Science and technology Observatory
Department of the assessment of the national research organizations

ANALYSIS OF THE SCIENTIFIC AND TECHNOLOGICAL PROFILE OF THE CNRS

European projects, publications and patents

Contribution to the assessment of the *Centre national de la recherche scientifique (CNRS)*

October 2023

Contents

Preamble

This report has been written by the « Science and technology Observatory » (Observatoire des sciences et techniques, OST) and the Department of the assessment of national research organizations of Hcéres, as a contribution to the assessment of the Centre national de la recherche scientifique *(CNRS) for the 2017-2021 period. It analyses three main types of CNRS "productions", namely European research and innovation projects funded by the European Commission, scientific publications, and patents filings. For the first two types, the report includes benchmark analyses, comparing CNRS with foreign and French institutions. The scope of the analyses presented in this report has been defined by Hcéres in agreement with the CNRS.*

Hcéres has signed the San Francisco Declaration on research assessment (Dora), has contributed to the preparation of the Paris call on research assessment in 2022 and has joined the Coalition on advancing research assessment (CoARA); therefore, it pays attention to carefully position the analysis of quantitative indicators within the overall research assessment. The assessment of the CNRS is a peer review, it is a qualitative *assessment that may make responsible use of quantitative indicators. Thus, the present analysis of quantitative indicators is not the assessment itself: it is a contribution to the assessment of the CNRS. While being fully aware that it does not cover all CNRS productions, nor describe all the impacts of its productions, Hcéres has elaborated this report with the objective to contribute information to both the assessment committee and to the CNRS itself.*

A first version of this report has been addressed to the assessment committee and to the management of the CNRS in April 2023, a few weeks ahead of the committee's visit to the CNRS. This final version includes additional information and remarks that have been considered useful following exchanges of Hcéres with the CNRS and with the assessment committee. Complements mostly relate to the data and methodology. In particular, an appendix dealing with the comparison of indicators calculated on different publication databases has been added, the aim being to check the robustness of results.

Acknowledgements

*The Scientific information resources unit of the Spanish national research Council (CSIC) has contributed with its tool GESBIB to the identification of CSIC publications in cooperation with OST. Similarly, the Research system and science dynamics research team from the German Centre for higher education research and science studies (*Deutsches Zentrum für Hochschul- und Wissenschaftsforschung: *DZHW) has contributed to the identification of Max Planck publications. OST acknowledges these contributions, which have increased the precision of the publications corpora built for these two institutions. The identification process for their publications has thus been quite close to the process implemented by OST for the CNRS itself, in collaboration with the CNRS bibliometric team. CSIC and DZHW did not participate to the data processing and are not responsible for the results presented in this report.*

Executive summary

This report analyzes the CNRS scientific and technological profile and international position between 2017 and 2021 on the basis of three types of data: participation in the European Framework programs for research and innovation, scientific publications, and patents. The analyses include comparisons with other research organizations in France and abroad.

The report emphasizes the specific position of CNRS in France: a large share of total French publications, namely 43% in 2020, stem from the research units of the CNRS, which are mainly joint research units with other French institutions. This is a distinct feature of CNRS in comparison with the other research institutions considered in this report.

CNRS participation in the European framework programs for research and innovation

Five European organizations are compared to the CNRS: CNR in Italy, CSIC in Spain, the *Max Planck Gesellschaft* (MPG), the *Helmholtz Gemeinschaft* and the *Leibniz Gemeinschaft* in Germany. Between 2017 and 2021, CNRS and Helmholtz are by far the two research organizations with the largest number of applications and funded projects. The three German research organizations have the highest success rates, between 18 and 20%, ahead of the CNRS (16%).

The CNRS and Max Planck receive a much larger share of their European funding from the European research council (ERC) than the other four research organizations. As a result, the ERC focus presented in Chapter 1 only compares these two institutions. Between 2017 and 2021, the CNRS has received twice as many ERC grants as Max Planck, but its success rate has been lower (15% vs 24%). The disciplinary profile is also quite different, with CNRS having its highest share of ERC grants in the physical and engineering ~~panel~~ domain (60%), while Max Planck highest share is in life sciences (46%).

CNRS scientific publications in the national and international context

Chapters 2 and 3 are dedicated to scientific publications. The analyses are mainly based on the OST-WoS database, Hcéres in-house version of the Web of science, and they use two different disciplinary classifications. The conclusions on CNRS profile and position in comparison with benchmark institutions are quite similar and this executive summary focuses on the results from Chapter 2.

Number and types of publications

The CNRS and the Chinese Academy of sciences (CAS) have by far the largest number of publications when compared to CSIC and Max Planck. Between 2017 and 2021, the number of CNRS publications has increased more slowly than that of the other three research organizations, especially CAS.

Open access publications represent a growing share of world publications, and CNRS is participating to this global trend. A normalized open access index is used to measure the share of open access publications, the neutral value for the world being 1. This open access index is 1.5 to 1.6 for most of the research organizations considered in this report, including the CNRS, MPG and CSIC; it is lower for CAS and for organizations conducting more applied research.

CNRS, CSIC and MPG publications also more often result from international collaboration than CAS publications. Overall, CNRS publications are highly collaborative, 65% being international co-publications and 25% domestic co-publications.

Disciplinary profile

A specialization index is used to discuss the profile of an institution: it measures the ratio between the share of a discipline or scientific field in its publications and the same share in world publications. Using the ERC classification, the CNRS is specialized in all the eleven panels in Physical sciences and engineering, in three panels out of nine in Life sciences and in one panel out of seven in Social sciences and humanities. CNRS three panels of highest specialization are mathematics, Universe sciences and study of the human past.

The comparison with benchmark institutions is also conducted using a classification in broad disciplines. The CNRS is much more specialized in mathematics than CAS, CSIC and Max Planck. MPG is the most specialized in physics; CAS is the most specialized in Earth and Universe sciences and in chemistry; CSIC is the most specialized in applied biology and ecology. MPG and CSIC are clearly specialized in fundamental biology. In social sciences or medical research, none of the four research organizations is specialized compared with the world average.

Measures of scientific impact

This report computes two types of indicators to analyze the scientific impact of publications. First, it uses an indicator of average impact, the mean normalized citation score (MNCS). Then, since the distribution of scientific publications according to their citations is highly skewed, this perspective is complemented with indicators on the distribution of publications among citation classes. Both types of indicators give strongly convergent results for the CNRS in comparison with benchmark institutions.

MNCS for CNRS 2019 publications is below the world average and the activity index in the top 5% most cited publications is more than 10% below world average. Among the four benchmarked institutions, Max Planck has the highest MNCS, close to 50% above the world average. CAS comes in second (close to +30%), before CSIC (close to +10%) and CNRS. By discipline, CNRS positions relative to the world average or to the benchmark institutions are consistent with its positions for total publications. CNRS average indicator is relatively better in Earth and universe sciences, applied biology and medical research

Robustness checks

The report includes several explorations in order to appreciate the robustness of the results. A specific appendix shows that CNRS disciplinary profile compared to all French publications is quite similar in the national archive HAL and in the OST-WoS database. In other words, while the OST-WoS database does not cover social sciences and humanities as well as life sciences or physical sciences, this does not generate a bias that would preclude international comparisons. Another appendix compares the results given in this report for the average impact indicator (MNCS) with the results for the same indicator calculated on another publication database. It concludes that the relative positions of the CNRS and Max Planck are the same; in other words, the observation of a substantial gap between the MNCS of the CNRS and Max Planck appears to be a robust result.

CNRS patent applications and co-applications

During the 2012-2019 period, the CNRS filed 4,384 priority patent applications; the annual number of priority applications has varied over the period, with a maximum in 2015.

CNRS patent fillings are even more collaborative than its publications: between 2014-2016 and 2017-2019 its co-filing at the European patent office (EPO) has increased from 91% to 95% and ~~it~~ this share tends to be higher at the United States patent and trademark office (USPTO). CNRS partners are mostly higher education institutions and research organizations. Foreign public institutions participate in 7% of the co-filings and foreign private institutions in nearly 3%. Between 2014-2016 and 2017-2019 the shares of higher education, R&D and healthcare institutions in co-applications have increased, while that of French and foreign private institutions have decreased.

CNRS first technology field is Pharmaceuticals, with nearly 20% of filings, both at the EPO and USPTO. The second field is Biotechnology, with 10% in Europe and 9% in the US. Measurement technology is the third field, with 7 to 8% in both offices. The following fields remain mostly the same at both offices.

The 4-year CNRS patent grant rate is 23% for the 2012-2017 applications, equivalent to the average rate for all applications at the EPO. The 6-year CNRS grant rate for 2012-2015 applications is 46%, slightly higher than EPO average. Grant rates vary greatly from one technology field to another. For CNRS first three patenting fields, the grant rate after 4 years is between 15% and 29%, and the grant rate after 6 years is between 41% and 54%. For both time windows, the CNRS grant rate is higher than the EPO average.

Introduction

This report analyzes the CNRS scientific and technological profile, and its international position during the 2017-2021 period on the basis of three types of data: participation in the European Framework programs for research and innovation, scientific publications and patents. For the first two types of productions, the analyses include comparisons with other research organizations in France and abroad.

Chapter 1 analyzes CNRS participation in the European framework programs for research and innovation. The indicators calculated for the CNRS are compared to those of other European institutions: CNR (National research Council) in Italy, CSIC (National research Council) in Spain, as well as the Max Planck society (MPG), the Helmholtz association and the Leibniz association in Germany. At the level of the whole framework programs, the indicators show the evolution of the number of projects, number of applications, success rates and amounts of EU funding, as well as the distribution of each institution's projects among the "pillars" of Horizon 2020 framework program: excellent science (pillar 1), industrial leadership (pillar 2) and societal challenges (pillar 3).

The analysis then focuses on the ERC (European research Council) projects. It shows the evolution of the number of projects and applications over the 2017-2021 period for the CNRS and for the ERC top-recipient countries. The CNRS profile by ERC panel is compared to that of MPG. Overall, the analyses of the ERC grants shows that the CNRS has a larger number of ERC grants than the five other institutions, even when an estimate of their size is taken into account, but MPG has a higher success rate than the CNRS.

Chapters 2 and 3 characterize CNRS scientific publications. Chapter 2 uses the OST publication database, an enriched version of the Web of science (WoS), to compare the CNRS with three foreign research organizations: Chinese Academy of sciences (CAS), CSIC and MPG. The perimeter for CNRS publications includes all publications of the joint research units (*unités mixtes de recherche*: *UMRs*) having the CNRS as one of their home institutions. The analysis describes the evolution of the number of publications over the 2017-2021 period, and evaluates the share of each of the four institutions in the publications of its own country, thus highlighting the unique national position of the CNRS. The CNRS high share of French publications is not even across disciplines, ranging from 70% in mathematics, physics and chemistry, to less than 30% in social sciences, and around 10% in medical research.

Chapter 2 provides a normalized index of open access, which is relatively high and quite close among the research institutions that are compared in this report, except for CAS. It also describes the share of international co-publications of the CNRS and the degree of affinity with partner countries. The scientific profile of the CNRS is then analyzed on the basis of two different nomenclatures, and compared with the profiles of CAS, CSIC and MPG. The chapter closes with measures of the scientific impact of the four institutions' publications on the basis of their citations. The impact indices of the CNRS appear to be lower than those of CAS, CSIC and MPG in most disciplines.

Chapter 3 provides a benchmark analysis of CNRS publications for nine scientific fields corresponding to nine of the CNRS Institutes – the Institute for humanities and social sciences is not covered in this chapter. In each field, comparisons are made with a specific set of French and foreign institutions, chosen jointly by Hcéres and the CNRS. This analysis essentially confirm the observations presented in chapter 2, while giving additional detailed measures and comparisons.

Chapter 4 characterizes the patents filed by the CNRS and their evolution. It gives a description of the geographical coverage of CNRS filings, and a detailed analysis of the technological profiles of the CNRS patent filings at the European and US patent offices. CNRS patents are mostly specialized in three technological fields: pharmaceuticals, biotechnologies and analysis of biological materials. Information on public and private co-applicants and on the grant rate of CNRS patent filings at the European patent office is also given.

The appendices describe the data sources used to compute the indicators and the methodology; they also give information on the benchmark institutions, as well as a list of acronyms. Two appendices focus on robustness checks. Appendix D compares the disciplinary profile of the CNRS observed on the basis of two complementary data sources: the French national open access archive HAL and the OST publication database. Appendix E deals with the comparison of indicators of scientific impact computed on two different databases

1. CNRS PARTICIPATION IN THE EUROPEAN FRAMEWORK PROGRAMS FOR RESEARCH AND INNOVATION

The analysis of CNRS participation in the European framework programs for research and innovation (FP), "Horizon 2020" (H2020) for the period 2014-2020 and "Horizon Europe" for 2021-2027, is presented hereafter with two different perspectives.

First, the participation of the CNRS in the framework program is compared with the participation of the five other research institutions included in the "G6 association": CNR (*Consiglio nazionale delle* ricerche, National research council) in Italy, CSIC (*Consejo superior de investigaciones científicas*, Superior council of scientific research) in Spain, and the *Max Planck Gesellschaft* (MPG), the *Helmholtz Gemeinschaft* (Helmholtz association) and the *Leibniz Gemeinschaft* (Leibniz association) in Germany. Since the CNRS share in the French participation is very high, the evolution of countries' participation is also analyzed.

Second, the CNRS participation in ERC (European research Council) projects is analyzed in more detail, and compared with the MPG participation.

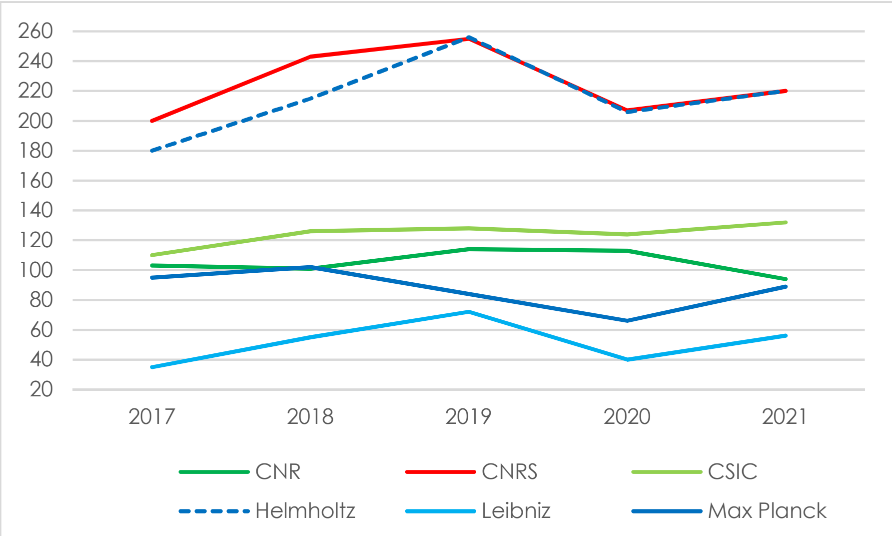
Key figures on each of the benchmark institutions are given in Appendix A. The data and methodology for the analysis of the European projects are described in Appendix B.

1.1. Participation in the European framework programs

a. Evolution over the 2017-2021 period

Figure 1 shows the evolution of the total number of projects obtained by each of the G6 organizations as direct beneficiary[1] for the whole FP between 2017 and 2021. CNRS and the Helmholtz association stand out with an average of around 200 projects per year; the CNRS had 10% more projects than Helmholtz in 2017 but the two institutions have the same numbers of projects in the second half of the period. CSIC comes next with about 120 projects per year. CNR and MPG have about 100, the Leibniz association about 60.

Figure 1. Number of FP funded projects, European G6 organizations, 2017-2021.*

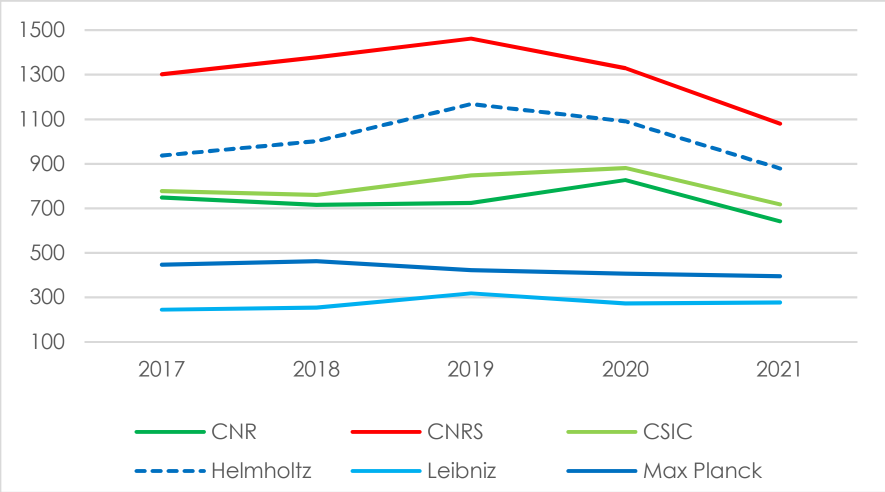


* All instruments, including Euratom.
Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Figure 2 displays the evolution of the number of applications for the 6 institutions. The CNRS submits significantly more applications than the Helmholtz association, even though the two institutions obtain an equivalent number of projects. The relative positions of the other institutions for the number of applications are equivalent to their positions for the number of funded projects.

[1] Projects in which an institution participates as a third party are not taken into account.

Figure 2. Number of FP applications, European G6 organizations, 2017-2021.



Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Table 1 provides the success rate in European programs. On average for the period 2017-2021, the Helmholtz association has the highest success rate with more than 20%, followed by Max Planck (19%) and Leibniz (18%). The CNRS has an average success rate of 16%, ahead of CSIC and CNR. All six institutions have higher success rates than the average for the entire FP, which is below 12%.

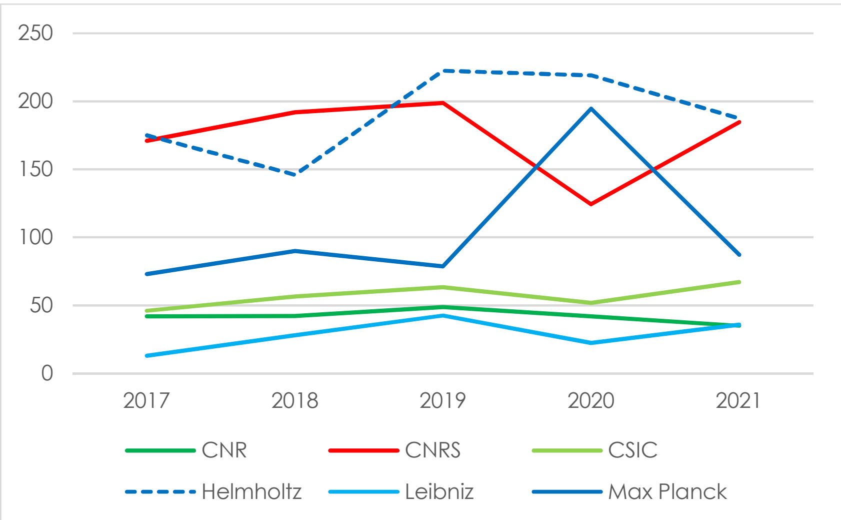
Table 1. Success rates at FP projects,* European G6 research organizations, 2017-2021.

	2017	2018	2019	2020	2021	Average 2017-2021	Evolution 2021 / 2017
CNR	13,1%	12,7%	14,2%	13,7%	15,6%	13,8%	1,2
CNRS	13,4%	15,9%	15,9%	15,0%	18,5%	15,7%	1,4
CSIC	13,0%	15,5%	13,2%	13,5%	18,1%	14,6%	1,4
Helmholtz	17,2%	20,9%	20,4%	18,6%	24,2%	20,2%	1,4
Leibniz	12,7%	19,3%	21,1%	14,7%	20,9%	17,9%	1,6
Max Planck	21,3%	20,1%	19,1%	15,7%	20,5%	19,4%	1,0
Total FP	**10,6%**	**12,3%**	**12,0%**	**10,4%**	**14,3%**	**11,7%**	**1,3**

* Ratio of the proposals admitted to the main list of the "proposals" database to the number of proposals in the same database (and not in the "grants" database). All pillars, including Euratom.
Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Figure 3 shows the evolution of the amount of funding received by the six institutions. It highlights that annual amounts, like numbers of projects, vary significantly from one year to another.

Figure 3. FP net contributions received by European G6 research organizations, 2017-2021 (M€).



Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

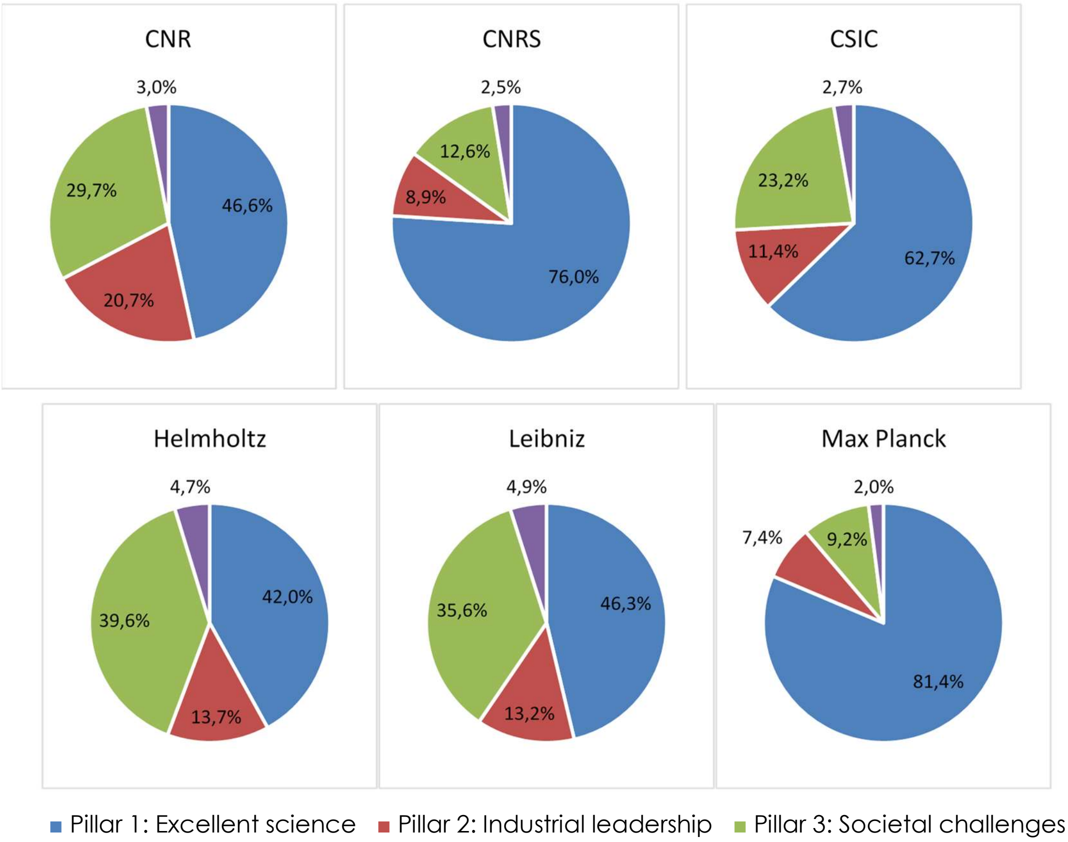
Between 2017 and 2021, the amounts slightly increase (about +7% or + 8%) for the two institutions receiving the largest funding, Helmholtz and the CNRS, while they vary significantly for the other four institutions, with an important growth for Leibniz and CSIC, a 20% increase for Max Planck, and a decrease for the CNR.

b. Distribution of the projects among the three pillars of H2020

It is interesting to analyze the distribution of funded projects between the three pillars of H2020: Excellent science (pillar 1), Industrial leadership (pillar 2) and Societal challenges (pillar 3). Pillar 1 includes the European research council (ERC) and the Marie Sklodowska Curie actions (MSCA), as well as the funding of research infrastructures and of "future and emerging technologies".

Figure 4 shows that the six institutions have different profiles. For the sake of simplicity, the comparison is limited to the H2020 projects, since a different definition of the pillars was adopted for Horizon Europe.

Figure 4. Distributions of H2020 projects by pillars, European G6 research organizations, 2017-2021.



Source: e-corda database (consulted in Oct. 2021), computed by OST.

Max Planck has 81% of its projects funded in Pillar 1, the CNRS 76%, and CSIC 63%. The three other institutions have less than 50% of their projects funded in Pillar 1, which suggests that they have a lower involvement in fundamental research. CNR has 21% of its projects funded under the "Industrial leadership" pillar. The Helmholtz association has 40% of its projects funded under the "Societal challenges" pillar; this share is also more than one third for the Leibniz association.

Over the period, ERC projects represent 37% of Pillar 1 projects in H2020 for the CNRS and 42% for Max Planck. These shares are even more important in terms of funding as the amount of ERC grants is about three times higher than the amount of MSCA grants on the average.

1.2. Analysis of the participation in the European Research Council

ERC is a funding organization established by the European Union (EU). in 2007 to support investigator-driven frontier research in all scientific fields. ERC grants are mainly awarded to individual researchers, but a number are awarded to teams. There are three basic principles: the ERC is open to any project in any scientific field, it is entirely bottom-up; the submitted projects must be innovative and push the frontiers of knowledge; the only selection criterion is scientific excellence. The ERC budget was €13.1 billion for the period 2014-2020 (H2020), and it is €16 billion for 2021-2027 (Horizon Europe).

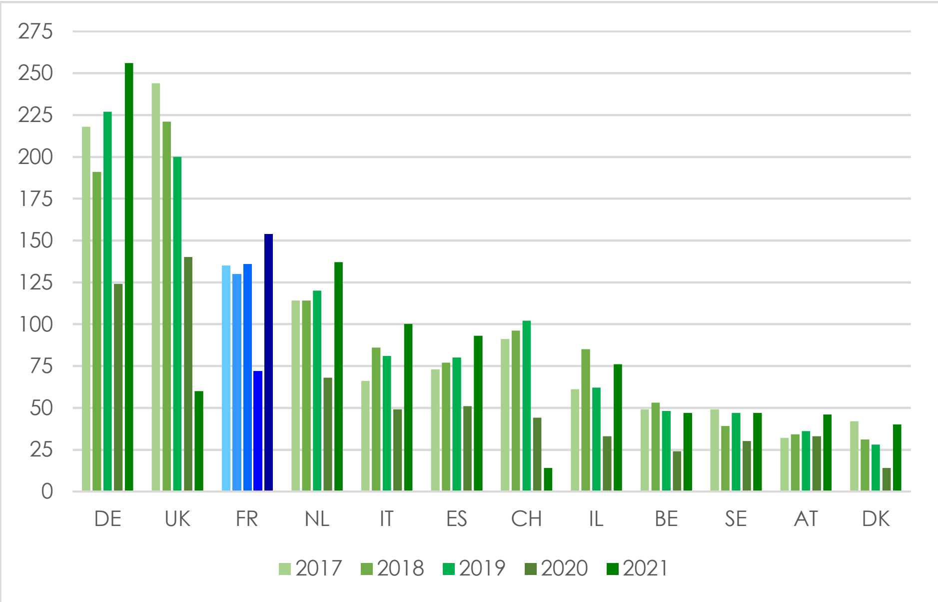
There are several types of ERC grants, with different eligibility criteria and funding levels. The numbers given below refer to the H2020 framework program.[2]

- ERC *starting grants* are for promising early-career researchers with 2-7 years of experience since completion of PhD and with a scientific track record showing great promise. They are awarded to a single principal investigator and are worth up to €2 million for a maximum period of 5 years. On average, around 400 ERC starting grants are awarded each year.
- ERC *consolidator grants* are for promising early-career researchers with 7-12 years of experience since completion of PhD and with a scientific track record showing great promise. They are awarded to a single principal investigator and are worth up to €2.75 million for a maximum period of 5 years. On average, around 300 ERC consolidator grants are awarded each year.
- ERC *advanced grants* are for established research leaders who have a track-record of significant research achievements in the last 10 years. They are awarded to a single principal investigator and are worth up to €3.5 million for a maximum period of 5 years. On average, around 250 ERC advanced grants are awarded each year.
- ERC *synergy grants* are for a group of two to maximum four principal investigators working together and bringing different skills and resources to tackle ambitious research problems. They can be up to a maximum of €15 million for a period of 6 years. On average, around 10 to 15 ERC synergy grants are awarded each year.
- ERC *proof of concept grants* allow laureates to explore the innovative potential of promising results from ongoing or recently completed ERC projects. The framework program provides for proof of concept grants of 150,000 Euros for a period of up to 12 months. On average, around 125 ERC proof of concept grants are awarded each year.

a. Evolution of ERC grants and applications over the 2017-2021 period

As highlighted in Figure 5, the distribution of ERC grants per country has changed significantly between 2017 and 2021, following the exit of the United Kingdom from the European Union and the changing position of Switzerland with respect to the FP. These two countries were among the top applicants and recipients of ERC funding. The decline in their shares of grants, especially since 2019, has unevenly benefited the other main ERC-funded countries.

Figure 5. Number of ERC grants, 12 top recipient countries,* 2017-2021.[3]



* Germany (DE), United Kingdom (UK), France (FR), Netherlands (NL), Italy (IT), Spain (ES), Switzerland (CH), Israel (IL), Belgium (BE), Sweden (SE), Austria (AT), Denmark (DK).
Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

If UK and Switzerland are taken apart, most countries appear to obtain more grants in 2021 compared to the beginning of the period. Table 2a shows the number of ERC grants for each of the top recipient countries and for the CNRS:[4] the left-hand side of the table gives the number of grants in 2017 and in 2021, whereas the right-

[2] Source: https://erc.europa.eu/apply-grant and https://erc.europa.eu/project-statistics/project-database.
[3] Grants are assigned to the closing year of the corresponding call for proposals, in compliance with the methodology used by OST to produce indicators on FP projects for the French government and parliament (see Appendix B).
[4] The number of ERC grants awarded to the CNRS as host institution is close to the number of ERC grants whose principal investigator (PI) is a CNRS employee, but these two numbers are not ensured to be exactly the same. As explained in Appendix B, the corpus of projects for each institution is determined with PIC codes (participant identification code). Normally, the PI of

hand side shows the ratio between 2021 and 2017. The table is in the decreasing order of the ratio in the right-hand column. It shows that the evolution of the number of ERC grants of the CNRS is close to Belgium's, Sweden's and Denmark's, with a slight decrease between 2017 and 2021, whereas each of the other top recipient countries increased its number of grants during the period.

Table 2a. Number of ERC grants* for the CNRS and top recipient countries, 2017 and 2021.

Countries and CNRS	Number of grants		Evolution 2021 / 2017
	2017	2021	
Italy	66	100	1,52
Austria	32	46	1,44
Spain	73	93	1,27
Israel	61	76	1,25
Netherlands	114	137	1,2
Germany	218	256	1,17
France	135	154	1,14
CNRS	72	70	0,97
Belgium	49	47	0,96
Sweden	49	47	0,96
Denmark	42	40	0,95

* From the "grants" database.
Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Table 2b shows analog data for the number of applications. Austria, and Belgium to a lesser extent, have increased their number of applications; all of the top recipient countries have more or less maintained their number of submissions (2021 / 2017 ratio between 0.96 and 1.02), with exceptions for Israel (0.94) and for France (ratio 0.86). The decrease in the number of applications is greater for the CNRS than for any of the top recipient countries.

Table 2b. Evolution of the number of ERC applications* for the CNRS and top recipient countries, 2017 and 2021.

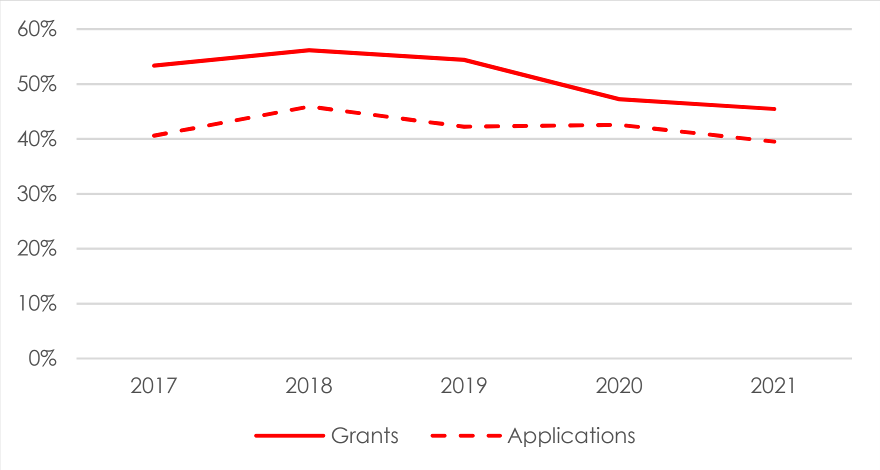
Countries and CNRS	Number of applications		Evolution 2021 / 2017	
	2017	2021		
Austria	222	278	1,25	
Belgium	311	329	1,06	
Denmark	264	269	1,02	
Germany	1276	1291	1,01	
Netherlands	665	672	1,01	
Spain	722	715	0,99	
Sweden	374	364	0,97	
Italy	890	852	0,96	
Israel	321	303	0,94	
France	962	827	0,86	
CNRS	391	327	0,84	

* From the "proposal" database.
Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Figure 6 focuses on the relative evolutions for the CNRS and France. It shows that the proportion of the CNRS ERC grants in the French total decreased from 53% in 2017 to 46% in 2021. The proportion of applications is more stable: it is close to 40% in 2017 and in 2021. The CNRS proportion of grants is higher than the proportion of applications.

an ERC application is employed by the host institution, but there may be cases where the PI's employer is a third party (see the Annotated grant agreement (draft version as of April 2023): aga_en.pdf (europa.eu)).

Figure 6. National share of CNRS for ERC grants and applications, 2017-2021.

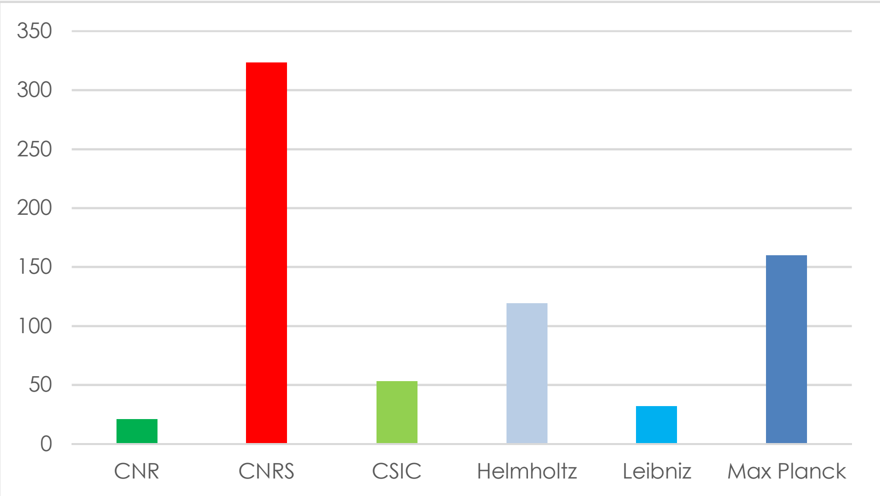


Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

b. Comparison with the G6 European research organizations

Figure 7 gives the numbers of grants of each institution for the 2017-2021 period. When examining this comparison, one should keep in mind that some of the institutions are more oriented to Pillars 2 and 3 of the FP, as shown on Figure 3 above, and do not devote as much effort as the CNRS and MPG to obtain ERC grants.

Figure 7. ERC grants, European G6 research organizations, 2017-2021.

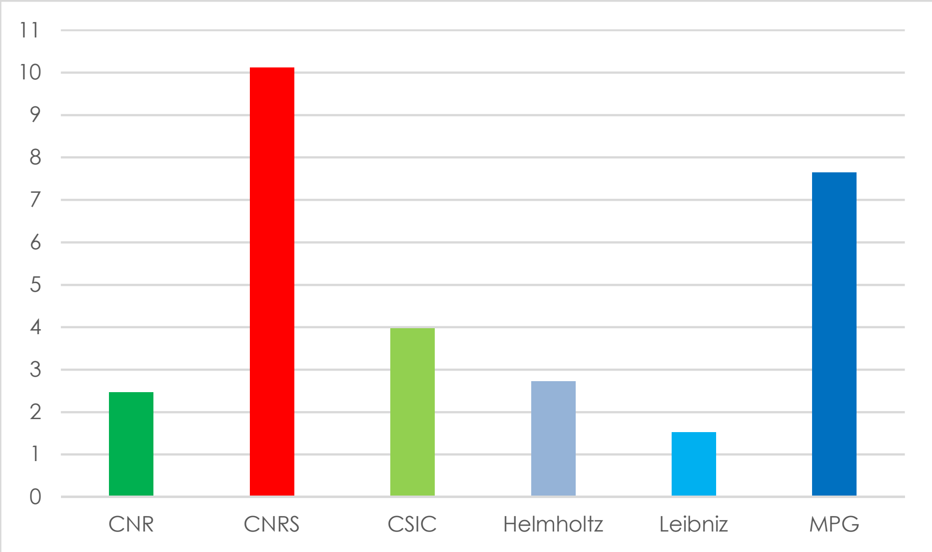


Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

The CNRS self-assessment report (SAR) presents a similar comparison concerning the ERC grants awarded to each of the G6 European institutions for the whole H2020 FP. Moreover, the ~~CNRS~~ SAR adds that the comparison "*should be refined by taking into account the number of researchers from the various institutions, likely to respond to ERC calls*". As a very first step towards such an analysis, Figure 8 shows a rough estimate of the number of ERC grants per 1,000 employees for each G6 institution for the 2017-2021 period.[5]

[5] This estimate is obtained as the ratio of the average annual number of ERC grants for the 2017-2021period (taken from Figure 7) to the total number of employees for each institution (given in Appendix A). As explained above, the number of ERC grants awarded to the CNRS is close to – but is not ensured to be exactly equal to – the number of ERC grants whose PI is a CNRS employee.

Figure 8. Rough estimate of the number of ERC grants per 1,000 employees, European G6 research organizations, 2017-2021.

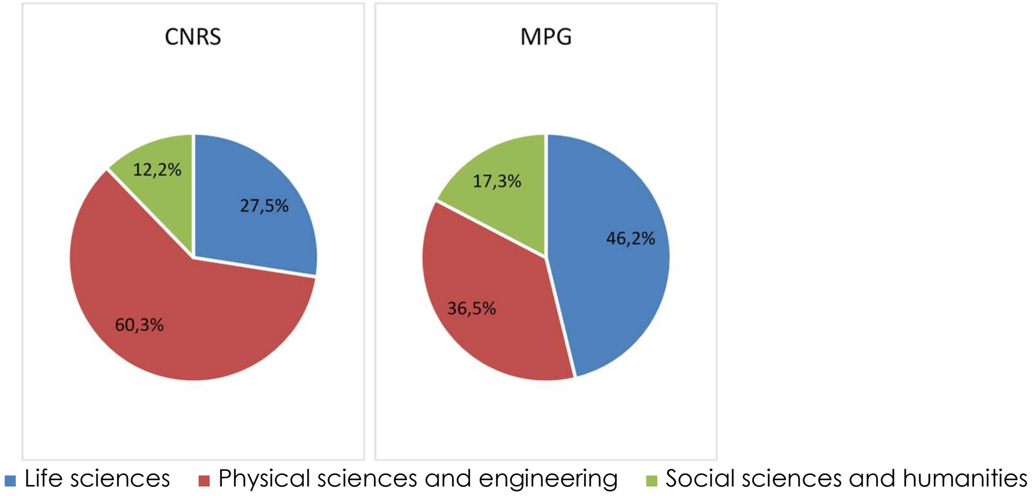


Source: data on grants from Figure 7 and data on employees (all staff) from Appendix A.

c. "ERC profiles" of the CNRS and MPG

In H2020, the ERC was organized in 25 disciplinary panels, belonging to three main domains: life sciences (LS), physical sciences and engineering (PE), social sciences and humanities (SH). Figure 9 shows the shares of the CNRS and MPG in these three domains.

Figure 9. Share of ERC grants of the CNRS and MPG by ERC domain, 2017-2020.



Source: e-corda database (consulted in Oct. 2021), computed by OST.

Table 3 reports the distribution of ERC grants by panel for the CNRS and Max Planck, for the 2017-2020 period. It also gives an "ERC specialization index", which is calculated as the ratio of the proportion of each panel for an institution (CNRS or MPG) to the proportion of the panel in the total ERC scope. The table thus highlights the profiles of the two institutions relative to Europe. The CNRS appears to be specialized in the Physical sciences and engineering (PE) domain, both in applications and in grants (index 1.3); the MPG applications are also specialized in the PE domain, but not the MPG grants (index 0.8). MPG is specialized in the Life sciences (LS) domain (index 1.5), whereas the CNRS is in the average of all ERC recipients for this domain (index 0.9). Both institutions are not specialized at all in the Social sciences and humanities (SH) domain, with the CNRS having a lower index than MPG (0.5 *vs* 0.7).

At the panel level, the CNRS appears to be specialized in Earth sciences (PE10) and Condensed matter physics (PE3), with indices 2.8 and 2.5. It is also clearly specialized in Cellular and developmental biology (LS3), Ecology, evolution and environmental biology (LS8) and Mathematics (indices greater than 1.8). MPG's specialization is even stronger in Cell and developmental biology (index 4.5); MPG also has indices above 2 in panels like Molecular biology, biochemistry and structural biology, in Neurosciences and neurology, as well as in Study of the human past and in the Universe sciences. Both organizations are slightly specialized in The human mind and its complexity (SH4), but not at all in panels HS1, 2, 3 and 5.

Table 3. Distribution and specialization index of ERC grants and applications, CNRS and MPG, 2017-2020.

Panels	Grants				Applications	
	Distribution	Specialization index	Distribution	Specialization index	Specialization index	
	CNRS		MPG		CNRS	MPG
LS1 Molecular biology, biochemistry, structural biology and molecular biophysics	3,5%	1,5	7,7%	3,2	1,8	2,5
LS2 Genetics, "omics", bioinformatics, and systems bio.	3,1%	1,2	2,9%	1,1	0,7	1,5
LS3 Cellular and developmental biology	4,4%	2,0	9,6%	4,5	2,1	2,3
LS4 Physiology, pathophysiology and endocrinology	0,4%	0,1	4,8%	1,3	0,3	0,7
LS5 Neurosciences and neural disorders	3,5%	0,9	8,7%	2,3	0,7	1,3
LS6 Immunity and infection	1,8%	0,7	1,9%	0,8	0,7	0,4
LS7 Applied medical technologies, diagnostics, therapies and public health	2,2%	0,4	1,0%	0,2	0,3	0,2
LS8 Ecology, evolution and environmental biology	7,0%	1,9	5,8%	1,5	1,6	1,5
LS9 Applied life sciences, biotechnology, and molecular and biosystems engineering	1,8%	0,5	3,9%	1,2	0,4	0,6
Life sciences (LS)	**27,5%**	**0,9**	**46,2%**	**1,5**	**0,9**	**1,1**
PE1 Mathematics	5,2%	1,8	0,0%	-	1,2	0,3
PE2 Fundamental constituents of matter	7,9%	1,7	6,7%	1,5	1,4	1,8
PE3 Condensed matter physics	10,9%	2,5	1,9%	0,4	2,4	1,5
PE4 Physical and analytical chemical sciences	3,5%	0,8	7,7%	1,8	1,1	2,0
PE5 Synthetic chemistry and materials	4,4%	0,8	1,0%	0,2	1,4	1,0
PE6 Computer science and informatics	1,8%	0,4	5,8%	1,2	0,6	0,9
PE7 Systems and communication engineering	5,2%	1,1	1,9%	0,4	1,2	0,3
PE8 Products and processes engineering	5,2%	0,8	3,8%	0,6	0,8	0,6
PE9 Universe sciences	5,2%	1,5	7,7%	2,3	1,5	3,9
PE10 Earth system science	10,9%	2,8	0,0%	-	2,1	0,9
Physical sciences and engineering (PE)	**60,3%**	**1,3**	**36,5%**	**0,8**	**1,3**	**1,3**
SH1 Individuals, markets and organisations	0,9%	0,3	0,0%	-	0,1	0,0
SH2 Institutions, values, environment and space	0,4%	0,1	0,0%	-	0,3	0,0
SH3 The social world, diversity, population	0,4%	0,1	1,9%	0,5	0,3	0,6
SH4 The human mind and its complexity	4,8%	1,1	4,8%	1,1	1,0	0,9
SH5 Cultures and cultural production	1,8%	0,4	0,0%	-	0,3	0,1
SH6 The study of the human past	3,9%	0,9	10,6%	2,5	1,1	1,1
Social sciences and humanities (SH)	**12,2%**	**0,5**	**17,3%**	**0,7**	**0,6**	**0,5**
Total panels	**100%**	**1**	**100%**			**1**

Source: e-corda database (consulted in Oct. 2021), computed by OST.

Table 4 compares ERC success rates of the CNRS and MPG by domain. The CNRS and MPG success rates are close in the PE domain, while MPG has higher average success rates in the LS and SH domains. As a result, the overall MPG success rate (24%) is substantially higher than that of the CNRS (15%). The difference in higher for ERC grants than for all Framework program projects (see table 1).

Table 4. ERC success rates of CNRS and MPG,* 2017-2020.

Domains	CNRS	MPG
Life sciences (LS)	**15.9%**	**32.4%**
Physical sciences and engineering (PE)	**15.4%**	**16.1%**
Social sciences and humanities (SH)	**11.5%**	**31.0%**
Grants that are not assigned to any panel	13.9%	36.2%
Total	**14.8%**	**24.3%**

* Ratio of the proposals admitted to the main list of the "proposals" database to the number of proposals in the same database (and not in the "grants" database).

Source: e-corda database (consulted in Oct. 2021 for H2020 and Aug. 2022 for Horizon Europe), computed by OST.

Overall thus, the CNRS has a larger number of ERC grants than the other G6 institutions, even when an estimation of their size is taken into account (Figures 7 and 8), but MPG has a higher success rate.

2. CNRS SCIENTIFIC PUBLICATIONS IN THE NATIONAL AND INTERNATIONAL CONTEXT

In this chapter and the next, CNRS publications over the 2017-2021 period are analyzed with indicators calculated on the OST publication database, which is an enriched version of the Web of science (WoS) database; its main characteristics are presented in Appendix C.[6]

This chapter compares CNRS publications with the publications of France as a whole and with the publications of three foreign interdisciplinary research institutions: CAS (Chinese Academy of sciences), CSIC in Spain and the Max Planck *Gesellschaft* (MPG) in Germany.[7] The characterization of CNRS publications first focuses on the evolution of their number and on their share in open access. Then it deals with the types of co-publications, and with the disciplinary profile of CNRS publications. Finally, it provides measures of the scientific impact of CNRS publications on the basis of their citations.

It is interesting not to analyze publications on the basis of a single database. This question is raised in the CNRS self-assessment report (SAR): the CNRS uses the WoS to provide several indicators, but it considers that the WoS is not adequate to analyze publications in social sciences and humanities (SSH) and in computer science. Indeed all databases used to produce bibliometric indicators have specificities in terms of coverage and quality of metadata; no single data source provides an exhaustive inventory of the scientific production, but each one can provide insight in line with its specificity, and the interpretation of the data must take this into account. This is why, as a complement to this chapter, additional analyses using other data sources are included. Box 1 and Appendix D deal with the comparison of indicators on CNRS and France publications that have been calculated from the French national open access archiv HAL, on the one hand, and from the WoS on the other hand. Moreover, Appendix E discusses the measurement of impact indicators with different publication databases.

Box 1. Analysis of CNRS scientific profile as observed in HAL and in the WoS.

Appendix D compares the typology and disciplinary distribution of the CNRS scientific production on the basis of two data sources, the French national open-access archiv HAL and the OST publication database. The main results are as follows.

- HAL includes more diverse types of production by researchers than the WoS. As a result, it counts a larger number of productions, including books and book chapters, contributions to conferences, reports, etc. ;
- When the corpus is focused on articles in scientific journals for which the evaluation process has been checked and on conference proceedings, the WoS counts more publications than HAL.
- The share of disciplines in the corpus (either for France or for the CNRS) varies with the types of productions. When the corpus is restricted to scientific publications in peer reviewed journals or conference proceedings, the share of physical and engineering disciplines somewhat increases and the share of life science disciplines strongly increases. Conversely, the share of SSH disciplines decreases.
- The disciplinary profile of CNRS scientific publications is quite similar when it is measured with WoS data or with HAL data. This is the case by broad domains and at the level of ERC panels.

The conclusions of this comparison are as follows. The WoS has a lower coverage of SSH disciplines than HAL, but a better coverage of life sciences; on the contrary, the coverage of computer science seems quite close in the two databases. Despite differences in discipline coverage, WoS data do not seem to be biaised when

[6] Appendix C includes specific information on the coverage of the WoS database for conference proceedings in the field of computer science, an issue that is raised in the SAR of the CNRS.

[7] Key figures on each of these institutions are provided in Appendix A.

it comes to compare the CNRS disciplinary profile with that of the whole of France provided that the construction of the corpora and of the indicators is carefully traced and explained. As a consequence, WoS data seem adequate to analyze CNRS publications, and in particular to compare them with publications from foreign research organizations.

2.1. Evolution of CNRS publications

Box 2 presents the constitution of the corpus of CNRS publications analyzed in this report.

Box 2. Constitution of the corpus of "CNRS publications".

Almost all CNRS research units are joint research units (*unités mixtes de recherche*: *UMRs*), that are common to the CNRS and to other institutions, mainly universities, *grandes écoles* and other national research organizations. A basic principle is that the publications of a given UMR are affiliated to each of the home institutions of the UMR, with no attention to which institution is the employer of which author of a publication.

As a consequence, the corpus of "CNRS publications" analyzed in this chapter and in the next includes **all publications of all CNRS *UMRs***. In other words, this means that, throughout this chapter and next, the words "*of the CNRS*" mean "*of the CNRS UMRs*"[8].

Some key figures about CNRS *UMRs* are given in the SAR: in 2021, the workforce of the CNRS *UMRs* was 109,800 persons, including 45,600 permanent researchers or university professors. These numbers are much higher than those of the sole CNRS employees in the *UMRs* (29,600, including 10,800 permanent researchers). The workforce of the CNRS *UMRs* is estimated to be about 44% of the workforce of the French public research[9].

a. Evolution of the number of publications

Table 5 and Figure 10 provide indicators on the evolution of the number of publications for the CNRS, CAS, CSIC and MPG. The number of publications are given in "*full counting*" and "*fractional counting*". Full counting relates to participations and gives a full weight of 1 to each publication, even if several institutions collaborated to produce that publication. Fractional counting relates to contributions of an institution to publications, taking into account the number of affiliation addresses. Since organizations are involved in many co-publications, their contributions represent only a fraction of their participations. For the CNRS, the ratio between contributions and participations is 0.5 on the average over the 2017-2021 period (90% of CNRS publications are co-publications, as described in section 2.2 below).

Table 5. Number of publications of CAS, CNRS, CSIC and MPG, full counting, 2017-2021*.

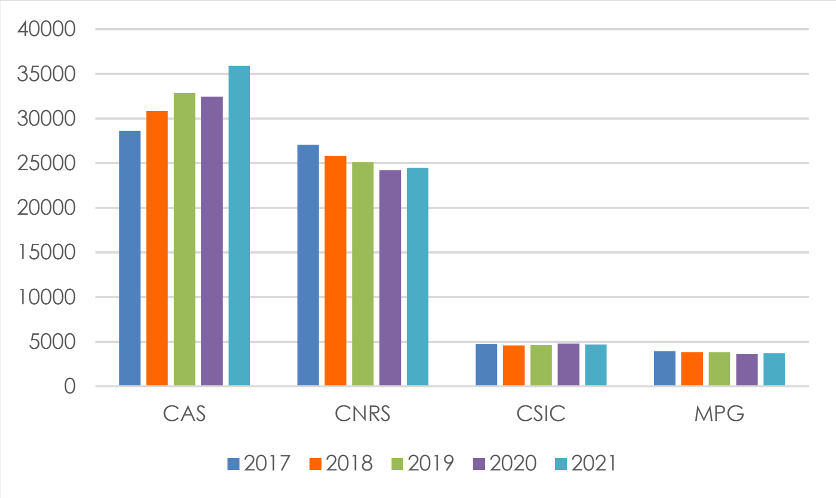
	CAS	CNRS	CISC	MPG
# 2017	51 486	55 220	12 408	12 050
# 2021	71 399	55 335	13 701	12 454
Evolution 2021 / 2017	+38.7%	+0.2%	+10.4%	+3.4%

* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

From 2017 to 2020, the number of CAS publications in fractional counting increases by 26%, while that of CNRS, CSIC and MPG decrease by nearly 10%, 2% and 6%, respectively.

[8] Each year, OST and the CNRS collaborate in order to precisely identify CNRS publications in the WoS database.
[9] Hcéres estimate (Department of the assessment of national research organizations) on the basis of *Bilan social 2019-2020 du ministère de l'enseignement supérieur, de la recherche et de l'innovation – Enseignement supérieur et recherche;* see https://www.enseignementsup-recherche.gouv.fr/fr/bilan-social-2019-2020-83381

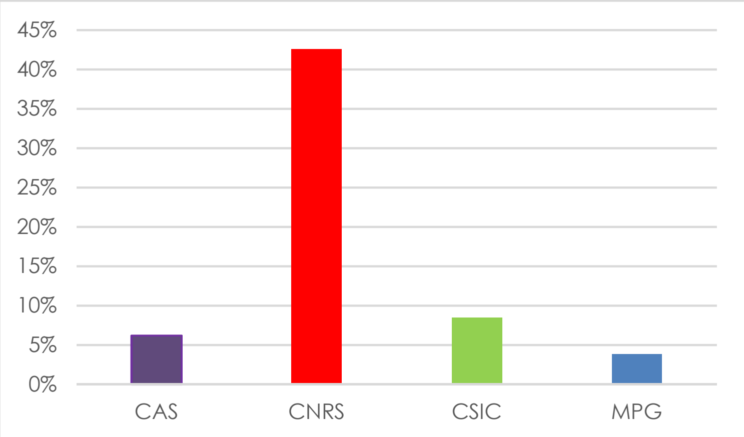
Figure 10. Number of publications of CAS, CNRS, CSIC and MPG, fractional counting, 2017-2021.*



* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

Figure 11 highlights the fact that the four institutions have very different domestic shares of publications. With nearly 43% of French publications (see Box 2 above), the CNRS has a much greater weight in France than the other three institutions in their own countries.

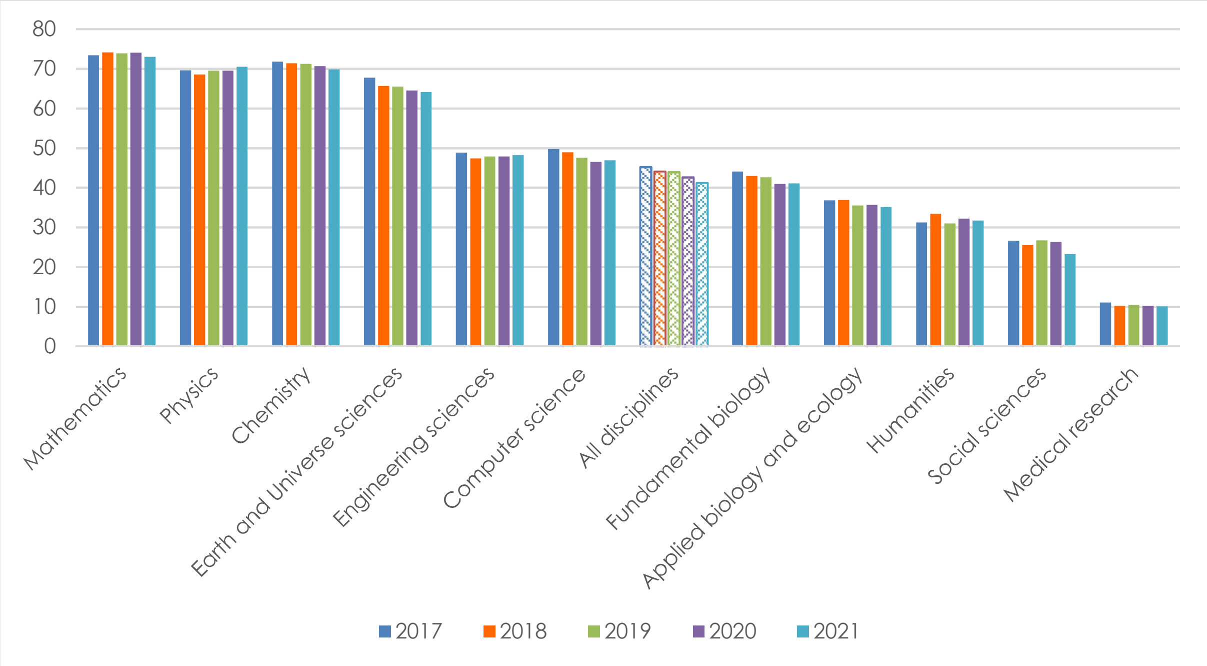
Figure 11. National share of publications, CAS, CNRS, CSIC and MPG, fractional counting, 2020.



Source: OST database, from the Web of Science.

Figure 12 provides the national share of CNRS publications by discipline (using OST nomenclature in 11 disciplines) and by year. It shows that the national share of the CNRS, all disciplines together, slightly declines between 2017 and 2021, from 45% to 41% of French publications. The CNRS share varies greatly from one discipline to another and does not decrease in all disciplines. It reaches 73% in mathematics and is almost stable. In physics, it increases from 69 to 71%. In chemistry and in Earth and Universe sciences, the CNRS share is close to 70% but is declining, while in engineering sciences it is stable at around 48%. In computer science, the national share of the CNRS declines from 49% to 47%. In fundamental biology, it decreases from 44 to 41%, and from 37% to 35% in applied biology and ecology. In the humanities, the CNRS share is stable at around 32%, while it decreases from 26 to 23% in social sciences. Finally, the CNRS share is the lowest (and stable) in medical research, around 10%.

Figure 12. National share of CNRS publications by discipline, fractional counting, 2017-2021, in %.*



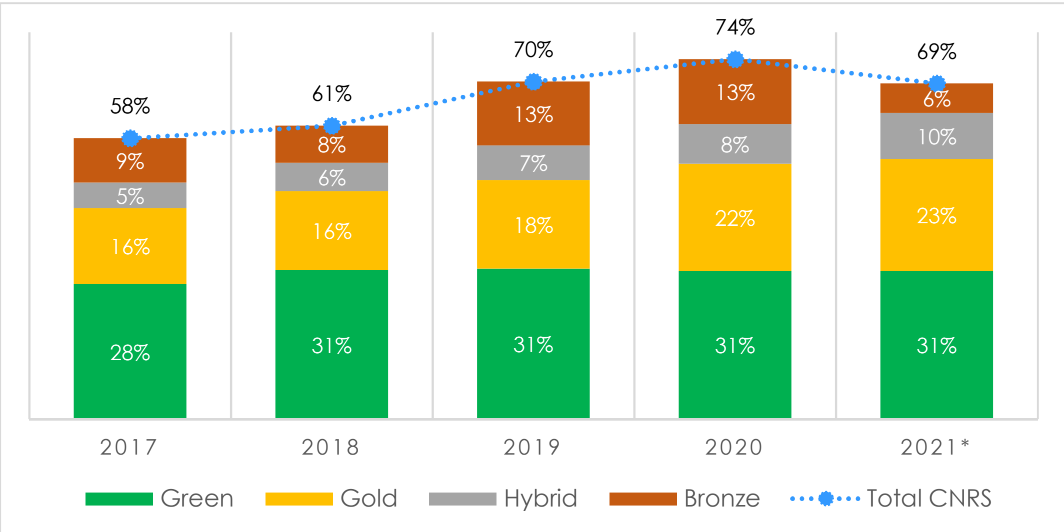
* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

Thus, despite slightly different nomenclatures, the disciplinary profile of the CNRS UMRs observed in the OST database is consistent with the ERC grants profile examined in section 1.2.d. This disciplinary profile is compared with those of CAS, CSIC and MPG below (section 2.3).

b. Open access publications

Figure 13 shows that the share of CNRS open access publications[10] has increased significantly over the 2017-2021 period. The decrease in 2021 is artificial insofar as bronze open access publications are accessible after a variable embargo period. If the bronze share in 2021 equals the one of 2019 and 2020 (13%), the 2021 open access share would be 76%.

Figure 13. Proportion of open access CNRS publications, full counting, 2017-2021.*



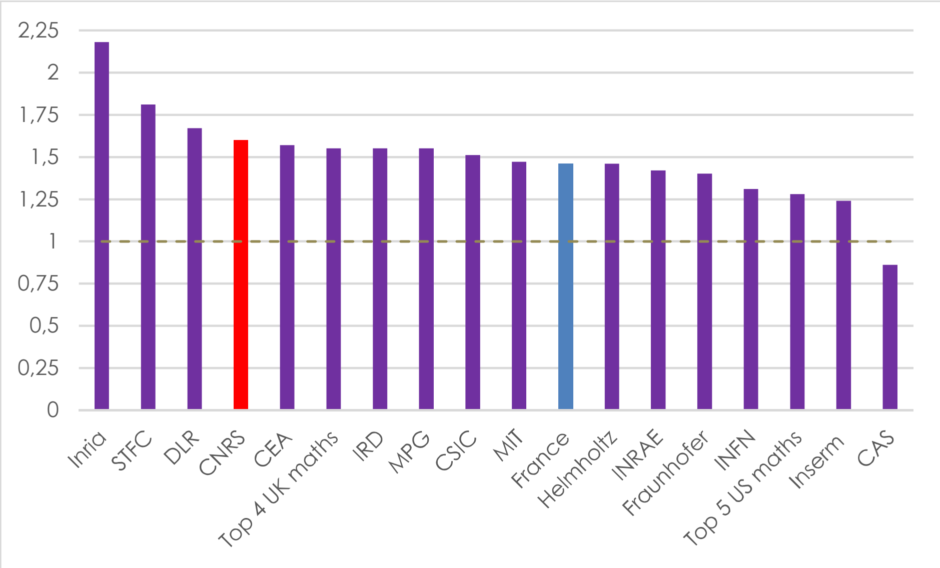
* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science and Unpaywall database.

OST also calculates an open access index. Since the practice of open access publishing is more or less developed according to scientific disciplines, the index normalizes the share of open access publications by the world average for each discipline; the indices are normalized and aggregated at the level of the institutions, which makes it possible to compare institutions even with different disciplinary profiles. The world average is 1, by definition; institutions with an index higher than 1 have more open access publications than the average.

[10] See Appendix C for definition and methodology.

Figure 14 provides the open access index for the CNRS and for all institutions considered in this report for comparison purposes (see Appendix A), as well as for France. The open access index is the highest for STFC and DLR (1.7 and 1.8 respectively). It reaches 1.5 to 1.6 for most of the institutions considered in this report, including the CNRS, MPG and CSIC, as well as for France as a whole. CAS has an index below the world average (0.9).

Figure 14. Open access index by institution and for France, 2019-2021.*

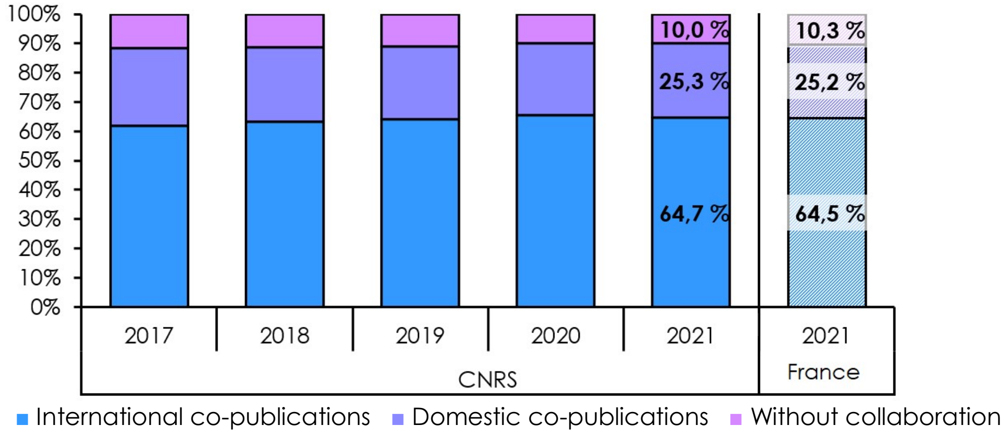


* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science and Unpaywall database.

2.2. Co-publications of the CNRS and partner countries

Figure 15 shows that 90% of CNRS publications are co-publications; the same applies for French publications. The share of CNRS international co-publications increases from 62% in 2017 to 65% in 2021.

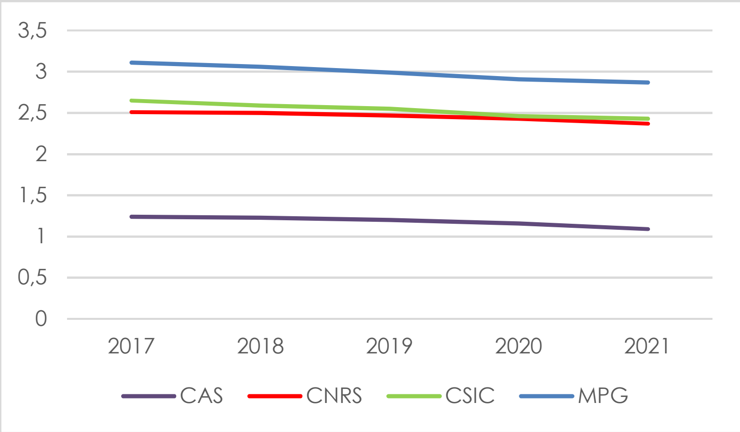
Figure 15. Proportion of co-publications for the CNRS and for France, full counting, 2017-2021.*



* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

The share of international co-publications varies by discipline: it is generally relatively low in medical research and high in the sciences of the universe. At the discipline level, an internationalization index can be calculated as the ratio of the share of international co-publications of an institution to the world average in the same discipline. These indices can be aggregated at the level of institutions, which are then comparable even if their disciplinary profiles differ. Figure 16 shows that the CNRS internationalization index is close to that of CSIC, at 2.4. It is lower than that of MPG and much higher than that of CAS. This relative position of the CNRS is not surprising insofar as internationalization tends to be positively correlated with the maturity of research systems and negatively correlated with the size of the country or of the institution. Figure 16 shows a slight decrease of the index for each of the four institutions, which may be related to the increase in the level of international co-publications worldwide, driven by emerging countries that participate more widely in international collaborations.

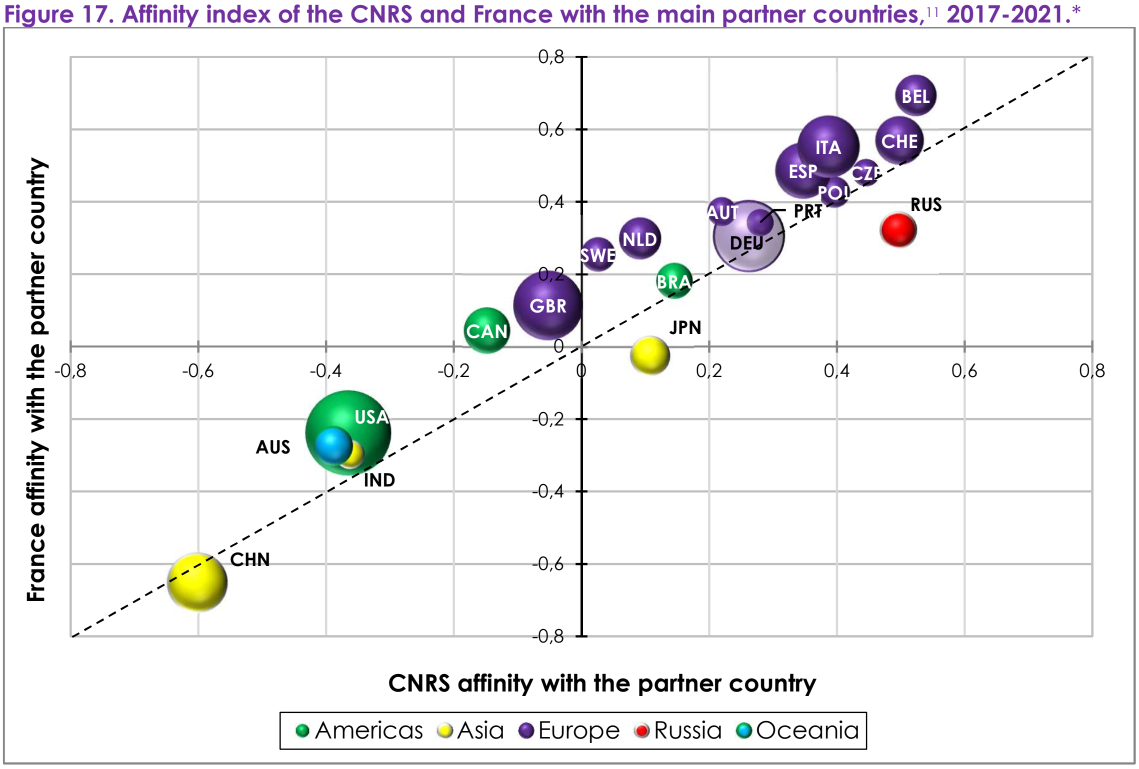
Figure 16. Internationalization index of publications for CAS, CNRS, CSIC and MPG, 2017-2021.*



* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

As is the case for many French research institutions, the United States are the CNRS leading partner country, followed by the EU countries, particularly the major European scientific countries. An affinity index can be calculated to go beyond this comparison of co-publications shares, by relating them to the shares of the partner country in the total of international co-publications in the world – in other words by neutralizing the size effect. The index varies from -1 to 1. Figure 17 shows that the CNRS affinity with the United States is low and even lower than that of France. The affinity of the CNRS with China is even lower, but higher than that of France. The affinity of the CNRS is also higher than that of France for Russia and Japan. The affinity of the CNRS with neighboring countries, especially French-speaking ones, is high but not higher than that of France.

Figure 17. Affinity index of the CNRS and France with the main partner countries,[11] 2017-2021.*



* Data for 2021 are approximately 95% complete. Source: OST database, from the Web of Science.

2.3. Disciplinary profile of CNRS publications

The OST database allows to analyze the scientific profiles of the publications with two different nomenclatures, which have both been built on the basis of the 254 "WoS-categories". The first nomenclature includes 11 broad disciplines (see Figure 12 above); each broad discipline aggregates some of the "WoS categories"[12]. The second

[11] ISO codes: Australia (AUS), Austria (AUT), Belgium (BEL), Brazil (BRA), Canada (CAN), China (CHN), Czechoslovakia (CZE), Germany (DEU), India (IND), Italy (ITA), Japan (JPN), the Netherlands (NLD), Poland (POL), Portugal (PRT), Russia (RUS), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR), United States (USA).
[12] Scientific journals are indexed in one or several categories. Articles published in multidisciplinary journals (like Nature or Science) are distributed among the categories according to their subject. More information is given in Appendix C.

nomenclature is based on the ERC panels; with 27 panels,[13] this second nomenclature can be seen as intermediate between the "coarse" version of the disciplinary nomenclature with 11 broad disciplines and the "fine" version of the disciplinary nomenclature with 254 WoS-categories. Both nomenclatures are used in this report.

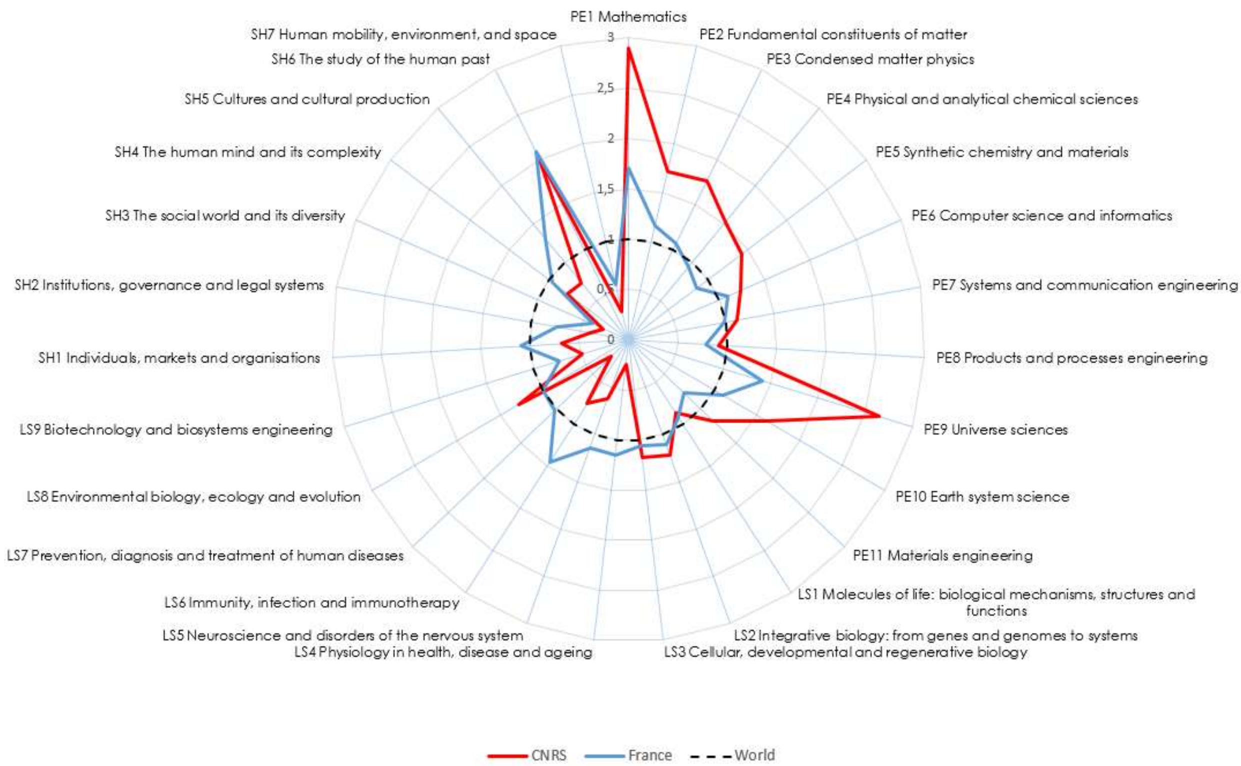
For each nomenclature, the specialization index of an institution in a given domain is the ratio between the share of the institution's publications in this domain and the share of the world's publications in the same domain. An institution (or a country) is said to be "specialized" in a domain if its specialization index is greater than 1. Appendix C gives more information on the nomenclatures and on the related methodology for the specialization analyses.

a. CNRS and France publication profile by ERC panels

Figure 18 presents the CNRS publications specialization index in relation to the world, in the nomenclature of the ERC panels, and compares it to France's specialization index. Compared to the rest of the world, the CNRS remains primarily specialized in the physical sciences and engineering (PE) field. It is more specialized than France in many of the PE panels, particularly in Mathematics (PE1: index 2.7), but also in Universe sciences (PE9), Condensed matter physics (PE3), Fundamental constituents of matter (PE2), Earth system science (PE10) and in the Chemistry panels (PE4 and PE5). It also appears to be specialized (as specialized as France) in The study of the human past (SH6: index 2).

The CNRS is specialized in all panels of the domain of physical sciences and engineering. On the other hand, it is specialized in only three panels of life sciences: Environmental biology, ecology and evolution (LS8), Integrative biology, from genes and genomes to systems (LS2) and Cellular, developmental and regenerative biology (LS3). In the other panels of life sciences, France may be specialized while the CNRS is not (Medical research) or none of the two is specialized. Similarly, in the SH panels other than SH6, France may be specialized while the CNRS is not (in SH1 and SH5), or neither of them is specialized (in SH2, SH3 and SH7).

Figure 18. Specialization index of CNRS and France publications by ERC panel, fractional counting, 2017-2021.*



* Data for 2021 are approximately 95% complete.
Source: OST database, from the Web of Science.

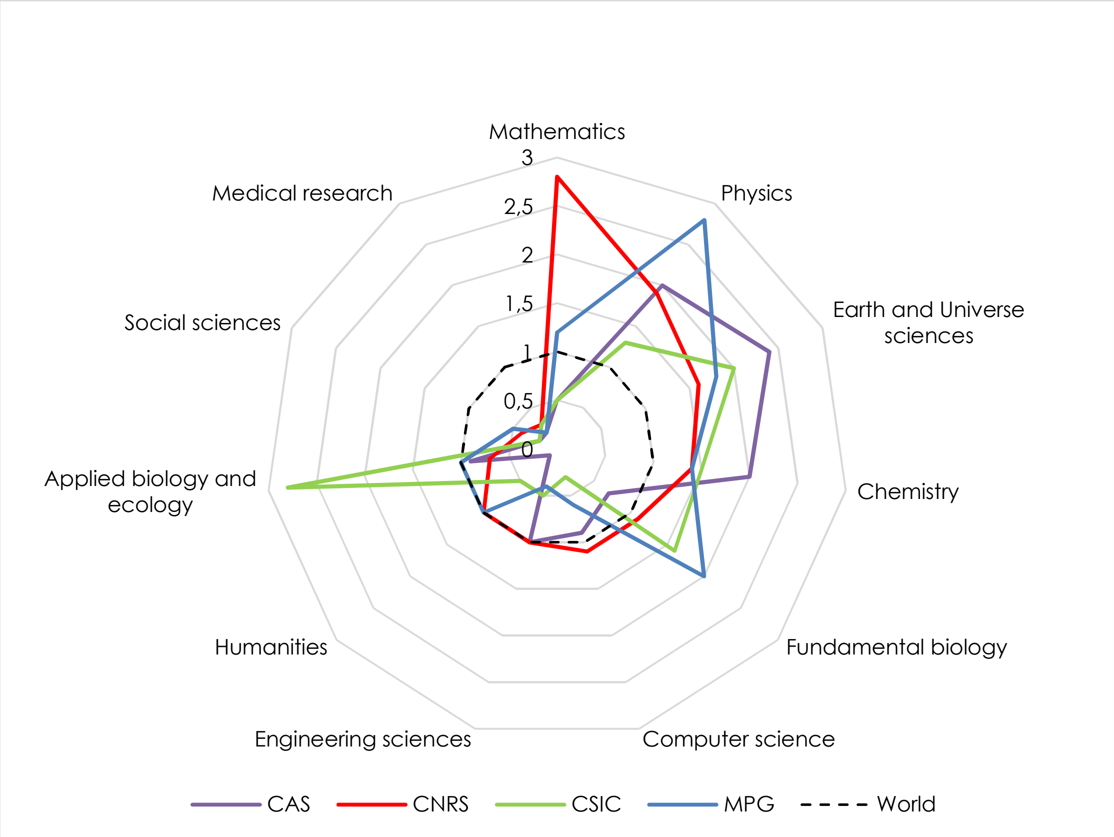
[13] The definition and the number of the ERC panels have been slightly modified in 2021 for Horizon Europe. The new nomenclature includes 27 panels, instead of 25 panels in H2020's nomenclature (see Table 3 above). The three domains that regroup the panels are unchanged: life sciences (LS), physical sciences and engineering (PE), social sciences and humanities (SH). The analysis presented in Figure 18 below uses the new nomenclature with 27 panels.

b. CNRS profile by discipline and comparison with foreign institutions

Figure 19 compares the CNRS profile by discipline with those of CAS, CSIC and MPG,. Compared with these institutions, CNRS appears to be much more specialized in mathematics (index 2.8). MPG is the most specialized in physics (index 2.8), CAS being at the same level as CNRS. CAS is the most specialized in Earth and Universe sciences (index 2.4) and in chemistry (index 2.2). CSIC is the most specialized in applied biology and ecology (index 2.8). MPG and CSIC are the most specialized in fundamental biology. In computer science and in engineering sciences, CNRS and CAS have a neutral profile (indices close to 1), while the two other institutions are not specialized. In humanities, the CNRS and MPG have a neutral profile while CSIC and especially CAS are not specialized; the previous analyses of ERC grants (section 1.2) and publications by ERC panel (section 2.3.a) suggest that the neutral position of the CNRS for humanities (specialization index 1) results from the aggregation of fairly strong areas of specialization (history, archaeology) and areas of non-specialization. Finally, in social sciences, none of the four organizations is specialized compared with the world average; MPG has the lowest index (0.5) in social sciences. The weak specialization of the four organizations is even more marked in medical research.

CNRS is at the world average of 1 or above in 8 disciplines. MPG is at the world average or above in 7 disciplines, CSIC in 5 and CAS in 4.

Figure 19. Specialization index of CAS, CNRS, CSIC and MPG, fractional counting, 2020.



Source: OST database, from the Web of Science.

2.4. Measures of the scientific impact of CNRS publications

Two types of indicators are computed in order to analyze the scientific impact of publications. A first set of comparisons is based on an indicator of average impact, the mean normalized citation score (MNCS). However, the distribution of scientific publications according to their citations is generally highly skewed: thus, the analysis complements this average perspective with a set of indicators that provide the distribution of publications among citation classes: the top 1% most cited publications, the next 4% most cited publications, etc.

a. Mean normalized citation score: comparisons with France and foreign institutions

The OST publication database makes it possible to calculate an impact index, which reflects the academic influence of the publications according to their normalized number of citations.[14] The index used here is the

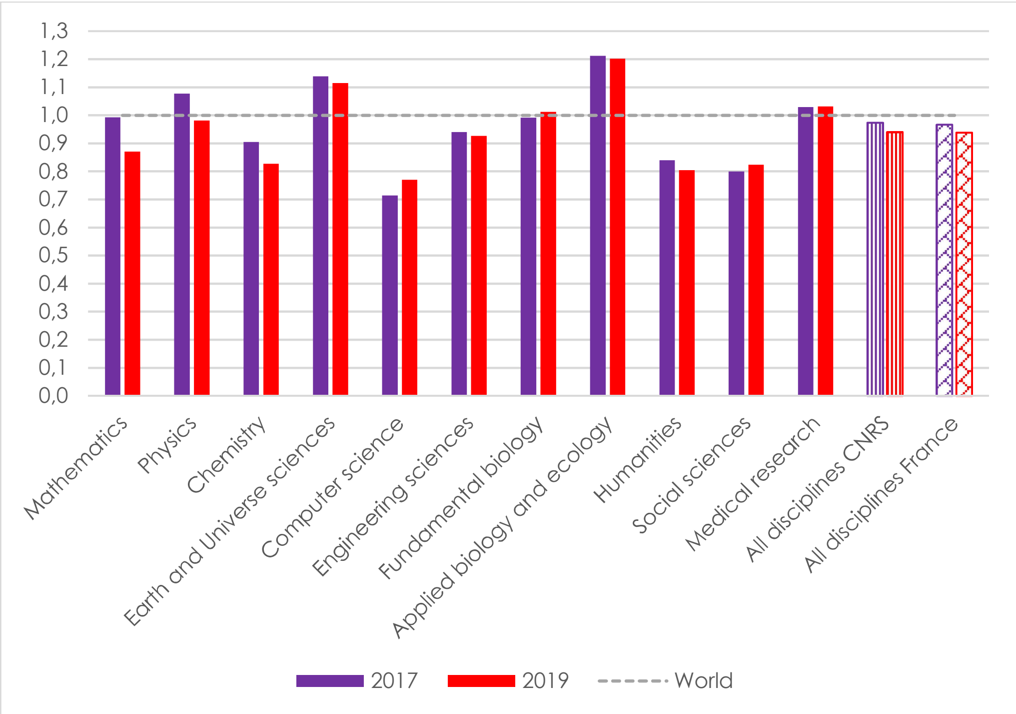
[14] The question of how to measure the impact and importance of research is a critical and delicate issue. The literature on citation motivation discusses a variety of reasons why scientists cite, and this has led some authors to argue that a citation may be an indicator of usage rather than impact. However, it is widely accepted, in view of the cumulative nature of science,

mean normalized citation score (MNCS). The MNCS of an institution in a discipline is greater than 1 if its publications are more cited on average than all world publications in the same discipline, year and type of document.[15] This normalized indicator allows comparisons between countries or institutions even if their disciplinary profiles differ.

In addition to the analysis presented in this section, Appendix E reviews studies showing that the levels of MNCS vary when measured with different publication databases, but that the relative positions of institutions remain similar.

Figure 20 displays the MNCS per discipline for CNRS publications of 2017 and 2019.[16] The values of MNCS vary by discipline, as do their evolutions between 2017 and 2019. CNRS MNCS is above the world average and stable in applied biology and ecology (1.2) and in Earth and Universe sciences (1.1). They are at the world average in medical research and fundamental biology. In mathematics, physics and chemistry, the indicator decreases between 2017 and 2019. It increases on the contrary in computer science, while remaining at the lowest level among all disciplines. The MNCS increases slightly in social sciences but decreases in humanities. Overall, the indicator slightly decreases between 2017 and 2019. The evolution is the same for the average of French publications.

Figure 20. MNCS of CNRS publications by discipline, fractional counting, 2017 and 2019.



Source: OST database, from the Web of Science.

Figure 21a gives the MNCS of publications published in 2019 for CAS, CNRS, CSIC and MPG by discipline. MPG score is the highest in 8 out of the 11 disciplines. The gap in favor of MPG is the largest in computer science, applied biology and ecology, and fundamental biology; it is also substantial in earth and Universe sciences. CAS has the highest MNCS in 3 disciplines: chemistry, mathematics and social sciences. In the latter however, CAS is not specialized and has relatively few publications.[17] CSIC highest MNCS are in applied biology and ecology (index 1.3) and in physics (index 1.2). The CNRS has the lowest MNCS of the four institutions in 9 of the 11 disciplines; it is below the world average in mathematics, physics, chemistry, computer science, engineering sciences, humanities, and social sciences.

Figure 21b summarizes the comparison: it shows the average MNCS of publications of the four institutions for all disciplines. MPG has the highest MNCS at 1.5, followed by CAS (1.3), CSIC (1.1), and CNRS (0.9). Besides, whatever the discipline, MNCS decreased between 2017 and 2019 for all institutions; it remained nearly stable for CAS and decreased by nearly 10% for MPG.

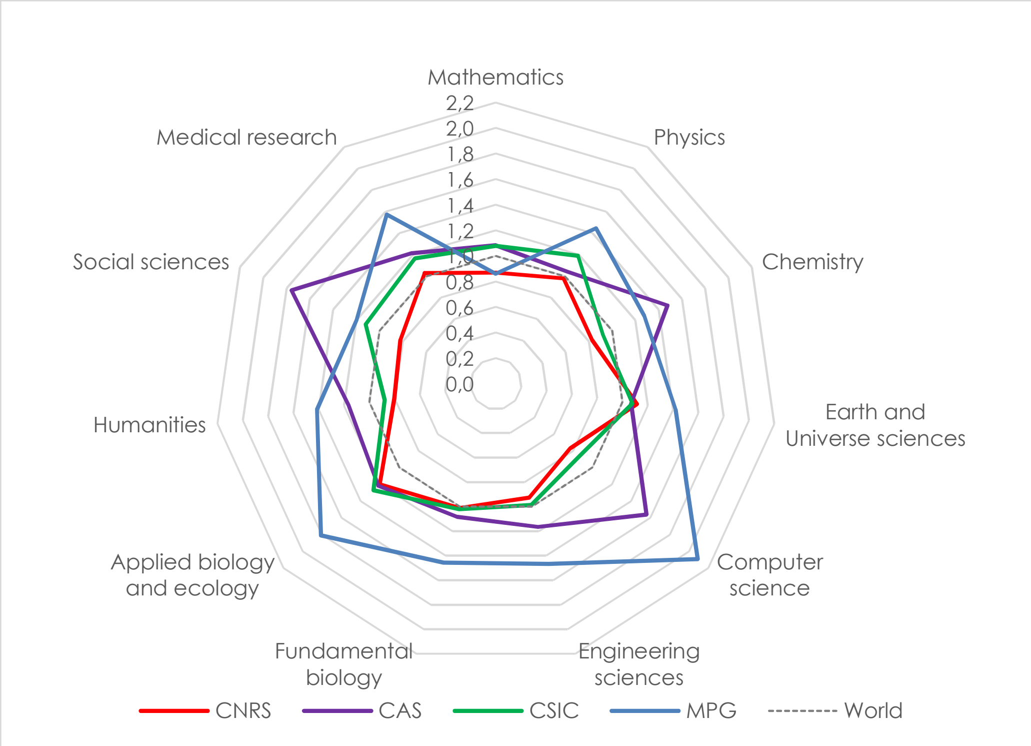
that citations in the aggregate can demonstrate the impact of work on the scientific landscape (Measuring research : what everyone needs to know, C. R. Sugimoto and V. Larivière, Oxford University Press, 2018).

[15] Citations are counted for each publication without reference to the journal, and no use of journal impact factors (see Appendix C for more details on the methodology).

[16] The impact index cannot be calculated over the entire period to allow a minimum time for citations to be recorded.

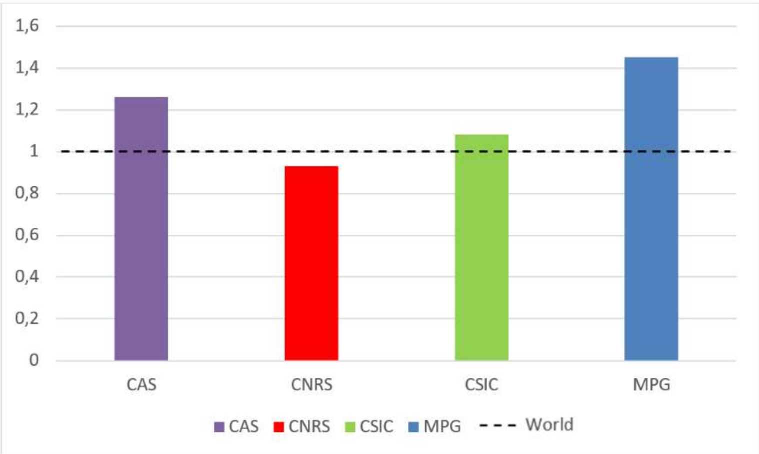
[17] It it would be interesting to examine CAS publications that have high impacts, in particular to see if they are international co-publications, typically more cited.

Figure 21a. MNCS of 2019 publications by discipline, CAS, CNRS, CSIC and MPG, fractional counting.



Source: OST database, from the Web of Science.

Figure 21b. MNCS of 2019 publications for CAS, CNRS, CSIC and MPG, all disciplines, fractional counting.



Source: OST database, from the Web of Science.

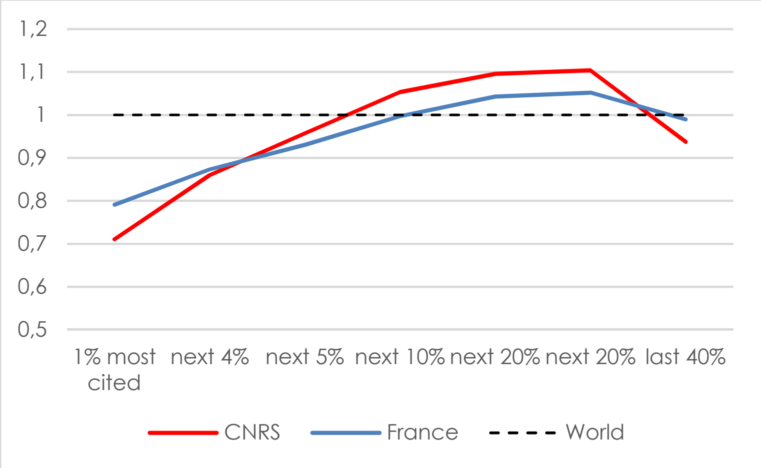
b. High impact publications

Most publications have few or no citations while a small number of them are highly cited. Since the average indicators cannot reflect this general characteristic, they are usefully complemented with the distribution of publications in the different citation classes, like the top 1% or 10% most cited publications.

The activity index of a given institution in a given citation class is the ratio of the share of the institution's publications in this class to the share of the world publications in the same class; by definition, the activity index of the world is equal to 1 for each citation class.

Figure 22 shows the CNRS activity indices for the publications published in 2019, and compares them with France's activity indices. The CNRS "activity profile" over the citation classes is less favorable than that of France for the most cited classes, and it becomes more favorable after the five most cited centiles. CNRS and France profiles are nevertheless similar, with activity indices below the world average (i.e., 1) for the publications in the most cited decile.

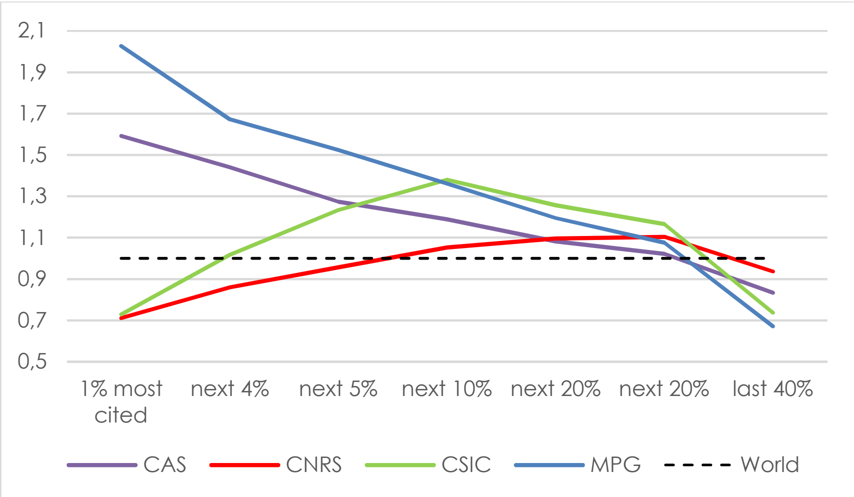
Figure 22. Activity index in citation classes for CNRS and France, 2019 publications.



Source: OST database, from the Web of Science.

Figure 23 shows that the activity profile of the CNRS in the citation classes is comparable to that of CSIC and less favorable than those of MPG and CAS: the two latter institutions are substantially more present in the most-cited classes. These profiles are consistent with the comparisons of the average impact indices of the four institutions (see Figure 21b).

Figure 23. Activity index in citation classes, CAS, CNRS, CSIC and MPG, 2019 publications.



Source: OST database, from the Web of Science.

Overall, the two types of indicators (MNCS and activity index) lead to a similar conclusion regarding CNRS position. The average indicator is below the world average and the activity index in the top 5% most-highly cited publications is more than 10% below world average. The position of CNRS relative to the benchmark institutions is also similar when observed with the two types of indicators.

3. INTERNATIONAL COMPARISONS OF CNRS PUBLICATIONS IN NINE SCIENTIFIC FIELDS

3.1. Introduction and scope of the analysis

This chapter presents a more detailed analysis of CNRS publications in nine scientific fields: nuclear physics and particle physics, chemistry, ecology and environment, physics, information sciences, biology, engineering and systems, mathematics, earth and space sciences.

These 9 fields bear the same names as 9 of the 10 CNRS Institutes (IN2P3, INC, INEE, INP, INS2I, INSB, INSIS, INSMI, INSU). After discussion between Hcéres and the CNRS, the field of social sciences and humanities (SSH) – corresponding to the 10th Institute (INSHS) – has not been analyzed. The CNRS considers that an analysis of the SSH field based on the OST database would have been irrelevant because the WoS does not include books or book chapters and has a partial coverage of journals in other languages than English (whereas books, book chapters and journal articles written in French are a substantial part of the CNRS production in SSH). While considering that such an analysis would have been worthwhile (see Box 1 above and Appendix D), Hcéres agreed not to extend the analysis to SSH in this chapter.

For each of these fields, the scope of the analysis presented below has been defined as a selection – proposed by the CNRS – of some of the 254 WoS categories. Each of the nine fields is defined as an aggregation of selected WoS categories, and the scope of each field-analysis is the corpus of *all publications of CNRS UMRs in the field.*[18] Table 6 below gives the number of categories selected in each field, and Appendix C2 gives the detailed list of categories for every field; it shows that there are some overlaps between the fields, since some WoS categories have been selected in several of the 9 scientific fields.

Table 6 also gives, for each field, the list of French or foreign institutions with which the CNRS is compared.[19] These benchmarked institutions have been chosen jointly by Hceres and the CNRS. They include the institutions that were already considered in chapters 1 and 2: CAS in China, CNR in Italy, CSIC in Spain, as well as Helmholtz, MPG and Leibniz in Germany. They also include other foreign institutions: DLR (*Deutches Zentrum für Luft- und Raumfahrt*, Aeronautics and space research center) and *Fraunhofer Gesellschaft* in Germany, INFN (*Istituto nazionale di fisica nucleare*, National institute of nuclear physics) in Italy, MIT (Massachusetts institute of technology) in the US, and STFC (Science and technology facilities council) in the UK. Moreover, in the field of mathematics, CNRS publications are compared with the publications of two sets of universities: "Top5 US", the set of the top 5 US universities in the Shanghai 2021 ranking in mathematics, which are all in the top 10 worldwide: Princeton (2nd), Stanford (6th), MIT (7th), NYU (8th), and UT Austin (10th)); "Top4 UK", the set of the top 4 UK universities in the same ranking, which are all in the top 20 worldwide: Cambridge (4th), Oxford (5th), Imperial College (17th) and Warwick (18th). Lastly, comparisons are also made in some fields with French national research organizations: CEA (Atomic energy and alternative energies commission), Inria (National research institute in digital science and technology), INRAE (National research institute for agriculture, food and environment), Inserm (National institute for health and medical research), IRD (National research institute for development).[20]

Table 6. Scientific fields, number of WoS categories and benchmarked institutions.

Scientific fields	Number of WoS categories	Benchmarked institutions
Nuclear physics and particle physics	2	CEA, INFN, STFC
Chemistry	29	CAS, CISC, MPG
Ecology and environment	21	CSIC, INRAE, IRD, MPG
Physics	20	CAS, CEA, CSIC, MPG
Information sciences	16	CEA, Inria, MPG
Biology	28	CAS, Inserm, MPG
Engineering and systems	59	CSIC, Fraunhofer, MIT
Mathematics	4	CAS, Top5 US, Top4 UK
Earth and space sciences	18	DLR, MPG, Helmholtz

Source: OST database, from the Web of Science.

Table 7 provides the average annual number of publications in each of the 9 scientific fields for the world, France and the CNRS, for the period 2017-2021. It also gives, for each field, the weight of the French publications in world publications. This weight varies from 2.3% in chemistry to 4.3% in Mathematics. Behind Mathematics; the French weight in the world is also above 3% in Nuclear physics and particle physics. One may recall here that for all disciplines, the world share of French publications is close to 2.5%.[21]

Moreover, the right-hand column gives the weight of CNRS publications in the French publications. This weight varies quite substantially, from 31% to 77%. It is below 50% in three fields, Biology, Ecology and environment and Information sciences. It is above 60% in Chemistry, Earth and space sciences, and Physics, and it is greater than 70% in Mathematics and in Nuclear physics and particle physics.

[18] For a given field, this scope is different from (and somewhat larger than) the set of publications of the sole CNRS UMRs that are attached to the corresponding Institute. For example, the scope of the field "Ecology and environment" includes some publications in vegetal biology of UMRs that are attached to the Biology Institute (INSB) and have no secondary attachment to the Ecology and environment institute (INEE).

[19] Key figures on each of these institutions are given in Appendix A.

[20] There are some overlaps between the set of the CNRS publications and the set of publications of these French institutions, since they have joint research units (*UMRs*) with the CNRS, or joint research teams within these *UMRs*.

[21] La position scientifique de la France à travers ses publications, 2021, Etat de l'enseignement supérieur, de la recherche et de l'innovation, MESR : https://publication.enseignementsup-recherche.gouv.fr/eesr/FR/EESR15_R_30/la_position_scientifique_de_la_france_dans_le_monde_a_travers_ses_publications/

Table 7. Average annual number of publications in each field for the world, France and the CNRS, fractional counting, 2017-2021.

Fields	World	France	France world share	CNRS	CNRS share in France
Nuclear physics and particle physics	11,935	392	3.3%	303	77.1%
Chemistry	397,936	9,286	2.3%	5,806	62.5%
Ecology and environment	189,344	5,201	2.7%	2,267	43.6%
Physics	271,790	7,180	2.6%	4,807	67.0%
Information sciences	223,492	6,350	2,8%	2,987	47,0%
Biology	363,069	10,681	2.9%	3,329	31.2%
Engineering and systems	804,118	19,728	2.5%	10,819	54.8%
Mathematics	60,595	2,593	4.3%	1,910	73.7%
Earth and space sciences	187,755	5,129	2.7%	3,319	64.7%

Source: OST database, from the Web of Science.

3.2. Field by field analyses

For each of the 9 scientific fields, the presentation of the bibliometric analysis follows the same two-step approach.

First, a table (Tables 8 to 16) compares CNRS publications in a given field with those of the other French or foreign institutions chosen for comparison purposes. The indicators presented in the Table are identical to those defined in chapter 2. Line 1 gives the number of publications in fractional counting for year 2020. Line 2 shows the evolution of the number of publications between 2017 and 2020, also in fractional counting. Line 3 provides the specialization index for each institution in the field (for 2017-2021).[22] The next three lines are about international co-publications: Line 4 gives the share of the international co-publications (for 2017-2021), Line 5 indicates the first partner country and Line 6 provides the share of this country in the international co-publications of each institution. The two last lines present impact indicators: Line 7 reports the MNCS of each institution in the field for year 2018 and Line 8 provides the activity index of each institution in the citation class of the top 10% most cited publications in the field, also for 2018.[23]

Second, for each of the 9 fields, a figure (Figures 24 to 32) displays the disciplinary profile of the CNRS and the benchmark institutions. The figure first shows the 5 WoS categories with the highest shares in the field for world publications. Then it shows, for each institution, the 5 categories that have the highest shares in the institution's publications, as well as the shares of the institution's publications in the categories that are among the world top 5. These figures are to be seen as a first illustration of the type of insights and questions that such bibliometric studies can bring to a large institution like the CNRS; a deeper analysis could be done only in close relation with the CNRS.

a. Nuclear physics and particle physics

Table 8 compares CNRS publications in the field of Nuclear physics and particle physics with those of CEA, INFN, and STFC. INFN has a higher number of publications than the CNRS, while CEA and STFC have a much lower number of publications. The evolution of the number of publications between 2017 and 2020 is quite contrasted, with a substantial decrease for the two French institutions, a significant growth for INFN and a very strong growth for STFC. The CNRS has a specialization index higher than 2 in this field. CEA and STFC are more specialized, and INFN is by far the most specialized.

STFC's international co-publication rate is close to 100%, but it is also very high for the other three institutions. For all four institutions, the first partner country, the United States, has quite a large share in the international co-publications.

The STFC's impact indicators are quite favorable but are based on a small number of publications. The MNCS of CEA and CNRS are below the world average of 1, while INFN's and STFC's are above the average.

[22] The specialization index is defined at the beginning of section 2.3 above.
[23] The MNCS is defined at the beginning of section 2.4.a. The activity index of an institution in a given citation class is defined at the beginning of section 2.4.b.

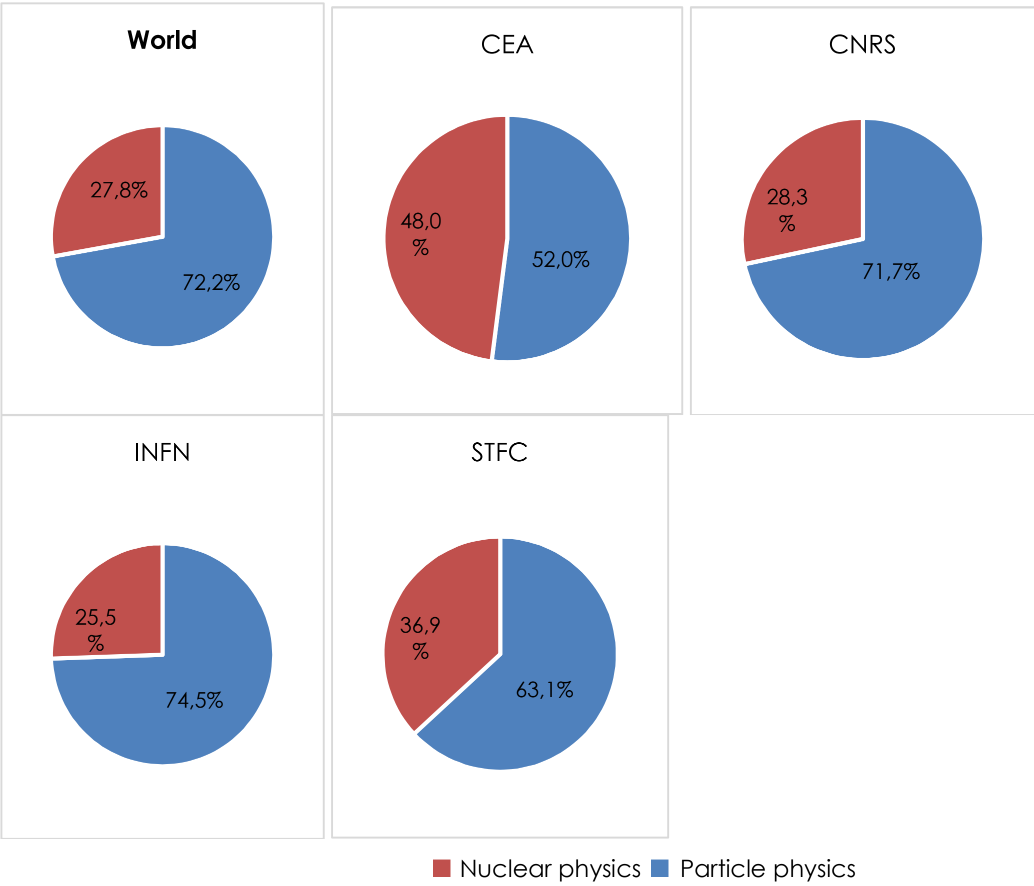
Table 8. Nuclear physics and particle physics: international comparison of CNRS publications.

	Indicators	CEA	CNRS	INFN	STFC
1	Number of publications	76	295	368	11
2	Evolution 2020 / 2017	-20%	-10%	+13%	+65%
3	Specialization index	8.3	2.3	60.9	9.7
4	Rate of international co-publications	89%	86%	79%	95%
5	First partner country	USA	USA	USA	USA
6	Share of this country	66%	51%	48%	89%
7	Normalized citation score (MNCS)	0.9	0.9	1.1	1.0
8	Activity index in the top10% most cited	0.9	1.1	1.1	1.4

Source: OST database, from the Web of Science.

This field has only two WoS categories: nuclear physics and particle physics. Figure 24 shows the shares of both categories for all world publications, and for each institution. The shares of the CNRS and of INFN are very close to world shares, while STFC and especially CEA have a higher share in nuclear physics.

Figure 24. Nuclear physics and particle physics: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

b. Chemistry

Table 9 compares the CNRS publications in the field of chemistry with CAS, CSIC and MPG. CNRS publications represent about 60% of those of CAS, and the two other institutions have a much lower number of publications. The number of publications slightly decreased between 2017 and 2020 for the CNRS and MPG. CAS is the most specialized institution of the four in the field, while CNRS specialization index is similar to that of CSIC and MPG.

The three European institutions have equivalent shares of international co-publications, while this rate is much lower for CAS (25%). The US share of co-publications is much higher for MPG and CSIC than for CNRS and CAS.

MPG and CAS have the highest impact indicators. The MNCS of the CNRS is slightly below the world average, and its activity index in the decile of most cited publications is at 80% of the world average.

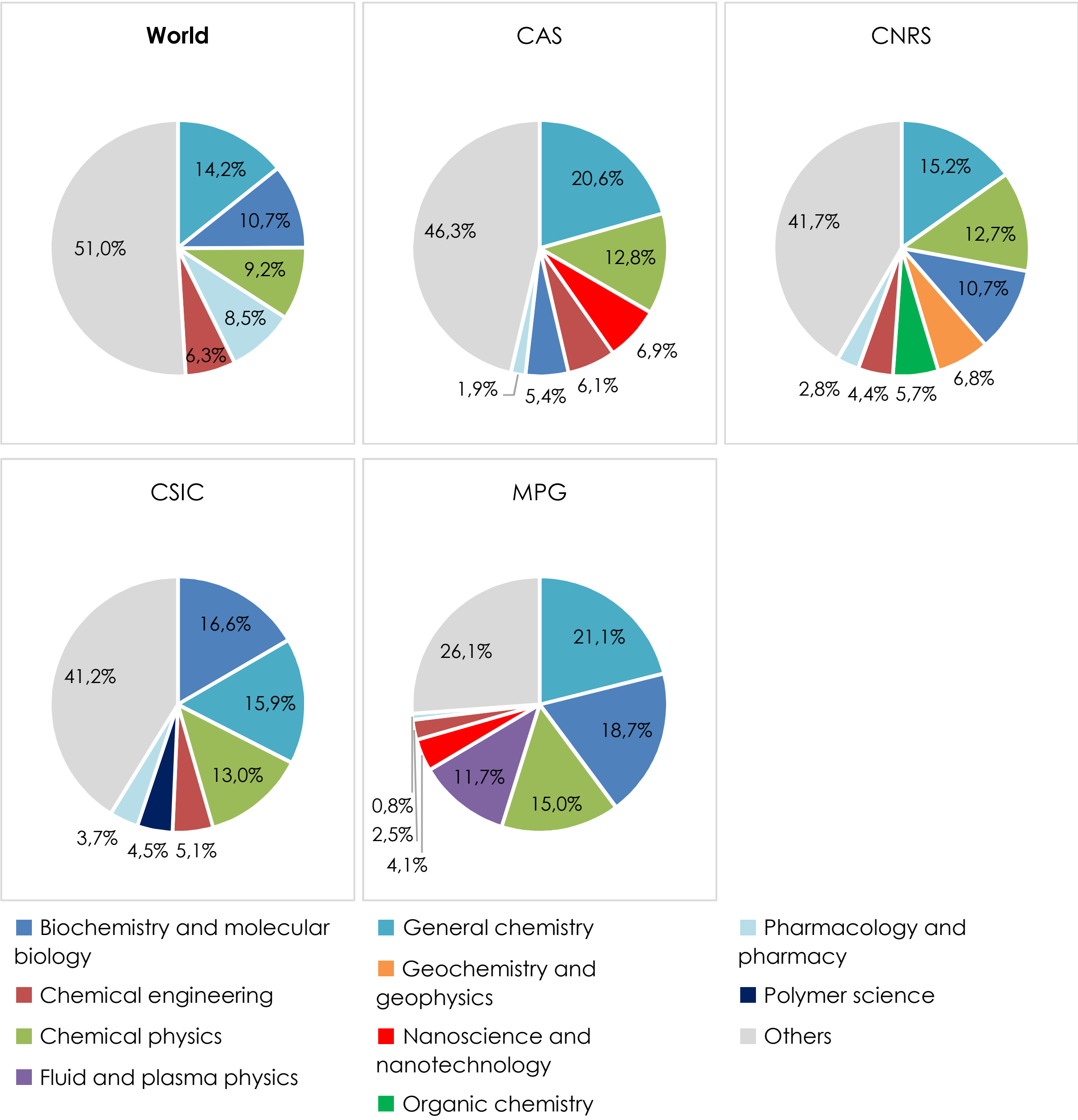
Table 9. Chemistry: international comparison of CNRS publications.

	Indicators	CAS	CNRS	CSIC	MPG
1	Number of publications	9,532	5,684	1,225	899
2	Evolution 2020 / 2017	+7%	-6%	+3%	-5%
3	Specialization index	1.7	1.3	1.4	1.4
4	Rate of international co-publications	25%	64%	61%	70%
5	First partner country	USA	USA	USA	USA
6	Share of this country	16%	18%	31%	41%
7	MNCS	1.4	0.9	1.0	1.5
8	Activity index in the top10% most cited	1.6	0.8	0.9	1.6

Source: OST database, from the Web of Science.

Figure 25 shows the shares of the WoS categories that have the highest numbers of publications, for the world and for each institution. Some categories are highlighted in the institutions' shares, such as Geochemistry and geophysics as well as Organic chemistry for the CNRS, or Nanoscience and nanotechnology for CAS.

Figure 25. Chemistry: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

c. Ecology and environment

CNRS publications in the field of Ecology and Environment are compared with those of CSIC, INRAE, IRD and MPG in Table 10. The CNRS is the institution that publishes the most because of its size, but it is not very specialized in this field (specialization index of 1.1). CSIC is the second most publishing institution and has a specialization index of 2.4. IRD and INRAE are logically much more specialized, with indices of 4.8 and 3.3 respectively. MPG has the smallest number of publications and the smallest specialization index. Between 2017 and 2020, the number of publications increased for all institutions except IRD.

The shares of international co-publications are relatively close among the five institutions, MPG having the highest rate. The first partner country is the United States for all five, with a much higher share for MPG.

The impact indicators of the five organizations are above the world average. MPG has the highest indicators: 1.6 for the MNCS and 2.1 for the activity index in the decile of most cited publications. The indicators for the French institutions and CSIC are close, between 1.1 and 1.2.

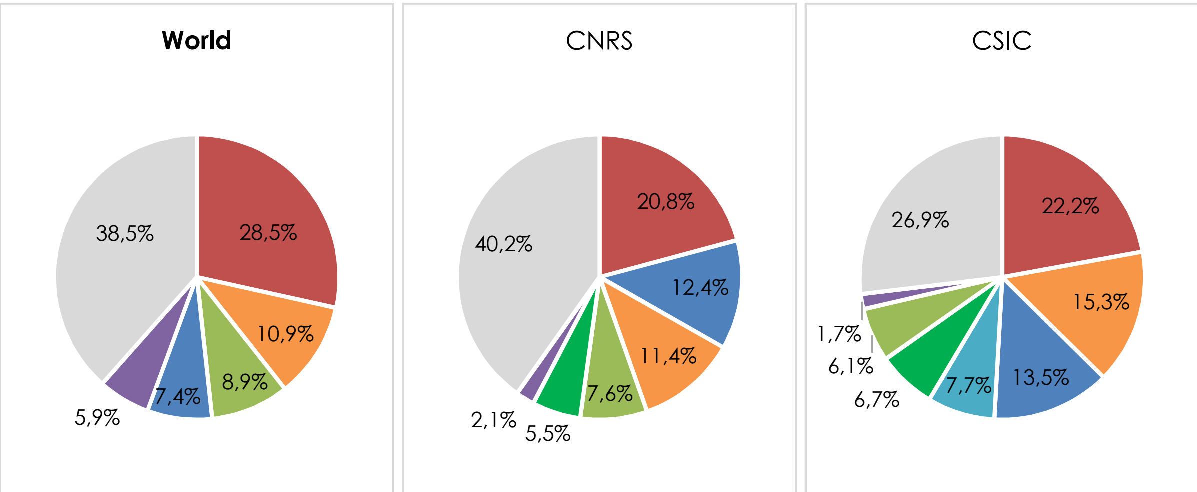
Table 10. Ecology and environment: international comparison of CNRS publications.

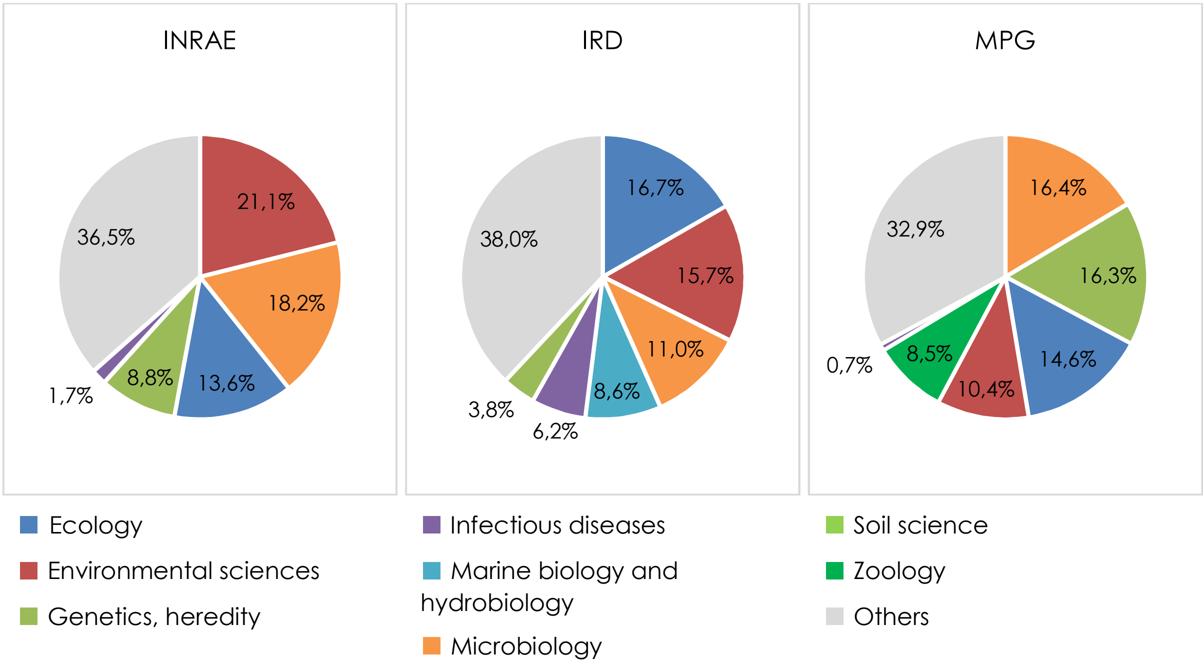
	Indicators	CNRS	CSIC	INRAE	IRD	MPG
1	Number of publications	2,310	975	707	584	277
2	Evolution 2020 / 2017	+5%	+11%	+16%	-2%	+5%
3	Specialization index	1.1	2.4	3.3	4.8	0.9
4	Rate of international co-publications	71%	68%	64%	75%	83%
5	First partner country	USA	USA	USA	USA	USA
6	Share of this country	26%	24%	23%	26%	42%
7	MNCS	1.1	1.1	1.1	1.1	1.6
8	Activity index in the top10% most cited	1.1	1.2	1.2	1.1	2.1

Source: OST database, from the Web of Science.

Figure 26 shows the shares of the WoS categories that have the highest numbers of publications in this field, for the whole world and for each institution.

Figure 26. Ecology and environment: top shares per WoS category for the world and for each institution.





Source: OST database, from the Web of Science.

d. Physics

Table 11 compares CNRS publications in the field of physics with those of CAS, CEA, CSIC and MPG. CAS has nearly 70% more contributions than the CNRS, which has nearly 5 to 7 times more publications than each of the other institutions. The number of publications increased for CAS between 2017 and 2020, while it decreased for CEA, the CNRS and MPG. CEA is the most specialized institution in Physics (index 3.5), while CNRS is less specialized than MPG and CAS, and more than CSIC.

Again, CAS has the lowest share of international co-publications and MPG has the highest. The first partner country is the USA for CAS, CNRS and MPG; it is Germany for CEA and CSIC.

MPG has the highest impact indicators, both the MNCS and the activity index in the decile of most cited publications. The CAS indices are between 1.2 and 1.3 and those of CSIC between 1.0 and 1.1. The CNRS and CEA indices are below the world average.

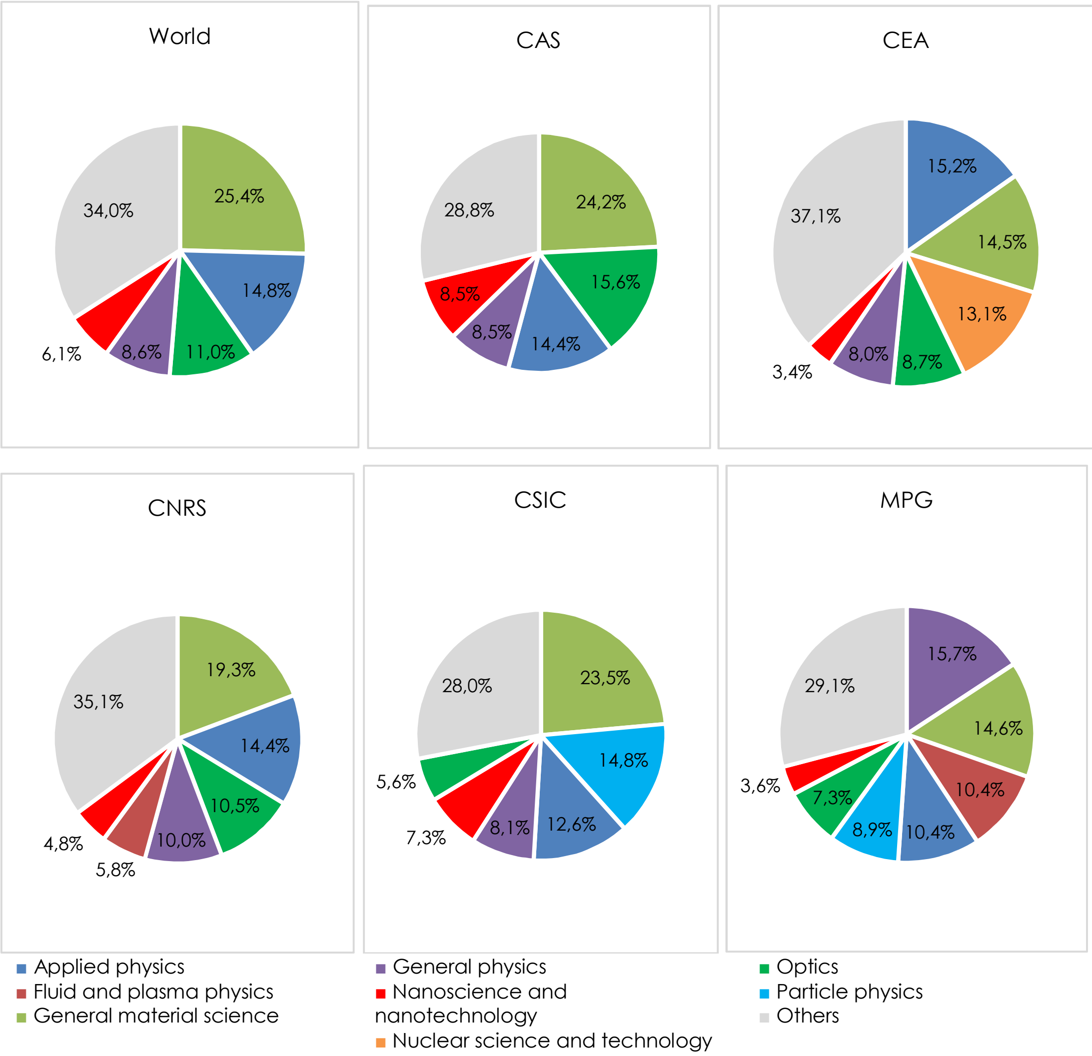
Table 11. Physics: international comparison of CNRS publications.

	Indicators	CAS	CEA	CNRS	CSIC	MPG
1	Number of publications	7,722	809	4,608	683	958
2	Evolution 2020 / 2017	+8%	-12%	-15%	-0%	-14%
3	Specialization index	2.0	3.5	1.6	1.2	2.3
4	Rate of international co-publications	28%	65%	67%	73%	78%
5	First partner country	USA	DEU	USA	DEU	USA
6	Share of this country	46%	41%	24%	32%	37%
7	MNCS	1.2	1.0	0.9	1.0	1.4
8	Activity index in the top10% most cited	1.3	1.0	0.9	1.1	1.8

Source: OST database, from the Web of Science.

Figure 27 shows the shares of the WoS categories that have the highest numbers of publications in this field, for the whole world and for each institution.

Figure 27. Physics: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

e. Information sciences

Table 12 compares CNRS publications in information sciences with those of CEA, Inria and MPG. The CNRS is the institution that publishes the most and it has a moderate specialization (index 1.2). Inria is a specialized institution devoted to this field (index 5.5). CEA and MPG are not specialized in information sciences and publish significantly less in the field. All institutions in the sample, more especially the CNRS and Inria, show a decrease in the number of publications between 2017 and 2020.

MPG is distinguished by a high share of international co-publications (71%) whereas the three French organizations have significantly lower shares of international co-publications. The first partner country is again the United States for all institutions.

MPG has again the highest impact indicators, at more than 2. Inria has indicators about 20% above the world average, both the MNCS and the activity index in the decile of most cited publications. The CNRS has an MNCS of 0.9 and an activity index in the decile of most cited publications of 0.8. The CEA's impact indicators are about 10% lower.

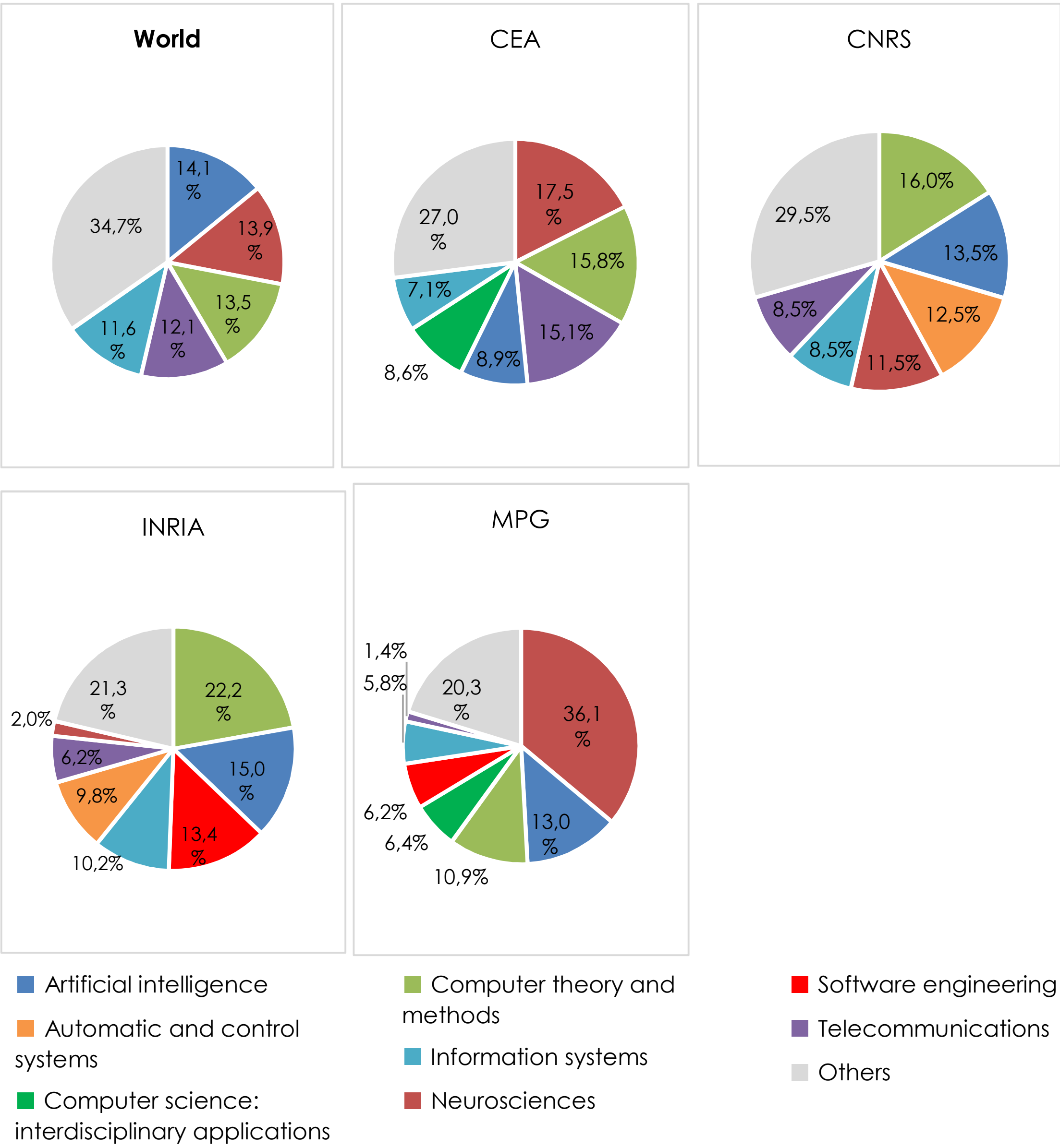
Table 12. Information sciences: international comparison of CNRS publications.

	Indicators	CEA	CNRS	Inria	MPG
1	Number of publications	121	2,544	518	283
2	Evolution 2020 / 2017	-15%	-27%	-28%	-17%
3	Specialization index	0.7	1.2	5.5	0.9
4	Rate of international co-publications	44%	56%	56%	71%
5	First partner country	USA	USA	USA	USA
6	Share of this country	34%	18%	25%	38%
7	MNCS	0.8	0.9	1.2	2.1
8	Activity index in the top10% most cited	0.7	0.8	1.2	2.2

Source: OST database, from the Web of Science.

Figure 28 shows the shares of the WoS categories that have the highest numbers of publications in the field, for the whole world and for each institution. Some categories are highlighted in the institutions' shares, such as Automatic and control systems for the CNRS and Inria, or Neurosciences for MPG.

Figure 28. Information sciences: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

f. Biology

Table 13 compares CNRS publications in the field of biology with those of CAS, Inserm and MPG. CAS and the CNRS are the institutions that publish the most, around 50% more than Inserm and four times more than MPG. Between 2017 and 2021, CAS and Inserm had the highest growth rates of the number of their publications, while the CNRS and MPG had a slight decrease. Inserm is clearly more specialized in biology than the other institutions: its specialization index is higher than 3, while that of MPG is 1.4 and those of CNRS and CAS are below the world average.

MPG again has the highest share of international co-publications (74%), followed by CNRS (63%), Inserm (57%) and CAS (35%). All four institutions have the United States as their first partner country.

MPG has the highest impact indicators (1.6 and 1.8), followed by CAS. In this field, the CNRS has impact indicators above the world average, around 10% lower than those of Inserm.

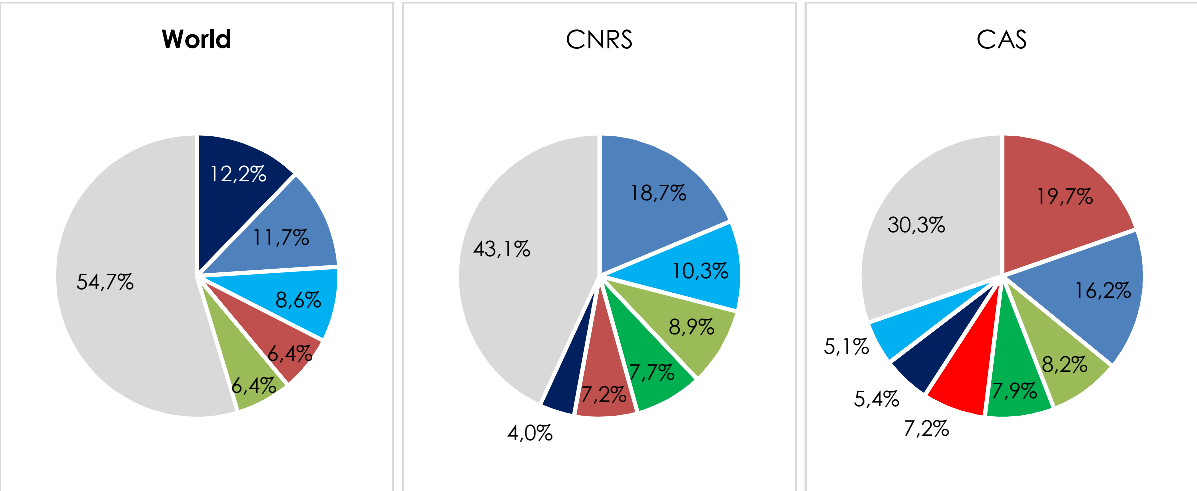
Table 13. Biology: international comparison of CNRS publications.

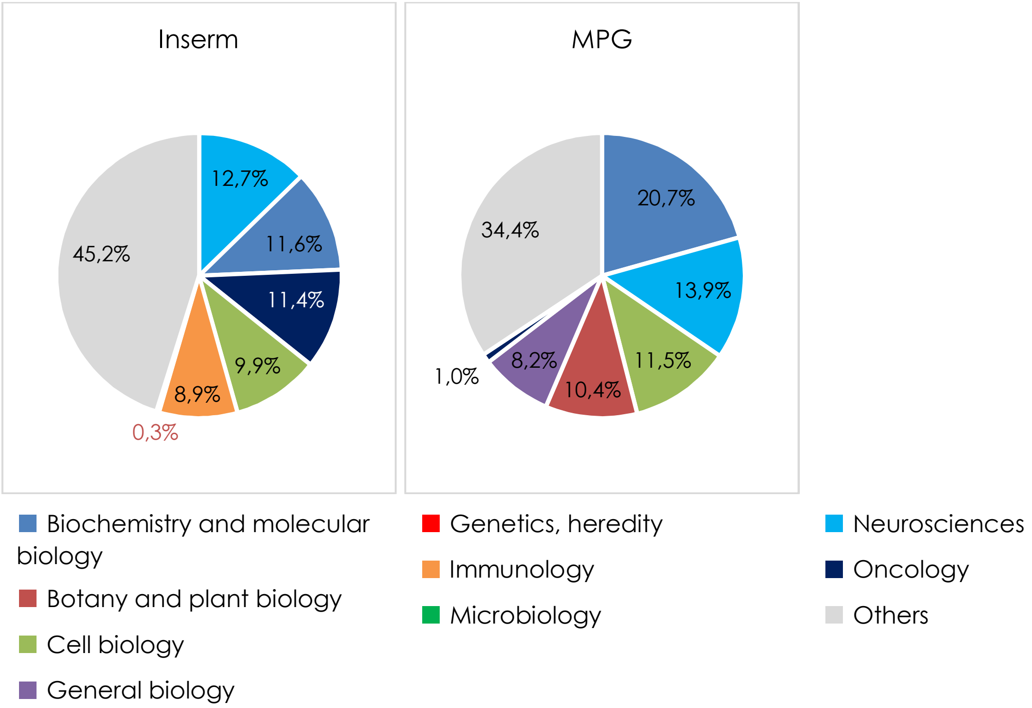
	Indicators	CAS	CNRS	Inserm	MPG
1	Number of publications	3,406	3,330	2,166	804
2	Evolution 2020 / 2017	+25%	-2%	+16%	-6%
3	Specialization index	0.6	0.8	3.1	1.4
4	Rate of international co-publications	35%	64%	57%	74%
5	First partner country	USA	USA	USA	USA
6	Share of this country	50%	30%	37%	38%
7	MNCS	1.1	1.1	1.1	1.6
8	Activity index in the top10% most cited	1.2	1.1	1.2	1.8

Source: OST database, from the Web of Science.

Figure 29 shows the shares of the WoS categories that have the highest numbers of publications in this field, for the whole world and for each institution. Some categories are highlighted in the institutions' shares, such as Microbiology for CAS and the CNRS, or Immunology for Inserm.

Figure 29. Biology: top shares per WoS category for the world and for each institution.





Source: OST database, from the Web of Science.

g. Engineering and systems

Table 14 compares CNRS publications in the field of engineering and systems with CSIC, Fraunhofer and MIT. CNRS is by far the most publishing institution, with over 10,000 publications, while CSIC and MIT have recorded about 1,500 contributions and Fraunhofer nearly 700. All four institutions have had a decrease in their number of publications between 2017 and 2020. Fraunhofer is the most specialized institution with a specialization index of 1.9; CNRS and MIT are moderately specialized, while CSIC's share of publications in this field is just at the world average.

Fraunhofer stands out with a low share of international co-publications (37%), while the other three institutions have a share between 57 and 67%. The United States is the leading partner country for the three European institutions, while the leading partner country of MIT is China.

MIT has the highest impact indicators at more than 2. Fraunhofer has indicators 20% above the world average, both the MNCS and the activity index in the decile of most cited publications. The CNRS has the lowest impact indicators, slightly below the world average.

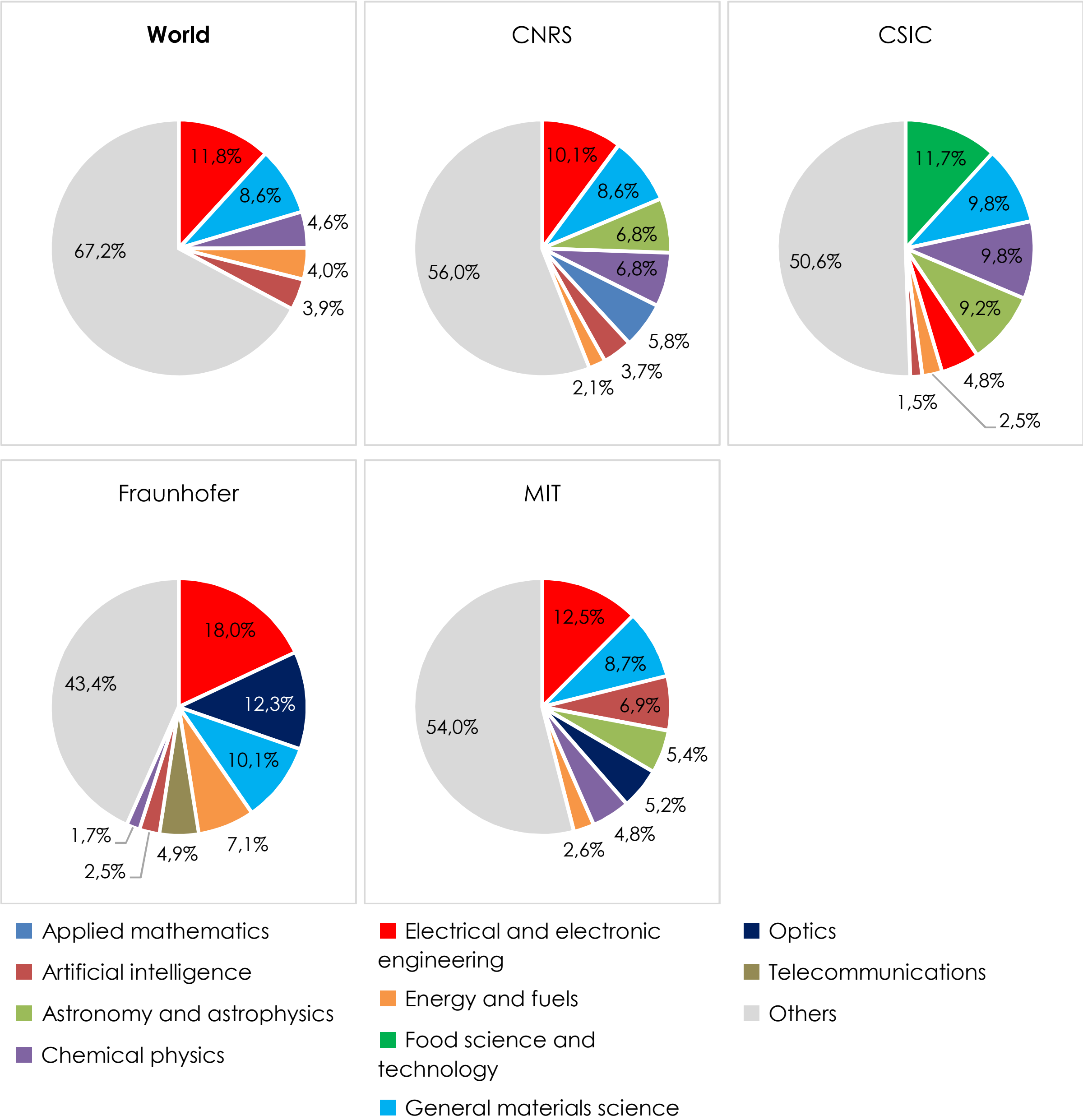
Table 14. Engineering and systems: international comparison of CNRS publications.

	Indicators	CNRS	CSIC	Fraunhofer	MIT
1	Number of publications	10,036	1,593	690	1,486
2	Evolution 2020 / 2017	-18%	-6%	-18%	-10%
3	Specialization index	1.2	1.0	1.9	1.3
4	Rate of international co-publications	65%	67%	37%	57%
5	First partner country	USA	USA	USA	CHN
6	Share of this country	23%	27%	23%	26%
7	MNCS	1.0	1.0	1.2	2.1
8	Activity index in the top10% most cited	0.9	1.0	1.2	2.3

Source: OST database, from the Web of Science.

Figure 30 shows the shares of the WoS categories that have the highest numbers of publications in this field, for the whole world and for each institution.

Figure 30. Engineering and systems: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

h. Mathematics

Table 15 compares the CNRS publications in mathematics with institutions in three countries: China, UK and the USA. The CNRS is the institution that publishes the most in mathematics in the sample, and it is clearly the most specialized in this discipline with an index of 2.8. CAS is not specialized at all in mathematics – which was already noted on Figure 19 in chapter 2; on the other hand, CAS is the institution with the largest increase in publications in this field between 2017 and 2020. The top 5 US universities also show growth, unlike the CNRS and the top 4 UK universities.

The share of international co-publications of the institutions are in line with the national averages: the countries with the most publications have a lower share of international co-publications. The first partner country is the USA for the non US institutions, and the group of the top 5 US universities has UK as their first partner.

The top 5 US universities have the highest impact indicators, both in terms of MNCS (1.9) and activity index in the decile of most cited publications (2.1). The top 4 UK universities have impact indicators 30% above the world average. CAS MNCS (1.1) is relativelow lower than its index in the decile of most cited publications (1.3). The CNRS has the lowest indices.

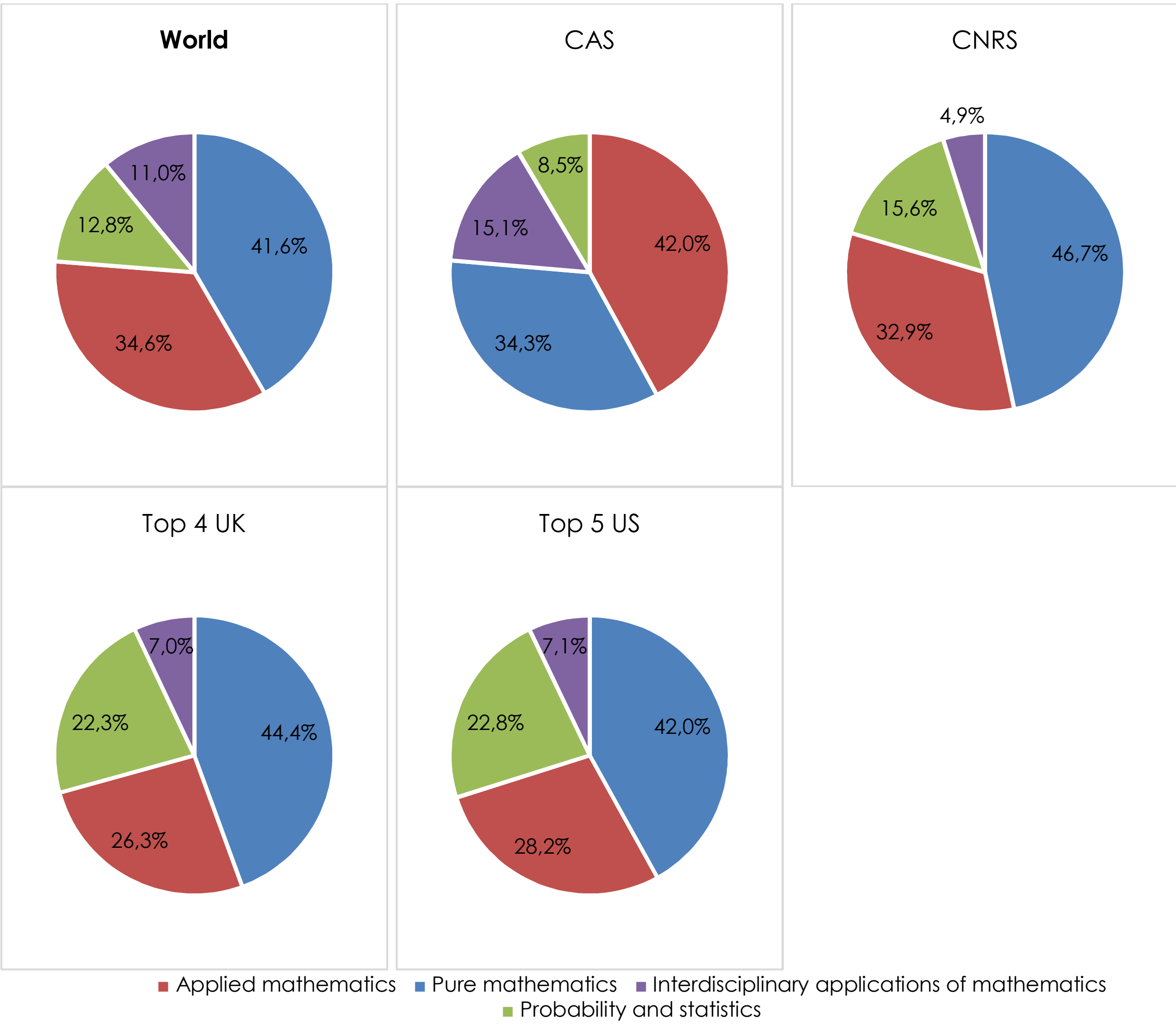
Table 15. Mathematics: international comparison of CNRS publications.

	Indicators	CAS	CNRS	Top4 UK	Top5 US
1	Number of publications	463	1,846	558	599
2	Evolution 2020 / 2017	+16%	-3%	-1%	+6%
3	Specialization index	0.5	2.8	1.4	1.4
4	Rate of international co-publications	31%	60%	64%	43%
5	First partner country	USA	USA	USA	GBR
6	Share of this country	41%	17%	34%	15%
7	MNCS	1.1	0.9	1.3	1.9
8	Activity index in the top10% most cited	1.3	0.9	1.3	2.1

Source: OST database, from the Web of Science.

This field aggregates 4 WoS categories: pure mathematics, applied mathematics, probability and statistics, interdisciplinary applications of mathematics. Figure 31 shows the shares of the four categories for each institution and for the whole world publications. CAS is the only institution having applied mathematics as the first publication category. US and UK universities have the highest shares in probability and statistics. The CNRS has the highest share in pure mathematics and the lowest share in interdisciplinary applications of mathematics.

Figure 31. Mathematics: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

i. Earth and space sciences

Table 16 compares the publications of the CNRS in the field of Earth and space sciences with three German institutions: DLR, Helmholtz and MPG. The numbers of publications are quite different from one institution to another: CNRS publishes the most in this field, ahead of Helmholtz, then MPG and DLR. However, DLR has recorded the highest growth between 2017 and 2020. DLR and Helmholtz are the most specialized institutions in the field with indices of 2.2 and 2.1 respectively; the specialization index is 1.8 for MPG and 1.6 for the CNRS.

MPG has a rate of international co-publications above 90%, while this rate is between 68 and 78% for the three other institutions. USA is the leading partner country for all four institutions.

All four institutions have impact indices above the world average. MPG has the highest indices, between 1.4 and 1.5; CNRS has the lowest, close to 1.1.

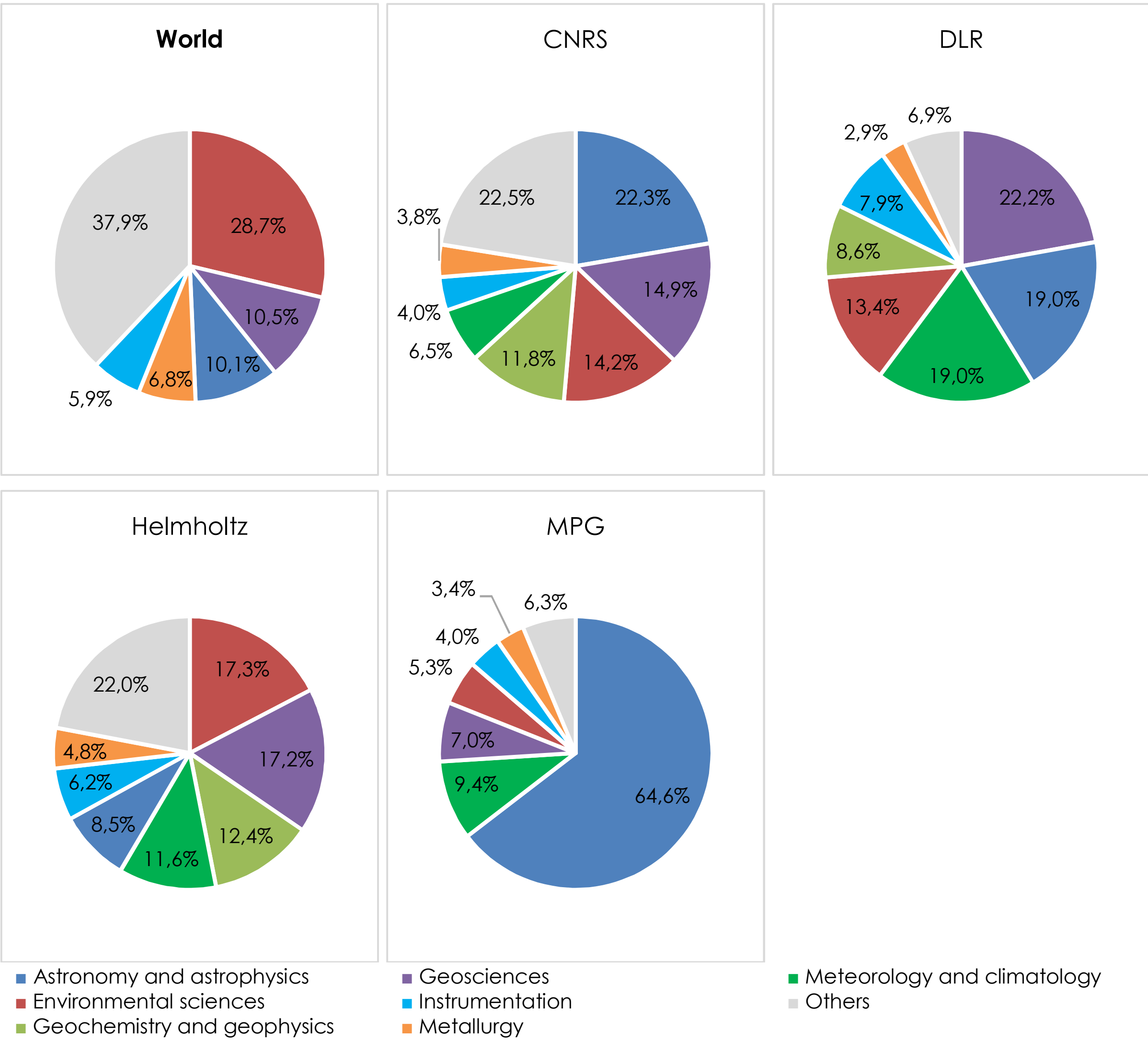
Table 16. Earth and space sciences: international comparison of CNRS publications.

	Indicators	CNRS	DLR	Helmholtz	MPG
1	Number of publications	3,292	143	1,125	531
2	Evolution 2020 / 2017	-1%	+8%	+1%	-5%
3	Specialization index	1.6	2.2	2.1	1.8
4	Rate of international co-publications	78%	68%	76%	91%
5	First partner country	USA	USA	USA	USA
6	Share of this country	35%	43%	35%	58%
7	MNCS	1.1	1.2	1.2	1.4
8	Activity index in the top10% most cited	1.1	1.3	1.3	1.5

Source: OST database, from the Web of Science.

Figure 32 shows the shares of the WoS categories that have the highest numbers of publications in this field, for the whole world and for each institution.

Figure 32. Earth and space sciences: top shares per WoS category for the world and for each institution.



Source: OST database, from the Web of Science.

j. Cross-fields analysis

Over the 9 fields considered, the CNRS is generally the institution that publishes the most, except for the fields of physics and chemistry where CAS publishes more and is more specialized than the CNRS. Since the CNRS has a much larger size than other institutions except CAS, it tends to publish more than the more specialized institutions – French or foreign – including in their fields of specialization.

A decrease in the number of CNRS publications between 2017 and 2020 can be observed in almost all fields, while several other institutions, except CAS, follow the same trend. The Ecology and environment field is an exception, corresponding to a dynamic field at the world level.

The rate of international co-publications evolves negatively with both the size of the country in which an institution is located and the applied nature of its research. From this point of view, the CNRS often has fairly high shares of international co-publications (above 60%), but lower than those of small, highly specialized organizations. However, MPG share of international co-publications is higher than that of the CNRS in all six fields where MPG is among the benchmarked institutions, except in Earth and space sciences where it is above 75% for both.

The CNRS impact indicators differ little between fields and are consistent with those observed in the previous chapter on the basis of the 11 disciplines (see Figure 20). CNRS impact index is below the world average in six of the nine fields; in three fields, Ecology and environment, Biology, Earth and space sciences, it is between the world average and 10% above. Compared to the other French institutions included in the comparisons, the CNRS impact indices are higher than those of the CEA, equivalent to those of INRAE and lower than those of Inria and Inserm. The CNRS impact indices are generally lower than those of the US, UK and German institutions included in the comparisons and they also tend to be lower than the CAS indices.

4. CNRS PATENT APPLICATIONS AND CO-APPLICATIONS

The OST patent database is built on the basis of the PatStat database from the European patent office (EPO). This database is supplemented with information from RegPat (Organization for economic co-operation and development) and from the French patent office INPI (*Institut national de la propriété industrielle*).[24] The analysis in this chapter is based on the spring 2022 version of PatStat

The PatStat database includes priority filings, as well as extension filings provided that the application has been published. The priority filing of a patent application is the first application filed to protect an invention with a patent office. The Paris Convention for the protection of industrial property, signed in 1883, provides a period of one year from the priority filing (priority date) for an applicant to extend the application to other countries that have signed the Convention. Each initial filing generates a family, which may consist of a single filing or of several patents filed in different patent offices. OST enriches the PatStat database and completes some missing information (inventors, applicants, technology fields) using information either from the family or from previous updates of the database. If the data cannot be completed, the application is not included in the final OST database and is ignored for the calculation of the indicators; this situation mainly occurs for Asian offices.

4.1. Priority applications and patent extensions

The CNRS has sent to OST its list of priority patents filed between 2012 and 2019 (initial filings). These data were analyzed by OST and entered into the OST database.

Box 3. Analysis of the data sent by the CNRS.

The initial list extracted from the CNRS patent database contained 5,187 priority filings after eliminating a few duplicates. The comparison with the OST database allowed to identify 4,804 priority filings. A more detailed analysis of these filings, particularly in terms of applicants, led to the inclusion of only 4,384 families in which the CNRS was identified as an applicant or co-applicant, *i.e.*, 85% of the initial list

Several factors may explain these discrepancies between the list transmitted by the CNRS and the patents that could be used for the analysis. Applications may not have been published and therefore not included in the spring 2022 version of PatStat; applications may have been withdrawn prior to publication by the office, and it may be uncertain whether the original filings can be traced; the application identification number may be inconsistent between the CNRS database and the OST database; the CNRS may appear neither as an applicant nor as a co-applicant in the OST database.

[24] The methodology is presented in Appendix F.

During the 2012 - 2019 period, the CNRS filed 4,384 priority applications. The number of priority applications varies between 499 and 578 per year with a maximum in 2015 (Table 17). The priority filings of the CNRS are quite stable over the period, the decrease observed in 2018-2019 is due to the fact that the version of the PatStat database used has not yet recorded all the priority filings and especially the extensions. Among these 4,384 families having the CNRS as an applicant, 3,643 have at least two filings in different offices; thus, around 83% of the CNRS priority filings are extended to at least another patent office. In total, there have been 13,780 extensions to priority filings between 2012 and 2019. The lower number of extensions in 2019 is also due to incomplete data.

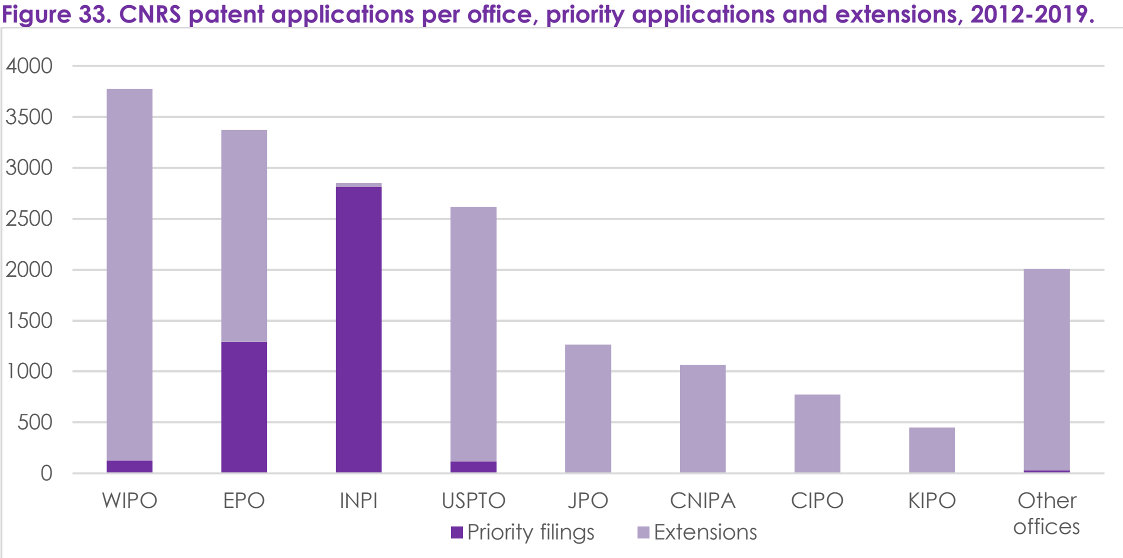
Table 17. CNRS priority patent applications, 2012-2019 and extensions, priority year.

	2012	2013	2014	2015	2016	2017	2018*	2019*	2012-2019
Priority filings	568	533	567	578	562	553	499	524	**4,384**
Extensions	2,048	2,125	2,090	2,057	1,815	1,802	1,342	501	**13,780**

** For 2018 and 2019, the version of PatStat that was used had not included all data, especially for extensions.*
Source: OST database, computed by OST using PatStat.

Figure 33 shows that filings are mainly with INPI (2,811) and EPO (1,292). The CNRS also has more than 100 filings at the WIPO (World intellectual property organization) and the USPTO (US patent and trademark office). Extensions are mainly made at WIPO (3,650 extensions), USPTO (2,800) and EPO (2,079). Other extensions are at the Japanese (JPO), Chinese (CNIAPA), Canadian (CIPO) and Korean (KIPO) offices.

Figure 33. CNRS patent applications per office, priority applications and extensions, 2012-2019.



Source: OST database, computed by OST using PatStat.

4.2. Technological profile of the CNRS patent filings

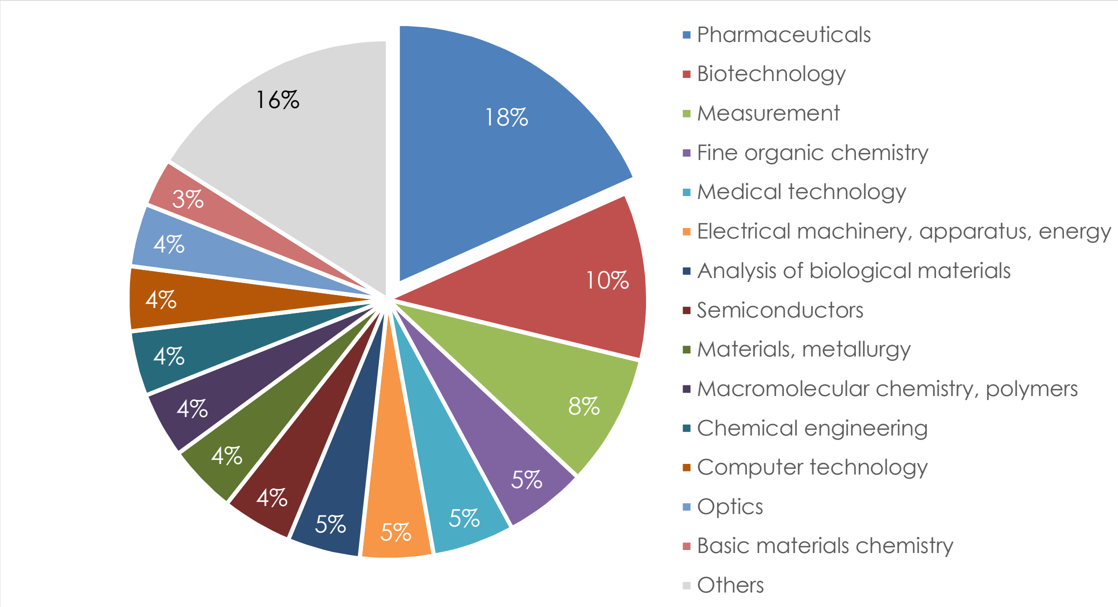
a. Distribution of the patent filings at the EPO and USPTO

This section analyses the distribution of the CNRS patent filings among the different technology fields. Only published filings for which information on technology fields and holders is available are taken into account. The analysis is carried out by year of filing.[25]

Figures 34a and 34b show the distribution of CNRS patent filings at the EPO and at the USPTO respectively. Whether at the EPO or at the USPTO, the technology field with the most filings is Pharmaceuticals, with nearly 20% of the total. The second field is Biotechnology, with 10% in Europe and 9% in the US. The third field, Measurement technology, is again the same in both offices with 8% and 7% respectively. The following fields remain mostly the same at both offices.

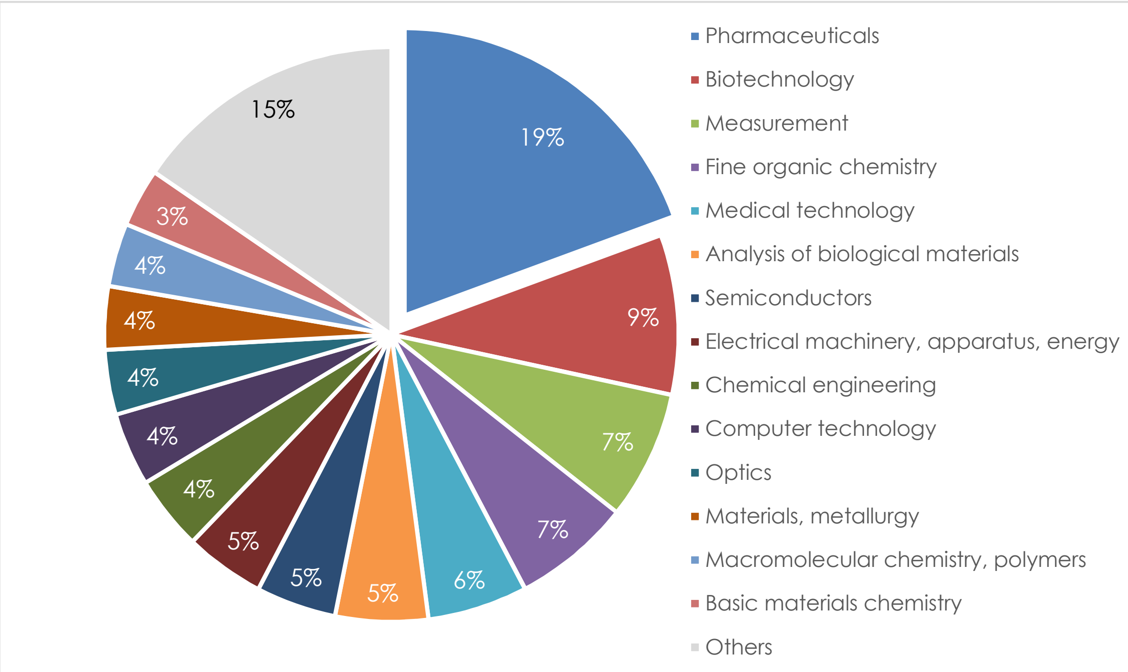
[25] Filings of extensions occur during the year following the first filing; the filings of the year 2012 mainly refer to the priority year 2011.

Figure 34a. Distribution of the CNRS patent filings at the EPO by technology field, 2012-2020.



Source: OST database, computed by OST using PatStat.

Figure 34b. Distribution of the CNRS patent filings at the USPTO by technology field, 2012-2020.



Source: OST database, computed by OST using PatStat.

The distribution of patents filings by technology fields were also calculated over two subperiods: 2014-2016 and 2017-2019. The results indicate, both at the EPO and the USPTO, a fairly high stability. The 14 fields with the highest number of filings remain the same over the two periods.

b. Comparison of the CNRS and France patent profiles

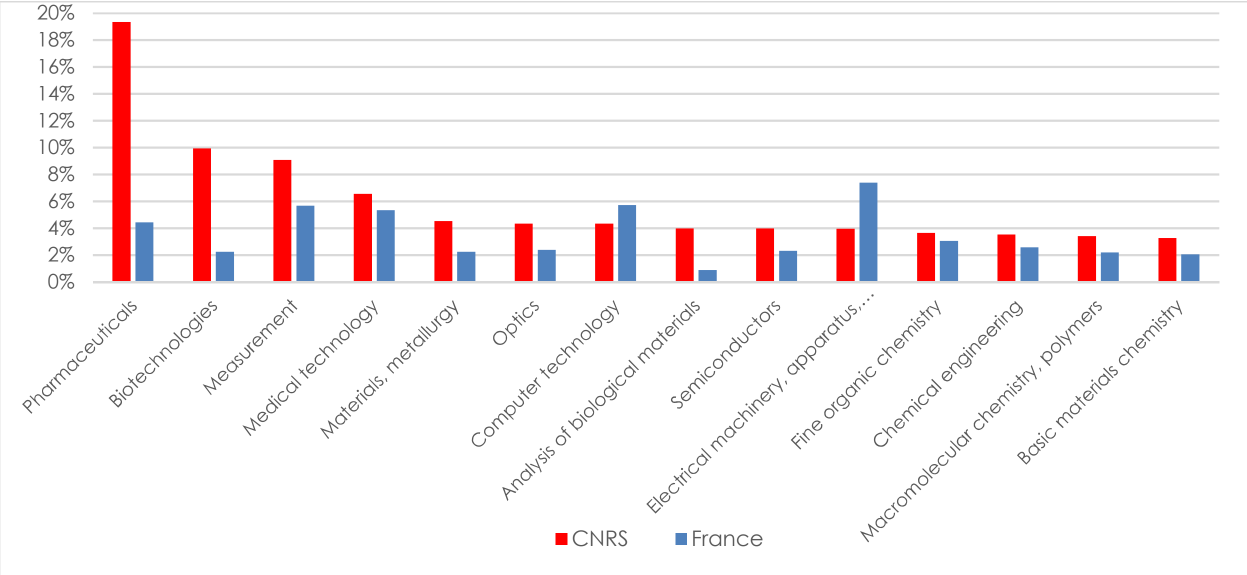
This section compares the CNRS patent profile describes above with France patent profile. The analysis is carried out for the 2017-2019 period.[26]

Figure 35 shows the distribution of the CNRS patent filings and of all French patent filings – i.e. all patent applications file or co-filed by a French public or private institution – at the EPO for the 14 major technology fields of the CNRS filings. The shares related to the first five technology fields – Pharmaceuticals, Biotechnology,

[26] Analysis of the full period 2012-2020 would have given very similar results.

Measurement technologies, Medical technology and Analysis of biological materials – are significantly higher for the CNRS than for France as a whole.

Figure 35. Distribution of CNRS and French EPO filings at EPO, CNRS main technology fields, 2017-2019.



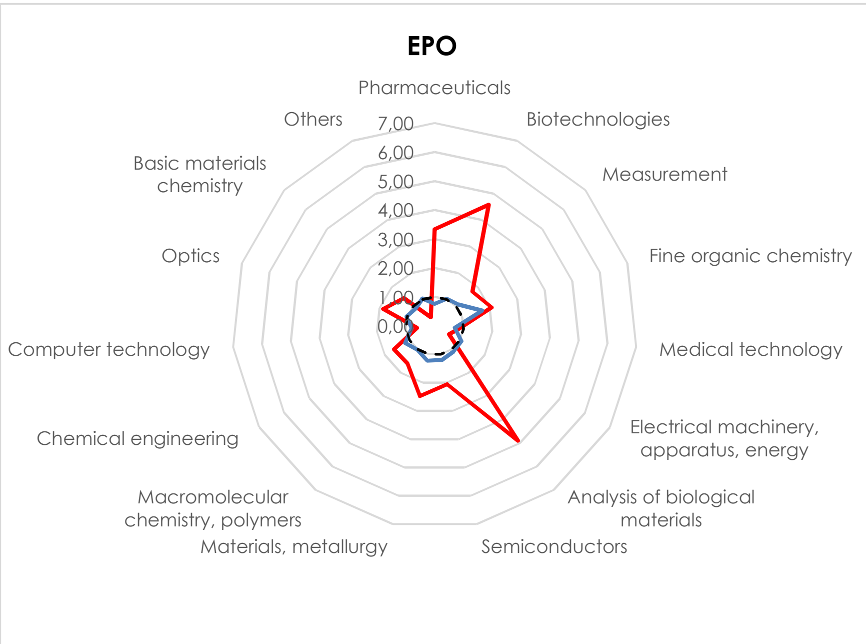
Source: OST database, computed by OST using PatStat.

Conversely, the fields of Electrical devices and energy and Computer science are more represented on average for the other French applicants than for the CNRS. In addition, patent filings related to the field of Transport do not appear in the first 14 technology fields for the CNRS, even though this is a field of specialization for France as a whole.[27]

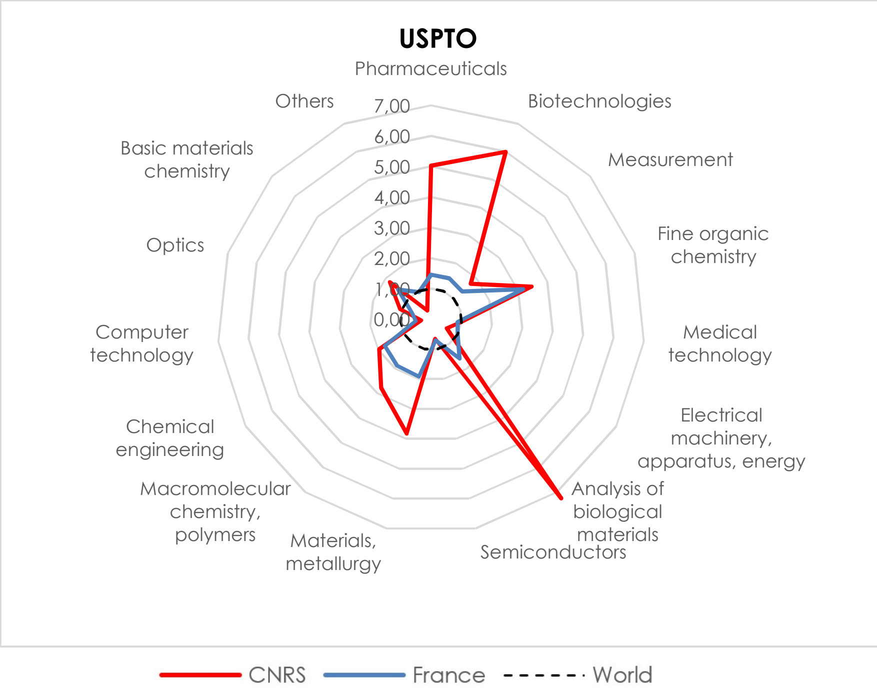
The same analysis has been carried out for the CNRS and French patent filings at the USPTO: it gives fairly similar results. The CNRS, like France as a whole, has a higher proportion of applications in Pharmaceutical at the USPTO than at the EPO. More generally, the first 14 technology fields of the CNRS concentrate a higher share of CNRS applications at the USPTO than at the EPO; this is also the case for France as a whole.

Figure 36 shows the CNRS and France's specialization index in the first 14 technology fields for the CNRS; this index measures the ratio of the CNRS share or the French share of the filings in a given technology field to the same share for all patent filings at the office (EPO or USPTO).

Figure 36. CNRS and France technological profiles at the EPO and USPTO, specialization index, 2017-2019.



[27] La position technologique de la France, 2022, in *Etat de l'enseignement supérieur, de la recherche et de l'innovation*, MESR, https://publication.enseignementsup-recherche.gouv.fr/eesr/FR/EESR15_R_33/la_position_technologique_de_la_france/



Source: OST database, computed by OST using PatStat.

The technology field with the highest specialization for the CNRS is Analysis of biological materials (index 5 at the EPO and index 7 at the USPTO), even if it is not the field with the highest number of filings. The Figure suggests that the fields of specialization of the CNRS and France are closer at the USPTO than at the EPO. The field of Transport, in which the CNRS does not file many applications whereas France is highly specialized at the two offices, does not appear explicitly; it is included in the "Other fields" group.

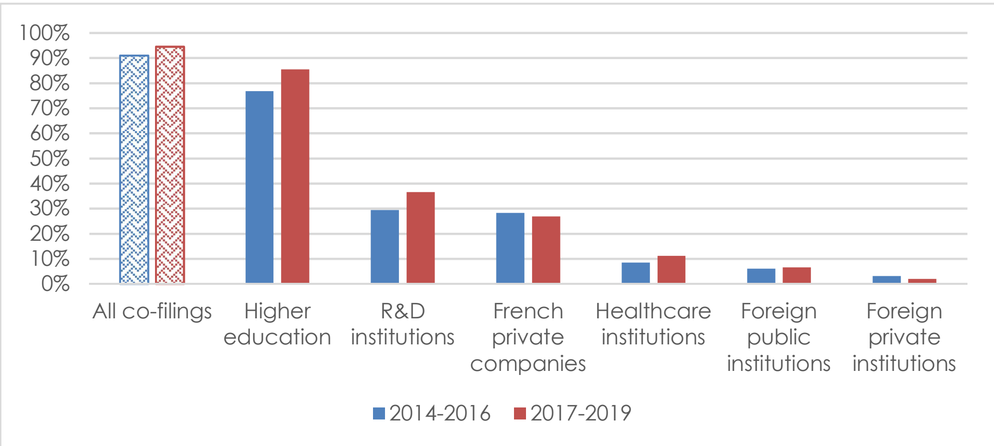
4.3. Patents co-applications

Between 2014-16 and 2017-19 the share of CNRS co-filing at the EPO has increased from 91% to 95%. CNRS share of co-filings, with one or several co-applicants tends to be even higher at the USPTO.

Figure 37 allows to analyze the typology of the partner co-applicants. OST has constructed a nomenclature to classify French applicants into institutional sectors: the “Higher education” category involves universities and *grandes écoles* while the “R&D institutions” category includes the national research organizations and private foundations such as *Institut Pasteur*. The data on foreign co-applicants is based on the PatStat nomenclature, which distinguishes public and private actors. Since co-filings can take place with more than one type of co-applicants, the analysis is carried out in full counting. As the results of the analysis show the same distribution of co-filings and the same dynamics for both the European and US offices, Figure 37 focuses on the EPO alone.

CNRS partners are mostly higher education institutions (80% over the period 2012-2020). R&D institutions participate in more than 30% of co-filings. Foreign public institutions participate in 7% of the co-filings and foreign private institutions in nearly 3%. Between 2014-16 and 2017-19 the shares of higher education, R&D and healthcare institutions in co-applications have increased, while that of French and foreign private institutions have decreased.

Figure 37. Typology of co-applicants of CNRS patents at the EPO, full counting, 2014-16 and 2017-2019.



Source: OST database, computed by OST using PatStat.

Table 18 gives the list of the main co-applicants. Inserm is the most important partner of CNRS for patent filing, notably due to the UMRs common to the CNRS and Inserm. The CEA is the second most important partner among R&D institutions and the fifth most important partner overall. However, most of the co-filings are made with universities, like Sorbonne University, City University of Paris and University of Montpellier in the top 5 institutions of the list. A health care institution, APHP (*Assistance publique des hôpitaux de Paris*), appears in the top 15 partner institutions. The first private company is Thales with 69 co-deposits, followed by TotalEnergies. The picture of initial co-applicants is very similar for USPTO filings.

Table 18. Top 25 co-applicants of CNRS at the EPO, 2012-2020.

Co-applicants	Number of co-applications	Share of co-applications (%)
Inserm	574	18.6
Sorbonne University	517	16.8
City University of Paris	321	10.4
University of Montpellier	248	8.0
CEA	247	8.0
Aix-Marseille University	210	6.8
Lyon 1 University	208	6.8
University of Bordeaux	204	6.6
Grenoble-Alps University	185	6.0
Strasbourg University	185	6.0
Université Paris-Saclay	179	5.8
Polytechnic institute of Bordeaux	127	4.1
Université Paul Sabatier Toulouse 3	103	3.3
University of Nantes	102	3.3
Assistance publique des hôpitaux de Paris (APHP)	100	3.2
University of Lille	99	3.2
École Supérieure de Physique et de Chimie Industrielles de Paris - PSL	98	3.2
INRAE	89	2.9
Institut Pasteur	73	2.4
Thales	69	2.2
École polytechnique	66	2.1
University of Rennes	63	2.0
TotalEnergies	56	1.8
PSL	55	1.8
Institut National des Sciences Appliquées de Lyon	54	1.8

Source: OST database, computed by OST using PatStat.

4.4. Grant rate of CNRS filings at the European patent office

Not every patent application results in a patent being granted. Some patents will never be granted, others will be abandoned during the process. In order to calculate a grant rate it is necessary to define "cohorts" of patents according to the year of filing and to specify a time window. For the period under study, only grant rates at 4 years and at 6 years after filing at the EPO can be calculated.

The 4-year CNRS patent grant rate is 23% for 2012-17 applications, equivalent to the average rate for all applications at the EPO (Table 19). The 6-year CNRS grant rate for 2012-15 applications is 46%, slightly higher than EPO average.

Table 19. Grant rate of CNRS patent applications at the EPO, 2012-2020, %.

Technology field	Grant rate at 4 years (%) (2012-2017 filings)		Grant rate at 6 years (%) (2012-15 filings)	
	CNRS	**EPO average**	**CNRS**	**EPO average**
Pharmaceuticals	16.5	12.3	40.8	38.1
Biotechnologies	15.2	12.5	42.4	39.6
Measurement techniques	28.8	23.1	53.8	43.7
Fine organic chemistry	24.4	23.9	46.7	49.9
Medical technology	23.1	20.8	47.0	43.8
Average for all fields	**24.1**	**23.5**	**45.9**	**44.2**

Source: OST database, computed by OST using PatStat.

Grant rates vary greatly across technology fields. For CNRS first three patenting fields, the grant rate after 4 years is between 15% and 29%, and the grand rate after 6 years is between 41% and 54%. For both time windows, the CNRS grant rate is higher than the EPO average. In measurement techniques, it is 10 percentage points above EPO average for the 6 year window. Table 19 also includes grant rates for CNRS next two filling fields, Fine organic chemistry and Medical technology.

Appendix A. Key figures on the French and foreign benchmark institutions

This Appendix gives some key figures on each of the French and foreign institutions that have been compared with CNRS in chapters 1 to 3 above. In alphabetic order: CAS, CEA, CNR, CNRS, CSIC, DLR, Fraunhofer, Helmholtz, INFN, INRAE, Inria, Inserm, IRD, Leibniz, MIT, MPG, STFC.

CAS: Chinese Academy of science (China)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields***	Number of HCR[28]	Annual budget (€ billion)***
1949	68,000	56,000	NA[29]	Chemistry Earth sciences Mathematics Physics Technology	204	22

* https://www.iybssd2022.org/en/dt_team/chinese-academy-of-sciences/
** https://golden.com/wiki/Chinese_Academy_of_Sciences-MX4JN
*** 2022 - https://cset.georgetown.edu/publication/chinese-academy-of-sciences-2022-budget/

CEA: Atomic energy and alternative energies commission (France)

Year of creation	Staff*	Number of researchers	Foreign researchers	Fields**	Number of HCR	Annual budget (€ billion)*
1945	19,000	8,300	NA	Defense and security Nuclear energy Renewable energy Technology for industry Health and life sciences Sciences of matter and Universe Climate and environment	3	5.5

* 2021: https://www.cea.fr/multimedia/Documents/publications/rapports/rapports-financiers/rapport-financier-cea-2021.pdf
** https://www.cea.fr/Pages/domaines-de-recherche.aspx

CNR: National research Council (Italy)

Year of creation*	Staff*	Number of researchers*	Foreign researchers	Fields**	Number of HCR	Annual budget (€ billion)*
1923	8,500	4,000	NA	Agricultural and food sciences Biomedical sciences Chemical sciences and materials Earth system science and environmental technologies Engineering, ICT, energy, and transportation technologies Humanities and social sciences, cultural heritage Physical sciences and technologies of matter	1	1

* 2021 - https://www.cnr.it/en/about-us#:~:text=This%20capital%20comprises%20more%20than,top%2Dpriority%20areas%20of%20interest.
** 2022 - https://www.cnr.it/en/cnr-in-figures .

[28] Highly cited researchers in 2022: https://clarivate.com/highly-cited-researchers/
[29] Not available.

CNRS: *Centre national de la recherche scientifique* (France)

Year of creation	Staff*	Number of researchers*	Foreign researchers**	Fields*	Number of HCR[30]	Annual budget (€ billion)*
1939	31,900 **Staff in *UMRs***** 109,800	11,400	18%	Nuclear physics and particle physics Chemistry Physics Information sciences Biology Humanities and social sciences Engineering and systems Mathematics Earth and space sciences	23	3.7

* Self-assessment report (SAR) of the CNRS, January 2023.
** Foreign scientists who have chosen France and the CNRS (in French), https://www.cnrs.fr/sites/default/files/download-file/DP%20scientifiques%20e%CC%81trangers%20CNRS.pdf

CSIC: National research Council (Spain)

Year of creation	Staff*	Number of researchers*	Foreign researchers*	Fields**	Number of HCR	Annual budget (€ billion)*
1939	13,330	4,345	4,5%	Agricultural sciences Biology Biomedicine Chemistry Science and food technology Human and social sciences Materials Natural resources Physics	18	0.8

* 2021 - https://www.csic.es/sites/www.csic.es/files/annual_report_csic_2021.pdf
**https://www.csic.es/en/research#:~:text=In%20order%20to%20achieve%20this,materials%2C%20natural%20resources%20and%20agricultural

DLR: *Deutches Zentrum für Luft- und Raumfahrt* (Germany)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields*	Number of HCR	Annual budget (€ billion)**
1969	8,000	4,700	NA	Aeronautics Digitalization Energy Security Space Transport	0	1.3

*2017 - https://www.dlr.de/content/en/downloads/publications/brochures/dlr-facts-and-figures.pdf?__blob=publicationFile&v=13
**2020 - https://www.dlr.de/DE/organisation-dlr/medien-und-dokumente/fakten/dlr-in-zahlen.html

[30] Number of CNRS HCR in 2022, as given in the CNRS self-assessment report.

Fraunhofer Gesellschaft (Germany)

Year of creation	Staff*	Number of researchers	Foreign researchers	Fields**	Number of HCR	Annual budget (€ billion)*
1949	22,000	NA	NA	Artificial intelligence Bioeconomy Digital healthcare Next generation computing Quantum technologies Resource efficiency and climate technologies Hydrogen technologies	4	2.9

* 2021 - file:///C:/Users/leravale/AppData/Local/Temp/MicrosoftEdgeDownloads/1bc503da-fc48-45e9-8066-136aeaaa4fe8/Fraunhofer-Annual-Report-2021.pdf
** https://www.research-in-germany.org/en/research-landscape/research-institutions/fraunhofer-gesellschaft.html

Helmholtz association of German research centers (Germany)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields*	Number of HCR	Annual budget (€ billion)*
1995	43,700	16,200	NA	Aeronautics, space, and transport Earth and environment Energy Health	15	5.35

*2021
https://www.helmholtz.de/system/user_upload/Ueber_uns/Wer_wir_sind/Zahlen_und_Fakten/2021/21_Jahresbericht_Helmholtz_Zahlen_Fakten_EN.pdf

INFN: National institute for nuclear physics (Italy)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields*	Number of HCR	Annual budget (€ billion)*
1946	8,200	2,000	NA	Aeronautics, space, and transport Earth and environment Energy Health	0	0.3

* 2013 - https://home.infn.it/images/materiale_istituzionale/brochure/INFN%20in%20numbers.pdf

INRAE: National research institute for agriculture, food and environment (France)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields**	Number of HCR	Annual budget (€ billion)*
1951	2,000	600	NA	Agroecology Biodiversity Bioeconomy Climate change and risks Food, global health	8	1

*2021 – https ://www.inrae.fr/sites/default/files/pdf/Presentation-INRAE2021-FrBdef.pdf
** https://www.inrae.fr/en

Inria: National research institute in digital science and technology (France)

Year of creation	Staff*	Number of researchers*	Foreign researchers*	Fields*	Number of HCR	Annual budget (€ billion)*
1967	2,700	1,500	21%	Applied mathematics Computer science	1	0,3

* 2020 - https://www.inria.fr/sites/default/files/2022-03/Livret-synthese-RSU-2020.pdf
** 2021 - https://www.inria.fr/sites/default/files/2022-07/Rapport%20d%27activites%20Inria%202021.pdf

Inserm: National institute for health and medical research (France)

Year of creation	Staff*	Number of researchers*	Foreign researchers*	Fields*	Number of HCR	Annual budget (€ billion)*
1964	5,100	2,200	15%	Biomedical research and human health	15	1.1

* 2021 annual report - https://www.inserm.fr/nous-connaitre/documents-strategiques/

IRD: National research institute for development (France)

Year of creation	Staff*	Number of researchers*	Foreign researchers	Fields*	Number of HCR	Annual budget (€ billion)*
1944	2,200	900	NA	Earth sciences Ecology Geography Health and public health Oceanography Sociology	1	0,25

* 2021 – Annual report, https://www.ird.fr/rapport-dactivites-2021-face-aux-defis-globaux-actuels

Leibniz association of German research centers (Germany)

Year of creation*	Staff**	Number of researchers**	Foreign researchers**	Fields*	Number of HCR	Annual budget (€ billion)**
1990	21,000	12,000	27%	Economics, social sciences, spatial research Humanities and education Life sciences Mathematics, natural sciences, engineering Environmental sciences	7	2,1

* https://en.wikipedia.org/wiki/Leibniz_Association
**2021 - https://www.leibniz-gemeinschaft.de/en/about-us/organisation/leibniz-in-figures

MIT: Massachusetts Institute of technology (USA)

Year of creation*	Staff*	Number of researchers*	Foreign researchers	Fields**	Number of HCR	Annual budget (€ billion)***
1861	16,300	4,000	NA	Architecture and planning Engineering Humanities, arts, and social sciences Management Science Computing	86	4.2

* 2022 - https://facts.mit.edu/employees/
** https://facts.mit.edu/enrollment-statistics/
*** 2022, report of the treasurer - https://vpf.mit.edu/sites/default/files/downloads/TreasurersReport/MITTreasurersReport2022.pdf

MPG: Max Planck *Gesellschaft* (Germany)

Year of creation*	Staff*	Number of researchers*	Foreign researchers*	Fields**	Number of HCR	Annual budget (€ billion)*
1948	20,900	6,700	55%	Astronomy and astrophysics Biology and medicine Environment and climate Humanities Material and technology	75	2.6

* 2021 - Annual report 2021, https://www.mpg.de/18802868/mpg-annual-report-2021.pdf.
** https://www.mpg.de/institutes?tab=institutes

STFC: Science and technology facilities Council (UK)

Year of creation*	Staff*	Number of researchers	Foreign researchers	Fields*	Number of HCR	Annual budget (€ billion)**
2007	1,700	NA	NA	Particle physics Astronomy and space science Nuclear physics		0.9

*2021- https://www.ukri.org/about-us/stfc/who-we-are/
** 2021 - Annual report and accounts 2021-22, https://www.ukri.org/wp-content/uploads/2023/01/UKRI-180123-AnnualReportandAccounts-21to22.pdf

Appendix B. Data and methodology for the analysis of European projects

Sources

The analysis is based on a complete version of the e-corda database, *i.e.*, both proposals and grants.[31]

As the 2017-2021 period considered for the CNRS assessment covers two research and innovation framework programs (end of Horizon 2020, beginning of Horizon Europe), two versions of the database were used:

- Version of 04/10/2021 for H2020 data (planned to last from 2014 to 2020, but ended in 2021),
- Version of 05/08/2022 regarding Horizon Europe data (started with a few months delay in 2021 and will last until 2027).

The 2021 data correspond to the sum of the H2020 and Horizon Europe proposals and grants for year 2021.

As the structure of the framework programs has evolved, everything cannot be compared over the entire period, especially at the pillar level. Nevertheless, the review of the ERC grants can be done over 2017 to 2021 since this component exists in the two successive programs (with however a change in the panel classification).

Attribution of projects to institutions

The corpus of projects of each organization was determined with the PIC code (Participant Identification Code).

In principle, the PIC code is a unique identifier for each legal entity. Given the legal structure of some G6 organizations, taking into account their scope requires the consolidation of several PIC codes. For the German organizations, OST has consolidated 38 PIC codes for Leibniz, 18 for Helmholtz and 3 for Max Planck. Table B1 provides a list of the codes used to identify each research organization.

For ERC, as a rule, the principal investigator (PI) of an application is employed by the host institution (HI), but there may be cases where the PI's employer is a third party.[32]

Counting projects by year

Projects have been assigned to the closing year of the corresponding call for proposals, in compliance with the methodology used in France for the indicators on EU research and innovation projects that are computed each year for the government and the parliament. Cancelled projects have not been taken into account.

Filters

A project is only counted for an institution if it appears in the list of participants as a beneficiary (and not if it is there as a third party). This choice has more consequences for the CNRS or CNR than for the other institutions considered in the analysis.

For reasons of homogeneity, the calculation of the success rate is carried out on the basis of "proposals", by filtering in the numerator the submitted proposals retained as "eligible" (formal criteria) and retained in the "mainlist", which is very close to the list of projects finally retained.

The financial amounts analyzed correspond to the net contribution received by the actors, i.e. deducted from what they paid to other actors.

---

[31] This complete version of the database is sent to OST by the French ministry in charge of research, to which the European Commission provides access.
[32] See the Annotated grant agreement (draft version as of April 2022): [aga_en.pdf (europa.eu)](aga_en.pdf) .

Table B1: List of PIC codes considered for each of the G6 institutions.

GENERAL_PIC	PARTICIPANT_LEGAL_NAME	pris en compte pour ..
999979500	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR
999997930	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	CNRS
999991722	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	CSIC
990797674	KARLSRUHER INSTITUT FUER TECHNOLOGIE	Helmholtz
999981731	DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV	
999980470	FORSCHUNGSZENTRUM JULICH GMBH	
999994632	HELMHOLTZ-ZENTRUM FUR UMWELTFORSCHUNG GMBH - UFZ	
999994341	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ	
999994729	HELMHOLTZ ZENTRUM MUENCHEN DEUTSCHES FORSCHUNGSZENTRUM FUER GESUNDHEIT UND UMWELT GMBH	
999990073	DEUTSCHES KREBSFORSCHUNGSZENTRUM HEIDELBERG	
999497507	ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG	
999986969	DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	
999507401	HELMHOLTZ-ZENTRUM HEREON GMBH	
974626416	DEUTSCHES ZENTRUM FUR NEURODEGENERATIVE ERKRANKUNGEN EV	
999470541	HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF EV	
986090458	HELMHOLTZ-ZENTRUM FUR OZEANFORSCHUNG KIEL (GEOMAR)	
999470347	HELMHOLTZ-ZENTRUM FUR INFEKTIONSFORSCHUNG GMBH	
999990461	MAX DELBRUECK CENTRUM FUER MOLEKULARE MEDIZIN IN DER HELMHOLTZ-GEMEINSCHAFT (MDC)	
999446000	HELMHOLTZ-ZENTRUM BERLIN FUR MATERIALIEN UND ENERGIE GMBH	
907336546	CISPA - HELMHOLTZ-ZENTRUM FUR INFORMATIONSSICHERHEIT GGMBH	
999995214	GSI HELMHOLTZZENTRUM FUR SCHWERIONENFORSCHUNG GMBH	
999464042	POTSDAM-INSTITUT FUR KLIMAFOLGENFORSCHUNG EV	Leibniz
999465885	LEIBNIZ-ZENTRUM FUER AGRARLANDSCHAFTSFORSCHUNG (ZALF) e.V.	
998028636	INSTITUT FUER WELTWIRTSCHAFT	
999468795	BERNHARD-NOCHT-INSTITUT FUER TROPENMEDIZIN	
997770131	DEUTSCHES PRIMATENZENTRUM GMBH	
999488486	LEIBNIZ-INSTITUT FUR VERBUNDWERKSTOFFE GMBH	
998615874	DWI LEIBNIZ-INSTITUT FUR INTERAKTIVE MATERIALIEN EV	
999619048	LEIBNIZ-INSTITUT FUER PHOTONISCHE TECHNOLOGIEN E.V.	
998401795	LEIBNIZ-INSTITUT FUR AGRARTECHNIK UND BIOOKONOMIE EV	
891893855	FERDINAND-BRAUN-INSTITUT GGMBH LEIBNIZ- INSTITUT FUR HOCHSTFREQUENZTECNIK	
999483442	LEIBNIZ - INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG	
989201927	Deutsches Rheuma-Forschungszentrum Berlin	
997836285	SENCKENBERG GESELLSCHAFT FUR NATURFORSCHUNG	
996854451	MUSEUM FUR NATURKUNDE - LEIBNIZ-INSTITUT FUR EVOLUTIONS- UND BIODIVERSITATSFORSCHUNG AN DER HUMBOLDT-UNIVERSITAT ZU BERLIN	
999562400	GESIS-LEIBNIZ-INSTITUT FUR SOZIALWISSENSCHAFTEN EV	
999451141	LEIBNIZ-INSTITUT FUR POLYMERFORSCHUNG DRESDEN EV	
995654076	FIZ KARLSRUHE - LEIBNIZ-INSTITUT FUR INFORMATIONSINFRASTRUKTUR GMBH	
999517295	ZEW - LEIBNIZ-ZENTRUM FUR EUROPAISCHE WIRTSCHAFTSFORSCHUNG GMBH MANNHEIM	
889869368	SCHLOSS DAGSTUHL - LEIBNIZ ZENTRUM FUR INFORMATIK GMBH	
997581272	DEUTSCHES INSTITUT FUR ERWACHSENENBILDUNG EV	
900611536	LEIBNIZ-INSTITUT FUR RESILIENZFORSCHUNG (LIR) GGMBH	
998962746	LEIBNIZ-INSTITUT FUR PLASMAFORSCHUNG UND TECHNOLOGIE EV	
999606438	IHP GMBH - INNOVATIONS FOR HIGH PERFORMANCE MICROELECTRONICS/LEIBNIZ-INSTITUT FUER INNOVATIVE MIKROELEKTRONIK	
998707054	IUF - LEIBNIZ-INSTITUT FUR UMWELTMEDIZINISCHE FORSCHUNG GMBH	
999514288	LEIBNIZ-INSTITUT FUER AGRARENTWICKLUNG IN TRANSFORMATIONSOEKONOMIEN (IAMO)	
998672231	LEIBNIZ-INSTITUT FUR LANDERKUNDE EV	
999544746	LEIBNIZ INSTITUT FUR FESTKORPER UND WERKSTOFFORSCHUNG DRESDEN EV	
972715807	RWI - LEIBNIZ-INSTITUT FUR WIRTSCHAFTSFORSCHUNG e.V.	
990193946	LEIBNIZ-INSTITUT FUR ALTERNSFORSCHUNG - FRITZ-LIPMANN-INSTITUT EV (FLI) LEIBNIZ INSTITUTE ON AGING - FRITZ LIPMANN INSTITUTE EV (FLI)	
999470056	LEIBNIZ-INSTITUT FUR ANALYTISCHE WISSENSCHAFTEN-ISAS-EV	
972468360	DEUTSCHE ZENTRALBIBLIOTHEK FUER WIRTSCHAFTSWISSENSCHAFTEN - LEIBNIZ- INFORMATIONSZENTRUM WIRTSCHAFT	
999450850	LEIBNIZ INSTITUT FUER TROPOSPHAERENFORSCHUNG e.V.	
999502357	WISSENSCHAFTSZENTRUM BERLIN FUR SOZIALFORSCHUNG GGMBH	
986304925	Zentrum für Zeithistorische Forschung Potsdam	
999482472	LEIBNIZ-INSTITUT FUR DIE PADAGOGIKDER NATURWISSENSCHAFTEN UND MATHEMATIK AN DER UNIVERSITAT KIEL	
998692601	LEIBNIZ-INSTITUT FUR ASTROPHYSIK POTSDAM (AIP)	
998595989	LEIBNIZ-INSTITUT DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GMBH	
991028049	HEINRICH-PETTE INSTITUT LEIBNIZ INSTITUT FUER EXPERIMENTELLE VIROLOGIE	
999990267	MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV	Max_Planck
998816761	MAX PLANCK INSTITUT FUER KOHLENFORSCHUNG	
998816858	MAX PLANCK INSTITUT FUR EISENFORSCHUNG GMBH	

Appendix C. Data and methodology for the analysis of CNRS publications

C1. Data and definition of indicators

The OST publication database

The analyses presented in chapters 2 and 3 are based on data from the OST publication database, which is an enriched version of the Web of Science (WoS) from Clarivate Analytics. In particular, the OST database includes additional data on nomenclatures (geographical and thematic) and institutional affiliations.

The WoS indexes the most influential scientific journals and conference proceedings at the international level. Its coverage is better for well internationalized disciplines than for some applied disciplines, disciplines with a strong national tradition, or disciplines the size of the community of which is small. For example, the database does not cover as well various humanities and social sciences disciplines in some non-English-speaking countries as it covers life and material sciences disciplines. Nevertheless, the coverage of the database is evolving and new journals are added each year; the coverage of refereed conference proceedings has also been substantially improved in the last years, in particular for disciplines where some refereed conference proceedings are considered important.

Counting of publications and calculation of indicators

The scope of the OST publication database includes the following indexes from the WoS: SCI (Science citation index expanded), SSCIS (Social sciences citation index), A&HCI (Arts & humanities citation index), CPCI (Conference proceedings citation index). Indicators are calculated by retaining only certain types of documents: original articles (including those from conference proceedings) and review articles. Documents for which some information is missing (research fields, country code, etc.) are not taken into account.

The attribution of CNRS and CSIC publications has been validated by these institutions. The attribution of MPG publications was done in partnership with the Research system and science dynamics research team from the German Centre for higher education research and science studies *(Deutsches Zentrum für Hochschul- und Wissenschaftsforschung*: *DZHW)*. For other foreign institutions, OST used the "organization enhanced" information from the WoS, checking their quality with Clarivate Analytics.

Apart from the co-publication and open access indicators, which use full counting, the indicators are calculated in fractional counting, which measures the "contributions": a publication is fractionalized in proportion to its number of affiliation addresses. As a robustness check, author fractionalization has been tested for CNRS and Max Planck corpora. The average impact indicators calculated this way does not reveal substantial differences with the address fractionalization used in the report.

If publications are indexed in more than a WoS category, they are fractionalized over categories (hence, in some cases over disciplines or sub-domains for ERC classification).

The **number of publications** gives the volume of production for an institution or country at a given level of the classification and for a given period. The year 2021 is incomplete in the version of the database (updated in July 2022) that was used for this report.

The **national share of publications** of an institution is defined as the ratio of its number of publications to the number of publications published in its country.

The scientific specialization of an institution in a discipline is defined by the share of the discipline in the institution's publications, divided by the share of the discipline in world publications. The higher the **specialization index** is above 1 (neutral value), the more the institution is considered "specialized" in the discipline.

For an institution, the **open access index** is the share of its publications in open access compared to the same share in the world at the disciplinary level. Open access publications are identified using information from the WoS and Unpaywall databases. The different types of open access are taken from these sources:

- gold open access publications are publications in an open access journal;
- hybrid open access publications are publications in a subscription journal that are open access with a license that allows the publication to be reused;
- bronze open access publications are publications in a subscription journal that are open access without a license that allows the publication to be reused;
- green open access publications are publications in a subscription journal that are open access not in the journal itself but in a repository.

A **co-publication** is a publication with at least two different affiliation addresses. For an institution, the **share of co-publications** is defined by the number of its co-publications compared to its total number of publications. A co-publication is national when all the authors' addresses are in the same country. A co-publication is international when it is signed by at least one author with an address in another country. The share of international co-publications is defined by the number of publications of an institution that are signed by at least one author with a foreign address, compared to the total number of publications of the institution.

All indicators related to **citations**, such as the **impact index**, are calculated with an open window. Normalization is made by year of publication.

The average impact indicator is **the Mean normalized citation score (MNCS)**. The citation number of each publication is normalized by WoS category, type of document and year of publication. The MNCS of an institution is then the mean of the normalized scores. If the MNCS is 1, the publications of the institution receive as many citations as the world average for similar publications. The MNCS is computed on the basis of fractional counting in order to avoid disciplinary biases and inconsistent results.[33]

The distribution of the publications is done in **citation classes** defined at the world level. They correspond to the breakdown of all publications into decreasing percentiles according to the number of citations received at the world level, for instance, one considers the class of the 5% most cited publications in the world, then the class of the next 5%, etc. The citation classes are disjoint.

The **activity index** for each citation class is equal to the ratio of the institution's share of publications in the class to the share of world publications in that class. By construction, the average value of the activity index is equal to 1 in each class.

Note: The indicators included in this report may or may not be size-dependent. Size-dependent indicators are those obtained from the absolute number of publications, while size-independent indicators are obtained by calculating the share of an institution's publications in a particular area or discipline. For example, the national share of publications and the number of highly cited publications of an institution are size-dependent indicators. The specialization index or the impact index are independent of size: these indicators, standardized by the same proportions worldwide, make it possible to position institutions or countries regardless of their size.

Nomenclatures

The classification into eleven disciplines in the OST database is the result of an aggregation of the 254 WoS 'categories'. The 11 disciplines are as follows:

Applied biology and ecology,	Humanities
Chemistry	Mathematics
Computer science	Medical research
Earth and Universe sciences,	Physics
Engineering sciences	Social sciences
Fundamental biology	

For example, the discipline "Mathematics" is the aggregation of four categories: mathematics, applied mathematics, probability and statistics, interdisciplinary applications of mathematics. Other examples of the WoS-categories are given in Appendix C2.

Disciplinary journals are linked to one or to several categories, and thus, to one or several of the broad disciplines; all articles published in these journals are then all linked to the category if the journal is linked to a sole category, or linked (in fractional count) to the different categories that the journal is linked to. Articles published in multi-disciplinary journals – such as Nature, PNAS (proceedings of the US National Academy of science) and Science – are individually linked to one or several categories according to their subjects; OST does this using an algorithm provided by Clarivate Analytics.

A focus on the Wos coverage for conference proceedings in computer science

The self-assessment report (SAR) of the CNRS highlights that, in the field of computer science, publications in proceedings of international conferences, with reviewing committees, are sometimes more important than publications in scientific journals, and it questions the coverage of the WoS database for the conference proceeedings.

To analyze this issue, OST has investigated the WoS coverage of the so-called "A and A* conferences" identified on the CORE portal (http://portal.core.edu.au/conf-ranks/). The analysis was limited to A and A* conferences and to the "Field of research" of Group 46 "Information and computing sciences" of the ANZSRC-2020 for years 2020 and 2021. 246 conferences meet these criteria, and the search in the WoS database allowed to retrieve papers for 226 of these 246 conferences, i.e. 92%. For some of these conferences, no papers were found in the WoS for certain editions, which may be due to the conference not being held that year, or not producing adequate proceedings, or to papers not being indexed in the WoS database. The list of the 20 conferences that brought back no papers in the WoS over the period 2017-2021 is provided below.

---

[33] On this issue, see Waltman, L., & Van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4),872-894.

Table C1: List of A and A* CORE-conferences in Information and computing sciences not found in the WoS, 2017-2021.

ACM CONFERENCE ON OBJECT ORIENTED PROGRAMMING SYSTEMS LANGUAGES AND APPLICATIONS
ANNUAL MEETING OF THE COGNITIVE SCIENCE SOCIETY
NORTH AMERICAN ASSOCIATION FOR COMPUTATIONAL LINGUISTICS
PRIVACY ENHANCING TECHNOLOGIES SYMPOSIUM (WAS INTERNATIONAL WORKSHOP OF PRIVACY ENHANCING TECHNOLOGIES)
EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING
INTERNATIONAL CONFERENCE ON EMBEDDED WIRELESS SYSTEMS AND NETWORKS (WASEUROPEAN CONFERENCE ON WIRELESS SENSOR NETWORKS)
MEASUREMENT AND MODELING OF COMPUTER SYSTEMS
UNIX SYMPOSIUM ON INTERNET TECHNOLOGIES
IEEE INTERNATIONAL SYMPOSIUM ON ARTIFICIAL LIFE
LOGIC PROGRAMMING AND AUTOMATED REASONING
CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING
THEORETICAL ASPECTS OF RATIONALITY AND KNOWLEDGE
EUROMICRO CONFERENCE ON REAL-TIME SYSTEMS
INTERNATIONAL WORKSHOP ON APPROXIMATION ALGORITHMS FOR COMBINATORIAL OPTIMIZATION PROBLEMS
COMPUTER SUPPORTED COLLABORATIVE LEARNING
INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS
WORKSHOP ON ALGORITHM ENGINEERING AND EXPERIMENTS
ADVANCES IN MODAL LOGIC
CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH
INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS

C2. The 9 fields analyzed in chapter 3

Each of the 9 scientific field mentioned in chapter 3 is constructed as the aggregation of several WoS-categories. This Appendix gives the detailed list of categories grouped in each of the 9 fields. For each category, the corresponding discipline is also indicated, which highlights the multidisciplinary character of some of the fields.

Nuclear physics and particle physics

This field aggregates 2 WoS-categories.

Categories		Discipline
Physics, nuclear	Physics, particles and fields	Physics

Chemistry

This field aggregates 29 WoS-categories.

Categories		Disciplines
Chemistry, analytical Chemistry, applied Chemistry, inorganic and nuclear Chemistry, medicinal Chemistry, multidisciplinary Chemistry, organic Chemistry, physical Crystallography	Electrochemistry Materials science, ceramics Materials science, coating and films Materials science, composites Materials science, paper and wood Materials science, textiles Nanoscience and nanotechnology Polymer science	Chemistry
Geochemistry and geophysics		Earth and Universe sciences
Engineering, chemical Green and sustainable science and technology Metallurgy and Metallurgical engineering	Microscopy Thermodynamics	Engineering sciences
Biochemistry and molecular biology Biophysics	Materials science, biomaterials	Fundamental biology
Pharmacology and pharmacy	Toxicology	Medical research
Physics, fluids and plasmas	Spectroscopy	Physics

Ecology and environment

This field aggregates 21 WoS-categories.

Categories		Disciplines
Biodiversity conservation Ecology Entomology	Ornithology Soil science Zoology	Applied biology and ecology
Marine and freshwater biology Geography, physical Limnology	Paleontology Water resources	Earth and Universe sciences
Anatomy and morphology Environmental sciences Evolutionary biology	Genetics and heredity Microbiology Parasitology	Fundamental biology
Anthropology	Archaeology	Humanities
Infectious diseases		Medical research
Environmental studies		Social sciences

Physics

This field aggregates 20 WoS-categories.

Categories		Disciplines
Crystallography Materials science, multidisciplinary	Nanoscience and nanotechnology	Chemistry
Microscopy Nuclear science and technology	Thermodynamics	Engineering sciences
Biophysics		Fundamental biology
Acoustics Instruments and instrumentation Optics Physics, atomic, molecular and chemical Physics, applied Physics, condensed matter Physics, fluids and plasmas	Physics, mathematical Physics, multidisciplinary Physics, nuclear Physics, particles and fields Quantum science and technology Spectroscopy	Physics

Information sciences

This field aggregates 16 WoS-categories.

Categories		Disciplines
Computer science, artificial intelligence Computer science, cybernetics Computer science, hardware and architecture Computer science, information systems Computer science, interdisciplinary applications Computer science, software engineering	Computer science, theory and methods Logic Medical informatics Robotics Telecommunications	Computer science
Automation and control systems	Operations research and management science	Engineering
Mathematical and computational biology Neuro-imaging	Neurosciences	Fundamental biology

Biology

This field aggregates 28 WoS-categories.

Categories		Disciplines
Biology	Plant sciences	Applied biology and ecology
Marine and freshwater biology		Earth and Universe sciences
Behavioral sciences Biochemistry and molecular biology Biophysics Cell and tissue engineering Cell biology Developmental biology Evolutionary biology Genetics and heredity Mathematical and computational biology	Materials science, biomaterials Microbiology Neuro-imaging Neurosciences Parasitology Physiology Reproductive biology Virology	Fundamental biology
Endocrinology and metabolism Cardiac and cardiovascular systems Hematology Immunology	Infectious diseases Oncology Psychiatry Substance abuse	Medical research

Engineering and systems

This field aggregates 59 WoS-categories.

Categories		Disciplines
Biotechnology and applied microbiology	Food science and technology	Applied biology and ecology
Chemistry, physical Electrochemistry Materials science, ceramics Materials science, characterization and testing Materials science, coatings and films	Materials science, composites Materials science, paper and wood Materials science, multidisciplinary Materials science, textiles Polymer science	Chemistry
Computer science, artificial intelligence Robotics	Telecommunications	Computer science
Astronomy and astrophysics Engineering, environmental Geosciences, multidisciplinary	Oceanography Water resources	Earth and Universe sciences
Automation and control systems Construction and building technology Energy and fuels Engineering, aerospace Engineering, biomedical Engineering, chemical Engineering, civil Engineering, electrical and electronic Engineering, industrial Engineering, marine Engineering, mechanical	Engineering, multidisciplinary Engineering, ocean Engineering, petroleum Green and sustainable science and technology Imaging science and photographic technology Mechanics Metallurgy and metallurgical engineering Nuclear science and technology Remote sensing Transportation science and technology	Engineering sciences
Cell and tissue engineering Materials science, biomaterials	Neuro-imaging	Fundamental biology
Mathematics, applied	Mathematics, interdisciplinary applications	Mathematics
Audiology and speech-language technology Medicine, research and experimental Primary health care	Radiology, nuclear medicine and medical imaging Sport sciences	Medical research
Acoustics Instruments and instrumentation Optics Physics, condensed matter	Physics, fluids and plasmas Physics, nuclear Spectroscopy	Physics
Social issues		Social sciences

Mathematics

This field aggregates 4 WoS-categories.

Categories		Discipline
Mathematics Mathematics, applied	Mathematics, interdisciplinary applications Statistics and probability	Mathematics

Earth and space sciences

This field aggregates 18 WoS-categories.

Categories		Disciplines
Astronomy and astrophysics Engineering, environmental Engineering, geological Environmental sciences Geochemistry and geophysics Geology Geography, physical Geosciences, multidisciplinary	Limnology Marine and freshwater biology Meteorology and atmospheric sciences Mineralogy Oceanography Paleontology Water resources	Earth and Universe sciences
Metallurgy and metallurgical engineering	Mining and mineral processing	Engineering sciences
Instruments and instrumentation		Physics

Appendix D. Analysis of CNRS scientific productions with two data sources

This appendix compares the typology and disciplinary distribution of the CNRS scientific production on the basis of two data sources, the French open access archive HAL and the OST publication database, an in house version of the Web of Science (see Appendix C).

The CNRS SAR considers that the Web of Science is not adequate to analyze publications in the SSH and computer science disciplines. As a consequence, it includes an appendix on CNRS productions in SSH disciplines on the basis of an internal information system dedicated to the activities in these disciplines, RIBAC. At the national level, HAL is the platform where researchers can signal their productions without the type of selection implemented by international data bases such as WoS or Scopus. HAL is thus a broader data base than RIBAC to observe publications in all disciplines; it is also open access while RIBAC is internal to CNRS. HAL has thus been chosen as a complementary data source in order to analyze the coverage of OST data base in different disciplines.

OST has built a corpus out of an extraction of the complete HAL database performed on September 14, 2022. The analysis focuses on the period 2017-2021 and distinguishes two corpora: the first corpus includes all records from French institutions (HAL), and the second includes all records with a CNRS affiliation (CNRS-HAL). The first section below (D1) describes the two corpora and the types of productions they include. The second section (D2) compares publications found in HAL and in the WoS for France and for the CNRS. The third section (D3) is dedicated to methodology and details the constitution of HAL corpora, data cleaning and the construction of a common disciplinary classification to compare HAL and WoS data.

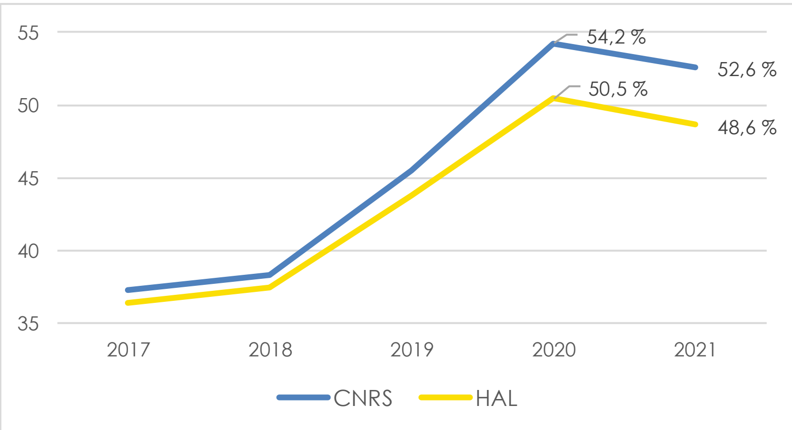
D1. CNRS productions reported in the national archive, HAL

HAL allows researchers and institutions to declare their productions with bibliographic notices or to deposit also the files corresponding to these productions. As it is a national database, it does not allow comparison with international data. HAL is purely bibliographic and does not include citation counts. Declarations are voluntary, which makes the completeness of the data variable according to institutions and research fields. In addition, in the deposit form, some fields are pre-filled, such as "editorial board" (by default to yes), "diffusion/dissemination" (as opposed to “scientific”, by default to no), "audience" (by default to international) or "proceedings" for communications in conferences (by default to no). The depositors do not necessarily modify the default value, even if it would be relevant. More details are provided in Appendix D3 below.

HAL data does not permit to account for the number of affiliated institutions of a co-publication and all observations in this appendix are based on full counts. That is, each publication is counted as 1 regardless of the number of authors and affiliation addresses.

HAL has a data field indicating the type of deposit that allows to distinguish a simple notice from an “archive”, *i.e.*, a file including the full document described in the notice. Figure D1 shows that the archive rate of the CNRS-HAL corpus is higher than the national average, while following the same evolution since 2017. The share of archive deposit increased significantly in 2019 and 2020 before declining. It exceeds 50% for the CNRS in 2020 and 2021.

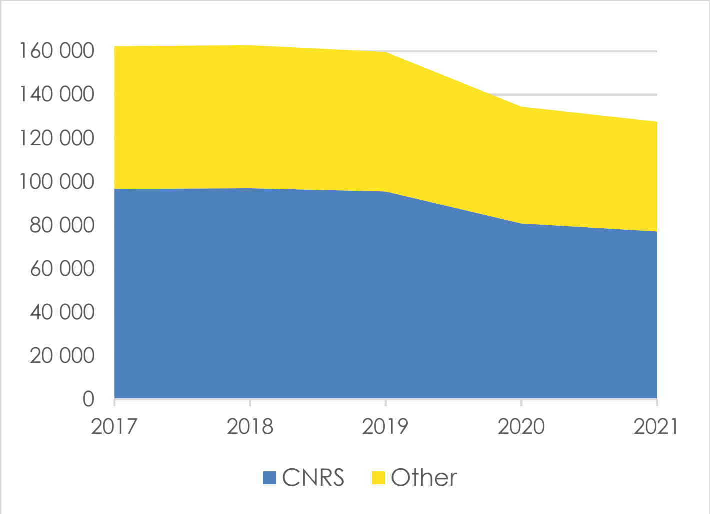
Figure D1. Share of archive deposit in HAL,* total and CNRS, 2017-2021, %.



* Definition: records with archives / total records
Source: HAL extraction 22/09/14, OST processing

Figure D2 distinguishes CNRS productions within the set of HAL records over the period 2017-2021. HAL has about 160,000 annual records from 2017 to 2019, then less records in 2020 and 2021. The evolution is similar for all French productions and for the CNRS: the number of productions declared in 2021 by about 20% lower compared to the beginning of the period.

Figure D2. Number of records in HAL: CNRS and other French affiliations, by year of production, 2017-2021.



Source: HAL extraction 22/09/14, OST processing

Table D1 shows the share of the different types of productions reported in HAL and in CNRS-HAL corpus. The typology of HAL is completed in two ways. First, within the type "article", the matching with different indices of publications aims at identifying the journals that implement an evaluation process (EP journals). This information is not available in HAL metadata[34] and the enrichment consisted in checking whether the journal is indexed in sources that control for the editorial processes: DOAJ, Bona Fide, Web Science (including the ESCI index), Scopus and Ulrich. Each of these sources is described in the methodological section (D3). Being indexed by at least one of these sources does not completely guarantee the implementation of a peer review process, but it does provide information on the existence of an evaluation process and for some of the databases, the process has to involve peer reviews. Second, for conference papers, the enrichment consisted in checking whether there are proceedings and, in this case, the assumption is that it is a publication rather than a simple oral communication.

Table D1 shows that "article" is the most frequent type, with 49% in HAL and 52% for the CNRS-HAL corpus. The proportion of articles that are published in EP journals is 78% in HAL and 84% in CNRS-HAL. Communications in conferences are the second most frequent type of production with 21% in HAL and CNRS-HAL. The share of communications published in conference proceedings is much lower, at 7%. The third most frequent type is formed by book chapters, with 9% in HAL and 8% in CNRS-HAL. The share of books is 1.5% in HAL and 1.4% in CNRS-HAL.

[34] Publication supports found in HAL are very diverse and include for example professional journals, newspapers and magazines.

Table D1. Distribution of records by production type, 2017-2021: HAL and CNRS-HAL.

Type of production	Share, 2017 - 2021		Change in number between 2017 and 2021	
	HAL	**CNRS-HAL**	**HAL**	**CNRS-HAL**
Articles				
All journal articles	49.1%	52.0%	-7%	-6%
of which published in journals with an evaluation process (EP journals[1])	*38.2%*	*43.8%*	*-1%*	*-4%*
Conferences				
All communications	20.9%	21.4%	-47%	-48%
of which with proceedings	*6.5%*	*7.5%*	*-55%*	*-51%*
Poster	2.0%	2.1%	-62%	-61%
Books and book chapters				
Book chapter	8.6%	8.0%	-26%	-18%
Book	1.5%	1.3%	-27%	-23%
Book editing	1.5%	1.4%	-37%	-33%
Other productions				
PhD. thesis	5.2%	5.5%	-32%	-33%
Dissertation, research paper	3.4%	0.1%	-10%	-33%
Other	2.9%	2.6%	-23%	-21%
Pre-print, working paper	2.7%	3.4%	106%	+110%
Other types[2]	0.9%	0.9%	-33.4%	-36.5%
Total	**100%**	**100%**	**-21.5%**	**-20.2%**

1. Journals indexed in at least one of the following sources: DOAJ, Bona Fide, WoS (including ESCI), Scopus, Ulrich "refereed".
2. For example: reports, HDR dissertation (*Habilitation à diriger des recherches*), images, sound, video, patents.

Note. Types are declared by the users and are not filled in homogeneously. For example, the type "book" is quite heterogeneous, and may concern reports or digital publications whose publisher is difficult to identify.

Source: HAL extraction 22/09/14, DOAJ, Bona Fide, WoS, Scopus, Ulrich; OST calculations

Between 2017 and 2021, all types of output tend to decline in number, suggesting that the overall trend in Figure D2 is not due to a single type. Nevertheless, the phenomenon is particularly pronounced for conference papers. Conversely, the only type whose number is increasing is that of preprints. These evolutions suggest that the Covid period had an influence on the activity of the researchers as it is reflected in HAL. This impact could nevertheless be combined with an evolution of the rhythm or the rate of deposit of the records in HAL – which, again, is voluntary and not automatic.

D2. Comparison of HAL and WoS corpora for France and the CNRS

Table D2 compares France and CNRS corpora in HAL on the one hand and in WoS on the other hand. It does so by type of productions as they have been described above (table D1). Table D2 shows that productions are more numerous in HAL for the broader corpora (lines 1 to 3). On the contrary, publications in EP journals and conference proceedings (line 4) are more numerous in WoS.

Table D2. Number of productions in HAL and in the OST database by type, 2017-2021.

Corpus considered in HAL	Indicators	France		CNRS	
		HAL	**WoS**	**HAL**	**WoS**
1. All productions	Number	746 628	570 221	446 851	292 265
	Ratio : HAL / WoS	1,31	1,00	1,53	1,00
2. All publications and contributions to conferences	Number	598 023	478 774	369 939	271 897
	Ratio : HAL / WoS	1,25	1,00	1,36	1,00
3. Journal articles and conference proceedings	Number	522 088	478 774	328 373	271 897
	Ratio : HAL / WoS	1,09	1,00	1,21	1,00
4. Journal articles EP and conference proceedings	Number	351 419	478 774	243 462	271 897
	Ratio : HAL / WoS	0,73	1,00	0,90	1,00

Sources: HAL extraction 22/09/22, DOAJ, Bona Fide, WoS, Scopus, Ulrich; OST calculations

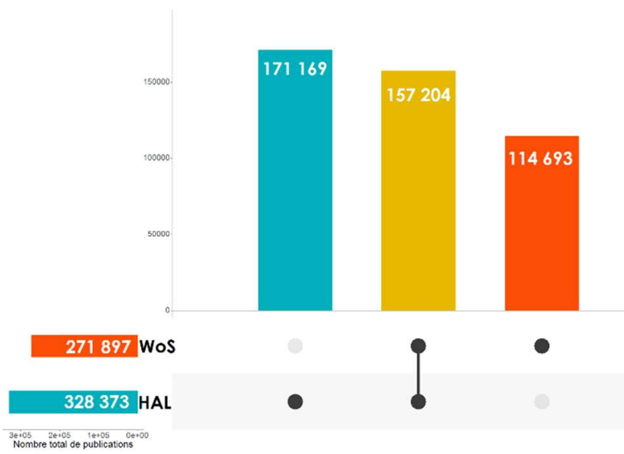
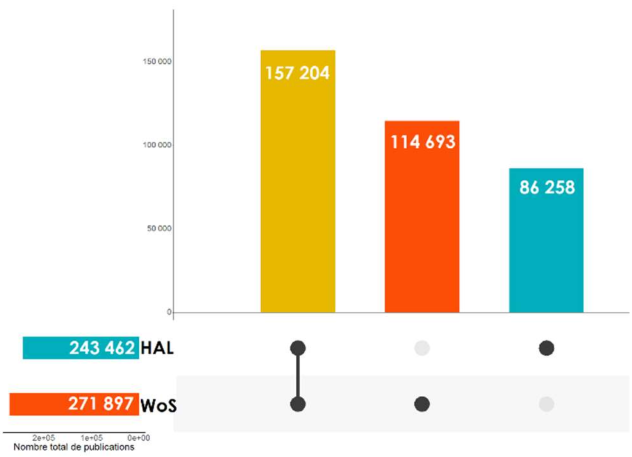
In order to draw more detailed comparisons between HAL and the WoS, the scope of the analysis is narrowed to scientific publications. Records related to the publications of book chapters or books in HAL do not distinguish dissemination from scientific contributions in a reliable way.[35] This distinction is moreover difficult and would require a specific expertise. In addition, the WoS covers very few books. Including the "book" or "book chapter" types would therefore not allow us to build a reliable corpus of scientific publications without extensive resources. The analysis therefore focuses on publications in journals and conference proceedings.

Figure D3 compares the number of CNRS publications in HAL and in the WoS by distinguishing three subsets: the intersection between the two data sources, publications indexed only in HAL and those indexed only in the WoS (It corresponds to lines 3 and 4 of table D2 for CNRS).

Column 1 compares the two sources for all journal articles and conference contributions published in proceedings (corpus 1). This corpus represents 272K publications in WoS and 329K in HAL. The intersection counts 157K publications, 115K are indexed only in WoS and 171K only in HAL. In other words, the two sources have a significant number of publications in common, but also publications that are indexed in only one of the two.

Column 2 restricts the corpus of articles to those published in EP journals as defined above. The WoS set is not modified insofar as the WoS selects journals notably on the criterion of an editorial process that includes peer review of papers. The intersection between the two databases is not modified either for the same reason. On the other hand, HAL corpus is reduced to 243K and counts less publications than WoS corpus. The reduction is explained by the articles published in journals for which no evaluation process could be identified. In this corpus, the intersection between the two data bases is the largest subset, the second one is that of publications present only in WoS and the third those that are present only in HAL (86K).

Figure D3. Number of CNRS publications in HAL, WoS and both sources, by types, 2017-2021.

Corpus 1. Journal articles and conference proceedings	Corpus 2. Articles in EP journals and conference proceedings
	

Source: HAL extraction 22/09/14, DOAJ, Bona Fide, WoS, Scopus, Ulrich; OST calculations

The CNRS disciplinary profile in HAL and WoS

Beyond the number of publications indexed by each of the two sources, it is important to analyze whether the disciplinary profile of the CNRS is similar or on the contrary different when observed with data from the WoS or from HAL. The deposit of records in HAL is voluntary, which can lead to an unequal representation of institutions and/or disciplines. Moreover, the information provided on different fields is of unequal reliability, which can also introduce biases between disciplines. In the case of the WoS, the journal selection process results in less coverage of certain applied disciplines and publications with a local or national audience. This is the case in particular for certain fields of engineering, health, social sciences and humanities.

In order to compare the distribution of types of productions by scientific field, it was necessary to elaborate a common disciplinary classification. The methodological section below explains the approach, which results in a classification close to the one ERC panels. The analysis is presented first by major field and then by discipline.

Table D3 presents the distribution of publications by major field and types of publications for both France and the CNRS.

[35] In particular the fields “diffusion/dissemination” being set by default to “no”, it implies that books or book chapters are automatically considered to be scientific contributions, while checks on samples show that it is often not the case.

Table D3. Distribution of publications by field* depending on source and types: France and CNRS, 2017-2021.

	HAL				WoS		Share HAL / Share WoS	
	France			CNRS	France	CNRS	France	CNRS
	1. Articles, comm., books and chapters	2. Articles, proceedings	3. EP articles, proceedings	4. CNRS EP articles, proceedings	5. Articles, proceedings	6. CNRS Articles, proceedings	7. France (3) / (5)	8. CNRS (4) / (6)
Physics, Engineering, Comp. sc. & Math.	35.9%	39.8%	47.0%	57.5%	48.7%	63.8%	1.0	0.9
Life Sciences & Health	23.2%	25.5%	32.1%	25.9%	40.6%	28.2%	0.8	0.9
SSH	39.4%	33.2%	19.8%	15.7%	10.4%	7.5%	1.9	2.1
Other, multi-disc.	1.5%	1.5%	1.1%	0.9%	0.4%	0.5%	2.8	1.8
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	-	-

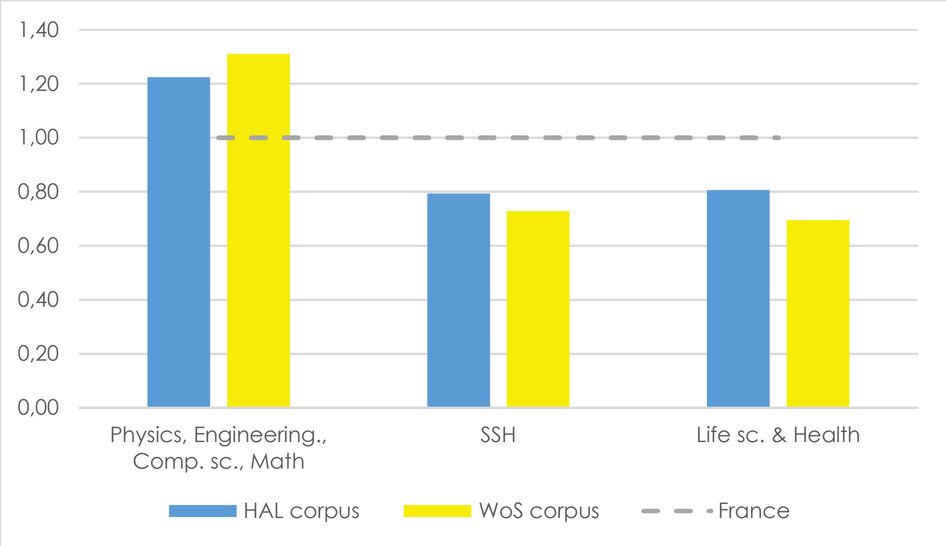
* Disciplinary fractional count so that the numbers of documents per area are summable.
Source: HAL extraction 22/09/14, DOAJ, Bona Fide, WoS, Scopus, Ulrich; OST calculations

The left part of the table shows the share of each field in the publications reported in HAL. Columns 1 to 3 refer to the entire database HAL. The proportion of the domain Physics-engineering-computer science and mathematics increases as the perimeter is restricted to EP articles and conference proceedings: it goes from 36 (col. 1) to 47% (col. 3). The increase is even greater for the Life sciences and health, whose proportion rises from 23 to 32%. The evolution is the opposite for Social sciences and humanities. The share of the field decreases by 6 points due to the withdrawal of books and book chapters, from 39% (col. 1) to 33% (col. 2), and then by 13 points when the scope is restricted to articles in EP journals.

The central part of the table provides the distribution in the WoS. For France, the share of Physics, Engineering, Computer science & Mathematics increases slightly again. It increases much more for Life sciences and health, from 32% (col. 3) to 41% (col. 5). Conversely, the share of the SSH falls from 20% to 10%. It thus appears that the corpus of EP journals is unequally covered by each of the databases. While the Physics-engineering-computer science and mathematics domain is slightly better covered by the WoS, the latter covers the Life sciences & Health domain better and, conversely, the SSH domain less well. These disparities of coverage are synthesized in the right part of the table which provides the ratio between the share of each domain in HAL and WoS. Columns 4, 6 and 8 provide the same indicators for the CNRS and the differences between domains are similar.

Based on the indicators in Table D3, Figure D4a compares the disciplinary profile of the CNRS in relation to France according to whether it is observed in HAL or in the WoS. For each of the three domains, the CNRS specialization index relates the share of a domain in CNRS publications to this share for France. The neutral value is 1. Figure D4a shows that the profile of the CNRS by domain is not different in HAL and in WoS despite the differences in coverage. This is explained by the fact that the differences in coverage of each of the two databases affect in a similar way the share of disciplines for France and for the CNRS.

Figure D4a. CNRS specialization index compared to France by data source, 2017-2021.



Source: table D3; HAL, col. 4/col.3, WoS, col. 6/col.5

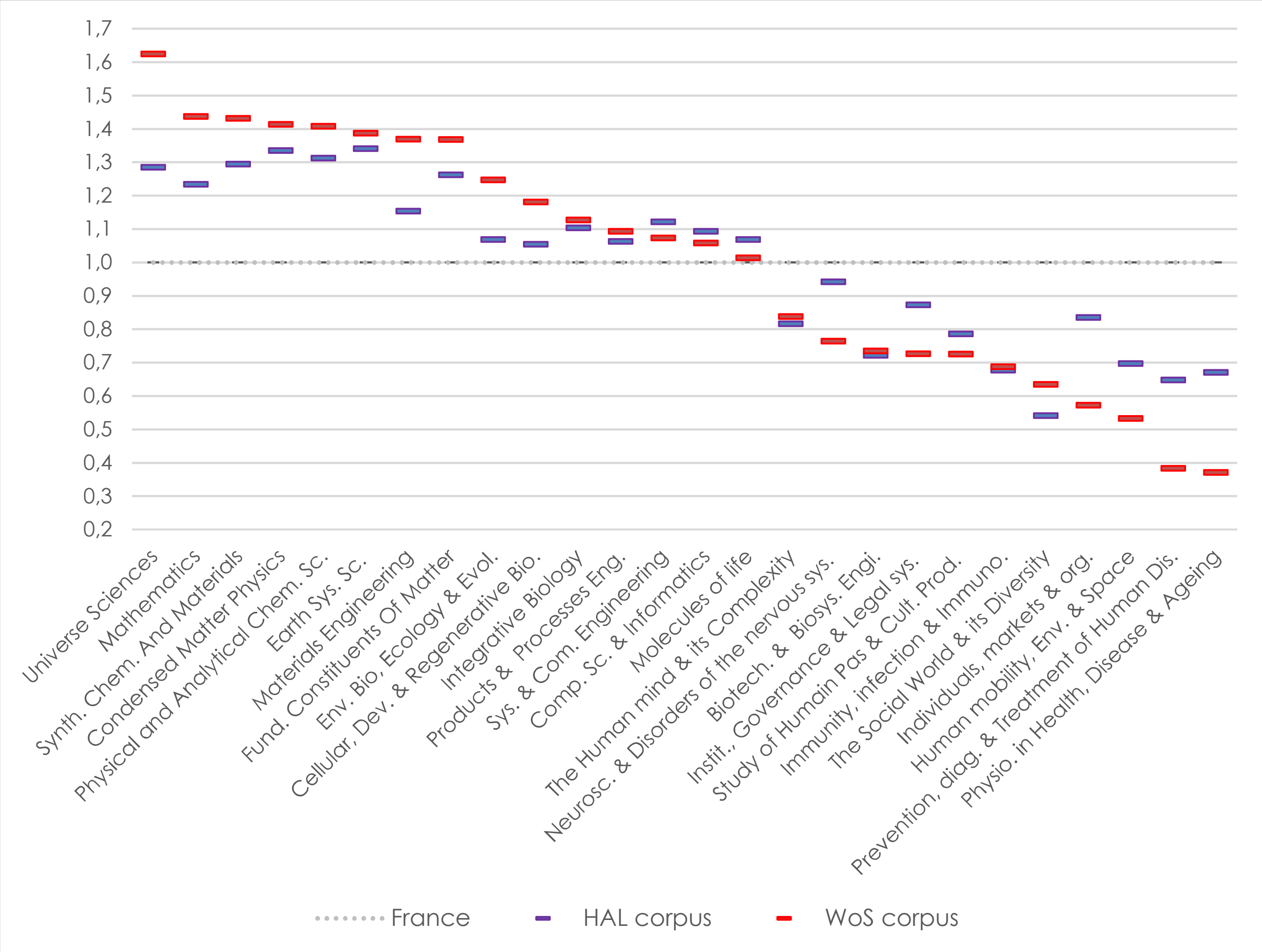
Compared to France, the CNRS is clearly specialized in Physics, Engineering, Computer science & Mathematics. Its specialization index is slightly higher in the WoS dataase (1.3 against 1.2 in HAL). Conversely, it is not specialized in SSH, with an index of 0.8 in HAL and 0.7 in the WoS. The CNRS is also not specialized in Life Sciences & Heath, its specialization index being 0.8 when calculated in HAL and 0.7 when calculated in WoS.

Figure D4b is the same as D4a at a finer grain of classification, with 26 disciplinary fields the scope of which is close to ERC panels (see the methodological section below). At this finer grain, the dispersion among fields is larger but the overall correlation between HAL and WoS specialization indices holds.

Figure D4b shows that the CNRS is specialized in all the Engineering, Computer science & Mathematics fields. Within this domain, CNRS is most specialized in Universe science, Mathematics and Synthetic chemicals and materials. It is least specialized in Systems and communication engineering and Computer science and information systems. Besides, for 7 fields out of 11, CNRS specialization indices are very close in the two databases: Condensed matter physics, Physical and analytical chemistry, Earth sciences, Fundamental constituents of matter, Product and process engineering, Systems and communication engineering, Computer science and information systems. The case of computer science may be noted since CNRS considers in its self-assessment report that the WoS does not cover correctly the proceedings in this field.

CNRS is moderately specialized in only 3 fields within the Life sciences & health domain: Environmental biology, ecology and evolution, Cellular development & regenerative biology, Integrative biology. It is on the contrary not specialized at all in the large fields related to medical research and health (indices below 0.7). In some of those fields, indices are different when calculated in HAL or WoS, but they are low in both cases. In 4 areas of the domain out of 9, the specialization indexes of the CNRS compared to France are very close between the two databases: Integrative biology, Molecules of life, Biotechnology & biosystems engineering, Immunity, infection & immunotherapy.

Figure D4b. CNRS specialization index relative to France, by field and data source, 2017-2021.



Sources: HAL extraction 5/9/22, DOAJ, Bona Fide, WoS, Scopus, Ulrich; OST calculations

Figure D4b shows that CNRS is not specialized relative to France in any of the 6 SSH fields. The specialization indices are close in 3 fields: The human mind and its complexity, The study of the human past and cultural production, The social world and its diversity.

In total, the distribution of the publications by field in the two data bases depends both on the share of the different types of publication supports that are indexed by each of the bases and on the propensity of the communities to report their productions in HAL. For example, fields in Life sciences & health tend to deposit a lower proportion of their publications in HAL, while the WoS covers this domain quite well. On the contrary, some SSH fields deposit a high proportion of their productions in HAL, while the WoS has a lesser coverage of these same fields.

The comparison between the disciplinary profile of the CNRS and that of the whole France does not appear however biased by the differences noticed between the two sources for the perimeter of the EP articles and the conference proceedings. These results obtained at the level of France suggest that the WoS can be used to compare the CNRS to foreign organizations. Indeed, the relatively low or relatively high share of certain disciplines in WoS will be found at the international level.

D3. Methodology for comparing data from HAL and WoS

Extraction of HAL data

The entire HAL database has been extracted via the API - *Application Programming Interface* (v3.0.) HAL (https://api.archives-ouvertes.fr/docs), on 14/09/2022.

The extraction contains all the HAL records, that is to say **3 112 267** records, of which **2 819 332** are unique (the same record can be present in duplicate or triple in the extraction). Only the records whose year of publication (not the year of deposit) is between 2017 and 2021 are taken into account: **857 845** records, including **788 953** unique ones.

The fields relating to authors' structures were used to identify CNRS records. The query is constituted by the union of the results of 4 queries: the acronym "INSTSTRUCTACRONYM_S" contains 'CNRS', the name "INSTSTRUCTNAME_S" contains the complete name; the identifier "INSTSTRUCTID_I" is the one associated with the CNRS; the name "INSTSTRUCTNAME" contains CNRS.

```
SELECT DISTINCT DOCID FROM BASE_DE_DONNEES_HAL
    WHERE DBMS_LOB.instr(INSTSTRUCTACRONYM_S, 'CNRS')>0
        OR DBMS_LOB.instr(UPPER(INSTSTRUCTNAME_S), 'CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE')>0
        OR DBMS_LOB.instr(INSTSTRUCTID_I, '441569')>0
        OR DBMS_LOB.instr(UPPER(INSTSTRUCTNAME_S), 'CNRS')>0
```

The CNRS HAL identifier "*441569*" alone returns **1,601,856 results**, while the query including acronyms **1,604,039** results. For the period 2017-2021, the total number of CNRS records with disciplinary information is **451,767**.

Data preparation

Different treatments have been necessary to validate the final set of data for the analysis.

Removal of duplicates

The DOI (Digital Object Identifier) of the records have been used to compare data from HAL with data from the WoS. In some cases, the authors of one publication create several bibliographic notices in HAL with different identifiers. Only occurrence has been kept in such cases. This concerns **16,596** records out of **785,693,** so the de-duplicated set contains **769,097** records. The same process has been performed on WoS data to exclude publications with the same DOI: only **266** publications were excluded.

Deletion of records with foreign addresses only

Although it is a national archive, HAL includes bibliographic notices without any French affiliation. In some cases it results from massive deposits of proceedings from international conferences that took place in France.[36] It can also be deposits of foreign institutions in HAL.[37]

The analysis excludes the records without any French address. The records which have no information on the country of the authors are kept by making the hypothesis that they are French (which risks to overestimate slightly the deposits of French authors). Among the **769,097** de-duplicated records, 3% have no French affiliation. The final set includes **746,628** valid records. The following table summarizes the data cleaning steps.

Step	Set of records	HAL		HAL - CNRS	
		#	%	#	%
1	Total HAL	785 893	100%	451 767	100%
2	HAL without duplicates whose information is available	769 097	98%	446 943	99%
3	No. 2, without addresses outside France	746 628	95%	446 851	99%

Disciplinary referential

HAL domain repository (https://api.archives-ouvertes.fr/docs/ref/resource/domain) and the table of correspondence between domain abbreviations and their labels have been extracted. The table contains 393 entries including the 3 disciplinary levels: level "0" in 13 domains,[38] level "2" in 166 domains[39] and level "1" in 214 domains.[40]

[36] For example: https://hal.archives-ouvertes.fr/search/index/?qa%5BjournalTitle_t%5D%5B%5D=Discrete+Mathematics+and+Theoretical+Computer+Science&qa%5Btext%5D%5B%5D=%28-en%29&submit_advanced=Recherche&rows=30
[37] For example: https://hal.archives-ouvertes.fr/search/index/q/*/contributorId_i/1158320
[38] https://api.archives-ouvertes.fr/ref/domain/?wt=xml&q=level_i:%220%22
[39] https://api.archives-ouvertes.fr/ref/domain/?wt=xml&q=level_i:%222%22
[40] https://api.archives-ouvertes.fr/ref/domain/?wt=xml&q=level_i:%221%22

Over 2017-2021, 0.4% of the records have no disciplinary information. The dataset without duplicate (or triplicate) HAL identifiers (DOCID) and with available disciplinary information contains **785,693** records.

Construction of the corpora according to production types

The analysis distinguishes between outputs (all types, including working papers, reports, images etc.) and scientific publications for which an evaluation process has been validated (EP). Three perimeters have been constituted.

All publications and conference presentations

This is a broad perimeter including records catalogued as "article", "communication", "book chapter" or "book". The types are declared by the users and are not filled in homogeneously. For example, a contribution in conference proceedings may be catalogued as an article, a book chapter or a paper. Based on samples, the type "book" appears to be very heterogeneous, and may concern reports or digital publications whose publisher is difficult to identify.

Journal articles and conference proceedings

This perimeter includes productions catalogued as "article" or "communication". The information provided on the journal or communication is declarative and may vary for the same publication medium. The "communication" type includes both communications in conferences with proceedings and those in conferences without proceedings (the default value for the "proceedings" field is "no"; its modification is not mandatory at the time of submission).

EP journal articles and conference proceedings

This perimeter aims at focusing on scientific publications for which the editorial process has included a peer review. In order to identify the relevant journals, the analysis consisted in observing whether the journals are indexed in at least one of the following databases, in decreasing order of rigor of the editorial process:

- Directory of Open Access Journals -DOAJ: https://doaj.org/
- Bona Fide (positive list of non-predatory journals): Academic libraries, guardians of the open access publishing domain - Quality Open Access Marker (qoam.eu); https://www.qoam.eu/bfj/journals
- WoS: https://clarivate.libguides.com/librarianresources/coverage , including the index of "emerging" journals - ESCI: https://clarivate.com/webofsciencegroup/solutions/webofscience-esci/
- Scopus: https://www.elsevier.com/?a=91122
- Ulrich (refereed reviews): https://about.proquest.com/en/products-services/Ulrichsweb

Communications published in proceedings have been identified from HAL field "Actes". The default value is "no" and it is up to the depositor of the record to change the value to "yes" if the communication has been published. Since this is not required at the time of filing and is not visible without prompting for the full metadata, the number of communications with proceedings may be underestimated. Nevertheless, in the random sample of "communication" records OST analyzed, nearly two-thirds did not have proceedings.

Construction of a common disciplinary classification HAL-WoS

HAL disciplinary field is not usable directly. First, there is no homogeneity of the disciplinary levels by record. Secondly, this field is declarative, which leads to variations. For example, in HAL documentation, there are 13 domains of level "0", but in the data extracted from the API, this level appears confused with other levels that have the same tag "0". There are 29 level 0 domains in the data.

In order to exploit the disciplinary information, OST-WoS nomenclature has been used to constitute groups of disciplines. An analysis of co-occurrence of the HAL domains with those of the WoS, in the set of publications that are present in both HAL and the WoS results in 10 broad domains. The steps of the construction of a common nomenclature are detailed below.

In the HAL extraction described above, 746 628 records have at least one level "0" domain. In the WoS, each publication is assigned to at least one scientific "category". OST has built a nomenclature that classifies these 254 WoS categories into the 27 ERC panels. The method consists in using these two classifications to build a common HAL-WoS nomenclature.

The intersection of the publications indexed in both HAL and WoS is identified with DOIs. It counts 238,732 publications. Each publication being indexed in both databases, it has been assigned a domain in HAL and one or more ERC panel in OST database.

A first step consists in calculating the number of publications for a given HAL domain-ERC panel pair. For example, 5,877 publications are classified in "Sciences of the Universe" in OST-ERC nomenclature, and their corresponding HAL domain is "Physics [physics]/Astrophysics [astro-ph]". Similarly, 4,285 publications are classified as "Environmental biology, ecology and evolution" in OST-ERC nomenclature, while their domain in HAL is "Environmental sciences/Biodiversity and Ecology". The number of co-occurrences is calculated for all HAL and OST-ERC disciplinary couples.

Once the co-occurrences between the nomenclatures are counted, a measure of association between disciplines is calculated using both the co-occurrence of disciplinary pairs and the total occurrence of domains. The "Association strength"[41] is calculated as follows:

$$S_A(c_{ij}, s_i, s_j) = \frac{c_{ij}}{s_i s_j}$$

With S_A the association measure, c_{ij} the co-occurrence of the HAL domain " i " with the OST-ERC domain " j ", s_i the number of occurrences of the HAL domain and s_j the number of occurrences of the OST-ERC domain.

From the proximity matrix between HAL and OST-ERC domains, a classification algorithm is used to cluster HAL and OST-ERC domains. Vosviewer software is used for clustering, which consists in maximizing the inter-cluster distance and minimizing the intra-cluster distance.

The classification is then generalized to the publications outside the HAL-WoS intersection by using the correspondence between HAL-WoS domains and the clusters. The clusters constitute the entries of the common HAL-WoS nomenclature used in the report.

[41] Eck, N, J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), Measuring scholarly impact: Methods (pp. 285-320). Springer.

Appendix E. Computation of mean normalized citation scores with different databases

In addition to the analysis presented in Section 2.4 above on the comparison of the MNCS of the CNRS with those of other institutions, a complementary analysis has been carried out by a member of the assessment committee, focusing on the comparison between CNRS and MPG. This analysis was made on the basis of the Dimensions publication database.[42] The assessment committee and Hcéres have investigated the comparison of the two analyses based on OST-WoS and Dimensions databases. The two main results from this comparison are as follows.

First, most of the disciplinary CNRS and MPG citation scores measured with Dimensions appear to be higher than citation scores measured with WoS. In particular, most of the "Dimensions-MNCS" of the CNRS are above the Dimensions-world average, whereas most of the "WoS-MNCS" of the CNRS are below the WoS-world average (see Figure 21a above).

- In most disciplinary fields, the "Dimensions-MNCS" of the CNRS are 10% to 30% higher than its "WoS-MNCS". We do not give precise figures, because it has not been possible to carry out a detailed comparison between the analysis presented in Section 2.4 above and the analysis made with Dimensions by a member of the CNRS assessment committee (in particular, the corpus of CNRS publications used for the analysis using Dimensions was 30% smaller than the corpus used in the analysis presented in Section 2.4 ; the methods used to assign publications to scientific fields or disciplines were also not identical in the two analyses).
- This result is consistent with similar observations from the bibliometric literature. A study comparing citation scores of the publications of German universities measured on the basis of four databases shows that "Dimensions-MNCS" are clearly higher that "WoS-MNCS".[43] This is related to the fact that the Dimensions database, which is less selective than the WoS database, includes a large number of publications that receive few citations. This explains why the world average number of citations is lower when measured with Dimensions than with the WoS, and why research institutions like the CNRS and MPG have higher "Dimensions-MNCS".[44]

Second, similar "gaps" between the citation scores of CNRS and MPG appear in most disciplines, be they measured with Dimensions or with the WoS.

- This result is also consistent with similar observations from the bibliometric literature. The above quoted study on German universities shows that the level on one institution's MNCS varies when measured with different databases, but the relative positions of different institutions are quite stable.

The exchanges between the assessment committee and Hcéres on these analyses have shed light on the factors that influence the computation of MNCS. The following general observations are worth noting:

- The precise identification of the corpus of publications of an institution is a difficult issue, especially in the case of France because of joint research units (see Box 2 on UMRs). Moreover, the difficulty depends on the quality of the metadata of each publication database.
- Information on citations are not 100% complete in any publication database.[45]
- The assignment of the publications to disciplinary fields is also a delicate issue, and it can contribute to variations in the field-by-field citation scores.
- The coverage of each publication database influences the measurement of citation scores. Hence, the values of the MNCS measured with different publication databases should not be directly compared.

---

[42] Dimensions was launched in 2015 by Digital science, on the basis of several sources such as Crossref, PubMed and arXiv. In addition to scientific publications, it includes data on patents, clinical trials, « policy documents », etc.

[43] See Scheidsteger, Haunschild & Bornmann (2023), How similar are field-normalized scores from different free or commercial databases calculated for large German universities?, 27th International Conference on Science, Technology and Innovation Indicators (STI 2023), https://dapp.orvium.io/deposits/6441118c643beb0d90fc543f/view.

[44] Similarly, "OpenAlex-MNCS" for German universities are higher than "Dimensions-MNCS".

[45] More than 10% of citation links mentioned in Scopus are not present in Dimensions (for a given corpus), whereas, 3% of the citations links mentioned in Dimensions are not present in Scopus (for the same corpus). See Orduña-Malea & López-Cózar (2018), Dimensions: re-discovering the ecosystem of scientific information, arXiv preprint 1804.05365.

A similar analysis comparing WoS and Scopus showed that (for a given corpus) 20% of publications have less citation links in the WoS than in Scopus, whereas 3% of the publications have more citation links in the WoS than in Scopus (Van Eck & Waltman, 2017, Accuracy of citation data in Web of Science and Scopus, Proceedings of the 16th International conference of the international society for scientometrics and informetrics, pp. 1087–1092).

Appendix F. Data and methodology for the analysis of CNRS patents

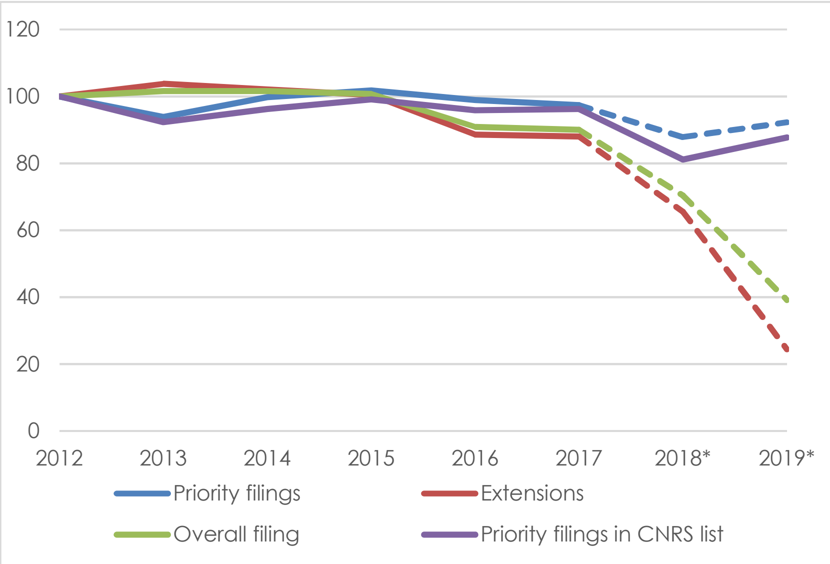
OST patent database

The patent data mobilize information from the OST patent database, built from PatStat and enriched by OST. The PatStat database was created by the EPO with the help of the OECD. The EPO updates and publishes the entire database twice a year (April and October). The extracted information is based on the spring 2022 version of PatStat, and takes into account all applications published until mid-December 2021. The PatStat data are used for the analysis of patent grants and extensions.

PatStat contains records of patent filings after publication of the application, i.e., 18 months after the date of the first filing (priority filing). It covers 80 national and regional patent offices worldwide.

The updating of the patent database by end of year 2021 implies that filing years 2018 and 2019 are not yet complete, namely the extensions of priority filing are not published and so are not accessible. The following figure with priority data in the OST database and in the CNRS list of priority filings can illustrate this. Priority data in OST database and in the patent list which was sent to OST are very similar.

Figure F1: CNRS priority filings, extensions and overall filings, 2012=100.



Sources: OST Database and CNRS, computed by OST using PatStat.

Patent of invention

The patent is a title of ownership that confers to its owner or successors, for a limited period of time and on a limited territory, an exclusive right of exploitation of the invention. In exchange for the exclusive right granted to him/her, the applicant of the patent has the obligation to make the invention public. The patent is therefore not only a legal title of property right but also a technical publication.

Patents can be considered as one of the results of R&D activity. As patents are one of the few sources of information on these R&D outputs, they are frequently used as an indicator of inventive activity and a measure of technological capability.

National and regional offices

The **INPI** is the **French office for intellectual property** (patents, trademarks, designs and models). It allows to file a patent application in order to protect an invention on the national territory. A large part of the patents filed by French actors are filed first at the INPI before being, if necessary, extended to other offices. The patent application filed at the INPI is published eighteen months after it is first filed, and a patent may not be granted until later.

The **European patent Office (EPO)** establishes a unified system for filing and granting patents in the European countries, signatories of the Munich Convention (1973), called "European patent system". Through a single filing and granting procedure, it is possible to obtain a "European" patent which has the same effects in each country designated by the applicant as a national patent filed in several countries that are signatories of the Munich Convention.

The **United states patent and trademark Office (USPTO)** allows any individual or company wishing to protect their invention in the United States to apply for a U.S. patent. This office has many specificities. For example, unlike the EPO, the patent is granted to the first inventor and not to the first applicant.

Another procedure for simultaneous applications in several countries has existed since 1978: the **PCT** procedure **(Patent Cooperation Treaty)** allows any applicant to file a patent application simultaneously in 184 countries. This procedure is managed by the World Intellectual Property Organization (WIPO). It has many advantages compared to the classical ways of application (one single step, lower cost, longer reflection period).

Many French institutions file their priority applications with the INPI before possibly extending the protection of their invention internationally. Numerous international extension processes (in particular according to the European or PCT procedures) are then possible.

Priority filings and extensions

The priority filing of a patent application is the first filing to protect an invention before a patent office.

The Paris Union Convention (PUC) for intellectual property provides for a period of one year (from the date of priority filing, the so-called priority date) to allow an applicant to extend his invention to other PUC contracting states.

Most French institutions file their priority applications with the INPI before eventually extending the protection of their invention internationally. Numerous international extension processes (in particular according to the European or PCT procedures) are then possible.

Two international extension processes are particularly used for a French applicant. They are illustrated below:

Case of an INPI filing extended to the EPO: When the priority application is filed at the INPI, the institution may wish to extend its application to the EPO in order to protect itself at the European level. During the filing procedure at the EPO, the applicant is asked to designate the European countries where to extend the protection. It is then possible to re-designate France, making the French priority application null and void, while keeping the initial priority date.

Case of an EPO filing extended to WIPO: When the priority application is filed at the EPO, the institution may wish to extend its application to the WIPO level in order to protect itself in offices such as the USPTO or the JPO. When the WIPO filing procedure enters the regional phase, it is possible to re-designate the EPO, with the new EPO application replacing the priority application.

Consequently, it happens in a significant number of cases that priority applications at the INPI or at the EPO are abandoned even before their publication because they are replaced by applications at the EPO or at the WIPO - non priority. Specific treatments allow to recover these abandoned priority applications in favor of EPO or WIPO applications, in particular through patent families.

Technology fields and sub-fields

In order to classify patents according to their technological content, the World Intellectual Property Organization (WIPO) created the International Patent Classification (IPC) in the Strasbourg Agreement (1971). This classification is very detailed and includes approximately 70,000 subdivisions. The same patent can be classified in different IPC classes. An IPC fractional counting or thematic fractional counting is therefore possible to take into account the relative weight of the different technologies contained in a patent, in the same way as the disciplinary fractional counting for publications.

For the last ten years, another classification has been developed jointly by the European Patent Office and the United States Patent and Trademark Office. This classification is based, in large part, on the IPC classification with some detail and some specific classes have been added, notably concerning environmental concerns. At the aggregate level IPC and CPC classifications are identical, so CPC codes can be used if IPC codes are missing in a patent.

An aggregated nomenclature was then developed by Schmoch (2008) for WIPO[46] to group IPCs into 5 technology fields, which are further subdivided into 35 sub-fields (see below).

Counting method

Computation of the total number of patents filed by the CNRS, the total number of their co-filings and the co-filings per co-applicant have been made in full counting in order to measure the participation of the partners.

The analysis of the sub-fields was performed using fractional counting. The calculation of the grant rate is done in full counting.

[46] Schmoch, U. (2008). *Concept of a Technology Classification for Country Comparisons - Final Report to the World Intellectual Property Organisation (WIPO)*. Karlsruhe, Germany.

Patent indicators

- Grant rate

The granting of a patent application at the European Patent Office (EPO) follows a long[47] process of examination of the application by experts judging its novelty, inventiveness and industrial applicability. Not every application will result in a patent being granted. Some will be refused by the examiners, others will be abandoned in the process by the applicants. The grant rate measures the number of applications actually granted to an actor in relation to the number of total applications of this actor for a given cohort of applications (for example, applications filed by the CNRS between 2012 and 2017 at the EPO). This indicator requires the use of a time window, calculated between the date of filing at the EPO and the publication of the grant, in order to be able to compare the grant rates for different filing years. In this study, we used a 4-year window and a 6-year window to be consistent with the studied period.

- Co-applications

The share of co-deposits is the ratio of the number of co-applications to total CNRS applications. Due to the use of full counting, the numbers and shares of co-applications are not summable.

- Sectoral classification of French applicants by OST

French applicants have been classified into sectoral categories, with the main classification criterion being the activity of the institution. The OST nomenclature has two levels. Only the first level is used in this report. CNRS co-applications with French applicants are identified according to these five categories.

Company	**R&D institution**[48]	**Administration &** NPO	Higher **education**	**Healthcare**
Establishments whose main activity is the production of market goods and services, whatever the origin of their equity capital, or the provision of R&D and innovation services to these companies (analysis, infrastructure, S&T skills)	Establishments whose main activity is research, financed entirely or partially by public funds	Administrative or cultural establishments financed entirely or partially by public funds, international organizations, NPOs with a national public service mission not elsewhere classified	Establishments whose main activity is teaching under private or public supervision (except university hospitals classified under "Care")	Private or public establishments whose main activity is in healthcare. The university hospitals are classified in this category.

For foreign institutions, OST used the classification made at the international level by PatStat. This classification is based on OECD data and other international organizations. OST has classified the applicants into foreign private institutions (generally companies but also foundations or private centers) and foreign public institutions (universities, foreign public hospitals).

[47] In recent years, the average time to issue applications at the EPO is 6 years.

[48] These institutions are usually classified in the *R&D Organizations* subclass of the *Government and NPIs class*, but it encompasses a wide variety of institutions that go well beyond research organizations in the classical sense, including NPIs that are associations or foundations. The objective is to be able to distinguish public research organizations within the *R&D Institutions* class, and to be able to aggregate them to the *Universities* subclass because of the presence of mixed research units and the fact that they belong to the ESRI system.

Technology nomenclature

The nomenclature of technology domains and sub-domains maintained by WIPO.

	Field / Sub-field	IPC Classes
I. Electrical engineering		
1	Electrical machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
2	Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K
3	Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q
4	Digital communication	H04L
5	Basic communication processes	H03#
6	Computer technology	(G06# not G06Q), G11C, G10L
7	IT methods for management	G06Q
8	Semiconductors	H01L
II. Instruments		
9	Optics	G02#, G03B, G03C, G03D, G03F, G03G, G03H, H01S
10	Measurement	G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, (G01N not G01N-033), G01P, G01R, G01S; G01V, G01W, G04#, G12B, G99Z
11	Analysis of biological materials	G01N-033
12	Control	G05B, G05D, G05F, G07#, G08B, G08G, G09B, G09C, G09D
13	Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G

Field / Sub-field		IPC Classes
III. Chemistry		
14	Organic fine chemistry	(C07B, C07C, C07D, C07F, C07H, C07J, C40B) not A61K, A61K-008, A61Q
15	Biotechnology	(C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S) not A61K
16	Pharmaceuticals	A61K not A61K-008, A61P
17	Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
18	Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K
19	Basic materials chemistry	A01N, A01P, C05#, C06#, C09B, C09C, C09F, C09G, C09H, C09K, C09D, C09J, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
20	Materials, metallurgy	C01#, C03C, C04#, C21#, C22#, B22#
21	Surface technology, coating	B05C, B05D, B32#, C23#, C25#, C30#
22	Micro-structural and nano-technology	B81#, B82#
23	Chemical engineering	B01B, B01D-000#, B01D-01##, B01D-02##, B01D-03##, B01D-041, B01D-043, B01D-057, B01D-059, B01D-06##, B01D-07##, B01F, B01J, B01L, B02C, B03#, B04#, B05B, B06B, B07#, B08#, D06B, D06C, D06L, F25J, F26#, C14C, H05H
24	Environmental technology	A62D, B01D-045, B01D-046, B01D-047, B01D-049, B01D-050, B01D-051, B01D-052, B01D-053, B09#, B65F, C02#, F01N, F23G, F23J, G01T, E01F-008, A62C
IV. Mechanical engineering		
25	Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66#, B67#
26	Machine tools	B21#, B23#, B24#, B26D, B26F, B27#, B30#, B25B, B25C, B25D, B25F, B25G, B25H, B26B
27	Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02#, F03#, F04#, F23R, G21#, F99Z
28	Textile and paper machines	A41H, A43D, A46D, C14B, D01#, D02#, D03#, D04B, D04C, D04G, D04H, D05#, D06G, D06H, D06J, D06M, D06P, D06Q, D99Z, B31#, D21#, B41#
29	Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22#, A23N, A23P, B02B, C12L, C13C, C13G, C13H, B28#, B29#, C03B, C08J, B99Z, F41#, F42#
30	Thermal processes and apparatus	F22#, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24#, F25B, F25C, F27#, F28#
31	Mechanical elements	F15#, F16#, F17#, G05G
32	Transport	B60#, B61#, B62#, B63B, B63C, B63G, B63H, B63J, B64#
V. Other fields		
33	Furniture, games	A47#, A63#
34	Other consumer goods	A24#, A41B, A41C, A41D, A41F, A41G, A42#, A43B, A43C, A44#, A45#, A46B, A62B, B42#, B43#, D04D, D07#, G10B, G10C, G10D, G10F, G10G, G10H, G10K, B44#, B68#, D06F, D06N, F25D, A99Z
35	Civil engineering	E02#, E01B, E01C, E01D, E01F-001, E01F-003, E01F-005, E01F-007, E01F-009, E01F-01#, E01H, E03#, E04#, E05#, E06#, E21#, E99Z

Appendix G. List of acronyms

C

CAS	Chinese academy of sciences
CEA	*Commissariat à l'énergie atomique et aux énergies alternatives* (French atomic energy and alternative energies commission)
CIPA	Canadian intellectual property office
CNIPA	China national intellectual property administration
CNR	*Consiglio nazionale delle ricerche* (Italian national council for research)
CNRS	*Centre national de la recherche scientifique* (French national center for scientific research)
CoARA	Coalition on advancing research assessment
CSIC	*Consejo superior de investigaciones científicas* (Spanish national research council)

D

DLR	*Deutsches Zentrum für Luft- und Raumfahrt* (German aeronautics and space research center)
Dora	Declaration on research assessment
DZHW	*Deutsches Zentrum für Hochschul- und Wissenschaftsforschung* (German center for higher education research and science studies)

E

EPO	European patent office
ERC	European research council

F

FP	Framework program for research and innovation

H

H2020	Horizon 2020
Hcéres	*Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur* (High Council for the evaluation of research and higher education)

I

INFN	*Istituto nazionale di fisica nucleare* (Italian national institute of nuclear physics)
INPI	*Institut national de la propriété industrielle* (French national office of industrial property)
INRAE	*Institut national de recherche pour l'agriculture, l'alimentation et l'environnement* (French national research institute for agriculture, food and environment)
Inria	*Institut national de recherche en sciences et technologies du numérique* (French national research institute in digital science and technology)
Inserm	*Institut national de la santé et de la recherche médicale* (French national institute for health and medical research)
IPC	International Patent Classification
IRD	*Institut de recherche pour le développement* (French research institute for development)

J

JPO	Japan patent office

K

KIPO	Korean intellectual property office

L

LS	Life sciences

M

MESR	*Ministère de l'enseignement supérieur et de la recherche*
MIT	Massachusetts institute of technology
MNCS	Mean normalized citation score
MPG	*Max Planck Gesellschaft* (Max Planck society)
MSCA	Marie Sklodowska Curie actions

O

OST	*Observatoire des sciences et techniques* (Science and Technology Observatory)

P

PCT	Patent Cooperation Treaty
PE	Physical sciences and engineering
PI	Principal investigator
PIC	Participant identification code

S

SAR	Self-assessment report
SSH	Social sciences and humanities
STFC	Science and technology facilities council

U

UMR	*Unité mixte de recherche* (joint research unit)
USPTO	United States patent and trademark office

W

WIPO	World intellectual property organization
WoS	Web of science

2 rue Albert Einstein
75013 Paris, France
T. 33 (0)1 55 55 60 10
hceres.fr