



HAL
open science

Grey Literature in Open Repositories: New Insights and New Issues

Joachim Schöpfel, Eric Kergosien, Hélène Prost, Florence Thiault

► To cite this version:

Joachim Schöpfel, Eric Kergosien, Hélène Prost, Florence Thiault. Grey Literature in Open Repositories: New Insights and New Issues. GL23 International Conference on Grey Literature, 6-7 December 2021, Amsterdam (The Netherlands), Dec 2021, Amsterdam, Netherlands. hal-03604697

HAL Id: hal-03604697

<https://hal.univ-lille.fr/hal-03604697v1>

Submitted on 10 Mar 2022

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

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

Grey Literature in Open Repositories: New Insights and New Issues

Authors

Joachim Schöpfel (corresponding author), University of Lille, ULR 4073 - GERiICO, France
joachim.schopfel@univ-lille.fr

Eric Kergosien, University of Lille, ULR 4073 - GERiICO, France
eric.kergosien@univ-lille.fr

Hélène Prost, CNRS, ULR 4073 - GERiICO, France
helene.prost@inist.fr

Florence Thiault, PREFICS Laboratory, University of Rennes 2, France
florence.thiault@univ-rennes2.fr

Keywords

Open science, open access, open repositories, research laboratories, grey literature

Abstract

HAL is the national open repository for documents and data from French scientists. The paper analyses how grey literature is represented on HAL. It presents original empirical results from a follow-up study to former research, based on the scientometric assessment of deposits on the French national HAL repository by more than 1,200 research laboratories. These laboratories are affiliated to ten large French research universities and cover the whole range of scientific domains. We assessed the distribution of document types, the degree of openness, the use of open licenses and the attribution of a DOI. The results are discussed under three aspects: the development since 2019, reuse rights and identifiers.

Introduction

Open repositories and in particular institutional repositories have been described as “home for grey literature” (Luzi, 2010). They are complementary channels for the dissemination of academic research output insofar they contain other versions of published articles, chapters or books (preprints) and, moreover, documents that have (and probably will) not be published via the usual channels of academic publishing. Part of the open science movement, open repositories are a major factor for the global and free dissemination of research results. Following the EU Open Science Monitor¹, they represent about 25% of the published research, with large differences between countries and disciplines.

Since more than ten years, we assess the deposit of grey literature in open repositories, especially in France, to provide empirical elements for a better understanding of this part of scientific communication and to make recommendations for a better findability and accessibility of grey

¹ EU Open Science Monitor https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/trends-open-access-publications_en

resources in repositories (see Schöpfel & Prost, 2010, 2014; Schöpfel et al., 2018, 2019, 2020; Stock & Schöpfel, 2009). Among our main findings:

- All open repositories contain grey literature².
- About one third of all deposits is grey literature.
- The part of grey literature is increasing over the years.
- The accessibility (degree of openness) of grey literature is higher than of commercial academic publications.
- The accessibility of grey literature varies between different repositories and between different document types, theses, reports and working papers being generally more open than conference papers.

In 2020, we presented original empirical results of 973,968 HAL deposits (30% of the total HAL content³) of more than 1,200 research laboratories from the ten most distinguished French research universities, including the University of Paris-Saclay ranked #13 by the 2021 Shanghai ARWU⁴. These laboratories cover the whole range of scientific disciplines, including medicine, law, economics and management. Based on this research, our paper contributes to a deeper understanding of the place of grey literature in open repositories through three complementary studies:

1. The evolution of the deposit of grey literature: is the part of one third stable over the time? How does the composition of grey literature change with the years?
2. The licensing of grey literature: what are the conditions of reuse of grey literature? Is open access to grey literature more “gratis” or more “libre”?
3. The existence of DOI for conference papers: which is the part of commercial publishing of conference papers?

In 2020, the major French research organisation CNRS decided the mandatory use of HAL for the reporting and assessment of the performance of the CNRS research laboratories and individual researchers. For this reason, the analysis of the 2021 deposits provides an exhaustive and reliable photography of the academic output and of the part of the grey literature.

Methodology

This paper presents results from an ongoing research project on open access strategies of French research laboratories⁵; it is a follow-up of two recent studies and employs the same methodology (Schöpfel et al., 2019, 2020). The sample consists of 1,213 research laboratories from the ten universities part of the French excellence initiative (IDEX). We used the laboratories’ identifiers in the HAL repository for the API search of each laboratory’s deposits (February 2021). The results were analysed based on the laboratory and deposit metadata (domain and discipline, university, deposit type, resource category). The API search retrieved 26 different HAL resource categories⁶; we merged

² Grey literature has been defined in various manner (Farace & Schöpfel, 2010; Schöpfel, 2011). Usually, grey literature is described as “unpublished”, “not peer reviewed” and “not in databases” and meaning most of the time reports and conference papers (Schöpfel & Prost, 2020).

³ HAL is the national open repository for French public research, with more than 2.6 deposits in August 2021 <https://hal.archives-ouvertes.fr/>

⁴ Academic Ranking of World Universities <https://www.shanghairanking.com/>

⁵ Project HAL/LO <http://gis-reseau-urfist.fr/hal-lo-valorisation-sur-hal-de-la-production-des-laboratoires-dans-lenvironnement-de-la-science-ouverte/>

⁶ These categories may change in the future; in April 2021, a working group recommended a revision of the actual typology, with some new document types, such as data papers and data management plans; cf.

and described them in the same way as in 2020 in order to simplify the description of the results (see Annex 1). In addition to the former assessment, we also analysed the availability of a DOI and the use of an open license.

Results

The part of grey literature

The API query retrieved 1,035,612 deposits which have been authored or co-authored by scientists affiliated to one of the 1,213 research laboratories of our sample. From all these items, 33.4% fall under the category of grey literature (figure 1).

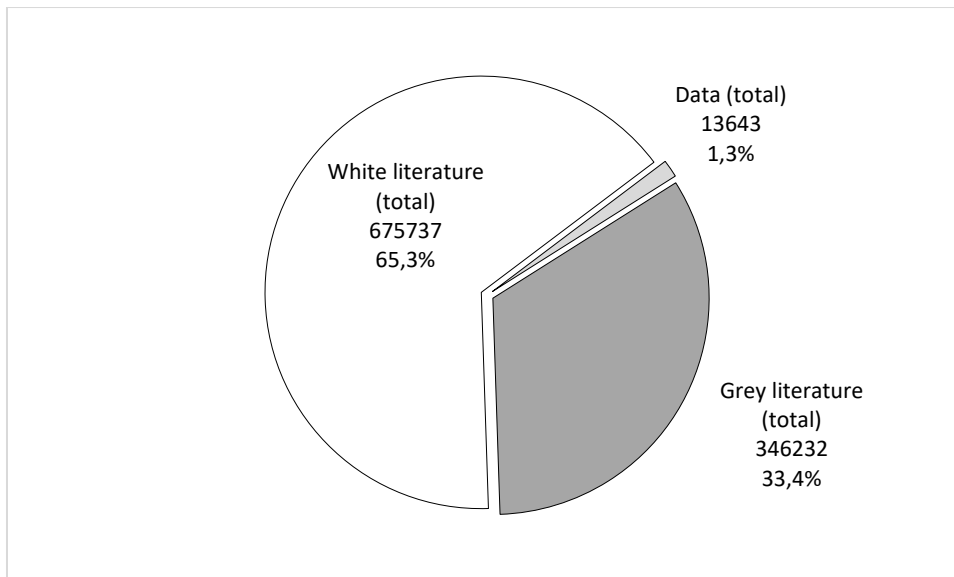


Figure 1. Part of grey literature (N=1,035,612 deposits)

Most of the grey literature consists of conference papers (70%), followed by PhD theses (11%) and working papers or preprints (8%). The different types of reports (project reports, activity or annual reports, short reports and report chapters) represent 5% while posters represent 4%. Other resource types are less important, such as BA and Master dissertations, habilitation theses or lectures, totalling together 2% (figure 2).

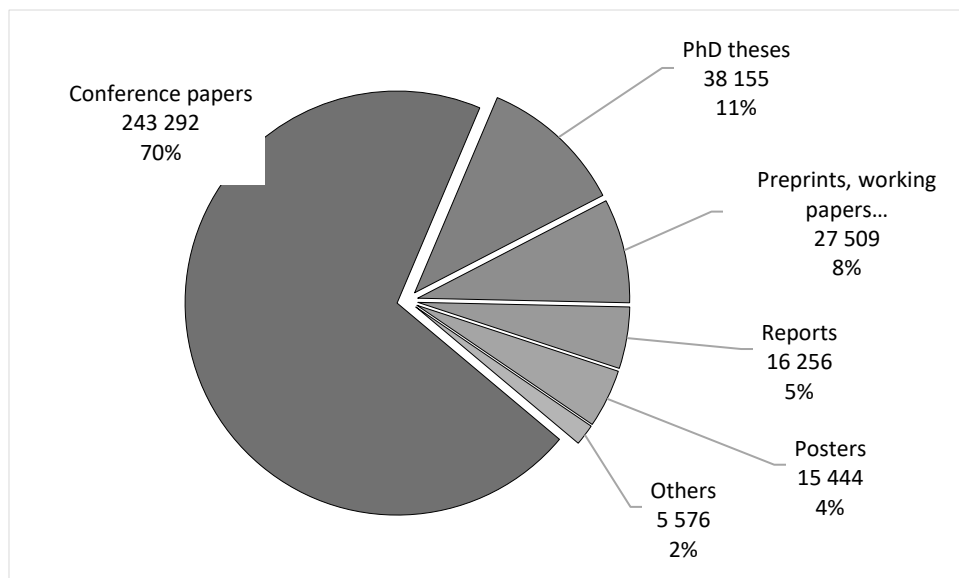


Figure 2. Types of grey literature (N=346,232 deposits)

Figure 1 shows that 1% of the deposits are neither white nor grey literature but datasets. HAL was designed as a document server but contains since 2010 the MédiHAL portal with deposits of visual and sound data (still images, videos and sounds), produced within the framework of scientific research. Also, there are some software deposits (codes) due to the partnership between HAL and the international Software Heritage project.

Degree of openness

The HAL repository contains deposits of full text and files as well as records, i.e., metadata without documents or data files. The part of deposits of full text in our sample is 32.3% which is similar to the overall percentage in HAL (32.9% in August 2021). This part of freely and openly available research output can be interpreted as degree of openness. If all deposits would consist of metadata and data (documents), this degree of openness would be 100%.

Figure 3 shows that the part of items with document and/or data files is significantly higher for grey literature (37.6%) than for white literature (28.2%).

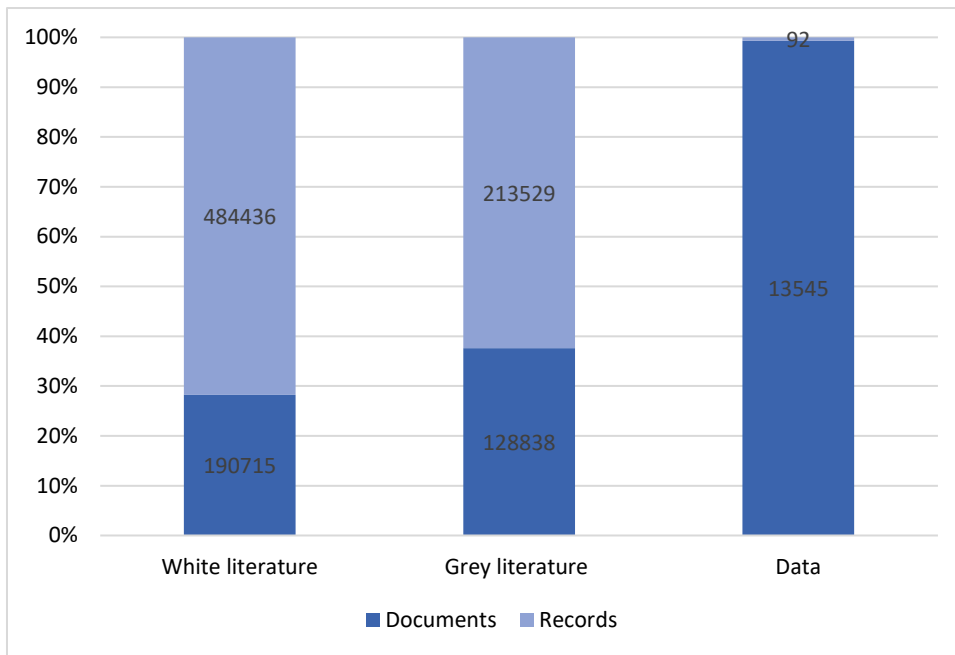


Figure 3. Deposits with and without document and data files (N=1,031,155 deposits, without annex files)

The differences between the document types are important. Figure 4 shows the part of full text for the different types of grey literature.

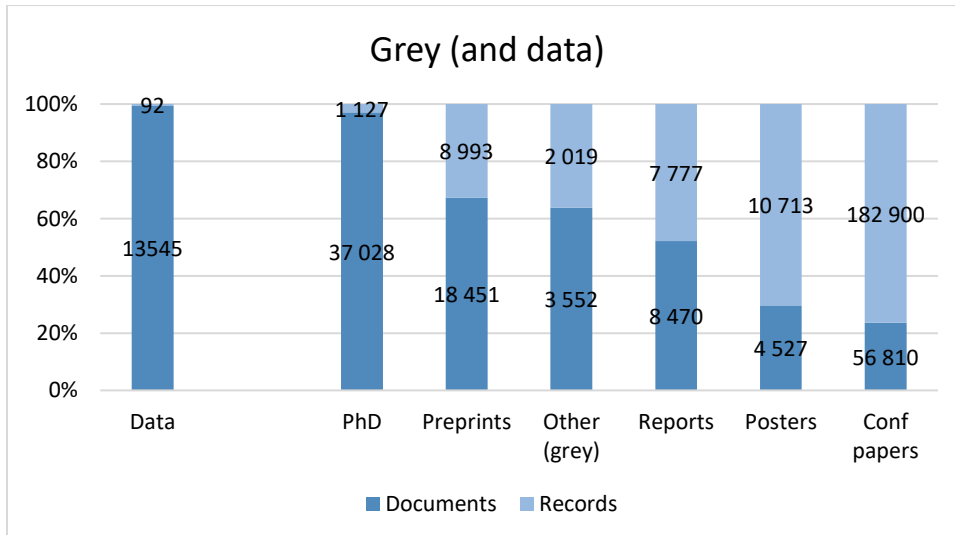


Figure 4. Grey literature deposits with and without document files (N=346,232 deposits)

The degree of openness of grey deposits is generally well above the average percentage, between 50% and near to 100%. The explanation of the exceptional part of openly available PhD theses (97%) is that the self-archiving of a PhD thesis on the HAL platform requires systematically the deposit of the text file.

The open part of the conference contributions, papers and posters, are lower, with 29% open posters and 23% open papers. This lower degree of openness is similar to the percentage of full text deposits of articles, books, chapters etc. (figure 5).

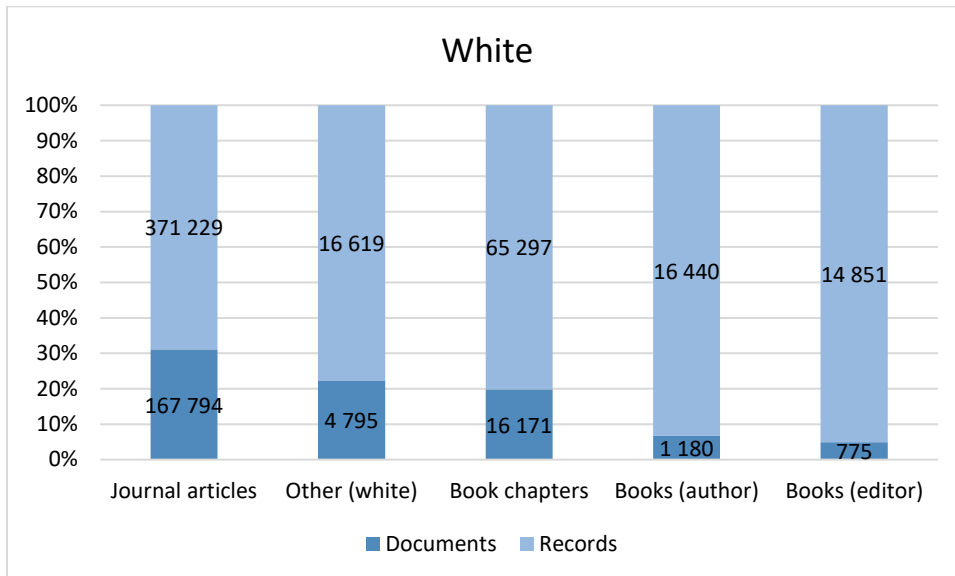


Figure 5. White literature deposits with and without document files (N=657,737 deposits)

31% of the journal articles are openly available on the HAL platform; the degree of openness of chapters, edited and authored books and other white resources is even lower.

Two other observations may be interesting. First, the part of grey literature of all deposits with full text is 40% which is higher than the overall part of grey literature (34%). Second, figures 3 and 4 include the degree of openness of datasets which is exceptionally high (99%); nearly all datasets have been deposited with the data files. Again, the reason is rather simple: MédiHAL, the data portal of HAL, requires the deposit of the data files for each data deposit. Only some software records have been created without the code files.

Disciplinary differences

Each research laboratory has been indexed with a large scientific domain and with a more specific research discipline. In the following, we present the analysis of the HAL deposits regarding four large scientific domains, i.e., science and technology (SciTech), life and medical sciences (BioMed), social sciences and humanities (SSH), and law, economics and management (Law, Econ). Figure 6 shows that there are significant differences between the four domains.

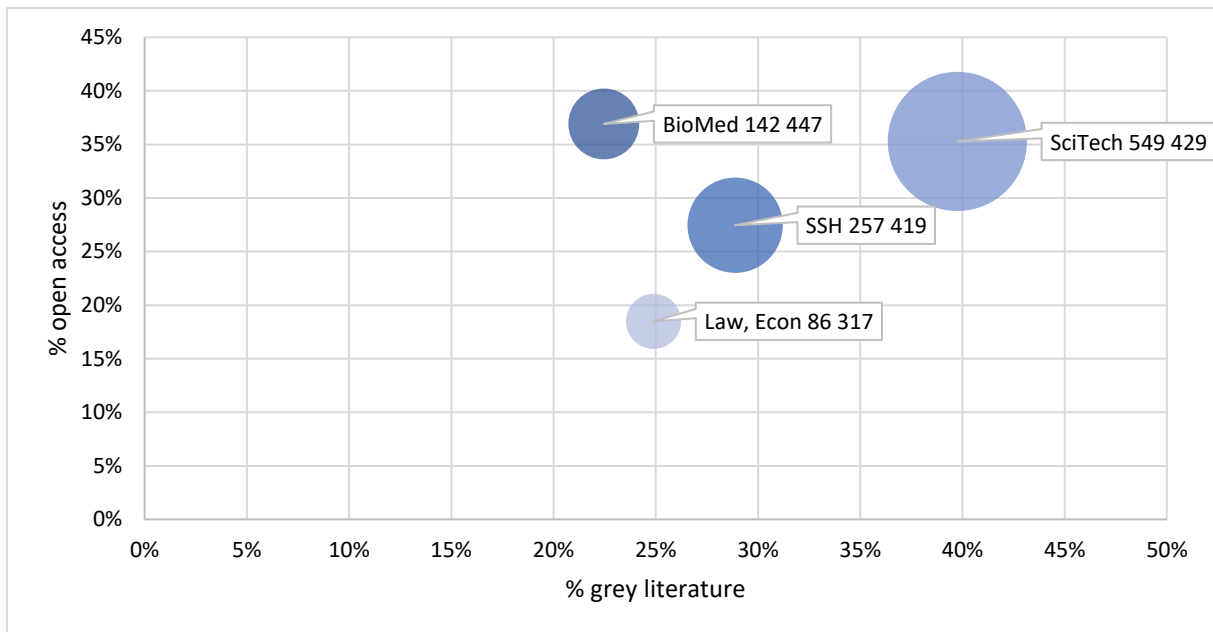
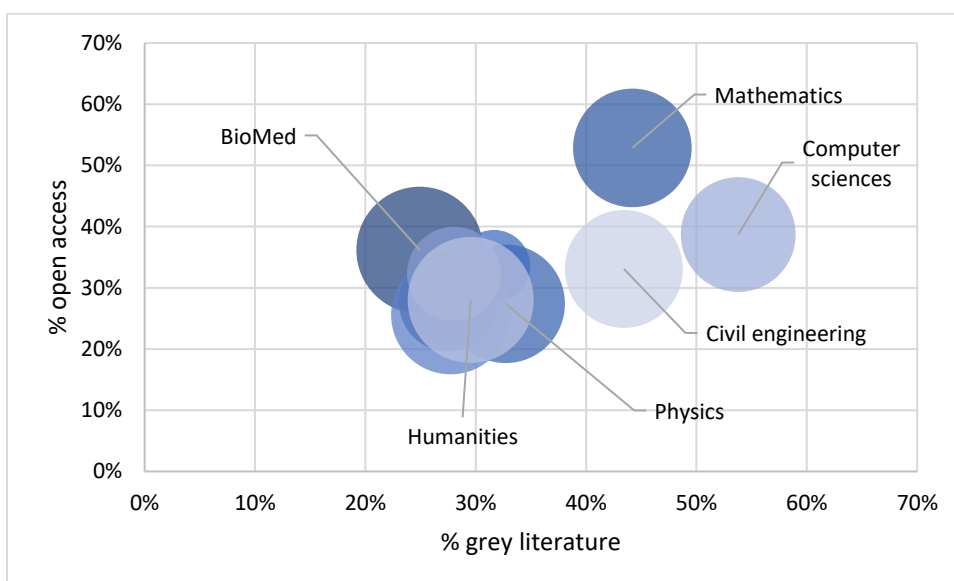


Figure 6. Degree of openness and part of grey literature in four scientific domains (N=1,035,612 deposits)

The deposits of the laboratories in life and medical sciences have the highest degree of openness (37%), followed by those in science and technology (35%). The same indicator is lower in social sciences and humanities (27%) and in law, economics and management (18%).

On the other hand, the laboratories in science and technology have the highest part of grey literature (40%), followed by those in social sciences and humanities (29%), law, economics and management (25%) and life and medical sciences (22%).

Figure 7 presents additional results with a more detailed distinction of ten scientific disciplines; the size of the bubbles represents the number of deposits (for the complete figures, see Annex 2).



*Figure 7. Degree of openness and part of grey literature per discipline
(N=1,035,612 deposits)*

For most disciplines, the grey part of the HAL deposits represents more or less 30%, varying from 25% (life and medical sciences) to 33% (physics). Three disciplines have significantly more grey literature: civil engineering (43%), mathematics (44%) and computer sciences (54%). Conference papers represent most of this grey literature.

Regarding open access, the part of openly accessible documents is about 30% for most disciplines, varying from 26% (social sciences) to 36% (life and medical sciences). The degree of openness is higher for computer sciences (39%) and particularly for mathematics (53%).

The mapping of document types against scientific disciplines reveals significant differences regarding open access (degree of openness) and relative importance (part of all deposits) (see table in annex 3). Some observations:

- The only grey document type that is really important are conference papers, especially in life and medical sciences (16%), mathematics (9%) and computer sciences (7%). Compared to the overall number of deposits, other grey resources like reports, preprints or working papers are much less important (even if they may contain unique and significant content).
- For conference papers, the degree of openness is higher in mathematics (39%), humanities (37%) and computer sciences (33%) than in physics (13%) or social sciences (12%).
- Regarding reports, mathematics, computer sciences and civil engineering are the disciplines with the highest degree of openness (>80%).
- Regarding working papers and preprints, humanities, mathematics and computer sciences are the disciplines with the highest degree of openness (>70%).

DOI

A digital object identifier (DOI) is a persistent identifier used to identify objects uniquely. The DOI system has been introduced in 1998 and standardized in 2012 (ISO 26324). Following the DOI website, approximately 257 million DOI names have been assigned to date⁷. While the initial focus on entities was documents/media (e.g., articles and data sets), the DOI system is now moving into parties and licences and extending to other sectors.

The initial initiative was taken by the three major international publishing trade associations in order to develop infrastructure for digital publishing. Today, a large part of DOIs is still attributed to articles, book chapters etc. by stakeholders of the traditional publishing economy.

The analysis of the HAL deposits provides the opportunity to assess the part of DOIs assigned to non-conventional literature, such as dissertations, reports, working papers and communications.

Of 1,179,145 analysed items, 486,474 have a DOI (41%). 90.5% of these deposits with DOI are white literature, mainly journal articles (88%) and book chapters (2%). 9.5% of the deposits are grey literature, mainly communications (9%), posters (0.2%) and preprints (0.3%).

⁷ DOI <https://www.doi.org/factsheets/DOIKeyFacts.html>

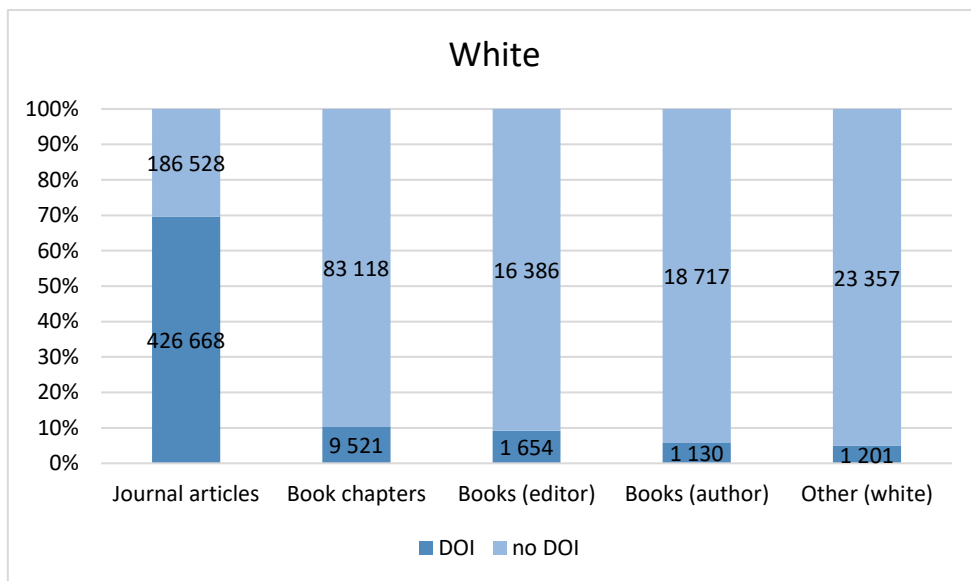


Figure 8. White literature with DOI (N=768,280 deposits)

While 70% of all deposited articles have a DOI (the percentage for book chapters is 10%, for edited books 9%) (figure 8), the part of communications with DOI is only 15% (posters 5%, preprints 4%) (figure 9).

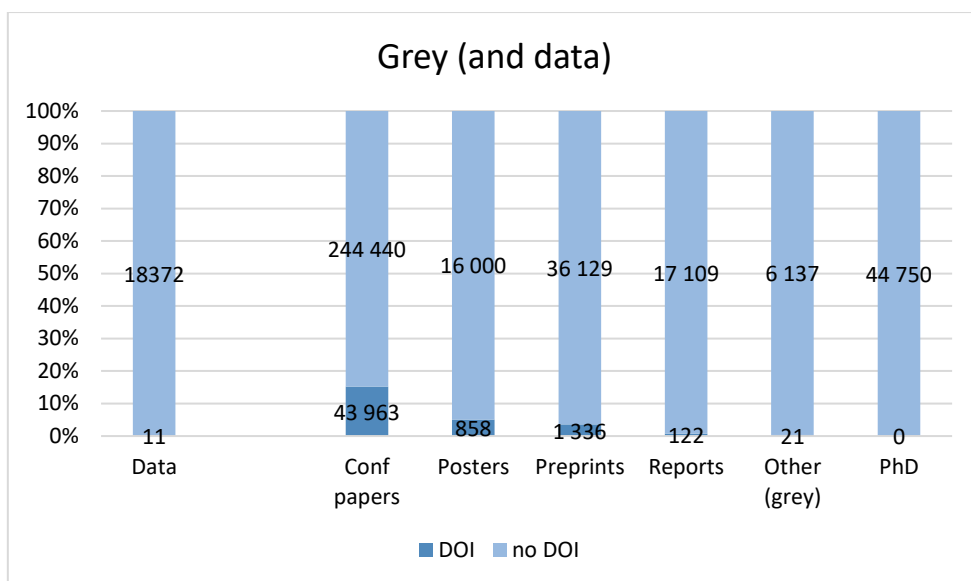


Figure 9. Grey literature with DOI (N=410,865 deposits)

Licensing

By default, the deposits of documents on HAL are published under the French IP law. However, HAL offers the possibility to publish with a license, under copyright or in the public domain, with ten different options. We analysed the legal regime of 351,672 published documents on HAL (table 1).

Legal conditions	Documents	%
By default (French IP law)	298 906	85,0%
Copyright	9 972	2,8%
CC-BY	19 309	5,5%
CC-BY-SA	2 178	0,6%
CC-BY-NC	4 003	1,1%
CC-BY-NC-SA	2 186	0,6%
CC-BY-ND	1 078	0,3%
CC-BY-NC-ND	10 394	3,0%
CC0 Public Domain Zero	82	0,0%
NC Public Domain Mark	29	0,0%
Etalab Open License	99	0,0%
Public domain	3 436	1,0%

Table 1. Legal regime of publishing (N=351,672)

85% of the documents are published under the French IP law, by default. 11% of the documents are published with a Creative Commons⁸ or Etalab⁹ open license, 1% are published in the public domain, with a CC Zero Public Domain Dedication (CC0), with a NC Public Domain Mark or with a simple public domain indication (figure 10).

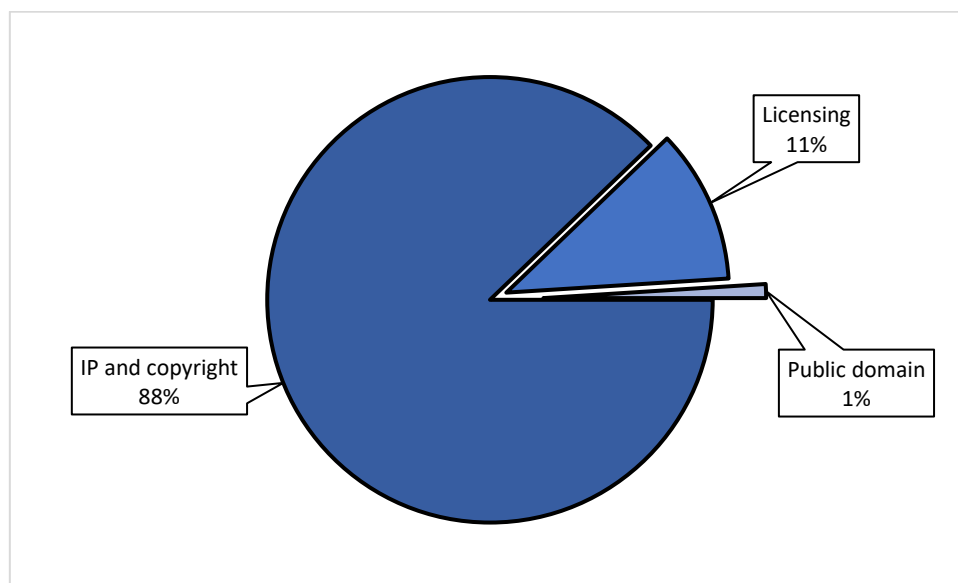


Figure 10. Legal regime of publishing (N=351,672)

39,148 documents are published with a Creative Commons license (11%, without CC0). Nearly half of these documents are published with a (liberal) CC-BY Attribution license, 26% are published with a (more restrictive) CC-BY-NC-ND license which doesn't permit any commercial use and any derivatives or adaptations of the work (figure 11).

⁸ Creative Commons <https://creativecommons.org/>

⁹ Etalab is the French public agency in charge of the open data strategy <https://www.etalab.gouv.fr/>

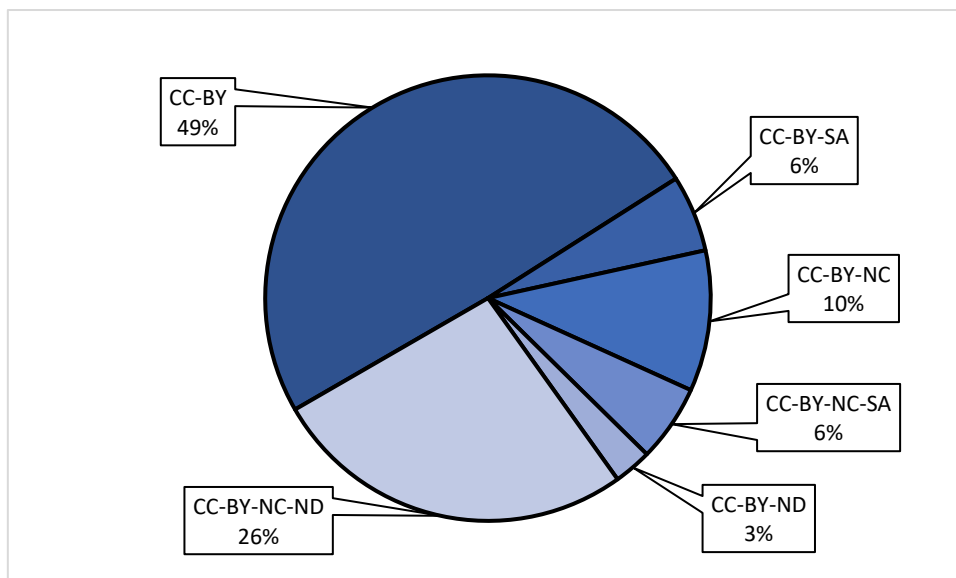


Figure 11. Creative Commons licenses (N=39,148)

For a small subsample of 22,508 deposited files from 70 laboratories we analysed the distribution of open licenses and document types. The results can be resumed as follows:

	Grey documents	White documents
CC-BY	1114	3676
CC-BY-SA	377	262
CC-BY-NC	205	501
CC-BY-NC-SA	271	352
CC-BY-ND	81	454
CC-BY-NC-ND	941	2316

Table 2. Publishing with open licenses (N=10,550)

- Grey items are less often published under an open license than articles, book chapters and other “white” documents. In fact, open licenses are most often applied to the sharing of research data and other materials, like images, AV files or software.
- Regarding different grey items, especially PhD theses are very seldom published under an open license, less often than posters, reports and other, miscellaneous documents.
- The most important licenses applied to grey literature are the liberal CC BY (Attribution) license (37%) and the restrictive CC BY-NC-ND (Attribution-NonCommercial-NoDerivs) license (31%). The main difference with articles, books and so on is that the part of the Wikipedia CC BY-SA (Attribution-ShareAlike) and the more restrictive CC BY-NC-SA (Attribution-NonCommercial-ShareAlike) licenses is more important for the publishing of grey literature (together 21%, against 8% for white items) (table 2).
- Publishing in open domain, for instance with a CC0 (Creative Commons Public Domain Dedication), plays nearly no role for grey items, similar to articles, books and so on.

Discussion

Changes

Since 2020, the number of grey items in our sample of more than 1,200 research laboratories increased by 2.5% (= 8,491 new deposits). This is less than the overall growth of all deposits (6.3%) and less than the articles, book chapters etc. (white literature) of which the number has risen by 8.1%. As a result, the part of grey literature decreased from 34.7% in 2020 to 33.4% in 2021. This is a general evolution in all domains and disciplines, even if in some disciplines (especially in civil engineering, computer sciences and physics) the observed decrease is a little bit more important than in others.

There are probably two main reasons for this slight relative decrease. The first is the new open access policy launched by the French research organisation CNRS in 2020; researchers from the CNRS laboratories must deposit their publications on the HAL platform for their individual annual assessment, and as this assessment focusses on published articles, their number and part increased in a significant way. The second reason is the secondary exploitation right introduced by the 2016 Digital Law which foster the deposit of published articles on open repositories and in particular, on HAL, and which doesn't include other document types.

However, this development should not be described as an erosion but is probably more as a kind of ceiling or capping of the part of grey literature at a level of about 33%.

Simultaneously, it can be observed that the part of open access among the grey literature slightly increased from 36.6% in 2020 to 37.6% in 2021, which is consistent with the open access strategy of the French government, the universities and research organisations. The differences of openness between the types of grey literature remained stable and unchanged, except for working papers and preprints where the degree of openness increased by 3%. In a more general way, grey items remain more accessible than articles and books, with one exception, like in the past: the openness of communications, i.e., the accessibility of papers presented at scientific events, remains low. Again, there may be at least two reasons for this lack of openness: first, for one part of the communications, there may be simply no written and publishable papers; second, another part of the communications is not really grey literature, as defined by the Luxemburg, New York or Prague definitions, insofar they are published (and controlled) by traditional academic publishing houses. The attribution of DOIs for 15% of the deposited communications is an indication for this distinction.

Reuse rights

Globally, IP and copyright protection is dominant. Only a small part of documents has been published with open licenses, mostly with the liberal CC-BY license which is the license preferred and fostered by French public research institutions and with the more restrictive CC-BY-NC-ND license. Very few documents have been published in the open domain. Regarding the FAIR principles, this means that the reusability of most of the HAL documents is not very high, in contrast to the research data and other materials deposited on HAL which are more often published under an open license. The results of a small subsample of 70 laboratory collections show that the two Creative Commons ShareAlike licenses are more often used for grey literature than for the sharing of journal articles, book chapters and so on, and that very few PhD theses are published under an open license. However, because of the small size of this subsample we should be careful and avoid generalizing interpretations of these results.

Identifiers

The attribution of DOIs is not limited to documents published by commercial academic publishing houses but in fact, it is. Up to now, very few academic or research institutions assign DOI names to their research output. Perhaps the growing DataCite network will change the game but for the moment, the DataCite DOIs are above all assigned to research data and software, not to other items, even if (at least in France) academic libraries started a discussion about the attribution of DataCite DOIs to documents like dissertation and theses, reports and so on.

Two aspects should be highlighted: As we have already said before, as long as DOIs are not massively assigned to grey literature like dissertation, reports, working papers, conference papers and so on, these documents will remain largely out of scope of altmetrics tools and open science monitoring, as both are mainly (exclusively) based on DOIs (Schöpfel & Prost 2016 and 2019). “Researchers wishing to assess impact of their own grey literature should endeavour to create DOIs where appropriate so persistent identifiers already used to assess impact by other mediums (Altmetric, n.d.) can be extended into the grey literature realm” (Bickley et al., 2021).

The fact that 15% of the deposited communications on HAL have a DOI can be seen as an indication that these deposits are probably not natively grey but white literature, published in proceedings and journals by publishers like Springer Nature, Elsevier, IEEE and so on. Here, we should be careful with the application of the concept of grey literature to documents which in fact are not (are no longer or have never been) grey items.

We already discussed the “greyness” of electronic theses and dissertations (ETDs) elsewhere (Schöpfel & Rasuli, 2018). In France, ETDs became mandatory in 2016. Since then, each PhD dissertation must be submitted to the French national ETD infrastructure STAR, will be available (if not confidential) via the national ETD portal theses.fr and the European portal DART-Europe and is preserved in a national and sustainable long-term dark archive hosted by the public agency CINES. And if not decided otherwise by the author, it will be accessible on the HAL platform (TEL portal) or on an institutional repository, embargoed or not. Thus, French theses have become significantly since 2016 less grey than before, and that is good news.

Conclusion

Our paper presents empirical results of a follow-up study on the deposits of more than 1,200 French research laboratories on the national open repository HAL. The figures show a significant increase of all deposits since 2020 (+6%), a growth which seems consistent with the global development of scientific performance and with the open science strategy of the French government and institutions.

As in 2020, the part of grey literature is about one third of all deposits, and this part is generally more accessible (with more deposits of document files) than articles and books. The general impression is that of continuity and stability, without significant changes since last year. However, it can be observed that HAL has become (a bit) more open and at the same time, (a little bit) less grey. We have provided some possible reasons for this development, such as the open science strategy of the French government, the secondary exploitation right for French public research and the mandatory open access policy of the CNRS.

In terms of FAIRness, the situation can be described as follows:

FAIR principle	General observation	Grey documents
Findability	Less than half of the deposits have a DOI	Weaker than articles

Accessibility	All deposits are retrievable on HAL with the standard OAI protocol but only for one third the document is accessible	Globally better than for articles and books
Interoperability	n/a	n/a
Reusability	Weak – only 15% items are released with a license	Similar

The comparison between white and grey deposits shows that the two FAIR principles findability and reusability remain a challenge for the production and the publishing of grey literature on HAL, because of the low levels of DOI assignment and of licensing.

As the deposit of publications on HAL becomes more mandatory and thus, more representative for the overall output of French research laboratories, it is possible that we can see here a kind of ceiling of grey literature, on different levels depending on disciplines and institutions. It will be interesting to assess the impact of new forms of scientific communication, such as webinars or preprints, especially in the aftermath of the Coronavirus pandemics. For the moment, we can't see any significant effect on the HAL platform, probably because these new forms are not relevant for the research assessment, career decisions and so on.

The development of conference papers in particular and of all kind of resources related to scientific events in general will be particularly interesting for the understanding of the future of grey literature. They are not only the most important part of the traditional "grey document types" but they are also somehow on the edge, partly grey, and partly controlled by commercial publishing.

The future will also show how the international (European) open science policy in favour of gold open access (article publishing in open access journals) will impact the development of open repositories like HAL. If more and more articles are freely available on journal platforms, maybe that those repositories will become (again) a home for grey literature, not controlled by commercial publishing. Perhaps this will also change our understanding of this part of scientific communication.

Acknowledgments

The study is part of the research project HAL/LO funded by GIS "Reseau URFIST" from 2019-2021 with 10,000 euros. We would like to express our gratefulness to the whole project team for rich discussions and especially to Bernard Jacquemin for data extraction from the HAL platform, and we would also like to thank Bénédicte Kuntziger from the CCSD for helpful advice.

References

- Bickley, M., Kousha, K., & Thelwall, M. (2021). A systematic method for identifying references to academic research in grey literature. *18th International Conference on Scientometrics & Informetrics Proceedings*, 121–132.
- Farace, D., & Schöpfel, J. (Eds.). (2010). *Grey Literature in Library and Information Studies*. Berlin, De Gruyter Saur.
- Luzi, D. (2010). Grey Documents in Open Archives. *The Grey Journal*, 6(3), 137–144.
- Schöpfel, J., & Prost, H. (2010). *Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement*. Université de Lille 3.
- Schöpfel, J., & Prost, H. (2014). Degrees of Openness. Grey Literature in Institutional Repositories. In *GL16 Sixteenth International Conference on Grey Literature. Grey Literature Lobby: Engines and*

Requesters for Change. 8-9 December 2014, Library of Congress, Washington D.C., USA (pp. 75–85). Amsterdam: TextRelease.

Schöpfel, J., & Prost, H. (2016). Altmetrics and Grey Literature: Perspectives and Challenges. *GL18 Eighteenth International Conference on Grey Literature*. 28-29 November 2016, New York Academy of Medicine, New York NY, USA.

Schöpfel, J., & Prost, H. (2019). The scope of open science monitoring and grey literature. *12th Conference on Grey Literature and Repositories*, National Library of Technology (NTK), 17 October 2019, Prague, Czech Republic.

Schöpfel, J., & Prost, H. (2020). How scientific papers mention grey literature: a scientometric study based on Scopus data. *Collection and Curation*, 40(3), 77-82.

Schöpfel, J., & Rasuli, B. (2018). Are electronic theses and dissertations (still) grey literature in the digital age? A FAIR debate. *The Electronic Library*, 36(2), 208–219.

Schöpfel, J., Prost, H., Fraise, A., & Chaudiron, S. (2018). Valoriser les publications d'un laboratoire universitaire dans l'environnement de la science ouverte : Retour d'expérience de la collection GERiICO sur HAL. In *ICOA 2018 3e colloque international sur le libre accès*, 28-30 November 2018, Rabat, Morocco.

Schöpfel, J., Prost, H., & Ndiaye, E.H.I. (2019). Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL. In *GL2019, 21st International Conference on Grey Literature. Open Science Encompasses New Forms of Grey Literature*, 22-23 October 2019, TIB, Hannover, Germany.

Schöpfel, J., Kergosien, E., Prost, H., & Thiault, F. (2020). The Grey Side of the Green Road. Empirical Assessment of Academic Publishing in the HAL Open Repository. *GL2020. 22nd International Conference on Grey Literature, Applications of Grey Literature for Science and Society*, 19 November 2020, Rome, Italy.

Stock, C., & Schöpfel, J. (2009). Grey literature in French digital repositories: a survey. In *GL10 Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*, Amsterdam, 8-9 December 2008 (pp. 39–53). Amsterdam: TextRelease.

Annex

Annex 1 - Resource categories and types

HAL category	Merged category	Resource type
Journal article	Articles	White
Book (author)	Books (author)	White
Book (editor)	Books (editor)	White
Book chapter	Book chapters	White
Other	Other (white)	White
Communication	Conference papers	Grey
Poster	Posters	Grey
PhD thesis	PhD theses	Grey
Report	Reports	Grey
Activity report	Reports	Grey
Short report	Reports	Grey
Report chapter	Reports	Grey

Other report	Reports	Grey
Undefined	Preprints, working papers...	Grey
Habilitation	Other (grey)	Grey
Lecture	Other (grey)	Grey
Master dissertation	Other (grey)	Grey
BA dissertation	Other (grey)	Grey
Lecture note	Other (grey)	Grey
Presentation	Other (grey)	Grey
Patent	Other (grey)	Grey
Image	Data	Data
Map	Data	Data
Software	Data	Data
Audio	Data	Data
Video	Data	Data

The HAL category “other” consists of book reviews, encyclopaedia entries, translations etc., most of them reviewed publications. The HAL category “undefined” consists mainly of working papers, preprints and other, non-reviewed and unpublished documents. Some of the HAL resource categories are specific to a particular, institutional workflow (i.e., ingestion of records from institutional partnerships), such as BA dissertations, lecture notes and report chapters; they represent very small figures (<100) and have been merged (reports) or included in the “other” category.

Annex 2 – Disciplinary differences

	Nb of deposits	% grey literature	% open access
BioMed	225 941	25%	36%
Chemistry	135 541	28%	28%
Mathematics	87 176	44%	53%
Physics	197 509	33%	27%
Agronomy, ecology	73 530	32%	34%
Social sciences	199 633	28%	26%
Earth sciences, space	124 108	28%	32%
Computer sciences	185 032	54%	39%
Humanities	223 067	30%	28%
Civil engineering	195 253	43%	33%

Annex 3 – Disciplinary differences and document types

For three grey document types: conference papers, reports and working papers/preprints.

Discipline	Nb of labs	Document type	Deposits	Records	Documents	% deposits	% open access
Agronomy, ecology	21	Conference papers	3 080	2 534	505	4,2%	16,4%
BioMed	356	Conference papers	35 324	25 778	8 816	15,6%	25,0%
Chemistry	38	Conference papers	3 523	2 763	675	2,6%	19,2%
Civil engineering	37	Conference papers	8 933	6 505	2 276	4,6%	25,5%
Computer sciences	24	Conference papers	13 780	9 047	4 526	7,4%	32,8%
Earth sciences, space	10	Conference papers	3 790	2 663	1 047	3,1%	27,6%
Humanities	24	Conference papers	2 242	1 316	821	1,0%	36,6%
Mathematics	12	Conference papers	7 732	4 634	3 027	8,9%	39,1%
Physics	31	Conference papers	8 038	6 877	1 078	4,1%	13,4%
Social sciences	13	Conference papers	1 109	962	131	0,6%	11,8%
Agronomy, ecology	15	Reports	255	155	100	0,3%	39,2%
BioMed	158	Reports	1 717	641	1 076	0,8%	62,7%
Chemistry	14	Reports	81	30	51	0,1%	63,0%
Civil engineering	20	Reports	147	29	118	0,1%	80,3%
Computer sciences	13	Reports	454	65	389	0,2%	85,7%
Earth sciences, space	8	Reports	277	148	129	0,2%	46,6%
Humanities	15	Reports	83	51	32	0,0%	38,6%
Mathematics	8	Reports	364	36	328	0,4%	90,1%
Physics	14	Reports	140	74	66	0,1%	47,1%
Social sciences	11	Reports	126	94	32	0,1%	25,4%
Agronomy, ecology	15	Preprints, working papers	261	180	80	0,4%	30,7%
BioMed	240	Preprints, working papers	3 369	1 082	2 272	1,5%	67,4%
Chemistry	26	Preprints, working papers	575	290	284	0,4%	49,4%
Civil engineering	32	Preprints, working papers	686	221	465	0,4%	67,8%
Computer sciences	20	Preprints, working papers	578	146	429	0,3%	74,2%
Earth sciences, space	10	Preprints, working papers	538	288	250	0,4%	46,5%
Humanities	23	Preprints, working papers	183	33	150	0,1%	82,0%
Mathematics	11	Preprints, working papers	1 003	288	712	1,2%	71,0%
Physics	27	Preprints, working papers	1 091	465	625	0,6%	57,3%
Social sciences	11	Preprints, working papers	219	126	92	0,1%	42,0%