respond appropriately to the challenges of assessment and international comparison, which are sometimes perceived as threatening.

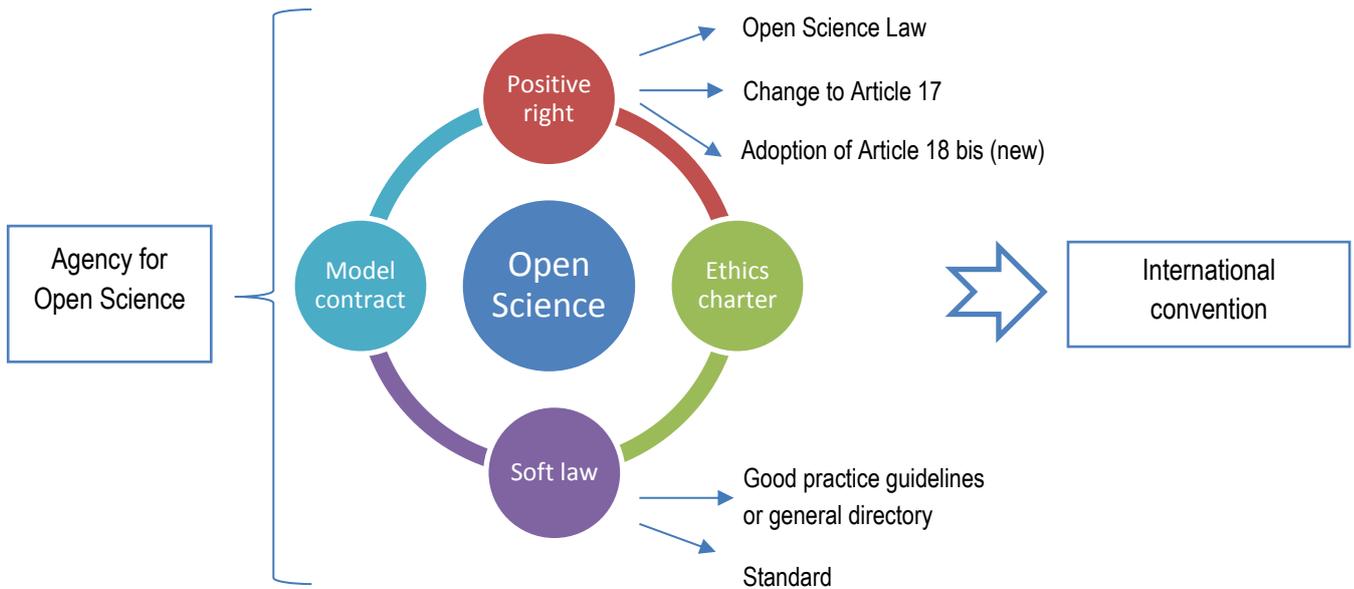
It is also necessary to provide institutions (research organisations, research universities and other such bodies) and assessment bodies with the tools and services they need to exploit this scientific production, assess the impact on knowledge and on society, and make more balanced comparisons.

Finally, the new uses of metrics should not only encourage the sharing and dissemination of bibliographic information, but also make it easier to exploit the content and thus provide a sounder basis for our understanding of the new research processes, and of practices in the production, use and exploitation of the results of research. TDM tools thus become the next generation of observation instruments, at the service of the activities of research and innovation, giving access to new worlds of results.

Daniel Egret, astronomer, Adviser to the President of Paris Sciences et Lettres University (PSL) on bibliometrics

2.7.2 Proposals

408. The full range of the proposals received can be summarised by the following diagram:



Acknowledgements

The authors of the White Paper would like to single out for their warmest thanks:

The CNRS's Scientific and Technical Information Office:

- Renaud Fabre, Director of the CNRS DIST
- Laurence El Khouri, Deputy Director of the DIST
- Charlotte Autard, ISTEEX Project Manager

The CNRS's Legal Affairs Department:

- Nicolas Castoldi, CNRS General Representative for Technology Transfer, previously Director of Legal Affairs

The Alain Bensoussan Consultancy:

- Alain Bensoussan
- Laurence Tellier-Loniewski
- Sarah Lenoir

Glossary

Article Processing Charges (APC)

Article processing charges are the sum of money required by publishers from the authors of scientific articles who wish them to be freely accessible to readers. They concern two types of journal: those that make some articles freely accessible and those where a certain number of articles are freely accessible (the hybrid model). (Source: INIST Glossary)

Gold Road

The Gold Road refers to the publication of articles in open access journals, regardless of their mode of funding. This is the second strategy recommended in the Budapest Open Access Initiative: "Open-access Journals: second, scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access." (Source: INIST Glossary)

Green Road

The Green Road describes self-archiving by researchers themselves or by a third party of articles in open archives. This is the first strategy recommended in the Budapest Open Access Initiative: "Self-Archiving: first, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called self-archiving." (Source: INIST Glossary)

Hybrid model

A journal can publish two types of article simultaneously: those that are freely accessible – in exchange for a fee paid to the publisher by the author or his/her funding source (see the "author-pays" model) – and those that are only accessible by subscription. This system is known as the hybrid model. (Source: INIST Glossary)

Institutional archive

An institutional archive is one belonging to an institution (university, *grande école*, research organisation, professional association) designed to contain, promote and preserve all of the latter's scientific production. (Source: INIST Glossary)

Metadata

Metadata are the set of structured data describing physical or digital resources. They are an essential link in the chain for sharing information and ensuring the interoperability of electronic resources. They are traditionally divided into descriptive, administrative or structural metadata. (Source: INIST Glossary)

Open access

Refers to permanent and free access for readers, over the Internet, to data from scientific research and education. (Source: INHA's InVisu)

Open archive

An open archive is a repository where data from scientific research and teaching are deposited, and to which access is open, i.e. there are no barriers. This opening is made possible by the use of common protocols that make it easier to access content from several repositories maintained by different data providers. (Source: INIST Glossary)

Open data

Refers to data that an organisation makes available to everyone in the form of digital files in order to permit their reuse.

Notes

1. Open data are not generally of a personal nature.
2. They are accessible in a format that makes them easy to reuse.
3. The reuse of open data may be subject to conditions. (Source: Vocabulary of informatics and law, Official Journal of the French Republic (JORF) n°0103 of 3 May 2014, page 7639)

Open process

Open process implies the right to freely observe data by the use of digital processing, analysis or exploration tools.

Open Science

This implies permanent free access over the Internet to data generated by scientific research and teaching, together with the right to observe these data with the use of digital processing, analysis or exploration tools. (Open Science = open access + open process)

Peer review

Peer reviewing refers to the validation of an article by a reading committee made up of scientists who are experts in the same disciplinary field as the content of the article. This process is intended to ensure the article's scientific quality. (Source: INIST Glossary)

Postprint/Author's accepted version

The postprint (post-publication) is the final version of a manuscript produced by one or more authors after peer review, with the modifications made by the peers but without the formatting provided by the publisher. (Source: INIST Glossary)

Preprint/Submitted version/Author's initial version

The preprint (pre-publication) designates any of the versions of a text produced by one or more authors before acceptance by an editorial board and possibly by peer review. (Source: INIST Glossary)

Public research

Public research is research undertaken in the public sector, including public institutions of higher education, public research institutions and health institutions, and in public companies. (Article L.112-2 of the Research Code)

Publisher's version

The publisher's version is the final, published version of a manuscript produced by one or more authors after peer review and with the formatting provided by the publisher. (Source: INIST Glossary)

Research data

All the data used to produce a scientific result.

Research results

Any scientific texts or data generated by a research activity and products based on the research data. These research results may have been published ("published result") or not ("unpublished result").

Results of public research

Results generated by public research, or by research financed at least 50% by public funds.

Scientific and Technical Information (STI)

"Scientific and technical information (STI) comprises the sum of information produced by research that is necessary for scientific and industrial activity. By its nature, STI covers all scientific and technical sectors and can exist in multiple forms: articles, reviews and scientific books, technical specifications describing synthesis processes, technical documentation that accompanies products, patent notices, bibliographic databases, grey literature, raw databases, open archives and data repositories that are accessible on the Internet, portals, etc."⁹³.

Scientific data

All scientific research data and results.

Scientific text

⁹³ <http://www.enseignementsup-recherche.gouv.fr/cid20438/les-missions-de-l-information-scientifique-et-technique.html>

Within the meaning of the French Intellectual Property Code, scientific text is considered as a work of the mind (Article L.112-2) and is therefore protected by copyright. Mainly refers to books, scientific articles and the proceedings of symposia or conferences, or reports.

Text and data mining (TDM)

Technique involving the automated processing of knowledge.

The "author-pays" model

This is the model that applies when authors or their institutions of affiliation or funding bodies pay the publisher an article processing charge to make the article openly and freely accessible to any reader. It is an alternative to the "reader-pays" and "sponsor-pays" models. (Source: INIST Glossary)

The "reader-pays" model

This is the traditional model in publishing, by subscription. Readers may only have access to journals and books for which they, or more often their institution, have purchased a subscription from one or more publishers. It is an alternative to the "author-pays" and "sponsor" models. (Source: INIST Glossary)

Annexes

1. Presentation of the White Paper

The White Paper: a public research approach in support of public research

409. At a time when the "Digital Republic" Bill is proposing to insert provisions relating to Open Access in the French Research Code, the French National Centre for Scientific Research (CNRS), alongside its partners in the ISTEEX project, as well as a large number of researchers and actors in the field of public research, are offering via this White Paper the results of their deliberations and analyses.

410. For several years now, the scientific community involved in public research has been arguing for the need to create a legal and organisational framework for access to scientific and technical data and information in the digital world, in particular data from its own research activities.

411. This White Paper gives an account of these reflections, on the practices of researchers with regard to the use of scientific and technical information and digital tools. The package of proposals for the creation of Open Science is the result of combined efforts and powerful testimonies from the world of research (1.3). The origin (1.1), objectives (1.2) and implementation approach (1.4) of this White Paper are presented below.

1.1 Origin of the White Paper

412. The plan to write this White Paper was conceived during discussions on securing the ISTEEX platform project (1.1.1) and the initial observation that the economic model of scientific publishing, a sector where prices have increased considerably, is no longer viable for education and research organisations as it stands (1.1.2).

413. Moreover, the call for Open Science is fully in line with the international Open movement in favour of sharing scientific knowledge (1.1.3) and with France's ambition, which has already been stated on several occasions (1.1.4).

1.1.1 The ISTEM Investments for the Future project: the first platform for open access to science

414. **ISTEX: a digital multi-use platform.** ISTEM, the Excellence Initiative of Scientific and Technical Information, is a project for a digital multi-use platform (database of databases), to the highest international standards, accessible remotely by every scientific community and offering "all the means currently available of consultation and analysis in all scientific communities"⁹⁴. This database of databases aims to:

- give researchers open and free access to all Scientific and Technical Information (STI) worldwide, contained in archives and current collections;
- provide researchers with high value-added services for the processing of scientific and technical knowledge and data.

415. The French Ministry of National Education, Higher Education and Research provides the following definition for STI:

- "Scientific and technical information (STI) comprises the sum of information produced by research that is necessary for scientific and industrial activity. By its nature, STI covers all scientific and technical sectors and can exist in multiple forms: articles, reviews and scientific books, technical specifications describing synthesis processes, technical documentation that accompanies products, patent notices, bibliographic databases, grey literature, raw databases, open archives and data repositories that are accessible on the Internet, portals, etc."⁹⁵.

416. In the framework of the legal underpinning of the ISTEM Investments for the Future project (ANR-10-IDEX-0004-02 - www.istex.fr), certain legal gaps and vacuums have become apparent.

417. **The issues.** An analysis of ISTEM's technical, economic and legal framework revealed several issues to which positive law offered no satisfactory answers in light of the needs of Science:

- the economic and legal model of scientific publishing no longer corresponds to the technical model of digital platforms;

⁹⁴ <http://www.istex.fr/>

⁹⁵ <http://www.enseignementsup-recherche.gouv.fr/cid20438/les-missions-de-l-information-scientifique-et-technique.html>

- access to and sharing of scientific data as working tools of scientific communities are confronted by clauses on exclusive transfer of intellectual property rights, as well as the principles of copyright and database rights;
- the ISTE platform includes value-added services such as the practice of "text and data mining": this enables researchers to use tools such as smart search, data cross-referencing, exploration and transdisciplinary searches. This practice has no legal framework and certain aspects conflict with copyright and database rights.

418. **The challenges.** ISTE fits more broadly into two challenges for STI in the digital age as reiterated by CNRS in its open strategy for STI of the future⁹⁶, i.e. to:

- "open access conditions" to STI;
- "provide a response to all requirements", in particular to take into account practices that differ according to the scientific communities.

419. This context, in the face of these findings (particularly the inadequacy of the legal-economic model of scientific publishing) and these challenges, led to the idea of drafting a White Paper identifying the needs of stakeholders in scientific research, and aimed at changing the legislation in force.

1.1.2 The imperative: to change the economic models of science in the digital age

420. **The existing models.** Several economic models co-exist in digital scientific publication. Their precise characteristics are described by the site www.openaccess.inist.fr and their principles are listed below:

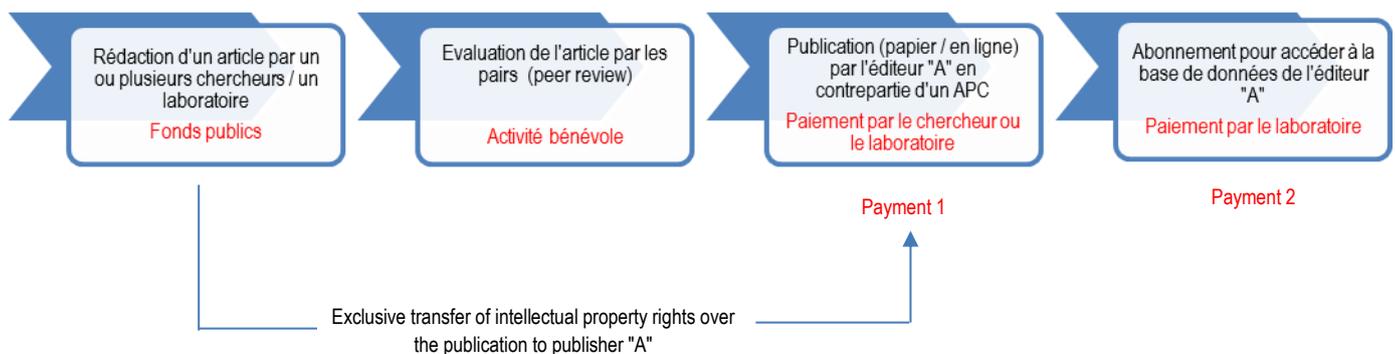
- The "author-pays" model: "when authors or their institutions of affiliation or funding bodies pay the publisher an article processing charge to make their articles openly and freely accessible to any reader";
- The "reader-pays" model: "traditional model in publishing, based on subscription. Readers may only have access to journals and books for which they, or more often their institutions, have purchased a subscription from one or more publishers";
- The "sponsor" model: "The journal is financed by a learned society, research organisation, foundation, etc.";
- The hybrid model: some publishers make articles published in their journals openly accessible in return for a fee paid by the authors or their funders (author-

⁹⁶ "An open policy for scientific and technical information of the future" CNRS page 9

pays model); readers must pay a subscription fee for access to the journals or books (reader-pays)⁹⁷;

- The "Green Road": The Green Road concerns self-archiving and centralised (such as HAL in France) or thematic (such as ArXiv in physics) institutional repositories, enabling the free access and use of scientific articles, on or shortly after their publication in a peer-reviewed journal.

421. Criticism of the hybrid model. Today, the model most widely used in practice is the hybrid model, which generates a double payment, most often by the laboratory. It can be summarised as follows:



Rédaction d'un article par un ou plusieurs chercheurs / un laboratoire Fonds publics	Drafting of an article by one or more researchers / a laboratory Public funds
Evaluation de l'article par les pairs (peer review) Activité bénévole	Peer review of the article Volunteer activity
Publication (papier / en ligne) par l'éditeur « A » en contrepartie d'un APC Paiement par le chercheur ou le laboratoire	Publication (paper / online) by Publisher "A" in exchange for an APC Payment by the researcher or the laboratory
Abonnement pour accéder à la base de données de l'éditeur « A » Paiement par le laboratoire	Subscription to access the database of publisher "A" Payment by the laboratory

422. The excesses of this hybrid model, author-reader-pay, have been widely reported by the scientists themselves:

- the emergence of predator publishers who "have polluted the global system of scientific publishing by taking advantage of Open Access in order to publicise pseudo-scientific journals"⁹⁸;

⁹⁷ Libre accès à l'IST [Open access to STI] – INIST CNRS: <http://openaccess.inist.fr/>

⁹⁸ <http://sciences.blogs.liberation.fr/home/2013/10/open-access-du-r%C3%AAve-au-cauchemar-bis.html>

- the payment by authors of a substantial APC: the study "Developing an effective market for Open Access APC" shows that the highest APCs are those of the hybrid model, around \$2727 (€2328) per article⁹⁹;
- as early as September 2012, the three French mathematics learned societies (French Statistical Society (SFdS), French Society for Applied and Industrial Mathematics (SMAI), French Mathematical Society (SMF)) published a declaration entitled "*Open Access : mise en garde et effets pervers du système auteur-payeur*" [Open Access: warnings and the perverse effects of the author-pays system];
- the Opinion of the CNRS Ethics Committee of 29 June 2012 on "Open access to scientific publications" warned against the dangers of this author-pays model;
- payment by the reader of a continually increasing subscription:
 - o "higher education and research institutions spend more than 80 million euros a year to gain access to electronic resources. Access fees have also continually increased: 7% a year over the past 10 years"¹⁰⁰;
 - o calls for a boycott of the major publishers were already being reported by French universities in 2012, challenging the price of subscriptions to scholarly journals paid by university libraries. "They often spend more than half of their budget on these purchases from three major commercial publishers: Elsevier, Springer and Wiley"¹⁰¹.

423. The march towards Open Science is therefore accompanied by a more general discussion:

- "on the sharing of values in the publishing chain, on the margins associated with the business of the global groups, on optimal economic publication models"¹⁰².

⁹⁹ <http://www.cnrs.fr/dist/z-outils/documents/Distinfo2/Distinfo4.pdf>

¹⁰⁰ http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs_4751847_3232.html

¹⁰¹ http://www.lemonde.fr/sciences/article/2012/04/25/harvard-rejoint-les-universitaires-pour-un-boycott-des-editeurs_1691125_1650684.html

¹⁰² STI Strategic Orientation Plan of CNRS, page 13 "The Economic Models of Publishing"

1.1.3 An international context that is broadly open to knowledge sharing

424. This White Paper is in tune with the international context that favours the sharing of scientific knowledge.

425. The move towards open access is a global phenomenon. In April 2012, the European Federation of Academies of Sciences and Humanities endorsed a declaration entitled "*Open Science for the 21st Century*", which advocates the sharing of research results and tools.

426. The accessibility of research data is also being debated in many international forums, including the OECD and UNESCO.

427. Examples have also proliferated at national level, with countries incorporating provisions in their legislation promoting Open Access and/or text and data mining.

428. French political discourse is also following this trend.

1.1.4 A new ambition for France

429. **Origin.** The debate on open access to scientific data, which emerged in the 2000s, has experienced a revival in France in both strength and scope in recent months, in the framework of the Digital Republic Bill, focusing on two main topics:

- the need to place scientific publications online along with the data underlying the scientific hypothesis;
- the need to enable researchers to conduct data processing, and text and data mining (TDM).

430. **Speech of January 2013.** During the speeches at the 5th Open Access Days on the theme of "Generalising open access to research results" (on 24 and 25 January 2013), Geneviève Fioraso, then Minister of Higher Education and Research, had already introduced the principle and the issues of Open Science by stating that:

- "Scientific information is a common asset that must be available to all".

431. **The Finance Bill for 2014.** In addition, the annex to the Finance Bill for 2014 entitled "*Rapport sur les politiques nationales de recherche et de formations supérieures*" [Report on the national policies of research and higher education] includes a Point 8 on scientific and technical information and documentary networks, mentioning in particular:

- "the development of open access to scientific publications".

432. **Digital Strategy of the Government.** In the Government's Digital Strategy of 18 June 2015, the action of "Fostering open science by the free dissemination of research publications and data" is also clearly stated as an emblematic measure of the digital plan.

433. The text specifies:

- "In order to ensure that our research is ever more competitive in the global arena, France is intensifying its commitment in the opening of publications and data from publicly-funded research";
- "The free movement of scientific knowledge and its free exploitation contributes to innovation, encourages collaboration, improves the quality of publications, avoids the duplication of effort, allows the exploitation of the results of previous research and promotes the participation of citizens and civil society."
- "Open access to research data, whose terms are being examined in ongoing work, will constitute an extension of open access to publications"¹⁰³.

434. **The versions of the draft bill.** Several versions of the draft Digital Republic Bill were unveiled before the official version that was submitted for public consultation.

435. A first version of the draft bill on France's digital ambition, the text of which was available online on 21 July 2015, included:

- a Section 3 "Open access to research work" creating the right to make scientific contributions, funded at least 50% by public sources, publicly available after an embargo period has been respected:

Article 39

Il est créé dans le code de la propriété intellectuelle un article L. 132-8-1 ainsi rédigé :

« Art. L. 132-8-1. – L'auteur d'une contribution scientifique, issue d'une activité de recherche financée au moins pour moitié par des fonds publics et publiée dans le cadre d'une collection paraissant au moins une fois par an, dispose du droit, même s'il a cédé un droit d'exploitation exclusif à l'éditeur, de rendre publiquement accessible la version acceptée de son manuscrit, au terme d'un délai de six mois pour les sciences et de douze mois pour les Sciences humaines et sociales à compter de la première publication, toute fin commerciale étant exclue.

¹⁰³ "Digital Strategy of the Government" Prime Minister - Gaïté Lyrique - Thursday 18 June 2015

Article 39

In the Intellectual Property Code, an Article L 132-8-1 has been created as follows:

"Art. L 132-8-1. - The author of a scientific article, arising from a research activity financed at least 50% by public funds and published in a journal appearing at least once a year, has the right, even if they have transferred an exclusive exploitation right to the publisher, to make the accepted version of the manuscript publicly accessible, after a period of six months for the sciences, and twelve months for the human and social sciences following its first publication, to the exclusion of any commercial purpose.

- a Section 4 "Exceptions to text and data mining and panorama" authorising the exploration of texts and data for public research needs, excluding any commercial purpose:

I. - L'article L. 122-5 du code de la propriété intellectuelle est ainsi modifié :

1° Après le neuvième alinéa, il est inséré un alinéa ainsi rédigé :

« f) Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de destruction des fichiers au terme des activités de recherche pour lesquelles elles ont été produites ; »

2° Après le vingt-et-unième alinéa, il est inséré un alinéa ainsi rédigé :

« 10° les reproductions et représentations, intégrales ou partielles, à l'exclusion de toute finalité commerciale, d'œuvres architecturales ou de sculptures, réalisées pour être placées en permanence dans des lieux publics. »

II. - Après le sixième alinéa de l'article L. 342-3 du même code, il est inséré un alinéa ainsi rédigé :

« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données dans un cadre de recherche, à l'exclusion de toute finalité commerciale. Ces copies et reproductions sont assurées par un organisme désigné par décret, qui garantit la destruction des fichiers au terme des activités de recherche pour lesquelles elles ont été produites. »

I. - Article L 122-5 of the Intellectual Property Code is modified as follows:

1° After the ninth subparagraph, a subparagraph shall be inserted as follows:

"f) Digital copies or reproductions made from a lawful source, with a view to exploring texts and data for public research needs, excluding any commercial purpose. A decree lays down the conditions under which the exploration of texts and data is implemented, as well as the terms for destruction of the files on conclusion of the research activities for which they were produced;"

2° After the twenty-first subparagraph, a subparagraph shall be inserted as follows:

"10° reproductions and representations, full or partial, excluding any commercial purpose, architectural works or sculptures, made to be placed permanently in public places."

II. - After the sixth subparagraph of Article L 342-3 of the same code, a subparagraph shall be inserted as follows:

"5° Digital copies or reproductions of the base made by a person with lawful access, in view of text and data mining in a research framework, excluding any commercial purpose. These copies and reproductions shall be made by an organisation appointed by decree, which ensures the destruction of the files on conclusion of the research activities for which they were produced."

436. Version 2 of the Draft Bill of September 2015 proposed the insertion of an article on open access in the French Research Code (in Chapter III "Exploitation of research results by research institutions and organisations").

Article 11 (39)
Open access
(arbitrage politique nécessaire)

Au chapitre 3 du titre 3 du livre V du code de la recherche, il est inséré un article L. 533-4 ainsi rédigé :

« I. – Les droits d'exploitation sous une forme numérique d'un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, sont cessibles à titre exclusif à un éditeur, dans les conditions mentionnées à la section première du chapitre II du titre III du Livre 1er du code de la propriété intellectuelle.

II. Lorsque un écrit scientifique est publié dans un périodique, un ouvrage paraissant au moins une fois par an, des actes de congrès ou de colloques ou des recueils de mélanges, son auteur, même en cas de cession exclusive à un éditeur, dispose du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier, au terme d'un délai de douze mois pour les sciences, la technique et la médecine et de vingt-quatre mois pour les sciences humaines et sociales, à compter de date de la première publication. Cette mise à disposition ne peut donner lieu à aucune exploitation commerciale.

« III. – Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. Elles ne s'appliquent pas aux contrats en cours. »

Article 11 (39)
Open access
(political arbitration necessary)

In Chapter 3 of Title 3 of Book V of the Research Code, an Article L. 533-4 shall be inserted as follows:

"I. - The exploitation rights in a digital form of a scientific text, arising from a research activity financed at least 50% by public funds, are transferable on an exclusive basis to a publisher, under the conditions mentioned in the first section of Chapter II of Title III of Book 1 of the Intellectual Property Code.

II. When a scientific text is published in a periodical, a publication appearing at least once a year, conference or symposia proceedings or compendia, its author, even in the event of exclusive transfer to a publisher, has the right to make available free of charge in digital form, subject to the rights of any co-authors, the latest version of his/her manuscript accepted by the publisher and excluding the formatting work which is the responsibility of the latter, at the end of a period of twelve months for the sciences, technology and medicine and twenty-four months for the human and social sciences, with effect from the date of first publication. This dissemination may not give rise to any commercial exploitation.

"III. – The provisions of this article are public policy and any clause to the contrary is deemed null and void. They shall not apply to contracts in progress."

437. This version 2 of the draft bill:

- proposes the creation of a right of deposit in the Research Code rather than in the Intellectual Property Code;
- doubles the embargo periods compared to the first version;
- removes the article on TDM altogether.

438. **Official versions.** The table below lists the official texts of the Digital Republic Bill and their developments:

- the text placed online for public consultation from 26 September 2015;
- the text from the public contribution as sent to the Council of State on 6 November 2015;
- the final version of the text adopted by the Council of Ministers on 9 December 2015;
- the text adopted by the National Assembly on 26 January 2016.

Text submitted for the public consultation 26-9-2015	Text resulting from the public consultation sent to the Council of State 6-11-2015	Text adopted by the Council of Ministers 9-12-2015	Text adopted by the National Assembly 26-1-2016
<p>Article 9 - Open access to scientific publications from public research</p> <p>In Chapter 3 of Title 3 of Book V of the Research Code, an Article L. 533-4 shall be inserted as follows:</p> <p>"Art. L. 533-4 – I. When a scientific text arising from a research activity financed at least 50% by public funds, is published in a periodical, a publication appearing at least once a year, conference or symposia proceedings or compendia, its author, even in the event of exclusive transfer to a publisher, has the right to make available free of charge in digital form, subject to the rights of any co-authors, the latest version of his/her manuscript accepted by the publisher and excluding the formatting work which is the responsibility of the latter, at the end of a period of twelve months for the sciences, technology and medicine and twenty-four months for the human and social sciences, with effect from the date of first publication. This dissemination may not give rise to any commercial exploitation.</p> <p>"II. – The provisions of this article are public policy and any clause to the contrary is deemed to be unwritten. They shall not apply to contracts in progress."</p>	<p>Article 14</p> <p>At the end of Chapter III of Title III of Book V of the Research Code, an Article L. 533-4 shall be inserted as follows:</p> <p>"Art. L. 533-4. – I. – When a scientific text arising from a research activity financed at least 50% by public funds, is published in a periodical, a publication appearing at least once a year, conference or symposia proceedings or compendia, its author, even in the event of exclusive transfer to a publisher, has the right to make available free of charge in digital form, subject to the rights of any co-authors, the final version of the manuscript accepted for publication, no later than six months for the sciences, technology and medicine and twelve months for the human and social sciences from the date of first publication, or at the latest when the publisher itself makes the text available free of charge in digital form.</p> <p>He/she is prohibited from exploiting the dissemination permitted under the first subparagraph in the framework of a commercial publishing activity.</p> <p>"II. – The research data legally made publicly available and arising from a research activity financed at least 50% by public funds, and which are not protected by specific rights, are 'commons', within the meaning of Article 714 of the Civil Code.</p> <p>"III. – The publisher of a scientific text mentioned in I shall not limit the reuse of research data made public in the framework of</p>	<p>Article 17</p> <p>At the end of Chapter III of Title III of Book V of the Research Code, an Article L. 533-4 shall be inserted as follows:</p> <p>"Art. L. 533-4. – I. – When a scientific text, arising from a research activity financed at least 50% by grants allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, is published in a periodical appearing at least once a year, in conference or symposia proceedings or compendia, its author, even in the event of exclusive transfer to a publisher, has the right to make available free of charge in digital form, subject to the rights of any co-authors, the final version of the manuscript accepted for publication, as soon as the publisher itself makes the text available free of charge in digital form, and, failing this, on expiry of a period running from the date of first publication. This period is six months for the sciences, technology and medicine, and twelve months for the human and social sciences.</p> <p>"He/she is prohibited from exploiting the dissemination permitted under the first subparagraph in the framework of a commercial publishing activity.</p> <p>"II. – Once the data from a research activity, financed at least 50% by funds allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, are no longer protected by specific rights, or special regulations, and they have been made public by</p>	<p>Article 17</p> <p>Chapter III of Title III of Book V of the Research Code is supplemented by an Article L. 533-4 inserted as follows:</p> <p>"Art. L. 533-4. – I. – When a scientific text, arising from a research activity financed at least 50% by grants allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, is published in a periodical appearing at least once a year, its author, even after having granted exclusive rights to a publisher, has the right to make available free of charge in an open format, in digital form, subject to the agreement of any co-authors, all successive versions of the manuscript until the final version accepted for publication, as soon as the publisher itself makes the latter available free of charge in digital form, and, failing this, on expiry of a period running from the date of first publication. This period is six months for a publication in the field of the sciences, technology and medicine, and twelve months in that of the human and social sciences. A shorter period may be provided for certain disciplines, by order of the Minister for Research.</p> <p>"The version made available in application of the first subparagraph may not be exploited in the framework of a commercial publishing activity.</p>

Text submitted for the public consultation 26-9-2015	Text resulting from the public consultation sent to the Council of State 6-11-2015	Text adopted by the Council of Ministers 9-12-2015	Text adopted by the National Assembly 26-1-2016
	<p>its publication.</p> <p>"IV. – The provisions of this article are public policy and any clause to the contrary is deemed to be unwritten."</p>	<p>the researcher, the establishment or the research organisation, they can be freely reused.</p> <p>"III. – The publisher of a scientific text mentioned in I shall not limit the reuse of research data made public in the framework of its publication.</p> <p>"IV. – The provisions of this article are public policy and any clause to the contrary is deemed to be unwritten."</p>	<p>"II. – Once the data from a research activity, financed at least 50% by grants allocated by the State, by regional or local authorities or public institutions, by grants from national funding agencies or by European Union funds, are no longer protected by specific rights, or special regulations, and they have been made public by the researcher, the research establishment or organisation, they can be freely reused.</p> <p>"III. – the publisher of a scientific text mentioned in I shall not limit the reuse of research data made public in the framework of its publication.</p> <p>"IV. – The provisions of this Article are public policy and any clause to the contrary is deemed to be unwritten."</p>
			<p>Article 17ter (new)</p> <p>The Government shall deliver to the Parliament, no later than two years after the promulgation of this Act, a report that assesses the effects of Article L. 533-4 of the Research Code on the scientific publishing market and on the circulation of scientific ideas and data in France.</p>
			<p>Article 18bis (new)</p> <p>The Intellectual Property Code is modified as follows:</p> <p>1° After the second subparagraph of 9° of Article L. 122-5, a 10° shall be inserted as follows:</p> <p>"10° Digital copies or reproductions made from a lawful source, with a view to exploring texts and</p>

Text submitted for the public consultation 26-9-2015	Text resulting from the public consultation sent to the Council of State 6-11-2015	Text adopted by the Council of Ministers 9-12-2015	Text adopted by the National Assembly 26-1-2016
			<p>data for public research needs, excluding any commercial purpose. A decree lays down the conditions under which the exploration of texts and data is implemented, as well as the terms for storage and communication of the files produced on conclusion of the research activities for which they were produced; these files constitute the research data;"</p> <p>2° After 4° of Article L. 342-3, a 5° shall be inserted as follows:</p> <p>"5° Digital copies or reproductions of the base made by a person with lawful access, in view of text and data mining in a research framework, excluding any commercial purpose. The storage and communication of technical copies resulting from processing, on conclusion of the research activities for which they were produced, are carried out by organisations appointed by decree. Other copies or reproductions are destroyed."</p>

1.2 Objective of the White Paper: a specific objective for public research

439. This White Paper aims to present the needs of public researchers in their research activity and to propose a legal framework able to enhance the competitiveness of French public research by equipping it with a pioneering and ambitious legislative arsenal:

- by promoting access to scientific data and results, and their reuse;
- by providing a legal framework for the actual existing practices and situations necessary to the scientific communities in public research, in order to secure them;
- by taking into account the imperatives of exploitation of innovation;
- by restoring a state of balance with the scientific publishers.

440. The objectives of this White Paper are in accordance with the intellectual property rights of authors as established by the Intellectual Property Code.

1.3 The key witnesses

441. In order to identify and give an overview of the practices and needs of researchers in the framework of science in the digital age, hearings were conducted with major witnesses, based on an interview guide. The minutes of these hearings, as well as the interview guide, are annexed to this White Paper.

442. The key witnesses interviewed made an essential contribution to this White Paper.

443. This White Paper is the result of sharing, mutual deliberation, interviews and collaborative work, taking place over more than a year, on open access and open process, with these key witnesses from and for scientific research.

1.3.1 Universities and the LERU

444. Figures from the academic world, university presidents and representatives and members of the League of European Research Universities (LERU) were interviewed:

- Alain Beretz, President of the University of Strasbourg and President of the LERU;
- Jean Chambaz, President of the UPMC and President of CURIF (Coordination of French Research-intensive Universities);
- Françoise Curtit, CNRS, Responsible for the "Open Science" mission at the University of Strasbourg;

- Jean-Pierre Finance, President of the Couperin Consortium, Permanent Delegate for the CPU in Brussels, former President of the University of Nancy 1;
- Paul-Antoine Hervieux, Deputy Vice-president for Partnerships with EPSTs and local authorities at the University of Strasbourg;
- Paul Indelicato, Vice-president for Research and Innovation at the UPMC.

1.3.2 CNRS Scientific Board

445. **Missions.** The CNRS Scientific Board ensures consistency in CNRS's science policy in conjunction with all the consultative scientific bodies of the National Committee for Scientific Research (CoNRS). In particular, it provides an opinion on:

- the major scientific policy orientations of CNRS;
- the common principles for evaluating the quality of research and the activity of researchers.

446. In addition, the framework "organic decree" governing CNRS, amended by Decree No. 2015-1151 of 16 September 2015¹⁰⁴, stipulates that under the scientific policy defined by the Government, in relation with the nation's cultural, economic and social needs and in conjunction with its higher education and research institutions, CNRS has the following missions:

- "participating in the analysis of the national and international scientific situation and its prospects for development, in view of shaping national policy in this area;
- ensuring the development and dissemination of scientific documentation and the publication of research work and data, particularly by making documentary platforms available to the scientific and academic community and contributing to their enhancement."

447. The Decree of 16 September 2015 has thus made it a national mission of CNRS to disseminate and enhance scientific and technical documentation, mainly through the digital tools known as platforms.

448. In light of this national mission, CNRS initiated and led the drafting of this White Paper.

¹⁰⁴ Decree No. 82-993 of 24 November 1982 on the organisation and functioning of the French National Centre for Scientific Research (CNRS), as amended by Decree No. 2015-1151 of 16 September 2015

449. **Working Group.** In the framework of its missions and following the interview with Renaud Fabre, the Scientific Board of CNRS decided to address the issues presented and proposed a working group made up of the following people:

- Bruno Chaudret, President of the Scientific Board of CNRS, Senior Researcher at CNRS, Laboratory of Space Studies and Instrumentation in Astrophysics (LESIA);
- Pierre Binetruy, Physicist, Director of the Astroparticle and Cosmology Laboratory (APC), Professor at the University of Paris 7;
- François Bonnarel, CNRS Engineer, Strasbourg Astronomical Data Centre (CDS);
- Claire Lemerrier, Senior Researcher in History at CNRS, Centre for the Sociology of Organisations (CSO), Paris;
- Sophie Pochic, Head of the PRO team (Professions, networks, organisations), Maurice Halbwachs Centre;
- François Tronche, CNRS Research Director, Paris-Seine Institute of Biology.

450. The deliberations of this working group, based on the securing of researchers' practices and needs, led to two documents, which have been annexed to this White Paper:

- a contribution accepted by the entire Scientific Board and to which this White Paper makes numerous references;
- a unanimous recommendation on Open Science.

1.3.3 ISTEK Executive Committee

451. This White Paper came from the initiative and shared reflection of the members of the Executive Committee of the ISTEK project under the impetus of Renaud Fabre, Director of Scientific and Technical Information at CNRS and leader of the ISTEK project.

452. The members of the ISTEK committee were closely involved in the reflections and analyses that presided over the development of this White Paper, and expressed their respective positions.

- Raymond Bérard, Director of the Institute for Scientific and Technical Information (INIST) and Laurent Schmitt, Head of the Projects and Innovation department;
- Grégory Colcanap, Coordinator of Couperin, a University consortium of digital publications, accompanied by Monique Joly, Head of the Studies and Forecasting department;
- Jérôme Kalfon, Director of the Bibliographic Agency for Higher Education (ABES);

- Jean-Marie Pierrel, Professor at Lorraine University, acting on behalf of the Conference of University Presidents (CPU).

453. Marie-Pascale Lizée (Scientific and Technical Information and Documentary Networks department (DISTRD), Sub-directorate for strategic management and territories, Section for coordination of higher education and research strategies) as well as Alain Abecassis (Head of the Section for coordination of higher education and research strategies of the Ministry of National Education, Higher Education and Research) followed the progress of the deliberations taking place around the ISTEEX project and the Digital Republic Bill.

454. The analyses and reflections, as well as the testimonies of public research stakeholders with respect to the needs and values of the scientific communities, were translated legally by the Legal Affairs Department at CNRS and by Nicolas Castoldi, General Representative for Technology Transfer at CNRS, in terms of proposed laws or regulations.

1.3.4 CNRS Ethics Committee

455. The CNRS Ethics Committee (COMETS) is an independent advisory body answering to the Board of Trustees. It considers the ethical aspects raised by the practice of research, taking account of its purposes and consequences; it identifies the ethical principles that relate to research activities, individual behaviours, collective attitudes and the functioning of the organisation's bodies.

456. In the framework of its missions, and following the interview with Renaud Fabre, the Ethics Committee decided to address the issue of the link between ethics and sharing of scientific data.

457. The COMETS published an Opinion on 7 May 2015 entitled "The ethical issues of scientific data sharing" (see Annex 4), whose findings were mentioned in this White Paper on many occasions.

458. The following were interviewed in the framework of the White Paper:

- Danièle Boursier, Senior Researcher at CNRS, lawyer and member of the COMETS;
- Michèle Leduc, Emeritus Senior Researcher at CNRS in the Kastler Brossel Laboratory at the École Normale Supérieure, Chair of COMETS.

1.3.5 The President of the French Digital Council

459. Author of the report "*Ambition numérique – Pour une politique française et européenne de la transition numérique*" [Digital ambition - towards a French and European digital transition policy] submitted to the Prime Minister in June 2015, the French Digital Council (CNNum) is an important player in the consultation on the Digital Republic Bill. The CNNum also published an opinion on the Digital Republic Bill on 30 November 2015, whose conclusions of interest to this White Paper have been included¹⁰⁵.

460. The President of the French Digital Council, Benoît Thieulin, accompanied by Yann Bonnet and Charly Berthet, were also interviewed in the framework of the White Paper in order to determine the CNNum's position on Open Science.

461. The contribution of the CNNum is annexed to this White Paper.

¹⁰⁵ <http://www.cnumerique.fr/avis-du-cnnum-relatif-au-projet-de-loi-pour-une-republique-numerique/>

1.3.6 Figures from the world of research and Open Access

462. Representative figures, recognised in the world of research and Open Access, were also interviewed. Possessing a unique view of the practices and needs of researchers, these witnesses expressed their commitment and their position in favour of the Open Science movement:

- Claude Kirchner, current President of the CCSD, Senior Researcher at the French National Institute for Research in Computer Science and Control (INRIA);
- Christophe Perales, President of the Association of Directors of University Libraries (ADBU);
- Christoph Sorger, Director of the National Institute for Mathematical Sciences (INSMI).

1.4 The approach

463. The preferred approach was a consensus approach in order to contribute to the emergence and sharing of mutual values by all the scientific communities.

464. To do this, the first step in the White Paper was to produce an inventory and snapshot of:

- uses of scientific and technical information by the scientific communities;
- exploitation practices, in particular as part of public/private partnerships.

465. Secondly, the results of these analyses were then compared with the existing French and international normative frameworks in order to identify the gaps and define the new emerging requirements for digital use of STI.

466. The third and final step was to develop proposals with the scientific communities in order to minimise this gap between the needs and the normative framework.

1.4.1 Emerging digital practices

467. Two types of practices were identified in the framework of this White Paper:

- researcher practices;
- exploitation practices.

468. **Researcher practices.** Two main methods were used to collect the practices of researchers with regard to the use of scientific and technical information by digital tools.

469. **CNRS survey.** The first method was a survey carried out by CNRS on the STI uses and needs of research units. This survey, conducted among CNRS Unit Directors in mid-2014, involved a 91-question questionnaire sent by the DIST to the Directors of 1250 CNRS units publishing articles. One third of them answered all of the questions: 432 complete responses were exploited.

470. As the units' responses were generally in proportion to the breakdown of units in the research fields considered, it can be assumed that the sample is representative of all public research.

471. **Hearings.** The second method used was the hearings. The close association with the universities and other organisations enabled hearings to be conducted with key witnesses and ensured pluralistic expression regarding the desired changes.

472. These hearings were conducted on the basis of an interview guide, which is annexed to this White Paper.

473. This guide proposed three open-ended questions, the aim being to encourage the interviewees to speak freely:

- about researchers' practices and needs in terms of access to and use of the data and results of public research, in particular in light of the privatisation of publications through intellectual property rights and publishing contracts;
- about the balance to be struck between the sharing of scientific data and the commercial side of scientific publishing; between the sharing of data and the issues of exploiting innovations. In other words, a distinction needs to be made between the misappropriation of scientific data and legal appropriation;
- about the need to define rules on data sharing and exploration: the place in which sharing and observation take place (the platform), the scope of the shared data (raw data, enriched data, results, publications, etc.) and the conditions under which they are made available, the quality of the associated metadata, and the status of the content created by users (user-generated content).

474. **Exploitation practices.** In accordance with the legal mission of exploitation of public research (Article L. 112-1 of the Research Code), and in a context of international competition, the proposals made in the framework of this White Paper need to take into account the issues of exploitation of research.

475. An analysis of the exploitation practices was conducted on the basis of practical examples and contracts entered into by CNRS, in particular with industrial companies:

- example from a standard research collaboration contract between CNRS and an industrial partner;
- example from a framework contract between CNRS and industrial partners.

1.4.2 Development of rules and rights

476. The inventory of these practices helped identify a number of needs of the scientific communities, which were echoed by the key witnesses.

477. These needs were compared with the existing orders:

- the legal order;
- ethics and the common values of the scientific communities;
- the economic order and the respect for a balance with the world of scientific publication and with industry;
- the inevitable and historic movement towards Open Science.

478. The review of practices, the definition of the discrepancies between the practices and the existing order, and the comparison of these two elements in particular with Article 17 of the Digital Republic Bill led to the formulation of normative or organisational proposals in order to reduce these discrepancies while maintaining the balance.

479. The proposals come from the deliberations of the working groups and the hearings, and reflect a consensus, as testified by the minutes from the hearings.

2. Contribution of the CNRS Scientific Board

Introduction

The place of digital technologies in scientific activity has today become capital, although of course, it is important to realise that they only partly freeze or take a "snapshot" of reality and research in a given state. Scientific activity has many other facets than the management of data. However, the digitisation of the data used by scientists and their publications enables automated processing, fast transfer, the harmonisation of access methods and descriptions; all these advantages help bring vast, rich and diverse resources within the reach of researchers, with much shorter lead times. By releasing the scientists from certain repetitive and time-consuming tasks, digital technology can therefore free up their reflexive and creative abilities.

As has been written many times, it is no doubt possible in this regard to compare the opportunities made available to research by digital technologies with those familiar to the scholars of the 16th and 17th centuries, with the invention of the printing press and the resulting acceleration in exchanges of knowledge.

Discipline by discipline

HSS

In the human and social sciences, with regard to recent scientific publications, while more and more French-language journals offer free access immediately or after a few years (mainly through HumaNum, BSN, OpenEdition), English-language journals are often confined to rather expensive platforms; the parallel submission of articles to open archives is relatively undeveloped. The scanning of printed sources used by a number of disciplines, whether it relates to the oldest scientific publications, or novels, journals, legal treaties, etc. is well on track; this is often available through open access, although some large companies (such as Gale) also produce databases at prohibitive prices, which are virtually inaccessible in Europe.

Platforms for exchanging numerical data, whether this relates to the most detailed scales of official statistics or data produced by research, have also been established (Quetelet Network, DIMESHS, etc.): they provide better circulation of data, compliance with the necessary constraints, such as anonymisation, and the documentation ("metadata") without which the figures would be unusable.

That said, the data from the human and social sciences, which vary greatly according to the disciplines (from history of art to economics, and including linguistics), are far from

being limited to copyright-free printed materials or figures. Platforms for sharing still need to be created, for example, for photographs from archives or photos of works of art taken with a scientific purpose (which raises the issue of the right of reproduction); they are still relatively undeveloped for data from qualitative field surveys (which pose complex problems of anonymisation, formatting and documentation). The problem here is that some of the data used by scientists in the HSS were not produced by them (this may concern a song, a company's annual report or the architecture of a monument): other natural or legal entities have rights over them. Data sharing and "data and text mining" techniques are thus unevenly spread depending on the types of data, mainly due to legal obstacles and a lack of human resources for the production and maintenance of quality metadata. Due to these constraints, for many types of data in the HSS, it seems difficult to imagine free sharing that would go beyond sharing for scientific use, with all the difficulties presented by the definition of this scope. Besides, there would indeed be a danger of appropriating data that may be highly sensitive. Moreover, for some types of data, the exploitation period before publication may be rather long, which argues for the embargo periods before they can be shared to be adjusted to take these specific characteristics into account.

Science of the Universe

In astronomy, and more generally in areas of the science of the universe or science of observation, the paradigm of the virtual observatory is becoming widespread. The data are freely accessible in astronomy for the entire community after expiry of a proprietary period. The preferred approach is the maximised reuse of data. To achieve this, the formats, descriptions and modes of access to archive data, metadata and the applications likely to be used to process them should be harmonised and standardised, in order to achieve interoperability. This interoperability extends to the linking of research data with online publications. The dangers of appropriation for commercial purposes have not been very pressing up to now, although things could change in the future (for example with space meteorology and the detailed observation of solar eruptions).

Biology

In biology, digital publishing is widespread and academic institutions have developed platforms to help researchers find articles, and access to abstracts is free. The most important (PubMed) is offered by the US National Institutes of Health (NIH). Access to all the articles is generally for a fee, with transfer of copyright to the publisher being the most common practice. It should be noted that the NIH has objected to this practice and proposes open access, via PubMed, to an unformatted version of any article published by a publisher describing work funded by the NIH. Open access has been developing over the past decade. The cost of publication is then generally paid by the authors on publication.

While text mining techniques are not a priority for most fields of biology as a discovery tool (but rather in terms of documentary collection), "data mining" itself is playing an increasingly important role. Free access to these data is widespread, as happened concerning the human genome. Many publishers, including Nature, also make publication of an article conditional on the depositing of mass data associated with a publication on a platform that is accessible to all, free of charge. It should be noted that this requirement goes beyond digital data and also concerns material produced within the framework of the corresponding research.

When a paper describes a particular material (cell lineage, microorganisms or genetically-modified mice, virus, antibodies, etc.), the publisher (Nature, etc.) asks the author to commit to donating this material to other academic researchers. International platforms exist for the storage and distribution of this material. Beyond the question of mass data, several publishers including Nature are considering implementing a system allowing access, via their sites, to the raw data that led to the development of the figures from an article. While this will help the reader ensure the correct interpretation of the results, the question arises of the ownership of these data and their eventual transfer.

Physics

With regard to physics apart from "major instruments", open access to raw data is not yet very widespread. In contrast, many digital libraries have been formed and made freely accessible by groups of researchers; regularly updated, they relate as much to the theoretical modelling of generic problems (electrical conduction, molecular dynamics) as to the development and management of experiments (interfacing of devices, libraries for processing data). Digital technologies also play an essential role in the dissemination of results, with the almost systematic use of pre-publication servers. Articles are deposited on these servers at the same time as they are sent to a peer-reviewed scientific journal, enabling readers to take early notice.

Chemistry

The field of chemistry is in fact really a bridge between the practices of the life sciences and those of physics. The rule is publications in paid journals from learned societies (American Chemical Society, Royal Society) or commercial companies (Wiley, Elsevier, etc.) and the timid development of "gold" type open access, paid for by the authors. There are in fact few differences between the two; negotiations with the ACS were for a time harder than with Elsevier. There is no pre-publication archive like ArXiv. Freely accessible databases are developing, especially the Cambridge Structural Database that contains all the published molecular structures.

Mathematics and Computer Science

In mathematics, databases relating to publications are very important for both individual and community work. A unique feature of this discipline is the importance of easy access to "old" publications (several years, decades or even centuries old). Long-term access to these publications is therefore crucial for research. Publication archiving platforms such as HAL or ArXiv thus respond in part to this problem and should be supported, along with metadata platforms (MathSciNet, Zentralblatt, etc.). As regards digital data, for issues of reproducibility, and comparison and interpretation of simulation and calculation methods, they need to be freely accessible and also maintained in the long term (archival, catalogues of datasets, etc.) and this concerns both software and computing code. Furthermore, mathematics plays an important role in the analysis, management and exploitation of masses of data (the issues surrounding Big Data). It is certainly very important for the data to be accessible, but when they become more and more massive, it must also be possible to exploit them effectively. In this area there are important challenges to be addressed for mathematical research.

The "publication of data"

An important question spanning all the disciplines is that of the "publication" of data. The requirement for free access is clear in the case of data associated with publications that have been duly validated by peer-reviewed journals. But what about data that may be placed on line before publication, for example for analysis and interpretation as part of a broad collaboration? This is a growing reality in a number of disciplines. This problem is especially acute since the definition of what constitutes published data is sometimes vague.

Dangers and legal safeguards

It can therefore be seen that, as with any type of progress, the digitisation of data and scientific results can have counter-effects. Scientific results play a key role in global economic competition by conferring sometimes considerable competitive advantages on their holders. In return, in order to develop and experiment, modern science needs the technologies often supplied by the world of production, which is governed by market forces. This is particularly true of the features that can be provided by scientific publishers.

Yet it is universally recognised that knowledge development occurs through the exchange of ideas, results and data between scientists. It is therefore vital to limit the appropriation of scientists' work by private interests and at the same time to provide a legal framework to free up as far as possible the exchange of data for scientific use.

Three principles:

The Scientific Board supports three important principles that would meet these objectives:

- The complete freedom of circulation and use of scientific data for reuse in the context of science, subject to a legally-guaranteed minimum embargo period enabling data producers to interpret and publish them. This requirement for free data circulation covers firstly publications and secondly data and texts that were not originally scientific but constitute the raw materials of much research, especially research in the human and social sciences.
- This requirement to make data available extends to added-value services (massive processing such as Big Data, data mining, relationship with metadata, interoperability) which must also be public and open-access to avoid any misappropriation. In the case of creation of services and platforms by publishers and more generally the private sector, this would imply legal guarantees of fair, non-discriminatory pricing.
- It also assumes clarification of the authors' rights to be able to use their scientific productions and publications in relation to publishers and other private actors. The scientists' intellectual property rights must under no circumstances be transferred to publishers free of charge, so that the free circulation of scientific results can be facilitated.

The Scientific Board also wishes to acknowledge the work carried out by the COMETS in its Opinion entitled "The ethical issues of scientific data sharing" and it endorses the recommendations contained in this text.

3. Recommendation of the CNRS Scientific Board

The Scientific Board of CNRS has been kept regularly informed of the discussions surrounding the preparation of the Digital Republic Bill, to be submitted to Parliament in autumn 2015. It has produced its own contribution to the organisation's White Paper on these issues of vital importance for scientific research activities.

The Board reaffirms two essential principles: (i) science is a common good of humanity that cannot be misappropriated by commercial interests, ii) any hindrance to open access to the results of scientific activity (publications, research data, metadata, value-added services) would compromise the development of science. This principle of free access is beneficial as much to authors and the scientific community, as to funding agencies and higher education more widely.

In light of this, the Scientific Board is concerned about any possible backtracking in the bill, in terms of the embargo period and open access to scientific publications. It reiterates that other countries, such as Germany, Canada, the United States and the United Kingdom, have been better able to resist the demands of private publishers by getting the principle of free access adopted in their legislation.

It reiterates its call to see current practices in access to scientific data consolidated by legislation, as is already the case in these countries:

- when the research activity has been partly financed by public funds, the transfer to a publisher of the rights over the data and the texts from this research cannot be exclusive;
- scientists must be able to make these data and results available for no fee, in digital form, a priori without any embargo period imposed by publishers;
- data-mining and similar services play a considerable role in the scientific exploitation of open access data and texts. They must not be hampered by commercial platforms for the dissemination of these data and texts.

4. Opinion of the Ethics Committee of 7 May 2015 "The ethical issues of scientific data sharing"

"Data Sharing" Group

The ethical issues of scientific data sharing

I- Internal referral

- 1- For the past two decades, data have acquired a central role in scientific production, regardless of the discipline. Researchers increasingly need vast data warehouses (*big data*) - but also datasets of more modest size (*small data*) - for exploring, viewing and comparing results, validating assumptions or formulating new ones, or even for automatically generating new knowledge (*machine learning*). Major infrastructures and shared platforms continue to be created for the archiving, storage and processing of information. Rapid advances in digital technologies have greatly improved the way in which data, information and tools can be disseminated, managed, used and *reused* between researchers, to constitute an ecosystem based around scientific publications. Movements favouring open access therefore become crucial to optimise the exploitation of vast deposits of data. No organisation has enough resources to conduct its work alone.¹⁰⁶ The effort required to exploit the digitised knowledge is immense, especially since it requires human intervention at a certain point in the process (text mining is necessary but not sufficient). Facilitating access to and reuse of these data has thus become a crucial issue for sharing and circulating research results more rapidly.

- 2- However, attitudes with regard to sharing and openness differ greatly depending on the types of data and the disciplines. In certain disciplines (in astrophysics or genomics, for example) the benefits of this data sharing have turned out to be considerable, and the disadvantages small enough to enable a trend for *data sharing* to develop. For these communities, the matching and comparing of data are clearly sources of new discoveries, and they consider that any obstacle to the circulation of scientific results is not only ineffective but contrary to the fundamental principles of widespread and open pooling of knowledge.

¹⁰⁶ See the article "*Dix laboratoires mondiaux partageront données et chercheurs*" [Ten world laboratories will share data and researchers] (*Le Monde* 4 February 2014). This project orchestrated by the NIH in particular asks public AND private laboratories to "not develop their own drugs from discoveries obtained before they have been made public".

However, for other disciplines (especially in the human sciences), data are often collected individually: these data, which are related to the subject of the research, may only be shared with the same embargo conditions as those for the publication of results.

- 3- Apart from the case of these "self-organised" scientific communities, government policies for *open data*¹⁰⁷ have in recent years aimed for the broad dissemination of data subsidised by public funds. Some of these data may be of interest to scientists and, conversely, the scientists' data may concern society. The *data sharing* movement must therefore be adapted to government open data policies that pursue significantly different objectives and are subject to different ethical and legal constraints.

- 4- The purpose of this opinion is to examine how the different scientific policies could be coordinated in a much broader field: the ethics of sharing research data. While many researchers support data sharing, most feel powerless or even reluctant in the face of this government obligation to disseminate (the open data), which may seem paradoxical, or even counterproductive: encouraged to disseminate widely, as confirmed by the European Horizon 2020 programme¹⁰⁸, they must at the same time apply legal restrictions to this public dissemination of data, in the name of respecting privacy, copyright, the duty of secrecy, confidentiality and security. Faced with these injunctions, which may seem contradictory, it becomes necessary to inform researchers about their various obligations and about the ethical implications of their choices relating to the data that they collect, share or reuse.

¹⁰⁷ Incidentally, we essentially consider scientific data in this text, which excludes from our analysis all types of data tracing individual activities that pose ethical problems of a different nature.

¹⁰⁸ Which enshrines the principle of "free access to research publications and data": see www.horizon2020.gouv.fr

II- Analysis

The scientific data (*research data*) considered here relate to all the data collected in the context of scientific research¹⁰⁹ i.e.:

- the primary data (empirical, observed, measured);
- the secondary data, derived from the primary data, annotated, enriched and interpreted, adding value to the initial data and possibly involving other actors;
- the metadata that structure, manage and facilitate the accessibility of the primary and secondary data.

These data may be text documents, graphs, pictures, multimedia or digital representations. The gap between data and publications also tends to be reduced with the concept of the open process, which consists in disseminating the knowledge and data used and created in the process of *thinking and writing* the scientific publication¹¹⁰.

1- Strategic context

Successive agreements and charters have marked the history of the data sharing movement. In 1996 for the first time, researchers involved in the sequencing of the human genome signed a series of agreements laying the foundations for the open sharing of prepublished data. Then the first definition of Open Data was given by the International Declaration on Free Access, in Budapest, on 14 February 2002, known under the acronym BOAI¹¹¹ (Budapest Open Access Initiative). Since then, many other initiatives have seen the light of day, with for example the Berlin Declaration of 2003 on open access to knowledge in the sciences and humanities¹¹², followed in March 2005 by a new Declaration called Berlin III aimed at strengthening the measures adopted within the framework of Berlin I. Most scientific organisations, including CNRS, have signed these declarations, thus legitimising this culture of open access. Several general recommendations are currently available (scientific consortia, OECD, etc.) and funding agencies rely on such principles to secure their requirements in this area, which constitute conditions for the granting of subsidies. Thus in 2013, the Horizon 2020 initiative defined a European policy of openness and sharing of scientific data to which national bodies must now comply when European funds are involved. Similarly, one of the most recent

¹⁰⁹ H. Tjalsma & J. Rombouts, *Selection of research data - Guidelines for appraising and selecting research data*, The Hague: Data archiving and Networked services 2011, pp.13, 14 at <http://www.dans.knaw.nl>

¹¹⁰ P. Uhler "Revolution and evolution in scientific communication: Moving from restricted dissemination of publicly-funded knowledge to open knowledge environments": http://www.communia-project.eu:communiafiles/conf202009_P_UhlerBS.pdf, 2008

¹¹¹ <http://www.budapestopenaccessinitiative.org/read>

¹¹² <http://openaccess.inist.fr/>

initiatives again comes from the field of human biology, with the launch in 2013 of the "Global Alliance for Genomics and Health"¹¹³.

The need for deliberation has already been included in the framework of the CNRS strategic plan "*A better sharing of knowledge, an open policy for scientific and technical information of the future*"¹¹⁴. This mobilising strategy must be translated into ethical obligations for researchers, the main producers and users of scientific data. Part of this programme also provides for the "establishment of a charter of ethics transcending instrumental considerations" and reaffirming the goals of public research.

2- Scientific context

Scientific activity relies increasingly on the shared creation and use of multi-source and multi-use data infrastructures. These recent transformations in the scientific process are related to three types of change:

- the technological development of measuring instruments and sensors of raw data for producing masses of data;
- computing capacity in terms of storage, archiving and analysis (birth of bio-informatics, for example);
- collaborative Internet and networking, enabling databases and platforms to be populated directly online by numerous stakeholders and enabling economies of scale.

This change is leading to a major upheaval in principles and practices, from *hypothesis-driven research* to the generalisation of *data-driven research*, i.e. to a process of construction based on data that are already formed. In this context, the data that are now being annotated, mined and analysed have become the essential components of the research activity. Multiple uses for data become the rule and the masses of data generated require infrastructures for multiple-use data to be created. The formation of data infrastructures used to underpin research operations (*research processes*), and no longer simply to archive results, becomes an important step in science. The capacities for data enrichment and annotation generate a need firstly for the corresponding databases to be monitored and developed, and secondly for metadata to be organised, to enable the effective use of the processed, aggregated and correlated data. Lastly, the data, which have been produced in vast quantities, independently of any particular hypothesis, often requiring large budgets that support a wide range of different research projects, become

¹¹³ *Nature* 498, 16–17 (5 June 2013) | doi:10.1038/498017a; The initiative <http://genomicsandhealth.org/> is in the process of drawing up an "*International Code of conduct for genomic and health-related data sharing*", currently available for comment on its website (<http://genomicsandhealth.org/our-work/work-products/international-code-conduct-genomic-and-health-related-data-sharing-draft-6>).

¹¹⁴ <http://www.cnrs.fr/dist/strategie-ist.htm>

what is called *community resources*, and for which maintenance and access are essential for the collective action of the group.

However, while 10% of the available data from experiments are provided via publications, 90% remain on the hard drives of computers. Data do not circulate quickly enough in the scientific world. Regardless of the disciplines, too many results remain unpublished and much of the data are under-used or lost¹¹⁵. The data from negative results are forgotten. How can researchers be encouraged to participate in the opening and dissemination of their data?

3- New responsibilities in the face of the global change in practices

The open sharing of scientific data intersects with another global movement that extends beyond the scientific field: the opening of public data, i.e. subsidised with public funds (*open data*). *Open data* policies, developed at State level, require authorities and public institutions to make their data accessible for sharing. However, although these two trends reinforce each other, they are not based on the same rationale.

Data sharing and the scientific commons movement

The dissemination of knowledge first took place through circles of scholars and then via exchanges between academies of science¹¹⁶. Scientists could debate directly via the accounts of their experiments and they thus became a community¹¹⁷. The raw data were shared and replicated because they were outside any exclusive appropriation. It was customary to say that the mousetrap was patented, but that the data from the experiment were not. The market for scientific data, raw or not, did not yet exist. But in the last few years, following the explosion of Big Data and the emergence of data-based science, the recognition of their value or even their monetisation, and the resulting legal guarantees, have prompted the scientific community to become organised in order to reaffirm the principles of openness and availability of data.

¹¹⁵ In early 2014, five articles were published in *The Lancet* on the theme Research: increasing value, reducing waste. See in particular: An-Wen Chan, Fujian Song, Andrew Vickers, Tom Jefferson, Kay Dickersin, Peter C Gøtzsche, Harlan M Krumholz, Davina Ghera, H Bart van der Worp, "Increasing value and reducing waste: addressing inaccessible research" in *Lancet* 2014; 383: 257-66.

¹¹⁶ "The Academy of Science owes its origins both to circles of scholars, who at the beginning of the XVIIth century would gather around a patron or scholarly figure, and permanent scientific societies that formed at the same time, such as the *Accademia dei Lincei* in Rome (1603), the Royal Society in London (1645), etc. Through its work and publications, the Academy makes an essential contribution to expanding scientific activity."

¹¹⁷ Evelyne Barbin (dir.), *Arts et sciences à la Renaissance*, [Art and Science in the Renaissance] Ellipses, 2007

There are many legal or technical obstacles to the dissemination of data: either the major databases are subject to rights of access, or the data are available in formats that are closed or require proprietary software. For this reason, in 2005 a community of researchers, aware of the resistance encountered during the implementation of open data policies, launched a global initiative to effectively create Science Commons, with tools and methods (access platforms, standard author contracts, etc.) to accelerate the circulation of results and reuse of the data on which they are based¹¹⁸.

Similarly, what has become known as the "Open Data Protocol" (which rather meant "data sharing") emerged in the research arena: this also effectively encourages the members of the global scientific community to pool their resources, regardless of their legal status. These shareable research platforms¹¹⁹ facilitate the development of new services for:

- *reusing research through policies and tools that help individuals and organisations make their work accessible and indicate, on their results and data, this option of reuse;*
- *immediate access to tools* (online calculations) through standard contracts that offer economies of scale to researchers, enabling them to duplicate, check and extend their research, and also to take part in the entire value chain through to *peer review*;
- *the integration of fragmented sources of information by giving researchers the means to find, analyse and use data from disparate sources by marking and integrating the information through a common, standardised language that can be translated into machine-compatible form.*

This movement to open and share data has been facilitated by open archive policies developed within scientific institutions (ArXiv, 1991). In France, HAL (Hyper Article onLine), was created in 2000 based on "the model of direct communication between researchers"¹²⁰ of their *preprints*: its management and missions, currently being overhauled, are still to be defined with regard to the archiving of scientific *data*, in order to integrate in the *design* the relationship with the embargo period and the open licence, which must be decided by the researcher alone at the time of filing.

In 2013 an initiative from the field of human biology launched the "Global Alliance for Genomics and Health"¹²¹. This is a unified movement of member federations from 170

¹¹⁸ Now Science at Creative Commons. See also: <http://sciencecommons.org/about/>.

¹¹⁹ This was the basis of the Science Commons project, see: D. Bourcier, Web, "*Science et Communication : l'exemple de Science Commons*" [Science and communication: the example of science commons], *Revue Hermès* 57, 2010, pp. 53-160.

¹²⁰ See the report by Serge BAUIN, "*L'Open acces à moyen terme : une feuille de route pour HAL*" [*Open access in the medium term; a roadmap for HAL*], DIST, CNRS, September 2014

¹²¹ *Nature* 498, 16–17 (5 June 2013) | doi:10.1038/498017a. The initiative <http://genomicsandhealth.org/> is in the process of drawing up an "International Code of conduct for genomic and health-related data

countries which decided to provide a platform for engagement for non-governmental organisations and create a powerful network designed to describe non-communicable diseases. Other disciplines such as physics and Earth and space sciences insist on other imperatives. Thus, the long-term observation of natural phenomena brings into play processes that may have long time constants compared to human life (the variations in the Earth's magnetic field, tectonics, climate change, the seismic cycle, etc.). They are in essence non-reproducible and underpin our knowledge of the world around us, its changes and the risks it poses to our societies. As nature is a common good, the archiving and free dissemination of these data is a public duty.

The digitisation of observations and the exchange of digital files offered new opportunities to those who wished to reinterpret or compare the data. These various movements have had a definite heuristic effect on the traditional scientific process that used to be described sequentially from design to the written results, whether digital or printed. Now, scientific discourse can no longer be described in a linear fashion but resembles a process where partial results evolve and are interdependent. These cognitive interactions are manifested in what is called *knowledge hubs*¹²², where several layers of more or less developed knowledge coexist. In this framework, the access to the primary data becomes the determining factor, by making it possible to check their quality and gauge the methodology and resulting interpretation. In addition, the open-access structure of knowledge has an influence on research itself, which is no longer an "independent variable"¹²³ of the development process, but the dynamic result of continuous brainstorming between researchers (*email science*).

As a result of this movement, publishers have become used to asking researchers, in addition to their scientific results, to place their data online¹²⁴ in order to verify the reproducibility of the experiment or the process. This has enabled them to first check the published results by comparing them with the data and therefore try and avoid plagiarism and fraud, which obliges them to retract articles. In the longer term, the possibility of these accumulated data eventually creating a "data market" for the benefit of publishers cannot be ruled out. This very disturbing phenomenon should be denounced and refused by researchers.

Scientific *data sharing* refers to a traditional community practice in science: making the data used for scientific research available to other researchers and creating *pools* of resources managed by the scientific community. Many agencies, institutions and journals

sharing", currently available for comment on the website (<http://genomicsandhealth.org/our-work/work-products/international-code-conduct-genomic-and-health-related-data-sharing-draft-6>).

¹²² H.D. Evers ["Knowledge hubs and knowledge clusters: designing a knowledge architecture for development"](#), 2008

¹²³ The motto of the UK Royal Society was "*nullius in verba*" (take no man's word for it)

¹²⁴ Toronto International Data Release Workshop Authors (2009), "Prepublication data sharing", *Nature* 461: 168-170)

have *supported* data sharing policies because openness and transparency were regarded as ethical principles inherent in scientific work.

However making this practice of sharing universal raises questions such as: to whom must these data be open? To which scientific community? Should they be open to citizens, to the public?

Sharing policies require researchers to be informed about the limits of this sharing. The data concerned may be unavailable because of their nature as non-anonymised personal data, or may be subject to special regimes such as that of national security and professional secrecy, or to restrictive contractual clauses or various commercial interests. Then, if researchers own the rights to these data and wish to share them, they are advised to place their protectable data under a free licence such as *Creative Commons*¹²⁵ in order to at least inform future users that most of the works are protected but have been "freed" under the author's conditions. Researchers need to be vigilant when they transfer their rights of exclusivity on their data or data banks to third parties¹²⁶, or vice versa when they use data generated by other researchers or by open government platforms.

Open data as public policy

Unlike the data sharing movement developed by the researchers themselves, public *open data* policies emerged outside the scientific community. Every day, a growing number of data are produced or collected by different actors operating in different business sectors, which differ according to their objectives and purposes. The State, first of all, is a major producer of data. Increasing quantities of statistical data are being produced, reproduced, collected, disseminated or re-disseminated by the public authorities in the framework of their institutional missions. These are mainly demographic, geographical, weather, economic, financial, cultural, tourist data, etc., which are designed to ensure the quality and continuity of public service – but which can also constitute new resources for researchers. Thus in Europe, following the Directive on the re-use of public sector information¹²⁷ and then the Directive establishing an infrastructure for geographical information¹²⁸, most countries adopted policies to promote the opening of public data. In the public sector, therefore, open data could be defined as the *open* and (almost) *free* provision of data, implying the option of reusing them with as few constraints as possible.

¹²⁵ www.creativecommons.fr, which is a project for sharing content and a platform for open licences, whose options extend from the most open licence (mention of the granting of rights) to the most "commercial".

¹²⁶ The status of data banks in Europe, defined as *sui generis*, cannot be transposed in most other countries. The United States, for example, does not recognise copyright over databanks.

¹²⁷ Directive 2003/98/EC of 17 November 2003 on the re-use of public sector information.

¹²⁸ Directive 2007/2/EC of 14 March 2007 establishing an Infrastructure for [Spatial Information](#) in the [European Community](#) (the INSPIRE Directive).

In France, the mission of Etalab¹²⁹, the service that manages public *open data* under the authority of the Prime Minister, is to communicate research data. Ideally, besides its network of experts it should include the ethical skills needed to define data of interest to research.

However, the policies that promote the opening of public data, i.e. promoting sharing and reuse, do not have the same objectives and the same targets as *data sharing*. One of the objectives of public-sector *open data* is to enable the exploitation or even the monetisation of these data by creating wealth for the companies that exploit them. In addition, the targets of this opening concern all the actors in the public sector, authorities and communities receiving public money. However, these policies all share the desire to promote the transparency of knowledge production methods and create deposits of data that are accessible and shareable.

The fundamental difference between *data sharing* and *open data* is that in the scientific world the movement emerged from the community itself and its purpose is ethical because it concerns values, i.e. defining the limits of what is good or bad for the community that applies it. In contrast, for government open data, the incentive was originally normative and applies *legally* to all public officials, including those working in public research.

One way to clarify the applicable regimes would be to differentiate scientific data and public data. But scientific data, most of which are produced with public funds, are all intended to become public, with just a few exceptions. And yet we find that not all researchers, depending on the disciplines, are favourable to the opening up of their data. In the human and social sciences, the embargo period from six months to one year, depending on time to publication, may be necessary for the primary data to be made available. This requirement varies in the exact sciences and in fact depends greatly on the fields: in many cases there is no embargo on the data; for example, in biology, data are generally provided at the same time as the publication. Regarding the exploitation of data from major instruments in physics or astronomy, there is a delay before the whole community can benefit, because the raw data must generally be processed before they can be exploited; furthermore in astronomy, very specific rules, laid down in advance, give a preference for a limited time (generally one year) for the exploitation of data by researchers having built an instrument involving large equipment (satellite, telescope). For researchers in these disciplines, this means that general policies on opening public data are not compatible with the customs of the community to which they belong, and they can only contribute to them by defining the limits of their practices.

On the other hand, researchers must benefit from public open data promoted by the State. Public data, especially in health, are destined to become scientific data that the

¹²⁹ Established by the Circular of 17 September 2013.

researchers can use. Thus the SNIIRAM¹³⁰ is defined as the world's largest database on health: for decades it has been populated by the information generated from the delivery of all healthcare and hospitalisations in France. These data are by definition sensitive: they have therefore undergone anonymisation procedures. However everyone – and particularly any researcher – knows that these procedures are not 100% reliable¹³¹. Responding to this request to open this deposit of (sensitive) data, which the report's authors call "common assets from research in public health", the French National Health Insurance Fund (CNAM) therefore provided a randomised sample of beneficiaries (one file out of 100) and not all the data, for the researchers and public bodies responsible for public health. Among the sometimes contradictory principles it wishes to apply, the CNAM decided to open access to batches of anonymous data by distinguishing the publication (free) and customised extractions or trend charts (payable). It thus intends to develop its policy of openness, whose criteria will be public interest (mainly research), the quality of the protocol, the need to access the data, the security of the procedures and the status of the applicant. It is therefore possible to understand how public data, even sensitive, can be useful for researchers and be the subject of specific negotiations with this community, while at the same time respecting the rights of the subjects concerned.

Ensuring the quality of the data and validating its processing in light of evidence-based scientific methods also represent another major challenge. Big Data – that refers to massive volumes of information that are complex and likely to be connected – can improve our understanding and prediction (by machine learning) of behaviours likely to affect health, and accelerate the cycle of knowledge dissemination. However "Big Error" can threaten "Big Data"¹³²: In this article, the authors ask for the systematic replication of epidemiological results and collaborative studies on a broad scale to test predictive tools and move from correlations obtained to *real* causalities.

Constraints concerning the processing of personal data

Researchers who use personal data – whether the person has been identified or is identifiable indirectly by profiling or targeting – are first confronted with heavy legal constraints. Yet in some cases the data processing model from which the privacy protection policy had been developed in the French Data Protection Act of 6 January 1978 may appear to be too restrictive in the new contexts of *big data* in relation to research objectives, or even obsolete as reported by the researchers questioned¹³³. Indeed, the rapid and open circulation of data between researchers disrupts the order of procedures

¹³⁰ Overview of the health insurance information system. See the proposal for opening and sharing these "public data": "*Rapport sur la gouvernance et l'utilisation des données de santé*" [Report on the governance and use of health data] by Louis Bras and André Loth, September 2013.

¹³¹ This implies that the researcher must also inform the patient.

¹³² M.J. Khoury & J.P.A. Ioannidis, "Big Data meets Public Health" in *Science*, 26 November 2014, vol. 346, 6213 p.1054-1055.

¹³³ Around twenty hearings were organised at CNRS.

and makes data flows relatively autonomous in relation to their sources or authors. It often becomes impossible to adhere to the principle of purpose (assumptions are not developed *a priori*), the principle of proportionality (it is not possible to know which data will be necessary before they are actually used) and the principle of non-conservation (the data are not destroyed at the conclusion of the research because of their open access and reuse). It should be noted however that all archived data are subject to an exception for research purposes, once the original research deadlines have passed.

There is the same need for resources in many disciplines. Take for example computer vision systems, whose purpose is to automatically recognise visual scenes. Facial recognition is one area of computer vision, which can be used in biometrics applications. The problem is then made more complex by rights of personal portrayal, as the identity of a person can be determined from their face. The case was raised recently in the framework of the organisation of a campaign to assess facial-recognition systems. After lengthy negotiations, the French Data Protection Authority (CNIL) gave its agreement on condition that these data were not kept beyond the duration of the project, unless an application for extension was submitted requiring a new case to be made, and the agreement of the CNIL. This therefore results in the paradox of prohibiting experiments of other systems on the same data in order to compare their performance, even though this is a normal scientific approach. The same paradox exists in the request made to Google by the G29 combining the CNIL's various European counterparts, to not keep for longer than three months the Google Street View images used in developing the algorithm for automatically "blurring" faces. It seems paradoxical to limit the use of these images, when the aim is to enable the development of the most effective face-blurring algorithms possible in operational mode. It therefore seems that this case is confusing the needs of research and the constraints of operational use, which are of a different nature. It would be useful to consider the introduction in French law of the concept of *fair use* found in common law in the field of copyright. It corresponds to reasonable or acceptable use. Transposed to the area of research, Parliament or the courts should then be asked to define a set of legal rules, which would try to take into account the concerns of both research and public interest, in authorising certain uses that would otherwise be considered illegal. Such a law enshrining *fair research use* would facilitate the development of research requiring the use of protected data under certain conditions.

More specifically, it is sometimes difficult to apply the basic principles of personal data processing, such as informing people about the fate and use of the data, or obtaining their consent. Thus, for example, the researcher's approach may require obtaining information without the knowledge of the person being investigated. It would be necessary to stipulate the principles to be observed in the absence of consent, such as a commitment to inform this person *a posteriori*. In other disciplines, the data belong to non-identifying datasets but if they are combined, this can lead to re-identifications that require procedures for the possible change of data "status": identifying or non-identifying. Similarly, the protection of

"anonymity" imposes other types of guarantee (commitment by institutions and researchers not to use the identifying characteristic of the data, in the event of re-identification). In any case, the researcher must inform the subjects of the impossibility of guaranteeing the strict anonymity of the data, and give them an assurance that all efforts will be made to ensure that measures are taken to protect their rights.

Lastly, the issue of the contribution of each party requires the unique, unambiguous and persistent identification of the researchers, which, by giving them credit for these contributions, will thereby also indicate their responsibilities (see the ORCID initiative, for example).

Call for researchers to be vigilant with shared data

In the context of the rapidly accelerating circulation of data encouraged by their supervisory authorities and by their community, researchers must:

- be aware of their individual, deontological¹³⁴ and ethical responsibilities, with respect to the community to which they belong;
- abide by the international undertakings of the institutions to which they report;
- participate in the definition of ethical principles specific to their discipline in *data sharing* and *Big Data* in general.

Data sharing had been launched by communities of researchers and fell within the scope of *soft law*, i.e. non-binding rules of conduct. Today, the institutional commitments to which the public researcher is subject have become binding since the aforementioned open data policies¹³⁵. The implications of these policies regarding the ethical dimensions of research need to be assessed on a case-by-case basis and the opening of the data should be applied *reasonably*.

As a general rule, public researchers must pursue an ideal of sharing and exchange between peers and take part in the dissemination of data obtained with public funds while respecting any exceptions of a contractual nature to which they may be committed. Conversely, standard consortium agreements involving public and private partners (in particular in competitiveness clusters) are often too restrictive with respect to the opening of data: they should now be negotiated in advance by public researchers in a way that does not lead to the confiscation of unexploited data by private partners at the end of the project.

¹³⁴ "Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences", *Committee on Responsibilities of Authorship in the Biological Sciences*, National Research Council. National Academy of Sciences.

¹³⁵ This PSI Directive is transposed by an Order supplemented by a Decree of 30/12/2005 pursuant to the CADA Act (*Commission on Access to Administrative Data*) of 1978.

Although subject to the principles of sharing and openness, the data are not *free*: whether or not they are structured in the data bank, they possess a market (economic) or non-market (ethical) value. Like published material, they are increasingly concerned by copyright. It is therefore necessary for the producer to explicitly define the restrictions or exceptions to researcher re-users. In addition, when they are sensitive, the data must strictly adhere to personal data protection policies throughout their processing, to avoid causing problems for subsequent uses.

Researchers have discovered that the opening of data – but also the software, ontologies and metadata that enable them to be exploited – implied a new responsibility: to take particular care over the quality of the information and data *offered* as well as the clarity of the accompanying documentation. To enable others to replicate or to reuse data, it is necessary to check the integrity and interoperability of the data, the identification of their sources, the dates they were collected or processed, and to conduct a detailed examination of the different steps leading to the constitution of the data deposits: collection, classification, standardisation, provision, reuse, storage, destruction. The issues relating to image rights, confidentiality and security also raise legal and ethical issues which, although they existed before *data sharing*, have become more difficult to interpret at a time of generalised international sharing of research results.

Therefore, the organisation, maintenance and accessibility of high-quality interoperable data become fundamental for ensuring the integrity of scientific data in the digital age and creating new legal and ethical responsibilities *between* researchers¹³⁶. Who owns the data? The laboratory? The researchers? The agencies? Are all the necessary means (software, algorithms) available to use and reproduce them? The change of scale also imposes real international data infrastructures, which further complicates their governance. Thus the rights concerning data and data banks are not homogenous, or even harmonised at European level. There is therefore a shift in the centres of gravity of scientific activity, which calls for continued deliberations, not only in terms of strategy but also in terms of ethics. As was rightly noted in an article on this subject in 2014, "the current trend towards the commercial exploitation of scientific results, with the emphasis on intellectual property, goes in the opposite direction to that of data sharing"¹³⁷. The intention of this opinion is to sound the alert against these practical contradictions facing the world of public research.

¹³⁶ "Ensuring the integrity, accessibility and stewardship of research data in the digital age", *Report of Committee of Science, Engineering and Public Policy*, Washington, the National Academies Press, 2009.

¹³⁷ M. Vito, "Partageons nos données" [Let's share our data], *Le Monde*, 28 May 2014.

III - Recommendations

1. The COMETS recalls that CNRS is a signatory to the Berlin Declaration (2003), like most major international research organisations. This commits researchers to the global movement of *open data sharing*. The COMETS invites all CNRS researchers to join this movement in accordance with the practices specific to each discipline.

2. The contribution to the work of data sharing must be recognised in assessments and decisions concerning the promotion of researchers. To facilitate this recognition, the COMETS recommends that appropriate indicators be created and that a section on these activities be added in the activity report and the annual activity sheet of researchers.

3. Researchers and personnel from the world of research must be trained in the ethical management of data (what is known as "privacy, accuracy, property, accessibility") and informed about the *rules of good practice*, as well as the legal rules concerning responsible sharing of data, including the fair and proportionate collection of personal data or data likely to re-identify individuals.

4. Data sharing practices should be encouraged in the publication policies of scientific journals and in the organisation of symposia, with regard to both authors and evaluators. The COMETS recommends that authors refuse to enable their data to be subject to special pricing by scientific publishers and/or separate subsequent exploitation by the latter (resale or *paywall*).

5. The COMETS advocates that the HAL open archive be preferred for depositing the data on which publications of research results rely and that the researcher be able to choose, by open licences such as Creative Commons, the conditions of their reuse.

6. It recommends that CNRS ensure the existence of sustainable infrastructures enabling management of the data platforms in the long term at the team, laboratory or network level. It suggests that CNRS encourage its researchers to participate in the establishment and activity of international bodies to process metadata using unique, lasting identifiers for these data.

7. The costs of sharing data, assistance with the creation and maintenance of data warehouses or databases, the construction and maintenance of multi-use platforms or open archives must be taken into account when the organisation is allocating the appropriate resources (grants, subsidies, etc.) to teams, without prejudice to any pricing of on-demand and customised data processing.

8. The COMETS recommends that a discussion be held with the French Data Protection Authority (CNIL) and the data protection representative at CNRS, as well as with ETALAB, in order to take account of the specificity of the data and their processing in

the world of research. It suggests the creation of an advisory committee for administration of scientific data, involving various disciplines in this debate.

9 Finally, it stresses the importance for the scientific communities of identifying in an open, collaborative way the legal obstacles to the ethical sharing of data (intellectual property data and *sui generis* status of data banks), in order to promote real scientific commons, to integrate the concept of *fair research use* and participate in the adaptation of data rights to the *legitimate interests* of research.

5. Interview guide for the hearings – the contribution of research to the themes of the Digital Republic Bill

I- Context

Digital technologies are profoundly transforming the modes of production and dissemination of scientific results: data, publications, analyses are now accessible on various platforms. This availability of scientific material contains a potential for knowledge exploitation and sharing for which the law must be able to define the conditions, terms and limits.

II- CNRS's proposals: conditions, terms and limits to the sharing of scientific information

A. Conditions

The main conditions for free access to scientific results are the abolition of limits that may be introduced by editorial legislation (publication rights, copyright), with a view to the exploration of digital corpora of publications or data.

Question 1: What is your opinion on the necessary adaptations to publication rights (publishers and/or authors) and the exploration of corpora (text & data mining techniques, APIs, etc.)?

B. Terms

The terms for sharing scientific results must, in the digital age, assimilate new constraints: the sharing of results between actors in public research on the one hand and between users and beneficiaries of public science on the other.

Question 2: How should the line be drawn between legitimate appropriation and misappropriation of results available on a public science platform, and how should these results be protected?

C. Limits

Science platforms today contain STI whose form, content and legal status are very heterogeneous. This lack of uniformity impedes the visibility of science platforms. In the same way that there is today a notion of general interest data (recent choice of the Minister of the Digital Economy), consideration should be given to the designation of public science platforms and, possibly, to a specific legal regime.

Question 3: Should we be moving towards a designation of "essential infrastructure" for major upstream research platforms, in cases where these platforms occupy a unique and irreplaceable function?

6. Minutes from the hearings

a. UPMC, Jean Chambaz and Paul Indelicato, 9 June 2016

Participants

For the UPMC:

- Jean Chambaz, President
- Paul Indelicato, Vice-president for Research and Innovation

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit

For the DIST CNRS:

- Renaud Fabre, Director
- Charlotte Autard, Manager in charge of the ISTEEX Investments for the Future project

Purpose

CNRS, in conjunction with the Alain Bensoussan Consultancy, decided to draft a White Paper on “Digital Science and Law” in response to issues raised by the legal framework covering the ISTEEX Investments for the Future project ANR-10-IDEX-0004-02 (www.istex.fr). This project raises many legal questions due to its potential in terms of TDM, interdisciplinarity, content aggregation and its aim of making its databases explorable.

This White Paper aims to **propose a legal framework for scientific data in order to address the concerns of the scientific communities** (data publication and uses, laws governing science platforms, text mining, etc.) and thereby **contribute to the French Digital Republic Bill**.

Initially, the Bensoussan Consultancy and the DIST will establish working groups and conduct hearings to **gather information about the community’s uses and the current state of exploitation practices**.

This information about uses and exploitation practices will enable the CNRS and the Bensoussan Consultancy to establish a **matrix concerning the relevance of analyses and practices which will be set against the current normative framework** in order to assess gaps in the law and make suggestions for the Bill.

Copyright and the law as applied to science

Currently, there is no legal text governing Science, apart from copyright law. Science has been bound by copyright as regards its legal framework, its evaluation and its dissemination. Copyright covers not only past science (articles) but also future science (analyses of results).

Copyright was designed to privatise the form of a piece of creative work by attaching it to its author, while enabling publishers to make money and contribute to culture through its dissemination. Later, things were reversed and now authors are somewhat under the thumbs of publishers, who collect 90% of the value of their creative work.

Today, copyright appears unsuited to digital science. The aim of the White Paper on Digital Science and Law is to propose solutions for the creation of a new Law governing Science.

It is extremely difficult to separate considerations concerning copyright for scientific data from the general issue of copyright, which has been the subject of intense debate for a long time. It may be more productive to focus on the notion of data and data rights, which is a distinct issue from copyright and can therefore be addressed without any misunderstanding.

Scientific data are constantly evolving, and although data science will never replace the scientific method, data and even more so the reuse of data are at the heart of this new approach.

Certain fields of research are based on the collection of large amounts of data on a specific subject or gathered from human activities (public, governmental, health data, etc.). These data have value for the research project in question but can also be reused for other research projects. It must be established who actually owns the data collected: the person who collected the data and labelled them, the organisation, or the laboratory that stores them?

The data must be considered a public asset which can be reused freely. If we could manage to define a specific status for scientific data (or data for scientific purposes), then copyright would cease to be useful.

The specific status of data

Jean Chambaz and Paul Indelicato differentiate between two types of data:

- Observable data
- Constructed data: collection of several pieces of data to which specific methods or algorithms are applied to allow observations to be made from a specific collection of data.

The intellectual process that makes sense of the data collected remains the intellectual property of the researcher and the funding organisation, but there is no reason the data produced during the research process should not become a public asset with a specific status allowing for its reuse. Creating a specific status for data could stimulate research activity for the benefit of society as a whole.

The way data is collected and used depends on the discipline concerned. For certain disciplines, collecting and sharing large amounts of data are crucial for consolidating the discipline, whereas in other areas, such as experimental science, reusing the data collected is not always relevant.

Paul Indelicato points out that, depending on the discipline, for scientific data to be of use it must be accompanied by the method and procedure that produced the data.

National (or European) Data Library

Today, the majority of data collected during research projects is escheated or deposited in dispersed data bases or directories. If data are to be deposited in a specific National (or European) Library, then the associated publications must be deposited with the data.

A European Library would have greater value for disciplines such as medical research, for which sharing data on a national level is not particularly relevant.

It could take a long time to set up a European Data Library; Jean Chambaz therefore recommends that this factor be included in the current debate concerning the JUNKER plan, and endorses the CNRS's conclusions on this point.

New legal aspects to be taken into account in the scientific process

Today, moral codes, religion and ethics have varying degrees of impact on new rights in the scientific field.

If we want to argue in favour of the creation of a global warehouse for free data, we must provide justification that the warehouse is a common universal asset. It is essential that scientists should be able to carry out TDM on scientific data. If this possibility depends on publishers, researchers will be subject to strict controls with no apparent limits.

Publishers are currently worried about the dangers that abusive TDM could represent concerning large community databases (data extraction, pirating, etc.). However, the solution to this problem involves educating users, not controlling access. Establishments, organisations and laboratories must educate and train their staff, researchers and students concerning the use of TDM as a scientific, economic and cultural tool. To enable TDM, the data must be labelled intelligently so as to be reusable and widely shared.

Paul Indelicato stresses that the conditions for reuse must be attached to all data subject to mandatory deposition.

Economic exploitation of the data

Raw data has no specific value in itself: but collecting the data is not a neutral act and is carried out within a specific framework, whereas the information itself does not contain all the ideas for how it can be exploited. It could be argued that the person who collected the data should benefit from all the potential exploitation of the latter. This fact would incite researchers to deposit data in a library and would give great added value to the establishments/organisations that finance the research. This would give value to the collection, processing and labelling of data in order to make them easier to exploit.

The notion of a scientific and/or economic embargo could be established for researchers who wanted to impose a delay before the data was deposited in the library (constructed data).

Jean Chambaz points out that the free reuse of data by the scientific community may achieve consensus, but the law must also provide for the reuse of data by companies and others who benefit from the use of public research. A balance needs to be found concerning the reuse of data for economic and scientific purposes.

For example: the collection of data linked to the behaviour of SNCF (rail) travellers could be used by a scientific team for a sociological study as well as by the SNCF to optimise new services.

Companies could have access to the data to develop their activities on the condition that a return for society as a whole is legally provided for. A clause concerning access to the data by private companies could provide for a fee on the exploitation of the data so that the National (or European) Data Library could finance itself, e.g. via a tax on data use.

Legal instrument for data

In terms of social acceptability, a data item may be freely available for reuse when this means reuse by the scientific community. As regards reuse by companies, a distinction should be made between observable data (free to reuse) and constructed data. As regards constructed data, research teams could use a legal instrument (patent, licence, etc.) to maximize ROI (return on investment) on the data they have collected and labelled.

In a limited number of cases, this legal instrument could justify an embargo period of 5 years (to help write off the costs). This period must be tested to ensure the system provides enough incentive to motivate researchers to deposit their data.

Ethical data

Processing and labelling data creates a new mission for public researchers. They must not only conduct scientific research but also deposit the data processed, which represents an additional cost in time and human resources.

Researchers must therefore not only conduct research but also ensure the digital transfer of data for the benefit of the community.

Alain Bensoussan notes that a new concept is currently coming to the fore; the notion of "ethical" data. This means data that has been "fairly" collected; data that allows for sustainable development.

Data for sustainable development should not be destroyed but tracked and deposited in a database.

Scientific communities should be educated and trained in adopting a sustainable development approach as regards the data they collect for their research projects.

Researchers must understand that processing and depositing their data does not represent an additional cost but new added value. The protocol concerning the

processing and depositing of data must be promoted as a valuable part of the researchers' work. This must become an integral part of their job, just like the assessment and publication of their research.

b. The CNRS Scientific Board: Bruno Chaudret, Claire Lemerrier, François Bonnarel, François Tronche, 30 June 2015

Participants:

For the Scientific Board's working group:

- François Bonnarel, CNRS Engineer, Strasbourg Astronomical Data Centre (CDS)
- Bruno Chaudret, President of the Scientific Board of CNRS
- Claire Lemerrier, Senior Researcher in History at CNRS, Centre for the Sociology of Organisations (CSO), Paris
- François Tronche, CNRS Research Director, Paris-Seine Institute of Biology, CNRS, INSERM, UPMC

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit

For the DIST:

- Renaud Fabre, Director
- Charlotte Autard, Manager in charge of the ISTEEX Investments for the Future project

Overview of the approach:

In the context of the ISTEEX Investments for the Future project, the CNRS Legal Affairs Department (DAJ) commissioned a legal consulting firm to provide project support, at the request of the DIST. The Bensoussan Consultancy was selected to support the Executive Committee in establishing a secure legal framework for the project. This collaboration has already enabled the studies conducted by ISTEEX to on the Digital Republic Bill to be taken into consideration.

This collaboration has led the Bensoussan Consultancy to support the DIST of CNRS in drafting a White Paper aimed at building a set of references, which have been largely reproduced in the Digital Republic Bill on Open Access to scientific data and the principle of Text and Data Mining (TDM).

The French government's digital strategy draws on CNRS's conclusions aimed at Higher Education and Research in terms of digital technology for science, and reasserts Open Access and TDM as key elements in this approach.

The Digital Republic Bill will be presented before the month of July. As such, the White Paper is of great importance. The conclusions drawn from the information collected about uses and practices from interviews with the Scientific Board and other partners (CNN, UPMC, COEPIA, University of Strasbourg, ISTEEX Executive Committee) will be included in the parliamentary debate concerning the Bill. The DAJ and the DIST, supported by the Bensoussan Consultancy, will work together to present the positions of the research community supported by the Scientific Board of CNRS.

The initial feedback from the Scientific Board must be sent to the DIST (for transmission to the Bensoussan Consultancy) before the end of August 2015 to ensure enough time to draft the White Paper before the filing of the Bill for the parliamentary session at the beginning of October.

The Scientific Board's contribution will help focus the debate on the White Paper during the next round of hearings. Minor modifications may be made to the text proposed by the Scientific Board following its next plenary meeting on 24 or 25 September.

Presentation of the objectives

The goal of the collaboration with the ISTEEX Executive Committee, guided by the DIST, the DAJ of CNRS and the Bensoussan Consultancy, is to find a legal framework for the ISTEEX project. A preliminary analysis was made during two CNRS Ethics Committee (COMETS) meetings.

Discussions concerning ISTEEX led to the creation of two projects:

- a "limited-scope" project concerning ISTEEX and the law governing platforms and TDM (more or less free use of scientific data),
- a "wide-scope" project concerning a more general legal framework of Scientific and technical information (STI): a law governing science. Both projects are based on a scientific approach for the use of data.

The Scientific Board's working group is now invited to give its opinion concerning these two "limited-scope" and "wide-scope" projects.

As regards the “wide-scope” project, the DIST and the Bensoussan Consultancy have already obtained the opinions of two leading experts: Laurent Cytermann, the Master of Requests (*Maître de Requêtes*) for the French Council of State (*Conseil d'Etat*) and Alain Abecassis, Head of the Department for Strategic Coordination and Regions.

The White Paper focuses on the opinions of the members of the Strategic Board in their capacity as scientists in order to understand the legal requirements linked to their fields of research.

Publications and data

Definition of data

Data can be sub-divided into 5 categories:

- **Raw:** available in their original state before any human processing
- **Instrumentalised:** obtained using a given instrument (e.g. a telescope in astronomy). The person who controls the instrument controls the data.
- **Analytically interpreted:** resulting from calculations or processing. Ownership must be established between either the person who supplies the data or the person who supplies the algorithm.
- **Scientific data:** interpreted by the human brain ($E=MC^2$). Establishing links between data. This often concerns data that have to be interpreted to support a theory.
- **Data about data, or metadata:** all the information about and relationships between pieces of data that make it possible to interpret the data itself.

At the centre of several uses:

- **Open Data** (open governance of data)
- **Open Access** (possibility of free access to data (or not) with embargo periods for certain fields of research)
- **Open Process**
- **Open Format**
- **Open Use**
- **Open Business**

Finding the right definition for the data can sometimes be complicated, for example in the case of a photo of a work of art.

At first glance, the photo is a piece of instrumentalised data but the object captured by the photo exists outside the photo itself and has its own rights, i.e. those of raw data. There can be legal conflicts regarding the reuse of raw data represented in picture form.

Numerous scientific fields work with data that are not produced by science, i.e. raw data; the uses of this raw data must therefore be precisely determined.

Differentiation between data and articles

An article that is written within the framework of State-subsidised research studies is financed using public funds. Based on the work conducted, researchers communicate using an international standard: the scientific article. Researchers thus become the owners of the articles in their capacity as authors, protected by copyright law.

Researchers pay to have their articles published and then pay again to have access to the journals in which their articles and those of their colleagues are published. Most often, they do not have the possibility of performing TDM.

This observation leads to the following questions: Is the existing system still suited to the opening up and sharing of science? Are private articles still the most appropriate way of disseminating science? Is copyright still adapted to the needs of science?

Compendium of uses

Practices and sharing: principle of Fair Use (TDM)

Practices and sharing: principle of Fair Use (TDM)

In biology, digital publishing has been generalised and article searches are done via platforms developed by academic institutions, such as the National Institutes of Health (NIH, USA), providing free access to article summaries. A fee must generally be paid to obtain access to the complete articles, with the transfer of copyright to the publisher being a common practice. It should be noted that the NIH has objected to this practice and proposes open access, via PubMed, to an unformatted version of any article published by a publisher describing work funded by the NIH. Open Access is growing rapidly.

The production of research articles published in digital format is not only accompanied by the production of raw digital data but also of equipment (collections of DNA, anti-bodies, cells, apparatus, software, etc.) and living organisms (micro-organisms, strains of mice – selected or genetically modified) used in the scientific process presented in the research article. Depending on the type of production, their ownership is subject to copyright or the practices of the supervisory authority. Let us take the example of a genetically modified mouse; the ownership depends on the practices of the supervisory authority and any potential revenue is divided between the inventors, the authority and the laboratory. When a published article describes research that produces raw digital data or biological material, the publisher generally asks the author to sign an undertaking to donate the data or the material to any researcher from a public organisation who might request it. This practice is essential for validating results by reproduction and is facilitated by the existence of international platforms providing free access to data and distribution of the biological material.

TDM on texts is not a priority in most fields of biology as a discovery tool, although it is as regards documentary collection. TDM on data is becoming increasingly commonplace. Free access to these data is widespread, as happened concerning the human genome.

Above and beyond the issue of mass data, several publishers, including *Nature*, are considering implementing a system allowing access to the raw data used to create the graphs in an article via their websites. While this will enable readers to verify that results have been interpreted correctly, it raises the question of who owns the data and how ownership is transferred. This suggestion that publishers could extend their control to include data is worrying and may represent a new obstacle to accessing science.

When we compare uses, **astronomy** would appear to be the best organised research field in terms of data and access interoperability.

The community reached a consensus for establishing an embargo period for disseminating and reusing instrumentalised data. This embargo period only comes into force on publication of the article, and applies after its publication according to a variable time period prior to free reuse.

François Bonnarel points out that astronomy is a research sector in which the commercial value of data is often zero, and this therefore facilitates data sharing. Moreover, the need to exchange data is driven by the wide variety of observation techniques (linked to different wavelengths) potentially available for a given object.

François Tronche points out that **in biology** data access is free for the academic community, sometimes with a one or two-year embargo period; this period does not depend on the publication date since the embargo begins when the data is discovered.

As it stands today, there is no specific document that stipulates the embargo period. **In astronomy**, the “ownership” period is defined based on the instrument or the complete project (several instruments).

The embargo period gives the researcher an exclusive right to his/her results and data for publication purposes.

As such, we are faced with a two-sided economy: a donation-based economy (data available freely to researchers in the same community) and a market-based economy (access to publications).

Claire Lemerrier notes that researchers must sometimes work with large groups such as Amazon to have access to large storage areas (example provided by Pierre Binetruy during a previous meeting) but always in exchange for something (payment, access to the data, etc.). Most of these large groups also have research laboratories.

If we plan to make the data available to the scientific community, how can we verify that they are not used by these private research laboratories to make a profit? How can we establish whether the reuse of the data is for scientific purposes, for profit-making or for other purposes (educational, for example)?

If a historian decides to use works of art for research or training purposes, these uses must be differentiated from the use of the data to publish “fine editions” of works of art (to make a profit).

In order to define the possible fields of data reuse, we could create a principle of *Fair Use*. There is great demand for the reuse of data by the scientific community and in education (the right to quote and reproduce for research and training purposes).

The concept of a right to practice TDM could therefore be extended into a principle of *Fair Use*.

Copyright

Exchange is a key part of research. If this is restricted by controlling access to data using logins (Application Programming Interfaces, or APIs) or payments (subscriptions, Article Processing Charges, etc.), scientific progress may be slowed down.

Fair Use provides for the right to use or quote a text on the proviso that the text is an extract, printed between quotation marks, with the author's name indicated. However, Fair Use is not valid for pictures. Pictures cannot be reproduced even with quotation marks, the source and the name of the author. This is also true for extracts of sound recordings.

This practice was relevant for analogue technologies but is not suited to the digital age.

Copyright remains sacrosanct in France. If we want to move things forward, we cannot seek to create an exception to copyright. For the moment, the focus must be on the data item and not on the article. In this way, science could one day return to an exception to copyright through the concept of "data right".

It is therefore crucial that a consensus be found between the different scientific communities in order to foster free data exploitation. If the data item is not free, it must be covered by the principle of Fair Use to allow for its reuse.

Claire Lemerrier notes that the principle of *Fair use* as regards sound extracts or pictures would only be possible for data produced by scientists. The problem in HSS is that this can involve the reuse of data that are not "scientific" in nature (for example, it is possible to study objects found during excavations, communications on company websites, or pop songs). Other data used in HSS cannot be made freely accessible because they are subject to private data protection or non-disclosure clauses. In certain cases, as in the field of public statistics, a solution was found by making the data anonymous.

Today, some publishers ask HSS authors to attach their data to their articles, particularly in economics; however, most HSS data cannot be published (personal data, psychological test data, photos of archive documents that archivists only allow researchers to take on the proviso they do not disseminate them, etc.).

The HSS community would favour the creation of exchange platforms but the legal issues need to be addressed for this to be possible.

For example, we could create a platform that allows free exchanges between researchers by making the data anonymous, as is the case for statistic exchange platforms (projects are under development for more "qualitative" data, but this is far from simple).

The platforms (datasets about a given project made available on a given computer) could be recognised as "essential infrastructures".

Legal amendments expected

Community expectations

The Bensoussan Consultancy, commissioned by the DIST, would like to interview the Scientific Board about these expectations with a view to drafting a text on Science Law and a text on Platform law.

Today, digital law concerning science is not clearly defined. Science produces platforms, but there is not yet a Platform law. Establishing a Science law is a major challenge as regards the recognition of scientific community practices. Today, digital science requires a specific law that allows it to evolve in terms of its uses and practices without being restricted by copyright or abusive publishing practices.

While the universal principle of science is well established, the advent of digital technology means we can no longer continue to have a donation-based economy that coexists with a market-based economy

With the CNRS itself, there is a paradox between the Open Access policy and the policy of the “technology transfer” service.

Claire Lemerrier points out that science must address the issue of misappropriation of results and data. Could this appropriation become legitimate? When is data appropriation considered abusive? Not only appropriation by publishers but also by other external players.

The White Paper's approach must be based on the rules provided by the scientific community.

Claire Lemerrier points out that, while this approach is valuable, particularly as regards preventing misappropriations and increasing access possibilities, there are two clear risks:

- Standardisation: what degree of standardisation would a Science law involve? The different practices of different scientific communities must be respected since these practices are generally there for good reasons.
- How non-scientific outcomes might be taken into account by Digital Science law: how could large groups such as Amazon and Google (and other companies in other fields, such as insurers, who also have their own research centres) be prevented from freely appropriating scientific content? There are certain types of

data that scientists dream of being able to share freely but which would pose serious ethical problems if exploited for profit-making purposes. A data item is not intrinsically scientific or non-scientific. It is the use that is scientific, but how can we establish what constitutes scientific use?

The Law governing Science could provide for:

- a principle of *Fair Use*
- the free exchange of all elements in a scientific process between scientists
- an embargo period before data publication/dissemination. Researchers would thus have priority access to the data for a given period (depending on article publication rates or exploitation initiatives).

Construction of the Digital Science Law

The Digital Science Law must be built; we can therefore propose content but this must be supported by scientific communities in order to be valid. If we create a text that is supported by the scientific community as a whole, its adoption will be quicker.

François Tronche points out that everything that is described in an article that is not patented can be reproduced by anyone. Many researchers do not file for patents so that their findings can be disseminated to a broad public.

Moreover, it would be possible to envisage not actually publishing, and thus not letting the article or data item become private property, while not restricting access either.

Publishers are starting to be interested in metadata. Metadata must also be protected against misappropriation by publishers.

If services to categorise and use metadata in a smart way were private or fee-paying, François Bonnarel believes researchers unable to access these tools would be less competitive than those who could afford to privatise their metadata.

Embargo period

The ownership period cannot be determined in numbers of years because the duration of projects can vary from a few months to a few decades. We could consider an embargo

period as being up to the date of publication of the article. This period would be linked to the notion of the project's end, as long as it is possible to determine the project period (start and end). Or the embargo period could be defined by an agreement.

Creation of an alternative publication model

Journals financed by public institutions

If we cannot change copyright, there are alternative models that could be developed to enable publishing without the problems of misappropriation.

In France, there are HSS journals (often financed by public institutions and based on an OA model), for example those present on the Open Edition platform, managed by Universities and CNRS. These new OA modes of publication, which are not subject to misappropriation, could become models for the future.

Impact factor

In France, the problem with OA journals is the impact factor. In many disciplines, the assessment of researchers in France and Europe is primarily based on their publications in journals with a high impact factor. This system is actively promoted for the assessment of research. Journals that attract articles are therefore those with a high impact factor. Since the impact factor depends on the journal's selectivity, the system can have perverse effects and certain OA journals have been able to increase the number of fee-paying articles published in the year during which their impact factor was good. This increased their profits, even if it lowered their impact factor the following years.

This type of assessment based on impact factor is changing, even in the United States. Scientific texts by French researchers are formatted with this assessment method in mind.

Information sharing

Although, for the moment, in most fields of biology, researchers only see the point of being able to consult articles, and not of being able to "mine" them, the community actually needs free access to articles as quickly as possible. In biology, sharing also involves sharing manufactured material. Many publishers want the community's access to the material described in the article to be guaranteed so that anyone can reproduce

the experiment. Furthermore, several publishers are currently implementing a system for accessing raw experimental data summarised in article tables or graphs.

Although the biology community appears to share the most, this sharing does not solely concern data but also the material produced within the framework of the article.

In HSS, the community needs to protect itself from data appropriation and to have free access to articles, thus reducing the cost of access to information. TDM is also important; this concerns access to a large body of resources. Data sharing is not yet sufficiently widespread; there is a vacuum in terms of culture and human resources as regards sharing. If a law governing public sciences and platforms is created, it will perhaps enable HSS communities to implement more widespread data-sharing platforms based on the model of statistical data platforms.

In HSS, it would also be useful to establish a principle for Open Data and sharing that would continue after the end of the project; however, for most data, this should probably be limited to sharing among scientists, with all the problems of definition that this entails.

In astronomy, the community shares a lot of data, but the reuse rate remains low. Algorithms must be linked to articles published in OA so as to make it possible to check the result and render it reusable.

The working group concludes that the common concern is to enable the free sharing of information between scientific communities. To achieve this, an intelligent means of sharing data must be found.

Software sharing

In astronomy, as regards scientific results, when we talk about OA we refer to the sharing of instrumentalised (or raw) data on which the algorithm (that makes the data usable) is built. Could the algorithm and the software also be shared?

François Bonnarel says that the sharing of software is not only desirable but already occurs. However, there is not yet a general, standardised system for exchanging software. Nonetheless, there is a whole range of interoperable software packages in the field of virtual observatories, although this represents only a small proportion of the software programmes used in astronomy.

François Tronche points out that in biology, not just the associated material but also the software can be shared, if it is described in the article. Most of the time, the software developed by the academic community is freeware. Certain large centres or academic organisations develop their own software in open access mode.

Specific forms of sharing practices and knowledge

Many platforms are created and then disappear according to needs.

EMBO (an international research and publishing organisation) was the first to implement a service providing access the raw or processed data used for graphs.

(Thomas Lemberger - <http://www.embo.org/>)

François Bonnarel notes that a recent paper, listed on the INSU website (National Institute for Earth Sciences and Astronomy) talks about articles that are no longer based on observations made within the framework of research studies but on archived data. The reuse of forgotten or unexploited data. This trend is growing in astronomy, where researchers sometimes do not make new observations but publish articles using open access archived data.

François Tronche adds that, in biology, a large proportion of a laboratory's experiments are often sub-contracted, within the framework of partnerships or service contracts. In practice, the person who acts as sub-contractor can do this for no fee if they are cited as co-authors of the article, or for a fee otherwise.

c. University of Strasbourg, Paul-Antoine Hervieux and Françoise Curtit, 10 July 2015

Participants

For the University of Strasbourg:

- Paul-Antoine Hervieux, Deputy Vice-president for Partnerships with EPSTs and local authorities
- Françoise Curtit, CNRS, Responsible for the "Open Access" mission at the University of Strasbourg

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit

For the DIST:

- Renaud Fabre, Director
- Charlotte Autard, Manager in charge of the ISTEEX Investments for the Future project

Overview of the approach:

The CNRS, in conjunction with the Alain Bensoussan Consultancy, decided to draft a White Paper on "Digital Science and Law" in response to the issues raised by the legal framework covering the ISTEEX Investments for the Future project ANR-10-IDEX-0004-02 (www.istex.fr). This project raises many legal questions due to its potential in terms of TDM, interdisciplinarity, content aggregation and its aim of making its databases explorable.

The new law on science has a specific exception concerning copyright: the legal status of scientific results and notably data and metadata.

This White Paper aims to **propose a legal framework for scientific data in order to address the concerns of the scientific communities** (data publication and uses, Science-platform law, text mining, etc.) and thereby **contribute to the French Digital**

Republic Bill. Science is preparing the terrain for the Digital Law, by working on tangible points upstream.

Initially, the Bensoussan Consultancy and the DIST will establish working groups and conduct hearings to **gather information about the community's uses and the current state of exploitation practices.** The close association with universities and other organisations will enable hearings to be conducted with key witnesses and ensure pluralistic expression regarding the changes being prepared as part of the Digital Republic Bill.

This information about uses and exploitation practices will enable the CNRS and the Bensoussan Consultancy to establish a **matrix concerning the relevance of analyses and practices which will be set against the current normative framework** in order to assess gaps in the law and **make proposals for the Bill.**

Contributions to the White Paper will be anonymous and the people interviewed will have total control over whether their names are quoted or not.

The DIST and the Bensoussan Consultancy will re-contact the people interviewed and will submit an initial text to the University Presidents for approval.

Today, the Bensoussan Consultancy (commissioned by the DIST) would like to record the opinion of the University of Strasbourg concerning potential draft proposals and recommendations.

Copyright

Paul-Antoine Hervieux points out that in academic communities there is an almost total lack of awareness of copyright and its implications, regardless of the type of document or data. The issues raised by the digital law represent a huge project in terms of training.

The White Paper aims to collect the opinions of scientific communities, which are aware of any deficiencies. The DIST and the Bensoussan Consultancy would like to know if the same rights concerning scientific results and data could apply equally to different scientific communities.

According to Paul-Antoine Hervieux, a law concerning data and results is becoming essential in the light of developments in the digital sector. Uses are changing and the paradigm for research data (BSN10) is undergoing a transformation. Researchers and academics are beginning to realise the value of their data. Once we start talking about value, we need to start thinking about rights. We are now in a world where private

businesses predominate and are increasingly interested in data with a view to commercialising its uses.

Legitimate appropriation and misappropriation

The University of Strasbourg is acknowledged for its research in the field of chemistry. The University owns many resources, such as experiment data and databases in which a considerable number of properties concerning chemical reactions are recorded. The latter can be of great interest for private companies; for example, the pharmaceutical industry is increasingly interested in the catalogues of chemical reactions, with a view to reusing this information for its own purposes.

These data must be protected. Scientific communities are not yet aware of this danger, notably in universities.

Renaud Fabre highlights that we must differentiate between legitimate appropriation and misappropriation.

A specific text concerning Open Science could define what constitutes misappropriation.

Business model for data

Paul-Antoine Hervieux highlights that the University of Strasbourg has data in its laboratories that could be very valuable for the private sector. A business model could be created using the data and could be used to support and finance fundamental research. The data should be open, even to the private sector; however, to do this, a business model would have to be defined, specific rules established (a guide to data) or an easy-to-use tool developed for researchers.

Research can be differentiated by:

- Financing through public funds
- Financing through private funds

These aspects must be used to define a business model for making the data available to private organisations.

One possible idea (as an outcome of the White Paper) would be to propose an A-to-Z, frequently asked questions aimed at researchers so they could find their way through the labyrinth of data law

Embargo period

Astronomical data are produced by specific instruments: instrumentation data. The community puts an embargo on its data for a one or two year period. Could the idea of an embargo period be shared with the University of Strasbourg ?

Paul-Antoine Hervieux answers that the key factor in science is the immediacy of research. Each data item has a certain lifecycle, but the fresher the better. For science to advance, the embargo period should not be an obstacle to the dissemination of results. However, depending on the scientific community concerned and for various reasons (e.g. competition, assessment process under way, etc.) an embargo period may be introduced.

Software and algorithms

To maximise potential data reuse, one must have access to equipment that can process it. When data is open and made available to communities, should the software and algorithms also be made available online?

Paul-Antoine Hervieux says that researchers have a philosophy of competition today. Each researcher must publish a certain number of articles a year to remain competitive. With this in mind, researchers will not be particularly willing to share the results of their work, or the calculation codes that are at the heart of their research. It is possible to share information about the structure of the code (but not about the algorithm itself, at least not for a certain period of time) in an article describing the methodology used or in a journal that publishes calculation codes. However, the majority of researchers are wary of immediately publishing their codes and algorithms.

Today, there are already journals such as *Computational Physics* that allow physicians to deposit their calculation codes.

It would also be possible to deposit the executable programme in open access mode with a copyright on the code.

If communities want to have an embargo period for the depositing of software and algorithms, a period of one to two years would be suitable.

Scientific publications and repositories

Paul-Antoine Hervieux points out that the relationship with publishers is central to the concerns of researchers, particularly as regards TDM. TDM will be the ultimate research tool in the years to come.

Publishers seek to protect themselves by controlling the possibility of TDM via Application Programming Interfaces (APIs), under contractual clauses (Creative Commons licenses, restricting the number of 'explorable' words).

It is obvious the principle of open access must comply with copyright, but a distinction must first be made between the exploitation of corpora and the reuse of data already compiled. We must differentiate between raw data and constructed data.

When an institution has an open archive, it has the opportunity to implement its own policy. At a time when university establishments are becoming increasingly independent and must establish their own strategies, open archives and TDM are a major issue.

The University of Strasbourg aims to build its own archive to store its scientific production and make it possible to locate it in a single place. This archive will give the university community in Strasbourg the right to consult and search through its own production with no lock-outs (unless there is an official lock-out, such as Sherpa, or an embargo).

According to Paul-Antoine Hervieux, the CNRS should ideally take the position that all scientific production (CNRS and partners) should be deposited in a common archive (not necessarily a single one) with similar rules.

As the contact person for the EPSTs at the University of Strasbourg, Paul-Antoine Hervieux points out that University/CNRS partnerships work well despite them having distinct policies. It therefore seems obvious that the same data could be shared with the same rights.

Possible actions

An article in the draft text proposed by the Bensoussan Consultancy could state that: “data from scientific research and the associated results can be freely used and made freely available to scientific communities involved in the publications, with no restrictions, except for software, which may be made available to the community after an embargo period of one to two years.”

Renaud Fabre points out that the purpose of the HAL platform is to incite CNRS researchers and organisations to deposit their publications.

Paul-Antoine Hervieux explains that the University of Strasbourg’s major project to build an open archive of the university’s knowledge also involves developing a connection between the university’s archive and HAL so that publications are automatically deposited on the platform. CNRS is supporting this project. The University of Strasbourg wants to focus on creating a repository of all the work produced by its researchers, essentially for promotion purposes, but also aims to comply with the CNRS and European policies via the automatic transfer of information to HAL.

Digital Republic Bill

Françoise Curtit raises the issue of the calendar for the digital law in France and that of European Directives that are currently being amended. Will European law not take precedence over the French digital law?

Alain Bensoussan points out that the President announced this regulation as creating a French law on digital issues in a similar vein to that of Human Rights. The Prime Minister supports this initiative and a text is already being studied by the Cabinet of the Secretary of State, Axelle Lemaire.

The Research Code deals with the organisation of science but not with that of scientific data. The Digital Law could include an article that would amend the code on scientific research and take into account the White Paper’s recommendations.

Renaud Fabre says that the government’s digital roadmap was made public by the Prime Minister on 18 June 2015 and that the provisions in terms of science promote the development of Open Access.

These provisions also cover the authorisation for text and data mining (TDM) and an amendment to the Intellectual Property Code is planned with this in mind.

In both cases, the road map clearly sets out the requisite elements for open access and free browsing through scientific literature.

Open Process

The issue of Open Process and of whether it should be possible to search freely through data highlights the question of legitimate and misappropriation. The DIST and the Bensoussan Consultancy aim to analyse this issue with the representatives of the communities interviewed. The hearings have already improved and clarified the definition of this practice. The Scientific Board of CNRS will give its opinion on these subjects in September.

It is also essential to gather the opinions of major Universities (such as Strasbourg and UPMC - University Paris 6) to have an overall view of the French research ecosystem.

Platform regulation

The question is as follows: should platforms be regulated and qualified as “essential infrastructures” for scientific research?

Paul-Antoine Hervieux believes it is crucial to establish a legal handbook for the use of platforms. This code of conduct would allow platform users to understand the regulations and uses and adapt their behaviour accordingly. The University of Strasbourg is already looking at drafting a document about platform use.

Renaud Fabre states that “essential infrastructures” are defined by two conditions: they must be essential and they can only be implemented through the resources of public authorities (e.g. airports).

Digital platforms for Science are essential and can only be a means for sharing knowledge.

For example, the TGIR Huma-Num (Very Large Research Infrastructure for the use of digital resources in the HSS) currently has no specific status but fulfils several needs of scientific beneficiaries. Labelling it as an “essential infrastructure” would establish it as an upstream structure of public science.

Paul-Antoine Hervieux agrees with Renaud Fabre about the need to recognise digital platforms as “essential infrastructures”.

Initially, rights must be defined for the platforms as regards the conservation, consultation and sharing of data between people and organisations involved in public science before the data are disseminated to all users and beneficiaries.

Françoise Curtit points out that her work on open science enabled her to discover the wide range of uses among different communities and the impact this has on the associated legal issues. A "wide-scope" model might not suit all disciplines. The variety of uses must also be taken into account when defining platform rights.

The DIST and the Bensoussan Consultancy propose to base the model on the generally-accepted practices of each scientific community in order to institutionalise the uses. This first requires compiling the practices of the different communities.

Paul-Antoine Hervieux and Françoise Curtit agree that this process is essential. The White Paper must take practices into account in order to define common orientations without dividing communities. The White Paper will define common guidelines that take different practices into account.

d. French Digital Council: Benoit Thieulin and Yann Bonnet, 1 September 2015

Meeting with the DIST and the Bensoussan Consultancy within the framework of the drafting of the “White Paper on Digital Science and Law”

The process:

Drafting of a White Paper (notably) with a view to contributing to the Digital Republic Bill.

- Project led by **CNRS** in conjunction with the **Bensoussan Consultancy**.
- In response to the legal issues raised by the **ISTEX Investments for the Future project** (www.istex.fr), which involves **two actions**:
 - A **vast programme concerning the acquisition of scientific resources**, in the form of national licences.
 - **The creation of a digital library** with remote access for all members of higher education and research establishments.
- This project raises many legal questions due to its potential in terms of text and data mining, interdisciplinarity, content aggregation and its aim of making its databases explorable.
- The White Paper aims to **propose a legal framework for scientific data in order to respond to the concerns of scientific communities** (data publication and uses, law governing Science platforms, text mining, etc.) and thereby **contribute to the French Digital Republic Bill**.
- The Bensoussan Consultancy and the DIST are establishing working groups and conducting hearings to gather information about the community's uses and the current state of the art regarding exploitation practices in order to establish a matrix concerning the relevance of analyses and practices which will be compared to the current normative framework in order to assess discrepancies with the law and make proposals for the Bill.

Responses to the interview guide

I. Context

Digital technologies are profoundly transforming the modes of production and dissemination of scientific results: data, publications, analyses are now accessible on various platforms. This availability of scientific material contains a potential for knowledge exploitation and sharing for which the law must be able to define the conditions, terms and limits.

The problem today is that there is no specific legal status concerning scientific information.

Scientific research studies and work are considered as written works, subject to copyright law.

- To develop their careers and reputation, researchers must publish the results of their work in certain journals; to do this, they must transfer all their rights to the publishers.

The publisher thus has the sole right to exploit, reproduce and disseminate the article.

→ This means that research organisations and the scientific community cannot have access to the study/article unless they conclude an agreement with the publisher.

- In addition to transferring all their rights, the research organisations must pay two types of costs:
 - Upstream publishing costs;
 - Downstream fees for consulting documents: higher education and research institutions spend more than 80 million euros a year to gain access to electronic resources. Access fees have also continually increased: 7% a year over the past 10 years.

→ This represents a burden for public finances and hinders the productivity of public research, which must already cope with intense international competition.

- As a result, there is a major imbalance between researchers and publishers, which has been aggravated over the last few years by the emergence of oligopolies in the scientific publishing sector (Elsevier, Springer, Wiley, etc.).

With this in mind, we must assert the right to open access and open science, in order to respond to the concerns of the scientific community.

Definition of open access. Open Access (OA) publications refer to articles that are accessible in digital format and that can be read free-of-charge via the Internet. Open access allows readers to read, download, copy, distribute, print, search or create a link to the full text of an article, to index it, recover it for computer processing and use it for any other legal use with no financial, legal or technical obstacles, while fully complying with copyright.

- **Objectives**
 - **Reduce the burden of costs** of digital journals in the budgets of public establishments;
 - **Facilitate access to scientific knowledge** for the research community and civil society;
 - **Provide companies with broader access to the results of scientific research**, notably small and medium-sized companies that could thus improve their innovation capabilities.
- The European Commission has invited Member States to "Define clear policies for the dissemination of and open access to scientific publications resulting from publicly funded research" (2012).

CNNum proposals on Open Access

- **The conservation and dissemination of research results are public-service missions.** A movement supporting open publication already exists through open archive warehouses belonging to universities and organisations, or via the HAL platform.
- Today, a specific legal framework would encourage this movement initiated by members of the scientific community.
 1. Recognise the right to secondary exploitation, as under German law. The author's version deposited in an institutional archive remains in open access, whatever publishing course is later taken as regards the work;
 2. Provide open access to scientific publications financed through public funds, after a short embargo period allowing for the publisher's commercial activity, either in open journals or in an institutional repository (as under German and Italian law). This obligation should not lie with the researchers but with the research organisations.

→ Care must be taken not to create a simple **option** to publish these documents in open access, which would render the rights ineffective (as is the case in Germany). It must be an **obligation**.

→ **The English example is highly instructive.** The HEFC¹³⁸, the organisation responsible for sharing out the overall sum of money allocated by the State between the different higher education establishments, has announced that any publication that is not available in Open Access as of next year will not be taken into account when assessing the activity of the establishments.

3. **Encourage researchers to give open access to raw, anonymous research data** as long as this does not involve any issues regarding ethics or personal data

II. **CNRS's proposals: conditions, terms and limits to the sharing of scientific information**

A. Conditions

The Position of CNRS: The main conditions for free access to scientific results are the abolition of limits that may be introduced by editorial legislation (publication rights, copyright), with a view to the exploration of digital corpora of publications or data.

Question 1: What is your opinion on the necessary adaptations to publication rights (publishers and/or authors) and the exploration of corpora (text & data mining techniques, APIs, etc.)?

Definition of text and data mining. *Text and data mining* refers to various extraction and analysis tools that allow automated exploration of digital content, which can include text, data, sound, images or other elements, or a combination of these elements, in order to find new knowledge or ideas.

Text and data mining techniques can help boost French research in the age of Big Data.

¹³⁸ Higher Education Funding Council.

- The MIT has qualified these techniques as one of the 10 emerging technologies that “will change the world” in the 21st Century. Examples of applications:
 - The Text2Genome project made it possible to map the human genome by automatically compiling 3 million publications.
 - As regards the press, data mining is undoubtedly one of the future business models as regards information.
 - TDM is what enables Amazon to generate 20% of its turnover.

TDM is not in itself a new activity.

- **It just involves reading and extracting information and meaning from documents.** It is not really so different to gathering information manually, which has been the way of research since the birth of science.

However TDM often requires the creation of copies and content storage, which constitutes reproduction in terms of copyright law. Carrying out computerised processing of content repositories, whatever its nature (text, data, static or moving images, music, sounds, etc.) means taking possession of the content and storing it in order to search through it, or carrying out substantial data extraction, as regards the rights of database producers.

The current normative framework does not allow TDM to fulfil its potential

- **The individual management of rights is not adapted to such a large mass of data.** In view of the many different sources and the large volumes processed, it would be impossible for the data miner to seek authorisations one by one.
- **The contractual solution is not satisfactory.** The agreements proposed by scientific publishers severely restrict the uses that are authorised and sometimes set significant constraints and prohibitions:
 - Researchers are required to declare their research, which goes against all scientific ethical codes (research confidentiality, etc.).
 - Researchers are dependent on scientific publishers who propose platforms with extremely limited services that are criticised by researchers. Contractual agreements make it difficult, if not impossible, to search between different repositories and datasets.

Proposals of the CNNum concerning text and data mining

Introduce an exception to copyright law, without any compensation, providing for the right to use the information for TDM purposes within the framework of research.

→ This does not go against the exceptions established by the EU concerning copyright. The English have such an exception through their “fair dealing” principle.

B. Terms

The Position of CNRS: The terms for sharing scientific results must, in the digital age, assimilate new constraints: the sharing of results between actors in public research on the one hand and between users and beneficiaries of public science on the other.

Question 2: How should the line be drawn between legitimate appropriation and misappropriation of results available on a public science platform, and how should these results be protected?

- **Position of the CNNum concerning the relationship between open resources and appropriation:**

- **The benefits of open dissemination are currently underexploited by society as a whole.** Often, it is the largest players, notably well-established web platforms, that seize these benefits by combining the Commons with their own resources; thus, there is a real risk of predation.

- **The response to this risk cannot be to turn back and abandon this movement in favour of openness.** The aim must be to help a large number of companies, associations, public organisations, researchers, media, etc. to develop their capacity to contribute and participate in the Commons, and particularly to use the resources.

→ **Possible use of non-commercial or share-alike licences:** there is a whole range of contractual solutions available for managing the reuse of research results. Creative commons has developed solutions that are both easy to understand and machine-readable.

→ **We must nonetheless consider the objective of public research: is it not publicly funded partly in order to drive the economy and society as a whole?** An increased number of licences may act as an obstacle to data exploitation and complicate the use of platforms. Ultimately, it is more of a political choice than a technical one as regards the public research model we wish to defend.

C. Limits

The Position of CNRS: Science platforms today contain STI whose form, content and legal status are very heterogeneous. This lack of uniformity impedes the visibility of science platforms. In the same way that there is today a notion of general interest data (recent choice of the Minister of the Digital economy), consideration should be given to the designation of public science platforms and, possibly, to a specific legal regime.

Question 3: Should we be moving towards a designation of "essential infrastructure" for major upstream research platforms, in cases where these platforms occupy a unique and irreplaceable function?

- **Platforms are becoming the gatekeepers of information. We need to rethink the concepts of competition law as regards "essential infrastructures" to impose an obligation of open access on these players.**
- This clearly raises the issue of independence. Do we want a private operator to control the access rules to public-research knowledge and to make money from this access? I do not believe so.

→ This is the very idea of the Commons.

- The importance of free access.

For knowledge and access to knowledge as a Commons, please see:

Charlotte Hess and Elinor Ostrom, (coord.) Understanding Knowledge as a Commons, MIT Press, Dec. 2006.

Gaëlle Krikorian and Amy Kapczynski (coord.), Access to Knowledge in the Age of Intellectual Property, Zone Books, 2010.

e. ISTEK Committee, Laurent Schmitt, Jean-Marie Pierrel and Grégory Colcanap, 24 September 2015

Participants:

For the ISTEK Executive Committee:

- For the COUPERIN Consortium: Grégory Colcanap, Coordinator, and Monique Joly, Coordinator of the Studies and Forecasting Department
- For INIST: Laurent Schmitt, Head of the Projects and Innovation Department, standing in for Raymond Bérard on the Executive Committee for the duration of his absence
- For Lorraine University: Jean-Marie PIERREL, Professor

For the Ministry of National Education, Higher Education and Research (MENESR):

- Marie-Pascale Lizée, Scientific and Technical Information and Documentary Networks Department (DISTRD)

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit
- Sarah Lenoir, Barrister, Intellectual Property Unit

For the DIST:

- Renaud Fabre, Director
- Laurence El Khouri, Deputy Director
- Charlotte Autard, Manager in charge of the ISTEK Investments for the Future project

Overview of the approach

As mentioned during the ISTEK legal seminar of July 2014 and in the Executive Committee meeting, the Bensoussan Consultancy, in conjunction with the DIST and the CNRS Legal Affairs Department, has undertaken to draft a White Paper entitled "Open Science in a Digital Republic", in response to issues raised by the ISTEK project.

This approach is based on the issues raised at the ISTEEX seminar in July 2014 on legal security, on the themes: objectives and management of databases, text and data mining (TDM), interdisciplinarity and content aggregation.

These themes are of course included in the Digital Republic Bill drawn up by Axelle Lemaire, Secretary of State for Digital Affairs.

The purpose of the White Paper is to contribute to an **examination of the legal framework for scientific data in order to address the concerns of the scientific communities** (data publication and uses, laws governing science platforms, TDM, etc.) and to **participate in drafting the Digital Republic Bill** by making useful proposals.

The approach proposed by the Bensoussan Consultancy is initially **to conduct hearings** in order to **survey the practices of the scientific community and learn how STI is currently exploited**.

By surveying and compiling these current practices it will be possible to draw up a **matrix of analyses and practices to be set against the current normative framework**, in order to assess discrepancies and **develop proposals**, which will be submitted to all those involved in drafting this White Paper.

TDM and the right of observation

The experience gained during the ISTEEX project shows that it is more difficult to set up tools and rights for TDM in France in the current legal framework and under the current relationship with the publishers.

Grégory Colcanap considers it indispensable to obtain an exception to copyright, as is the case in England, to allow TDM practices to develop in France, together with a right to read.

In France, publishing is mostly considered to be a cultural issue: an exception that had an adverse effect on publishing would be seen as an attack on French culture.

Jean-Marie Pierrel recalls that TDM is merely automated reading and therefore basically no different to normal reading, except that it is done by a machine.

The right to perform TDM is also the right to produce research data that become accessible to the entire community.

Beyond the possibility of TDM, it is necessary to ensure the preservation of intermediate copies, i.e. annotated and modified documents. At present, some laboratories are

already performing this preservation work but must now develop the tools for disseminating and sharing these data, as well as metadata.

Depending on the disciplines and practices, metadata can be divided into two categories:

- Information associated with the document (metadata),
- Information associated with the way the document is used (usage data).

User-generated content (UGC), a by-product of the use of article repositories, which is useful and recognised within the industry, must henceforth be included in these discussions in order to assess the added value of these usage data and determine how best to incorporate them in the assessment of scientific work.

Renaud FABRE indicates that TDM is widely known and recognised as a practice, and has inspired considerable writing and led to strong positions being taken by such organisations as the ADBU, academic libraries, and all the directors and STI associations of research institutions.

In the framework of the consultation on the Digital Bill, TDM as a tool for investigating the immense quantities of data and publications via electronic processing is a crucial issue for the future of research. Most of the major countries in research (Germany, Canada, United States, United Kingdom, etc.) have adopted such legal provisions: France cannot impose measures that set it apart from the international scientific community.

Jean-Marie Pierrel indicates that researchers acquire TDM tools to search through texts or datasets in order to explain phenomena in extreme detail and to be able to use electronic processes to bring them to light.

Researchers now need to be able to:

- acquire primary data considering current market conditions,
- access these primary data for unlimited observation.

TDM must therefore involve not just a right to read but a right to observe. TDM is a right of observation of scientific objects, indispensable for science. If scientists have a universal and fundamental right of observation in the interests of scientific progress, then an exception to TDM is no longer necessary because TDM becomes a fundamental right.

The right to observe scientific publications is necessary in order to be able to synthesise a multitude of observations across common areas. IT is just a specific device for observing data.

The monopoly held by publishers aims to limit this right of observation of digital scientific databases by imposing the use of their own TDM tools and by controlling their use (APIs, limiting searches to a certain number of words, etc.).

When acquiring the right to read content, this should not include any limitation on the right to observe.

Critical infrastructures and trusted third parties

As TDM rights are recognised in ISTEEX, this platform could be used as a model of a trusted third party for the depositing/archiving of intermediate copies.

ISTEEX could also be recognised as an “essential infrastructure”.

Other organisations, such as the French National Library (BNF), could potentially act as trusted third parties.

Embargo period and deposition

Jean-Marie Pierrel considers that ideally there should not be any embargo period on scientific publications but only on scientific data.

This is because an embargo period enables teams to benefit from the full use of the scientific data throughout the entirety of a research project or a thesis.

In addition, the metadata must be accessible from the outset in order to ensure the sharing of information in the framework of ongoing projects in order to avoid duplicating costs, for example in projects using the same set of corpora.

Marie-Pascale Lizée points out that the embargo period recommended by the European Commission (6 months after the date of first publication and 12 months for the HSS) is supported by the MENESR, but that opposition from publishers in the framework of the Digital Bill has led to a retreat in the Bill, with a proposal for an embargo of 12 months, and 24 months for the HSS.

Jean-Marie Pierrel states that for some areas of research that require an embargo period on data of 2 to 4 years this should be possible (if justified).

The 24-month embargo period for the HSS clearly expresses the fears of French publishers, but mainly concerns books and not articles. An embargo period on articles by the publishers of scientific journals in HSS would not affect them. French publishers must be educated in this respect to defuse the situation.

Grégory Colcanap suggests that, concerning research data that are to be made available, some provision could be introduced in the funding mechanism stating that when data are produced in the framework of a project funded by a public body they must become public at the end of the project (or of an embargo period).

The embargo period on articles by the publishers of scientific journals must be removed, but not that on books.

f. ABES hearing: Jérôme Kalfon, 5 October 2015

Participants:

For the ABES:

- Jérôme KALFON, Director of the ABES

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit
- Sarah Lenoir, Barrister, Intellectual Property Unit

For the DIST:

- Laurence El Khouri, Deputy Director

Overview of the approach

As mentioned during the ISTEEX legal seminar of July 2014 and in the Executive Committee meeting, the Bensoussan Consultancy, in collaboration with the DIST and the CNRS Legal Affairs Department, undertakes to draft a White Paper entitled "Open Science in a Digital Republic", in response to issues raised by the ISTEEX project.

This approach is based on the issues raised at the ISTEEX seminar in July 2014 on legal security, on the themes: objectives and management of databases, text and data mining (TDM), interdisciplinarity, and content aggregation.

These themes are of course included in the Digital Republic Bill drawn up by Axelle Lemaire, Secretary of State for Digital Affairs.

The purpose of the White Paper is to contribute to an **examination of the legal framework for scientific data in order to address the concerns of the scientific communities** (data publication and uses, the law governing science platforms, TDM, etc.) and to **participate in drafting the Digital Republic Bill** by making useful proposals.

The approach proposed by the Bensoussan Consultancy is initially **to conduct hearings** in order to **survey the practices of the scientific community and learn how STI is currently exploited**.

The objective is to define the status that the scientific communities wish to see attributed to open science and, more particularly, what status should be attributed to data, metadata, articles or scientific results.

Data and publications: the ideal solution

For Jérôme Kalfon the ideal solution for opening up access to data and publications is as follows:

Above all it is necessary to avoid appropriation.

For data: making data open is the basic principle, which must include justifiable (and usually temporary) exceptions. Precautions should be taken concerning the opening up of data, by defining the ethical requirements and the rules concerning confidentiality. The barriers to appropriation must be:

- ethical security;
- confidentiality;
- the definition of a temporary monopoly and an embargo period on certain data, with the period varying on a case-by-case basis.

While scientific publishing used to be only moderately profitable, the strategy of the major publishing groups has made it one of the most profitable sectors by manipulating two levers:

- Scientific publishing is a vital and strategic area for any academic structure, whether for researchers or for the institution or institutions on which they depend: as each new product is irreplaceable and exclusive by nature, the institution will be forced to accept any rise in the price of publishing, even if it is unrelated to changes in the costs of production;
- The technological transition has upset the way things were traditionally organised and encouraged the concentration of actors to the extent that there is no alternative but to use them.

For publications: the very purpose of a publication is to make content publicly available, and therefore no filter or embargo period should prevent or slow down publication. Publication should not be subject to appropriation of any kind. The bulk of the economic investment in publishing (writing, revising, basic editorial tasks, etc.) has always been

provided by the scientific community. The publisher's role – marginal if compared with the work involved in literary or artistic publishing – is limited to type-setting, layout, printing and distribution (managing subscriptions).

Distribution mostly meant exchanging publications and distributing reprints. Publishers were essentially service providers, and a share of the potential profits from paying subscriptions went to the learned societies who owned the journals. The signing of contracts between authors and publishers is a recent development and has become more widespread with digital distribution.

If it were necessary to invent a system of scientific publication today, we would never choose a system as costly, slow, inefficient and complex as the one in force now. At a time when rarity (physical limits to the number of copies and their distribution) is giving way to natural abundance, it is absurd and paradoxical that we should continue to create an artificial shortage.

This anachronism has lasted for too long. It is based on an economic model involving the transfer of exclusive rights by the author to the publisher (a recent phenomenon that developed at the end of the 1990s – there were few written contracts between authors and publishers before that time).

Consequently, **publications in the world of science must be completely open, with no embargo period, subject only to the right of authorship.**

However, if "we quit wishful thinking and return to reality" we need to define the conditions for a transition to a model that meets the requirements for building know-how and knowledge.

A temporary embargo period could be established to cover a transition period, without losing sight of the ultimate objective of making publications available immediately. Embargo periods should be kept as short as possible and through negotiations with the publishers may be set at a maximum of 6 months, and 12 months for the HSS.

Remuneration for publishers

Scientific publishing needs to find a new economic balance.

Publishers must be paid for the work they perform, for the value they add, on a 'jobbing' basis.

Publishing is financed by the organisations that fund research. The contractual relationship between the researcher (the community) and the publisher must be translated by remuneration for a **delivery of service**. This service must be independent of copyright.

Literary and Artistic Property (LAP) must be kept distinct from rights regarding science:

- on the one hand, academic authors receive no remuneration from publishing (it is now extremely rare for authors to receive royalties – on the contrary, through the APC, they are now increasingly likely to have to pay to be published). Authors are paid for their academic activity, of which publication is an integral part.
- on the other hand, the LAP system protects the authors of literary and artistic creations and provides them with remuneration. This traditional arrangement must not apply to science.

For scientific publishing, there must be transparency regarding the added value for which the publisher receives remuneration.

There is nothing to prevent scientists from publishing some parts of their work under LAP and other parts under open access.

There are several legal considerations surrounding intellectual property today, which vary depending on the media and the type of product (audio, video, music, patents, trademarks, etc.). The failure to distinguish scientific publications from literary and artistic property is currently a problem for science, but the scientific publishing model is bound to evolve in a way consistent with its requirements. If no distinction is made, it is likely to be the LAP that comes under pressure.

Defining rights specifically for science

It is therefore necessary to **define rights specific to science** (as there is for the filing of patents, for databases, etc.).

Why do we need a specific law? Because:

- science is a very special form of 'trade';
- research activity, of which publishing is an integral part, enjoys specific tax regimes (tax deductibility) in almost all countries;

- science is a factor of collective enrichment and development: scientific data, especially publications, must be treated as open data or content;
- the use of this open content contributes to economic development;
- the creation of wealth associated with the opening up of such content is much greater than that induced by their appropriation;
- science is a particular field that follows a specific approach: researchers need to have access to and discuss texts and the results of research, and to repeat experiments (which in no way precludes filing patents);
- copyright includes a right of withdrawal, which is incompatible with the ethics of publishing and scientific debate.

Distinction between open access and open process

Just because a publication is in open access, this does not mean that there must be a transfer of ownership from the author to the publisher.

Open Access means giving access to the publication on the publisher's website. It consists of a right to read on the publisher's media. In a scientific article:

- the idea must be entirely free to read;
- the scientific article can be read on the publisher's website (place chosen by the publisher for Open Access). Publishers restrict the exchange of ideas to a right to read.

Open Access on the publisher's website must be distinct from deposition in an open archive. This deposition is based on the concept of filing the final version submitted to the publisher (and what then becomes of any differences with the final published version?).

To overcome the problems related to intellectual property, the following proposals are made:

- Establish the principle that publishers must deposit works systematically in an open archive (with remuneration for the useful work performed by the publisher); this open archive system is a necessary transition but must remain transitional.
- Science as a set of specific rules for the development of knowledge should be devoid of authorship rights, only the moral right must be preserved.
- STI must be defined as commons by its very nature.

The definition of a new ecosystem

The following could define a new model:

- publishers would be service providers: their services would include organising the peer review, labelling, assessment, and release for free access and free processing; the publisher would provide different services to the different communities;
- the publisher would be paid for this service;
- the final service proposed would be unlimited access and reuse of publications in any space, public or private, and including for commercial purposes: this would be "gold open access" but without reversibility and without transfer of authorship rights;
- the financing of the dissemination of knowledge would be integrated from the outset in the financing of the academic activity (from a macroscopic viewpoint, this is a zero-sum game: the sums spent on purchasing publications – mainly by libraries – finance the act of publishing, which is an integral part of an academic research project);
- scientific publications would be copyright-free, without ownership rights; they would be a different kind of object: a "commons".

This system keeps publishers as professional third-party service providers within the dissemination chain of STI. Publishers cease to hold ownership rights over the content they help disseminate. The model is cleansed of any appropriation for the sole benefit of one of the actors, who previously gained economic and technical control of the entire value chain.

All other systems are compromises. Embargo systems, hybrid journals, and recourse to TDM under contractual rules must be purely transitional solutions.

g. CCSD, Claude Kirchner, 15 October 2015

Participants:

For the CCSD Steering Committee:

- Claude Kirchner, Current President of the CCSD Steering Committee, Adviser to the President of INRIA and Senior Researcher

For the Bensoussan Consultancy:

- Alain Bensoussan, Barrister specialising in law concerning advanced technologies
- Laurence Tellier-Loniewski, Barrister, Director of the Intellectual Property Unit
- Sarah Lenoir, Barrister, Intellectual Property Unit

For the DIST:

- Renaud Fabre, Director
- Charlotte Autard, Manager in charge of the ISTEEX Investments for the Future project

Overview of the approach

The idea by the CNRS to draft a White Paper entitled "Open Science in a Digital Republic", in collaboration with the Alain Bensoussan Consultancy, is a response to the issues raised by the legal framework covering the ISTEEX Investments for the Future project ANR-10-IDEX-0004-02 (www.istex.fr), which raises many legal issues by its potential in terms of TDM, interdisciplinarity, and content aggregation and its aim to make its databases searchable.

The new law on science has a specific exception concerning copyright: the legal status of scientific results and notably data and metadata.

The purpose of the White Paper is to **propose a legal framework for scientific data in order to address the concerns of the scientific communities** (data publication and uses, the law governing science platforms, TDM, etc.) and to **participate in drafting the Digital Republic Bill** by making useful proposals. Science has not waited for a Digital Act and is proceeding with practices that now need to be formalised.

The Bensoussan Consultancy will initially organise working groups and conduct hearings in order to **survey the practices of the scientific community and learn how STI is**

currently exploited. The close association with universities and other institutions will make it possible to hear key witnesses and multiple views on the changes that are being prepared in the framework of the Digital Republic Bill.

Surveying these practices and the state of the art will enable CNRS to draw up a **relevance matrix of analyses and practices to be set against the current normative framework** in order to assess discrepancies and **develop proposals.**

Contributions to this White Paper will be anonymous and the people interviewed will have total control over whether their names are quoted or not.

The DIST and the Bensoussan Consultancy will re-contact the people interviewed and will submit an initial text to the University Presidents for approval.

Public consultation on the Digital Republic Bill

In the framework of the public consultation on the Digital Republic Bill, Claude Kirchner met with the Secretary of State for Digital Affairs and the Director of her Cabinet and informed the representatives of the MENESR about the following points in particular:

- All scientific data must remain under the control of scientists.

Scientific data include texts, articles, web pages, calculation data, lab books, programmes, etc.

But control does not mean possession. The objective is not to own the data but to have control in the sense of being able to use them in any way. Scientists must be able to access the complete text and the data in its entirety, so as to be able to read and reuse it, repeat and reproduce experiments, run programs and be able to reference all or part of the appropriate data, while for all these actions complying with the ethical rules that apply to the areas of science concerned.

- The services concerning data must be open to competition: there must therefore be complete access for TDM, in particular.

For example, the CCSD is considering setting up services for data mining and analysis. Researchers must be able to access, and especially to verify data, whereas today the private sector withholds the information necessary for performing such verification.

Opening suitable services would give the scientific communities access to the best possible tools and enable them to avoid losing their scientific sovereignty: the ability for a

given discipline, or laboratory, or country to develop the best science at the international level.

The competing interests of publishers in the management of scientific data might not be sufficient to force them to deliver all the necessary data to researchers.

However, being a researcher means reproducing experiments, verifying data and the way the data were exploited.

Today the possibility of publishers biasing search queries is plain to see for researchers and computer experts in algorithms and there is an awareness that while some of the strategies applied to algorithms can be shown to be fair, others may be biased. The recent case of Volkswagen has shown how important close supervision of access to data and algorithms can be.

There is no proof at present that publishers knowingly bias access to their data, but the possibility of verification would remove any doubt, now or in the future.

Article 4 of the Digital Republic Bill on the "Creation of a public service concerning data (to guarantee the quality of public reference data)" provides for this right of evidence, which is so vital for scientific publishing.

The right of transparency expressed in Article 4 must be extended to include the queries themselves and the information revealed by the queries used and the work being done by those conducting the searches.

The queries are themselves a form of data and must be accessible when they relate to scientific objectives. As long as ethical considerations are observed, the data from scientific discussions on social networks should also be accessible and free for dissemination.

Concerning Article 9 of the Digital Republic Bill, the following remarks were made (the version of the article given below was the current one on the date of this interview):

"I. When a scientific text arising from a research activity, which has been financed at least 50% by public funds, is published in a periodical, a publication appearing at least once a year, or in conference or symposia proceedings or compendia, its author, even in the event of exclusive transfer to a publisher, has the right to make available free of charge in digital form, subject to the rights of any co-authors, the latest version of his/her manuscript accepted by the publisher and excluding the formatting work which is the responsibility of the latter, at the end of a period of twelve months for the sciences,

technology and medicine and twenty-four months for the human and social sciences, with effect from the date of first publication. This dissemination may not give rise to any commercial exploitation. The text of the Digital Republic Bill – Article 9"

This wording prompts the following comments.

On the section "*... even in the event of exclusive transfer to a publisher, ...*": as Roberto Di Cosmo states in his comment on the proposal of Olivier Morin for "full and mandatory free access" in the framework of the public consultation on the Digital Republic Bill, if researchers do not transfer their rights to a publisher, they cannot sign the contracts currently written by publishers that impose the transfer of rights for the publication of articles in prestigious journals. It is therefore essential that the Act guarantees free dissemination, online, elsewhere and immediately, without contract assignment.

"*... the latest version of his/her manuscript accepted by the publisher and excluding the formatting work which is the responsibility of the latter*": Claude Kirchner first insists on an important element of the process of publication in a journal. A scientific paper submitted for assessment by a journal's Editorial Board has three main successive versions that it is important to identify correctly. The **author's initial version**, also called the **author-submitted version**, is the version of the document as sent for assessment to the Editorial Board. The **author's accepted version** is the one declared accepted by the Editorial Board after the peer review. The **publisher's version** is the one that, based on the author's accepted version, has been edited by the publishing house to improve the presentation and style, and put it in the right format for the journal if the authors have not done this themselves. This is the version that will be published in the journal.

Since only the peer reviewers have worked on the author's accepted version, it is inappropriate to apply an embargo period: the draft Article is therefore incorrect here. The version on which an embargo period could be applied is the publisher's version, that is to say the one that the publishing house has helped to format.

These reflections also led Claude Kirchner to point out that the current transformation that scientific communication is undergoing reveals three fundamental facts:

- **Publishing** a text, a result, a study, or data involves making these items public by definition. Publishing may be preceded or followed by qualification and validation as explained below, but it is fundamental to note that today, anyone can publish. People can tweet, post texts on their blogs or their web pages, publish articles in a

magazine, give papers at a conference, post opinions or texts on a scientific social network, etc.

- **Qualification:** A document, a text or a dataset, whether published as explained above or not, can be reread to assess the quality, originality, the contribution compared to the state of the art, the referencing, the quality of writing, etc. This qualification can be carried out by a journal's Editorial Board or by the Programme Committee of a conference, in which case we speak of peer reviewing. But this qualification is also often carried out elsewhere: in correspondence exchanged between a small group of people, on scientific social networks, in discussion forums specific to a community or in workshops. It can also be done during the assessment of research proposals for the French National Research Agency (ANR), the European Research Council or Horizon 2020 for example. In the framework of citizen science, it may be the result of different processes, such as voting.
- **Validation or certification:** Following a qualification process, a community of researchers such as, for example, the Editorial Board of a journal or the Programme Committee of a conference may decide to validate (or "certify") the document or the dataset. In the case of a scientific text, this validation by an Editorial Board may result in the acceptance of the article in a journal. Because before the digital revolution it was difficult or impossible to publish outside a magazine, there is still confusion between validation and publication. It is vital to note that these two elements are fundamentally different.

In this framework, quite independently of the consultation and following a long process of reflection that started at the very beginning of the 2000s, the President of INRIA very recently distributed a note that applies to all the scientists conducting their research in an INRIA project team. The note required that all works be published in full text in an open archive such as HAL and that for assessments of the Institute, its project teams and its researchers, only these publications will be taken into account. In the event that these works are also made public in a journal or a conference, the note requested that the version deposited in HAL, either the initial version of the article or (not exclusively) the author's accepted version. The note also stated that if a publisher seeks to oppose this deposit on HAL, challenging for example the free posting online of the author's accepted version, INRIA undertakes to assume this responsibility.

Availability; the phrase "*This dissemination may not give rise to any commercial exploitation*" inspires the following comment. The content of a scientific text is of course

potentially a source of innovations with considerable commercial benefits. The transfer of scientific progress, and therefore of the texts that describe it, is one of the fundamental missions of scientists in research organisations and universities. To prohibit the commercial exploitation of the contents of a scientific article by its authors and their employers would therefore be contrary to the fundamental missions of schools, research institutes and universities.

"II. – The provisions of this article are public policy and any clause to the contrary is deemed to be unwritten. They shall not apply to contracts in progress"; At the present time, the current contracts with the major publishers are signed for periods up to 2018. It would therefore not be possible to apply any new Act until the end of this period.

The Act must formally state that it applies to contracts currently in force or this clause must be removed to leave the door open.

Legitimate appropriation and misappropriation

Claude Kirchner stated that all the scientific data, at least before their publisher's version, must be made available to all comers, whether public or private.

During the previous hearings a consensus could be seen emerging on free access to data, but not concerning the embargo period. This can differ depending on the practices of each discipline and the researchers' need to be able to exploit their research prior to its dissemination and sharing.

According to Claude Kirchner, today the essential element when a result is made public is to have determined beforehand whether or not it should be protected.

Before a theorem, a process, an article, data, a program or an algorithm is posted online, INRIA takes careful heed of the willingness or otherwise of the researcher and the co-authors to publish the material in question.

In some cases:

- The Institute may issue an unfavourable opinion concerning publication but the author has the final decision and all the elements necessary to make that decision.
- The Institute may prohibit dissemination if the content fails to comply with its ethical rules. There are in fact a number of publications or programs that cannot be made public directly, such as for example a program capable of breaking encryption or hacking into bank accounts.

The different entities that oversee research activity should be organised in such a way as to allow the verification of publications in OA (compliance with ethics, possibility of exploitation) and to inform researchers prior to deposition in an open archive.

Services and innovation

In order to ensure the control of data, the CCSD via the platform HAL wants to create services of international quality. If the government supports this Bill and accepts this text for Article 9 on open access, "A shorter embargo period, no hindrance to TDM (text and data mining) and no prohibition of commercial exploitation" as proposed by the DIST, it will also be necessary to set up public-sector services capable of competing with commercial offers.

These public services could, for example, take the form of a recognition of some science platforms as "essential infrastructures".

7. Table summarising foreign legislation: Open access and TDM

Country	Principles	References
<p>European Union (since 2012)</p>	<p>Discussions on the introduction of TDM and Open Access in European regulations.</p>	<p>Recommendation of the European Commission "Access to and preservation of scientific information" of 17 July 2012, C(2012) 4890.</p> <p>In April 2014, a group of EU experts published a report entitled "Standardisation in the area of innovation and technological development, notably in the field of text and data mining".</p> <p>In December 2014, the "OpenAire" project was set up and the "Open Access Guidelines for Research Results Funded by the ERC" were amended.</p> <p>On 9 July 2015, the European Parliament adopted the "Reda Report" on the implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, whose rapporteur was Julie Reda. It advocates:</p> <ul style="list-style-type: none"> - That it is essential to properly assess the enablement of automated analytical techniques for text and data (e.g. 'text and data mining' or 'content mining') for research purposes.
<p>Germany (2013)</p>	<p>Recognition of a right for the authors of scientific texts to make their manuscripts available to the public on expiry of a period of 12 months.</p>	<p>Art. 38 (4) of the German Copyright Act¹³⁹, 2013.</p>

¹³⁹ "The author of a scientific contribution which is the result of a research activity that is at least 50% publicly funded and which has appeared in a collection which is published periodically at least twice per year has the right, even if he/she has granted the publisher or editor an exclusive right of use, to make the contribution available to the public in the accepted manuscript version upon expiry of 12 months after first publication, unless this serves a commercial purpose. The source of the first publication shall be indicated. Any deviating agreement to the detriment of the author shall be ineffective."

Country	Principles	References
United Kingdom (since 2013)	1/ Introduction of an exception for data searches	1/ Copyright, Designs and Patents Act, Article 29A "Copies for text and data analysis for non-commercial research", October 2014 ¹⁴⁰ .
	2/ Dual solution set up combining "green" and "gold" roads.	2/ Publication of a report on Open Access by the British Parliament in September 2013.
Spain (2011)	If research is financed mainly by the Government, a copy of the final version of the researcher's article must be filed in an institutional or theme-based archive, as rapidly as possible and no later than 12 months after publication.	Act on Science, Technology and Innovation, Article 37, 2011.
Italy (2013)	Work by researchers whose research is financed at least 50% by public funds must be published in open access journals or the final manuscript must be deposited in an institutional or theme-based archive within a time limit fixed by law.	Act on the exploitation of culture, 8 August 2013.
United States. (since 2008)	1/ Introduction of legal provisions on the public dissemination of research work financed by the National Institutes of Health (NIH). This Act provides that all articles published in journals as the result of work funded by the NIH must be deposited in the NIH's own online open archive, the "National Library of Medicine's PubMed Central". Contracts with the publishers must allow it explicitly.	1/ Consolidated Appropriations Act, 2008.

¹⁴⁰ "(1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that:

(a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and
 (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).

(2) Where a copy of a work has been made under this section, copyright in the work is infringed if

(a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or
 (b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.

(3) If a copy made under this section is subsequently dealt with

(a) it is to be treated as an infringing copy for the purposes of that dealing, and
 (b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.

(4) In subsection (3), "dealt with" means sold or let for hire, or offered or exposed for sale or hire.

(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable."

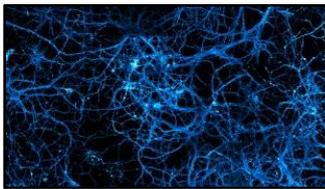
Country	Principles	References
	2/ Presentation of the "Fair Access to Science and Technology Research Act". 3/ Recognition that TDM operations can be covered by the "fair use" exception.	2/ The "Fair Access to Science and Technology Research Act" (FASTR) was submitted to Congress in February 2013. 3/ Case of Authors Guild v. Google (14 November 2013), in the framework of the implementation of a vast programme to digitise books and build up a universally accessible digital library.
Canada	Recognition of the principle that TDM operations can be covered by the "fair dealing" exception.	"Fair dealing" exception
Japan (2009)	Introduction of an exception for data searches	Article 47 septies of the Japan Copyright Act ¹⁴¹ introduced in 2009.
Israel	Introduction of the concept of "fair use" in Israeli legislation and reflection on the application of the "fair use" exception to TDM operations.	Amendments to the Israel Copyright Act in 2007.
Argentina (2013)	Creation of institutional repositories by research institutions, for depositing the results of research financed by public funds.	Act Initiated by the Ministry of Science, Technology and Innovation on the creation of open digital and institutional archives, Act No.26899, 2013.

¹⁴¹ "For the purpose of information analysis ('information analysis' means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other such information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use of a person who makes an information analysis."

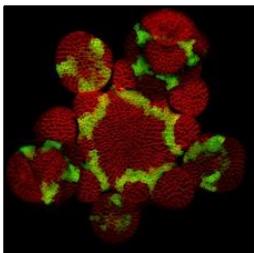
Credits

English translation by Jackie Godfrey, John Kerr and John Whitaker– Coup de Puce Expansion (27 Allée Edouard Branly, F-31400 Toulouse - www.coupdepuce.com) under the supervision of Jean-François Nominé - INIST-CNRS (2, Allée du Parc de Brabois, F-54519 Vandoeuvre-lès-Nancy – www.inist.fr)

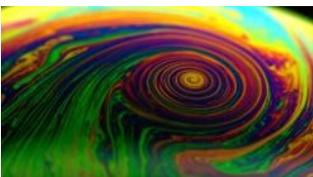
Picture credits



20120001_1920
© Sébastien MARAIS/Daniel
CHOQUET/Bordeaux Imaging
Center/CNRS Photo library



20090001_0466
© Pradeep DAS/ENS Lyon/CNRS
Photo library



20140001_0904
© Hamid KELLAY/LOMA/CNRS
Photo library