



HAL
open science

Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL

Joachim Schöpfel, Hélène Prost, El Hadji Ibrahima Ndiaye

► **To cite this version:**

Joachim Schöpfel, Hélène Prost, El Hadji Ibrahima Ndiaye. Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL. GL21 International Conference on Grey Literature, 22-23 October 2019, Hannover (Germany), Oct 2019, Hannover, Germany. hal-02300017

HAL Id: hal-02300017

<https://hal.science/hal-02300017>

Submitted on 12 Nov 2019

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.

Going Green

Publishing Academic Grey Literature in Laboratory Collections on HAL

Authors

Joachim Schöpfel (corresponding author), GERiiCO Laboratory, University of Lille, France

joachim.schopfel@univ-lille.fr

Hélène Prost, associated member of GERiiCO Laboratory, CNRS, France

helene.prost@inist.fr

El Hadji Ibrahima Ndiaye, University of Lille, France

elhadji.ndiaye.etu@univ-lille.fr

Keywords

Open science, open access, open repositories, research laboratories, grey literature, reports, conference papers, working papers, theses and dissertations

Abstract

HAL is the national open repository for documents and data from French scientists. The deposits are organized in institutional portals and collections from research laboratories and projects. The paper analyses how grey literature is represented in the collections of French research laboratories on HAL. We assess the grey literature deposits on HAL from the 66 research laboratories affiliated to the University of Lille, covering all STM and SSH fields. The focus is on conference papers, reports, working papers, theses and dissertations. The study distinguishes between deposits of documents and records without documents, compares deposits from different disciplines, different laboratory collections and different document types. Typical strategies (or lack of strategies) on the local level of research laboratories are identified. Conditions and variables that may explain these differences are discussed, together with potential effects on the visibility, impact and evaluation of the laboratories' research output.

Introduction

Last year, in 2018, the French Minister of Higher Education, Research and Innovation announced a National Plan for Open Science¹. The plan defines open science as the practice of making research publications and data freely available, open to all, without hindrance, without delay, without payment. The first commitment of the plan is to generalise open access to publications through open access platforms, whether in journals or books or through an open public repository such as HAL². This commitment includes the confirmation of the main role of the HAL open repository in the French ecosystem of open science infrastructures. Launched in 2001 by the Centre for Direct Scientific Communication (CCSD) of the CNRS, the multidisciplinary HAL archive (= "Hyper Articles on Line") has

¹ National Plan <http://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html>

² Hyperarticle en ligne <https://hal.archives-ouvertes.fr/>

become over the years one of the most important platforms of the "green road" to open access to scientific information.

Green road means the self-archiving of scientific publications by the authors themselves, on a dedicated open platform or repository (Harnad et al., 2004). The green road strategy is set up on two stakeholders: the authors, insofar as they hold the intellectual rights to deposit their own publications (article, chapter, communication, thesis, etc.), and their institution (research organisation, university, school, etc.), insofar as it has the possibility of encouraging or enforcing self-archiving and since it also has the resources and legitimacy for an institutional archive (Lynch, 2003). In France, the green road is facilitated by the 2016 Law for a Digital Republic and the creation of a secondary exploitation right for French researchers (Article 30, cf. CNRS-DIST 2016).

HAL is the "green heart" of the French open access infrastructure. Currently (September 2019), the repository contains more than 1.9 million items, mostly articles (55%) and conference papers (30%) but also book chapters (9%), dissertations (5%) and other text and data files in all disciplines. 32% of the items are document deposits, the other 68% are records, i.e. metadata without text or data files.

Like its American model, the arXiv e-prints service, the HAL repository was initially designed on the principle of direct communication among researchers, to facilitate and accelerate the exchange of scientific results even before they are published in a journal or book. With time, in particular after the signature at the Academy of Sciences on 2 April 2013 of the "Partnership Agreement in favour of open archives and the shared HAL platform" between French universities and research organizations, HAL has become a kind of national institutional repository, a "shared national infrastructure hosting institutional archives or towards which other institutional archives are firmly invited to release their content" (Bauin 2014).

Several hundreds of universities, research organisations, business and engineering schools and research laboratories have created their own portals or collections on the HAL platform as an institutional repository or as a digital showcase of their scientific output. Our focus is on the academic laboratory collections. In France, academic research is organized via university-based laboratories which are the researchers' working environment. The laboratories are the basic level of university research; research projects are organized around the laboratories, and researchers are evaluated by the High Council for the Evaluation of Research and Higher Education (HCERES) within the framework of their laboratories.

An open repository collection can contribute to the visibility and the impact of the laboratories' scientific production; it can produce some basic scientometrics (number and typology of scientific papers etc.) and altmetrics (views and downloads), and it can supply data for further assessment (internationality, network analysis etc.).

A preliminary study was conducted in 2018 on the HAL collection of the GERiICO laboratory³ in order to analyse some of these basic metrics and to assess not only the interest but also the required investment and potential shortfalls (Schöpfel et al. 2018). The following study takes this research a step further, with a sample of 66 laboratories covering the whole range of scientific disciplines, focussing on grey literature and open access strategies and comparing the 2019 situation with survey data from 2008 and 2009.

³ A research laboratory in information sciences, communication and cultural studies, affiliated to the University of Lille <https://geriico-recherche.univ-lille3.fr/>

Methodology

The 66 research laboratories of the University of Lille were selected based on the university's public list⁴. For each laboratory, we determined the name, the acronym and/or number, the type of research unit (university structure or mixed governance with research organisations), the field of research (arts, social sciences and humanities; science and technology; medical sciences and public health; law, economics and management) and the discipline. This was done with information from the university's and the laboratories' websites.

25 labs (38%) receive funding (budget, equipment, staff) from the University of Lille only while 41 labs (62%) are located on the campus of the University of Lille but partly funded by other French research organisations, e.g. CNRS (multidisciplinary, fundamental research), INSERM (medical science) or INRIA (applied computer science).

These 66 labs cover the whole range of scientific disciplines. Best represented are medicine (20), biology (7), engineering (5), chemistry (4) and pharmacology, physics, informatics and economics (each 3).

In a second step, we determined for each laboratory if it had an institutional identifier in the HAL system⁵ and if it had created a collection on the HAL repository⁶.

Third, we assessed for each laboratory the number of deposited documents (documents with full text) and of records (metadata without documents) through a direct search in the HAL repository⁷. We also assessed the number of documents and records for specific types of documents, i.e. articles, communications (conference papers), book chapters, preprints, reports and PhD dissertations. This selection was based on previous research on open repositories and on a pragmatic definition of grey literature, considering preprints, reports, PhD dissertations and partially also conference papers such as grey (non-commercial, unconventional) literature and articles and book chapters generally such as white (commercial) literature. These figures were analysed with descriptive statistics (uni- and bivariate analyses). Chi-squared tests were performed on a $p=.05$ level.

Finally, we tried to get usage statistics from the people in charge of these laboratory collections, i.e. download figures for the last year (2018) as an indicator of impact for the whole collection and for the different types of documents.

Results

Collections and deposits

Except for a research unit at the Faculty of Medicine, all 66 research labs of the University of Lille are represented on the HAL repository. However, only 16 labs (24%) have their own collection, while 49 (74%) have but an institutional identifier which allows to link to related documents and metadata from their researchers. As figure 1 shows, the collections of research labs in arts, social sciences and humanities are significantly overrepresented, while there is no collection from labs in medical science and public health.

⁴ <https://www.univ-lille.fr/recherche/laboratoires/> (accessed April 30, 2019)

⁵ Accès unifié aux référentiels HAL <https://aurehal.archives-ouvertes.fr/structure/index> and index of research units <https://hal.archives-ouvertes.fr/browse/structure> (accessed June 10, 2019)

⁶ Index of collections <https://hal.archives-ouvertes.fr/browse/collection> (accessed June 10, 2019)

⁷ In the following, "document" means deposited item with full text, "record" means deposited metadata without full text, "deposits" means both documents and records.

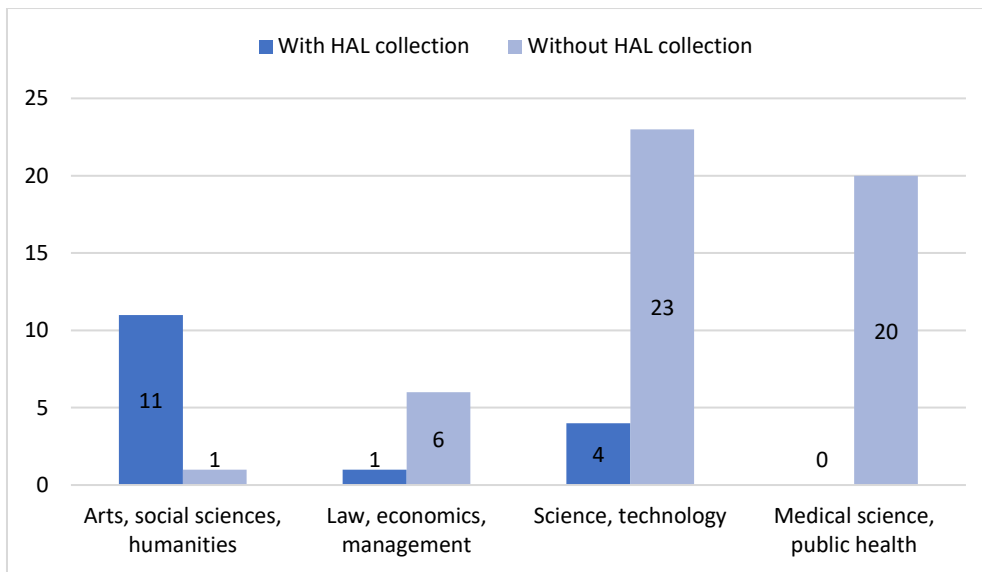


Figure 1. Number of laboratories with and without a collection on HAL (N=66 research labs)

This difference between scientific domains is even more significant on the level of deposits, i.e. the total of deposited documents and records without documents. The 66 labs of the University of Lille have 41,701 deposits on HAL; 26,158 deposits (63%) are produced by the 16 labs with a HAL collection (24%) while 15,543 deposits (37%) are produced by the other 50 labs without collections (76%). As figure 2 shows, the most significant differences are the distribution of deposits in art, social sciences and humanities (nearly no deposits outside of collections), in law, economics and management and in medical science and public health (both domains with few or no collections and a higher number of deposits outside of collections).

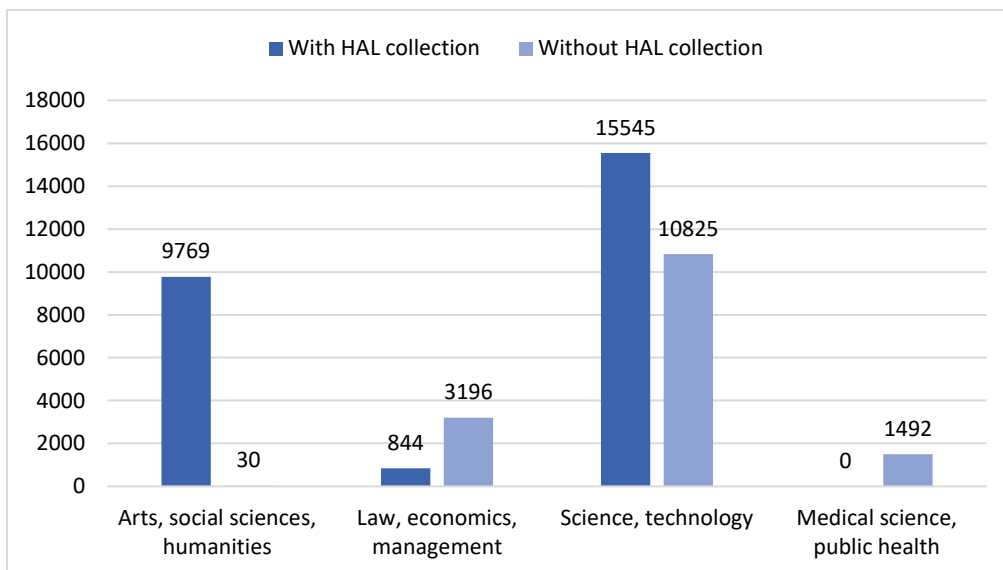


Figure 2. Number of deposits of laboratories with and without a collection on HAL (N=41,701 for all 66 labs)

The relationship between the creation of a collection on HAL and the number of deposits is significant. As figure 3 reveals, except for medical science which is without any collection on HAL, the median number of deposits per research lab is significantly higher for those with a collection than for those without (in brackets: only one lab). For all labs of the University of Lille, the median for those with a HAL collection is 764 deposits, i.e. 50% of the labs with a collection on HAL have deposited more than 764 documents and/or records. For those without a collection on HAL the median is 57 deposits, i.e. more than ten times lower. Figure 3 shows the situation for the four scientific domains.

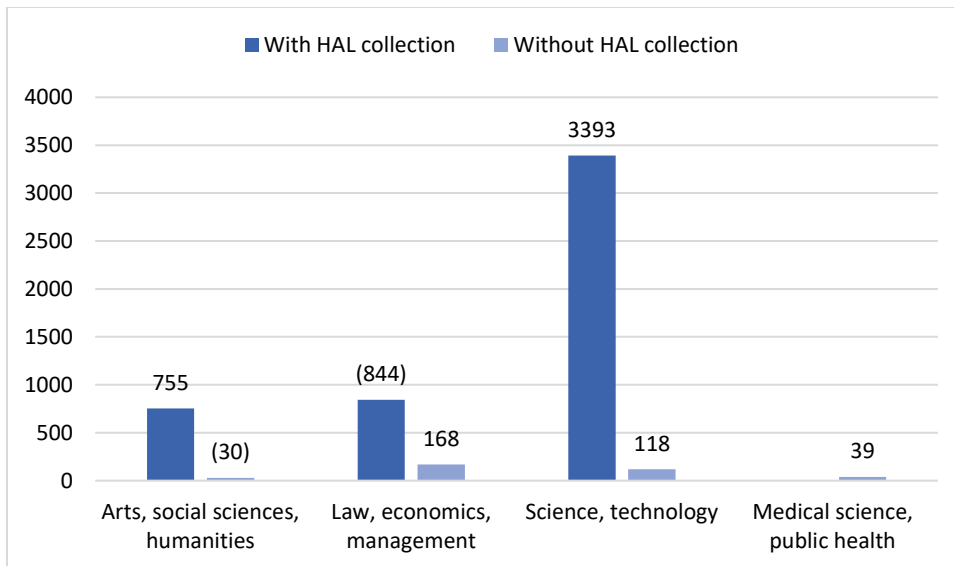


Figure 3. Median number of deposits per laboratory (N=66)

The high median number of deposits in the domain of science and technology is the result of the strong, institutional commitment to open access by two large research structures on the Lille campus, i.e. CRISTAL in the field of applied computer science and IEMN in the field of electronics, microelectronics and nanotechnology.

Documents and records

As mentioned above, the HAL repository contains two types of deposits, i.e. documents (with metadata) and records (metadata without documents). 32,247 deposits of the Lille laboratories are records without documents (77%) while 9,454 deposits are documents (23%). This distribution is rather similar to the overall distribution of the HAL deposits, even if the part of deposits with the document is lower at Lille than the general average (figure 4).

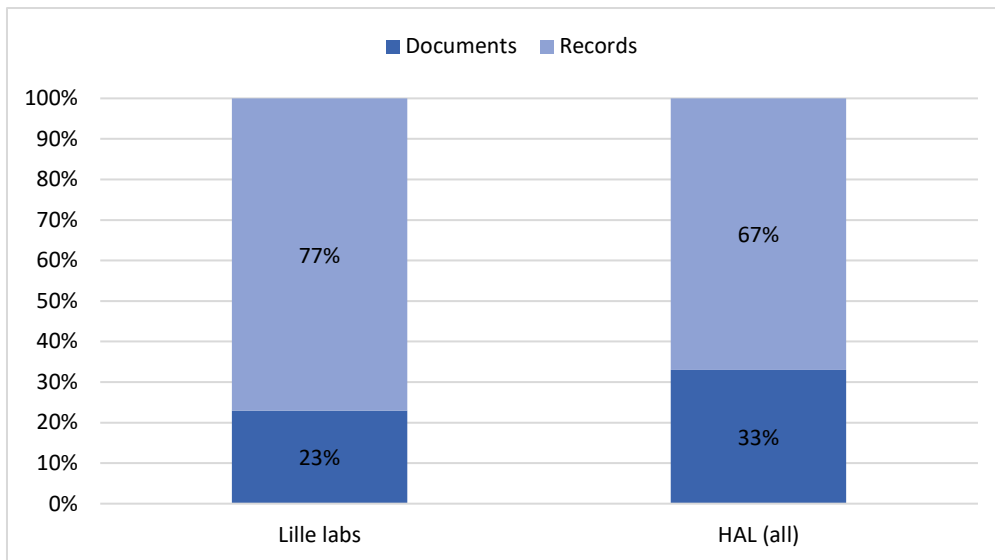


Figure 4. Documents and records of the Lille labs (N=41,701 deposits)

The numbers of deposits with and without full text are weakly correlated ($r=.32$).

The differences between the research domains are not significant (about 20% documents and 80% records), except for the small number of deposits from medical sciences and public health where half of the deposits are documents (figure 5).

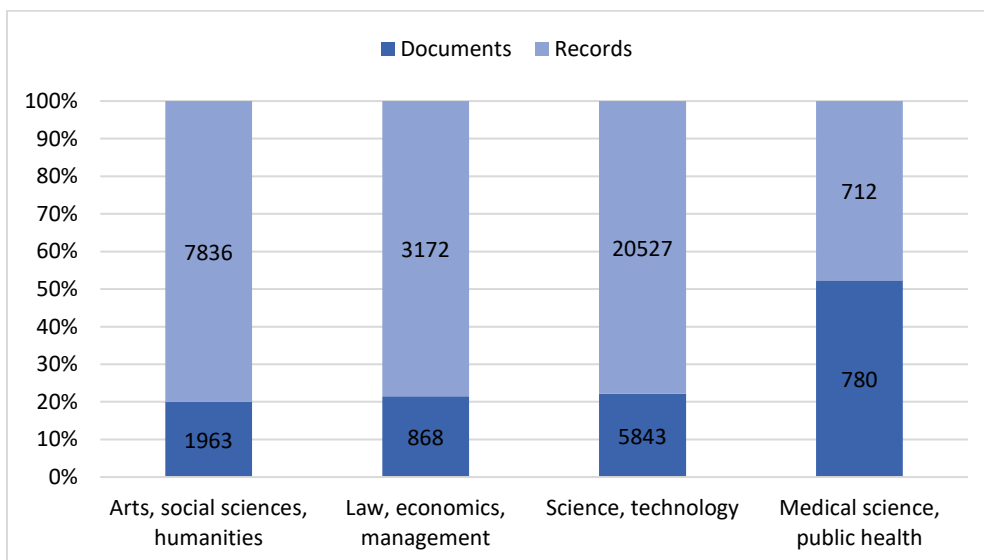


Figure 5. Documents and records per domain (N=41,701 deposits)

Now, what is the difference between the laboratories with and without collections on HAL, if there is any? The overall difference is not very significant: all collections together contain 21% documents, while the part for the labs without collections is slightly higher at 25%. However, there are some interesting differences between domains, as figure 6 shows.

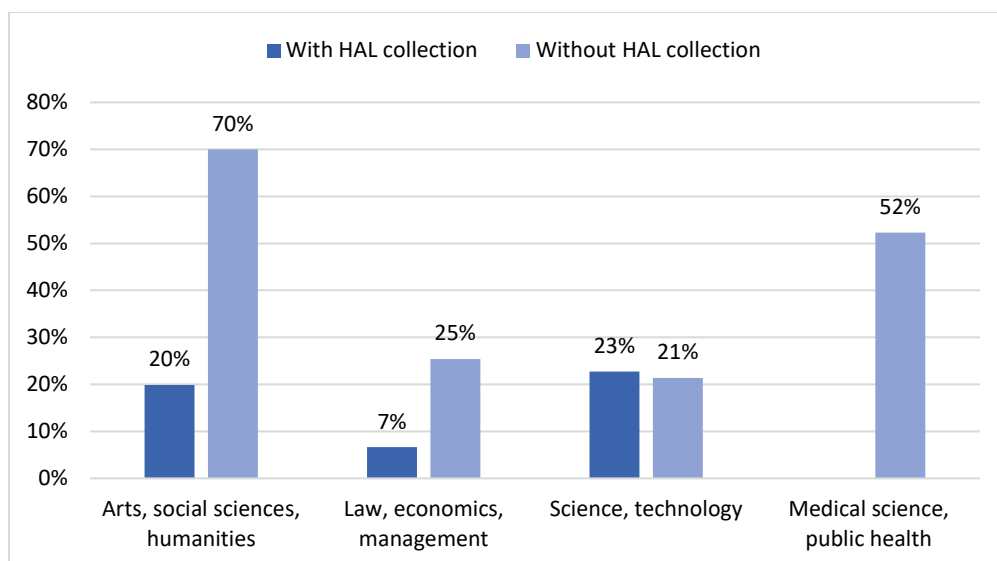


Figure 6. Part of documents (in %) of all deposits, per domain (N=41,701 deposits)

The collections of the research labs in arts, social sciences and humanities and in law, economics and management contain a relatively large part of records without documents, compared to the deposits from those labs without collections. At first sight this may appear paradoxical; yet, the explanation is easy (see below): when a research lab creates a collection on HAL, it probably will use the collection to make the academic production of its members visible, and this often means, to deposit records without the text or data files.

Grey literature

The HAL repository indexes 18 categories of publications, unpublished papers, academic works and research data. Some categories are clearly “white” or commercial literature (e.g. journal articles, books or book chapters), others can be described as grey literature (e.g. preprints, reports or dissertations) while other categories are mixed (e.g. conference papers), borderline (e.g. posters) or no literature (e.g. image files). Therefore, we didn’t try to determine the exact number of grey documents among the deposits but we limited ourselves to evaluating some representative categories. Figure 7 shows the distribution of the selected categories, with their numbers and percentages; the most important categories of the deposits are journal articles (42%) and conference papers (communications) (33%), followed by book chapters (9%). Together, the selected categories represent 90% of all deposits from the Lille laboratories (N=37,546).

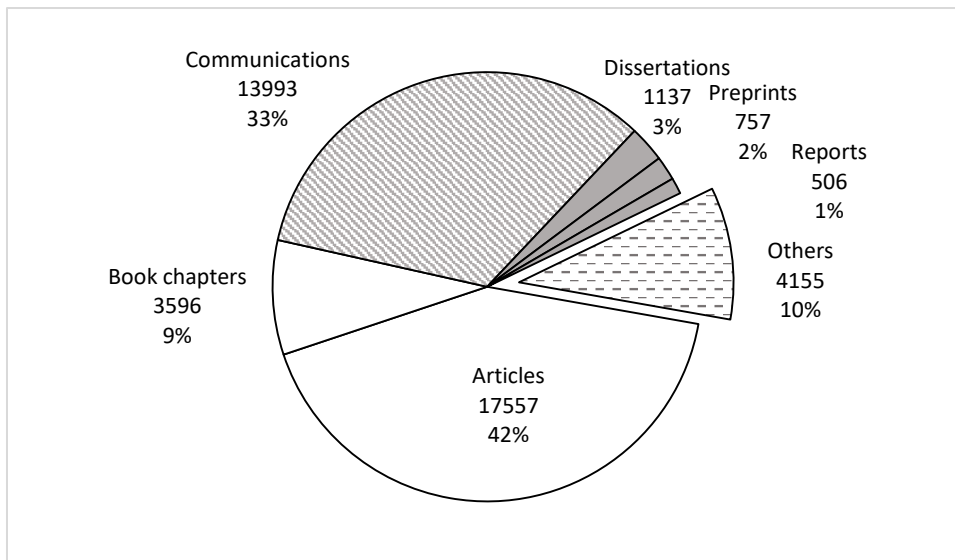


Figure 7. Deposits of selected categories (N=41,701)

Including conference papers (communications), the share of grey literature is 39%; without conference proceedings, grey literature represents 6% of all deposits. The problem with conference papers is that some of them are published in serials or books published by regular publishers while others are just PowerPoint presentations and will never be published, and it is difficult to say which percentage is really grey and which is not.

The next figure confirms the difference between conference papers on HAL and the other grey items. Figure 8 shows the relative share of documents and records (deposits without documents) for each category. While for most of the book chapters, communications and journal articles the full text is missing, 69% reports, 78% preprints and nearly all dissertations have been deposited with the document file.

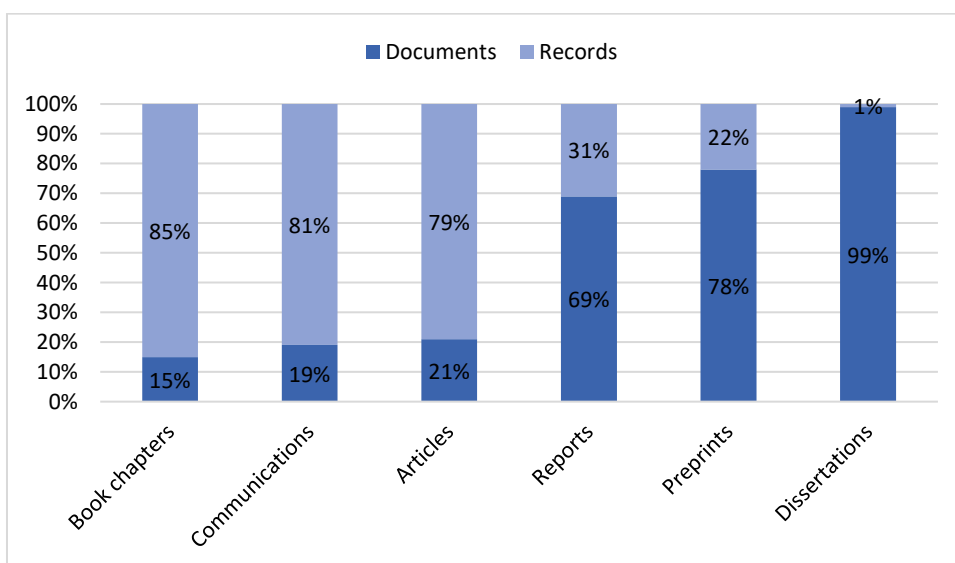


Figure 8. Records and documents per category (in %, N=37,546)

29% of the grey items is open access; without the communications, the part is 86%. For the “white” items (book chapters and articles), the percentage in open access is 20%.

Figure 9 shows a significant correlation between the number of grey deposits (preprints, reports, dissertations, conference papers) and the other items (articles and book chapters), with $r=.84$.

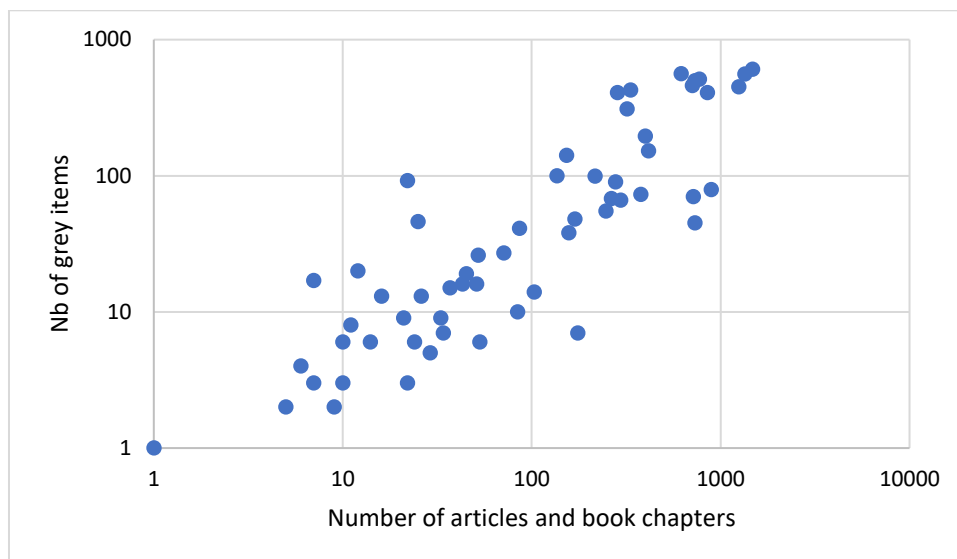


Figure 9. Correlation between grey and other deposits (N=66 labs)

It seems that the researchers or the technical staff who deposit the documents or create the metadata do not distinguish between different document types or prefer one or another. The difference is just that there are in average 3x to 4x more articles and chapters than grey items. Also, we cannot identify any significant differences between labs with and without collections, or between scientific domains.

The next figure shows the strong relationship between grey items and the total number of deposits (figure 10). In order to compare these statistics with former results (Stock & Schöpfel 2009), we calculated the z-scores for both variables.

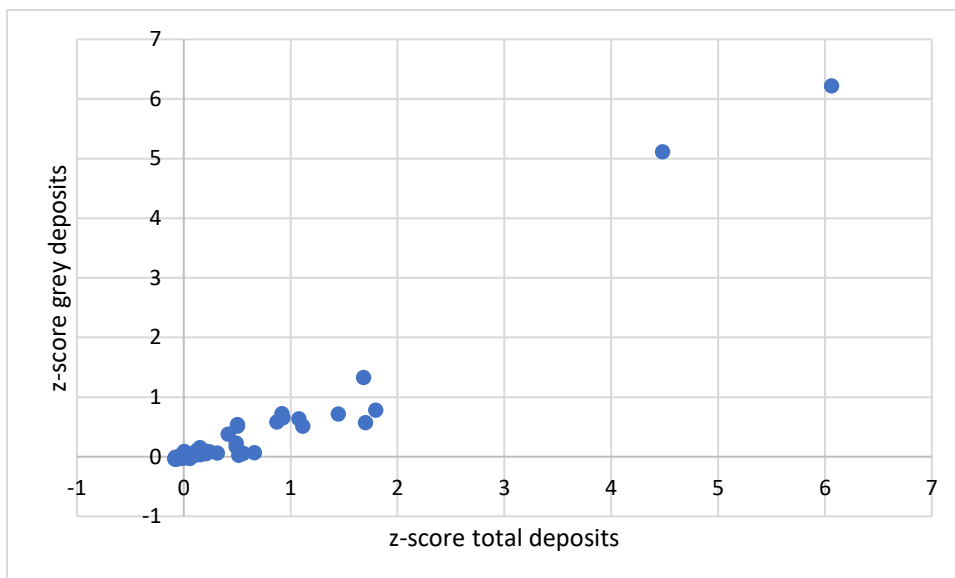


Figure 10. Grey deposits and total number of deposits (z-scores, N=66 labs)

The correlation between both variables is very high ($r=.96$). The higher the total number of documents, the higher the number of grey items. The two labs with the highest number of deposits are from engineering and informatics and have both created a collection on HAL. But, as above, we can't identify significant differences between labs with and without collections, or between scientific domains.

Clusters of laboratories

Finally, we'll try to distinguish different groups of laboratories regarding four variables: the total number of deposits, the degree of openness (deposits with full text), the existence of a collection, and the number of grey deposits.

Figure 11 compares the total number of deposits with the percentage of open access, i.e. of documents deposited with their full text.

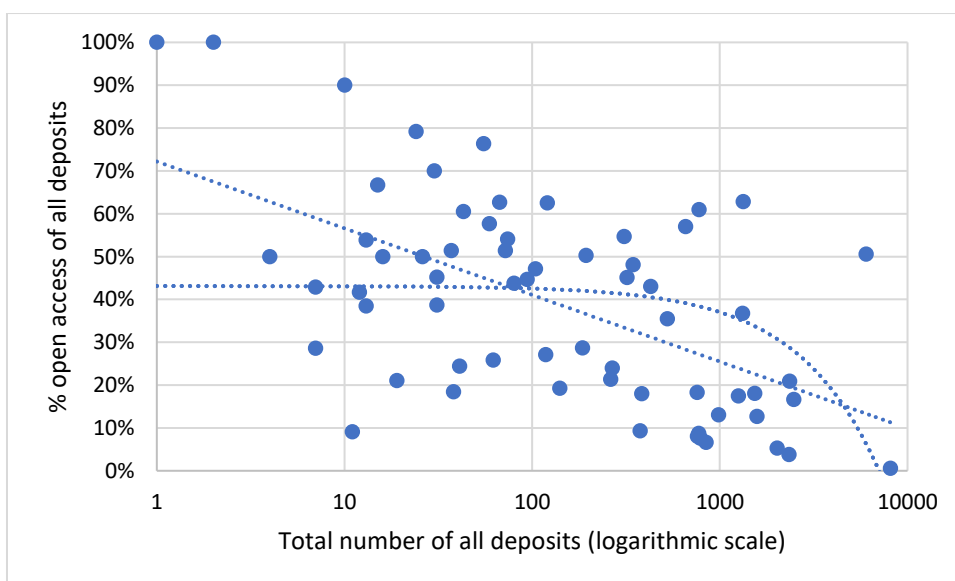


Figure 11. Total number of deposits and % open access (N=66 labs)

Figure 11 shows a weak negative correlation between the total number of all deposits (documents and records) and the percentage of documents in open access ($r=-.34$). In other words, we can observe a weak tendency in that the higher the total number of documents and records, the lower the percentage of freely accessible documents.

- Cluster 1. More than half of the laboratories (N=38) have less than 100 deposits on the HAL platform. Except for two, they did not create a collection on HAL. Probably, these laboratories do not have any significant strategy regarding open access and repositories, and the deposits are due to the scientists' personal choice, without "institutional curation".

These laboratories are clearly visible on figure 12 which presents the same figures but limited to those laboratories without a HAL collection.

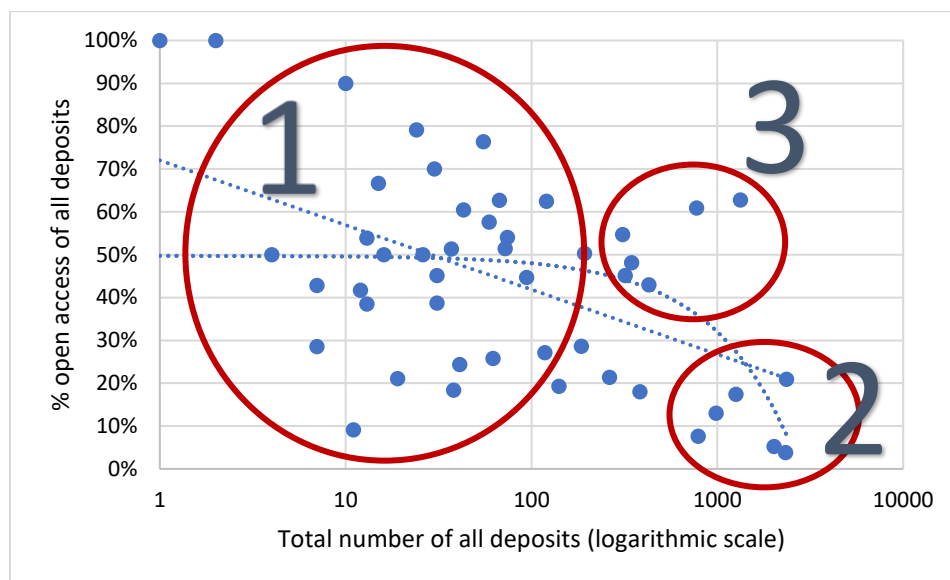


Figure 12. Total number of deposits and % open access for labs without HAL collection (N=50 labs)

Figure 12 shows two other groups.

- Cluster 2. A small group of laboratories (N=6) have a relatively high number of deposits (>700) but only less than 20% provide open access to the full text.
- Cluster 3. Another small group (N=6) have a higher number of deposits (>300), with more than 40% documents in open access.

The next figure shows the results for those laboratories which created their own collection on the HAL platform (figure 13).

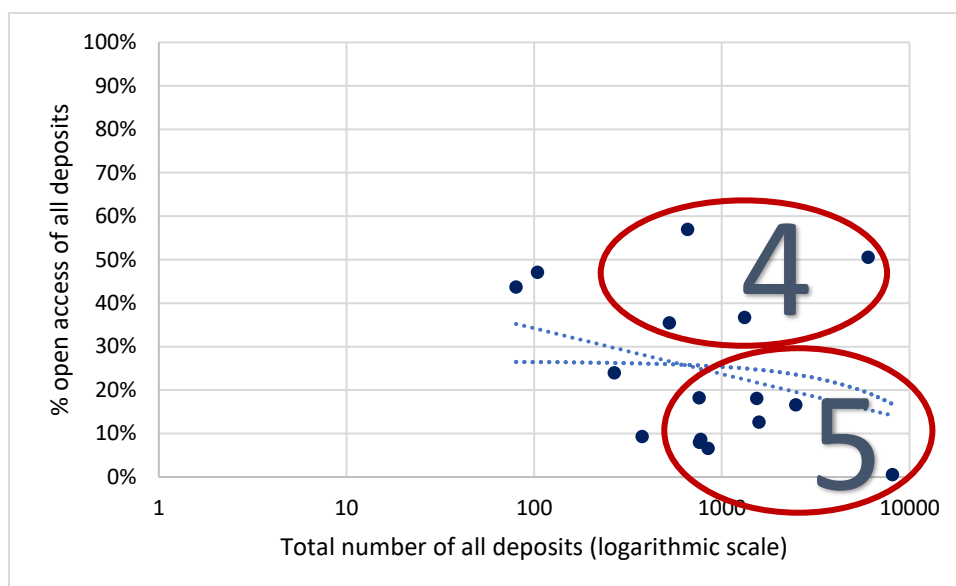


Figure 13. Total number of deposits and % open access for labs with a HAL collection (N=16 labs)

Following figure 13, we can distinguish two groups of laboratories.

- Cluster 4 (N=4). Four laboratories have a relatively high number of deposits (>500) and a percentage of open access above the average (>35%).
- Cluster 5 (N=8). A fifth group consists of laboratories with a large collection (>700 deposits) but a small percentage of open access (<20%).

As mentioned above, there are nearly no laboratories with less than 100 deposits.

Discussion

The scientometric assessment of the HAL deposits produces a lot of data, and the information based on these data may appear complex and difficult to understand. We'll try first to synthesize some significant results and then discuss them in terms of different strategies and policies. The main results:

- All research labs of the University of Lille are represented on the national HAL repository, except for one medical unit. Together, they have 41,701 items.
- One out of four labs have created its own, customized collection on the HAL server (24%).
- Together, these labs account for 63% of all items. The median number of items is higher for labs with collections than those without.
- Most collections have been created by research labs in arts, social sciences and humanities.
- The part of grey literature is 39%; most of these grey items are conference papers.
- The number of grey deposits is strongly correlated with the total number of deposits.
- 23% of the deposits are open access, the rest is metadata.
- This percentage is higher for grey literature (29%) than for articles and books (20%); without conference papers and presentations, this percentage is much higher (86%).
- Regarding the part of open access, there is a weak tendency in that the higher the total number of deposits, the lower the percentage of freely accessible documents.

The creation of a "lab collection" on the national repository HAL is a political and strategic decision to expose the scientific production of the laboratory's researchers elsewhere than on the laboratory

website, on individual web pages or on the institutional repository. Often, this decision is part of a larger strategy to increase the visibility, outreach and impact of the laboratory's research results, which is important and relevant for individual and institutional evaluation, for funding, and for networking and cooperation.

In the following, we'll try to provide some possible explanations for the observed results.

Institutional policy

Institutional policy, i.e. an explicit and assumed open access strategy, can explain one part of the results, especially the overrepresentation of arts, social sciences and humanities. The Lille campus of social sciences and humanities (the former Lille 3 University) decided a couple of years ago to launch its institutional repository as a university portal on the HAL server and encouraged the campus-based research units to create their own laboratory collection on HAL. This encouragement included not only helpful advice and assistance but also, partly, the creation of metadata records.

There was no similar approach to open access in the other former Lille universities Lille 1 (science and technology) and Lille 2 (medicine, law and politics); they did not launch an institutional repository neither on HAL nor elsewhere, and they did not foster the creation of laboratory collections on HAL. This may explain why still today, nearly two years after the merger of the Lille universities, we can observe large disciplinary differences regarding the number of size of collections and deposits.

On the national level, the French national research institute for digital sciences (INRIA) has been promoting for a couple of years now an open access policy and has launched its own institutional repository on HAL⁸, with today more than 60,000 documents, mostly conference papers. The INRIA scientific output is systematically entered into HAL and the INRIA researchers are encouraged to deposit their documents on HAL. One of the INRIA regional institutes is located in Lille, and one of the INRIA research laboratories (CRISTAL) is based on the university campus, which explains the large collection of computer sciences (14% of all Lille deposits are from CRISTAL) and also the relatively high part of documents with access to the full text (51%).

Research laboratories

Research laboratories have quite different ways to deal with the HAL repository, and up to now they are more or less free to decide if and how to manage their publications on HAL or elsewhere. To simplify, we can distinguish at least three different approaches:

No strategy (58%): no collection on HAL, and a low number of deposits (<100) or a high percentage of documents in open access (>50%), which probably indicates individual practice (=deposit of publications by the authors) but no collective, coordinated action. This is the cluster 1 mentioned above.

Reference management (29%): a higher number of deposits (>100) but a low percentage of OA (<25%). Half of these laboratories have more than 1,000 deposits on HAL, up to 8,000, well above the average. Half of them have a collection on HAL, the others have not. Probably, at least those with the collection but may also be those without. These laboratories make use of HAL for the monitoring of their scientific output, helpful for reporting and follow-up and especially for the national research assessment exercises. For instance, the Lille IEMN laboratory in civil engineering (mechanics) which represents 19% of all HAL deposits of the University of Lille is a pluri-disciplinary federation of 12 research units, affiliated with different research organisations and the University of Lille. Obviously, this federation makes use of the HAL repository as a public and free reference management tool. In fact, only 1% of

⁸ HAL-Inria <https://hal.inria.fr/INRIA>

the 8,103 items from IEMN are documents, the rest of the items (99%) are metadata without access to the full text. This makes no sense for an open repository, but this makes sense if the repository is diverted for reference management and output monitoring. Clusters 2 and 5 are part of this category.

Showcase (9%): a collection on HAL, a higher number of deposits (>100) and a higher percentage of open access (>30%). These laboratories have decided to create a collection, and obviously they make efforts to deposit their output and to increase the part of freely available documents on HAL. Obviously, their use of HAL is to show their papers and to provide a representative catalogue of their production. The impact of community seems evident – the laboratories are from psychology, literature, informatics, astronomy and information science (GERiCO). These laboratories are mainly in cluster 4.

A small group of three laboratories (physics, medicine, law/political science) appears to be quite similar to this last group but did not create a collection on HAL. They are in cluster 3.

The reason to take one decision or another can be motivated by various factors, e.g. institutional policy or community practice, and also personal awareness and attitudes towards open access; sometimes the lack of human resources – one part of the university laboratories do not have their own librarians – may be a major obstacle to moving forward, i.e. creating and curating a HAL collection.

We added in brackets an estimated percentage. However, more information is needed to confirm the different approaches and the underlying reasons and objectives.

Disciplinary practice

Scientific disciplines do not adopt the new open science policy at the same speed and in the same way. The European Open Science Monitor, for instance, shows large differences between fields of science and technology regarding the percentage of open access publications⁹. Disciplinary practice and research culture may explain, for instance, the low percentage of OA deposits of chemistry laboratories (6%), sociology (19%) or law (37%) compared to computer science (50%), astronomy (61%) or mathematics (63%).

The CRISTAL collection in computer science shows another community effect beyond the institutional strategy of valorisation and evaluation. More than other academic communities, the researchers in computer science disseminate their results by means of conference papers. More than half of the CRISTAL collection is composed of conference papers, and more than half of these papers are available in open access, which is probably due to a combined effect of institutional strategy and disciplinary practice.

Grey literature

About ten years ago, we assessed the development of open repositories in France, particularly of institutional repositories (Stock & Schöpfel 2009, Schöpfel & Prost 2010). The three samples (and the sampling methods) are too different to allow for a direct, statistical comparison; however, it is possible to make some general observations.

First, the part of grey literature in the Lille sample is higher than in the national samples ten years ago. It was then between 16% and 18%; in our new study it is 39%, more than two times higher. This difference is essentially due to conference proceedings.

⁹ https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en

The same studies provided some data about the distribution of the different types of grey literature (figure 14).

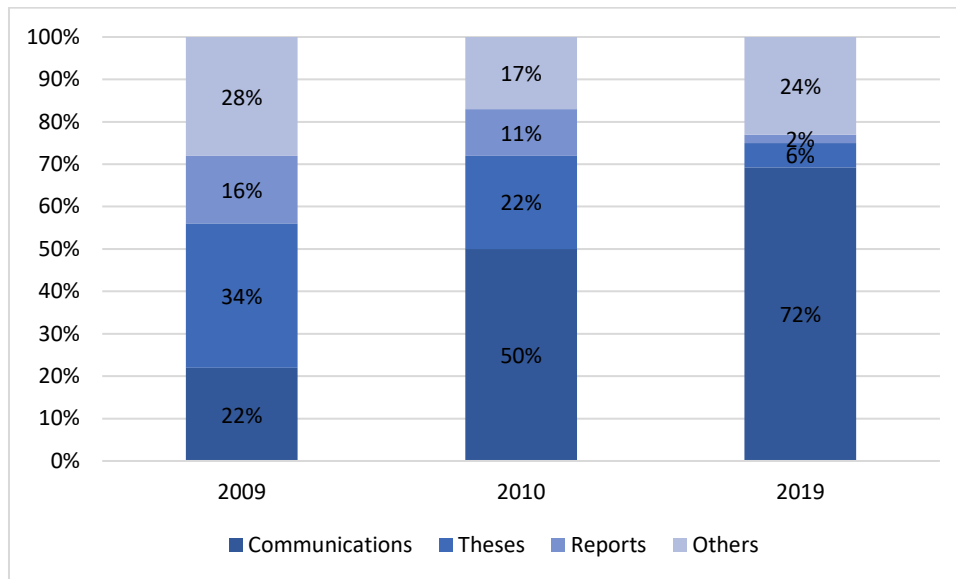


Figure 14. Distribution of grey literature categories in three samples (in %)

Again, because of the different sampling methods and characteristics, it is not possible to describe the differences in terms of development over time. Yet, it seems possible to say that the Lille sample is particular insofar as it contains a relatively large part of conference papers, as compared to theses and reports. The reason is probably the importance of the CRISTAL collection in informatics, with many conference papers, together with a high number of metadata records in civil engineering (IEMN) and from two laboratories in chemistry. In other words, the reason is probably a combination of community practice and laboratory strategy.

Five years ago, we assessed the part of freely accessible items in terms of degrees of openness in an international sample of 25 large institutional repositories, together with 2,068,622 deposits (Schöpfel & Prost 2015). We made two observations.

First, even if the correlation between the repositories' size and their degree of openness was weak, all large repositories had degrees of openness below the median while those repositories with higher degrees of openness (above the median) were smaller. As mentioned above (figure 11), the same observation applies for the Lille sample: we can observe a weak tendency in that the higher the total number of documents and records, the lower the percentage of freely accessible documents.

Second, the comparison between different types of documents reveals different degrees of openness, in a consistent way: except for conference papers, the percentage of freely accessible documents (open access) is higher for grey literature than for books or journals. Again, the same observation applies for the Lille sample (figure 15).

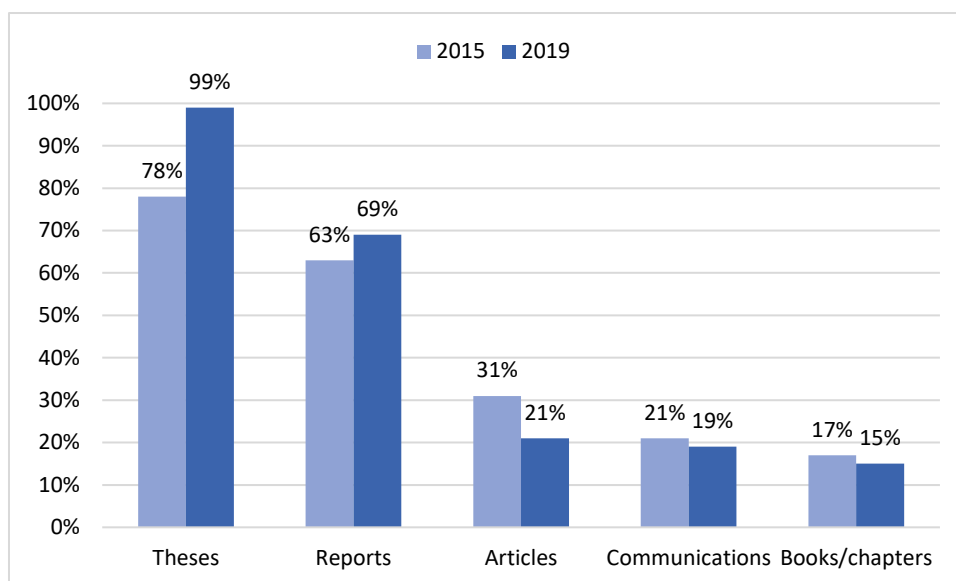


Figure 15. Degrees of openness for different document types

Again, we cannot make direct comparisons because of the differences of sampling, size and so on. But the similarities are obvious. In the international sample of large repositories as well as in the collections and deposits of the Lille research laboratories in HAL, the degree of openness of theses and reports are significantly higher than of articles, books (chapters) and communications. The reason why in the Lille sample nearly all theses and dissertations are freely accessible (99%) is simple – up to now, HAL has not accepted the deposit of PhD theses without the full text. The explanation for the small percentage of conference papers in open access is more complicated. There are probably two reasons – many papers are published in (white) conference series or as special journal issues, and HAL does not consider the deposit of a conference presentation as a full text document but as a record with supplemental material.

Conclusion

The scientometric analysis of the deposits and collections from the University of Lille research laboratories provides rich statistical material and some insights into differences between document types, disciplines and laboratories. Grey literature represents nearly two out of five deposits, consisting mainly not only of conference papers but also of theses and dissertations, reports and other types, such as working papers, preprints and courseware. The degree of openness (% of open access) is significantly higher particularly for theses and reports than for articles and books, and also for conference papers.

The study provides a kind of static photo, taken at a given moment (Spring 2019) in a dynamic and quick moving environment. In fact, the situation may change quickly, for at least three reasons – local decisions (such as an institutional mandatory policy), national (and European) research policy with new laws and other rules, and new agreements with academic publishers such as BioMed Central or Elsevier, allowing for automatic feeding of the HAL repository from the publishers' platforms.

In order to better understand the open access behaviours and strategies on the level of research laboratories, and to provide more insight into the impact of the global ecosystem of open science on the individual and collective decisions on the local level, more research is needed, and a different kind of research. For this reason, the GERiICO laboratory is undertaking a new research project on

laboratory collections and strategies on the HAL repositories, with a representative sample of the ten highest ranked French research universities, accounting for several hundred research laboratories, and applying a combined quantitative (scientometrics) and qualitative (surveys, interviews) methodology¹⁰.

Acknowledgments

A significant part of the data collection was realized during an internship at the GERiiCO laboratory. We would like to thank Patrice de la Broise and Delphine Spileers from GERiiCO for hosting and funding the internship between January and May 2019.

Data availability

The dataset is available on EASY (DANS) at <https://doi.org/10.17026/dans-xsq-bus7>

References

- Bauin, S. (2014). *L'open access à moyen terme : une feuille de route pour HAL*. CNRS Direction de l'Information Scientifique et Technique, Paris. http://coristshs.cnrs.fr/Rapport_HAL_DIST_2014
- CNRS-DIST (2016). *Livre blanc : une Science ouverte dans une République numérique*. CNRS Direction de l'Information Scientifique et Technique, Paris. <http://books.openedition.org/oep/1548>
- Schöpfel, J., & Prost, H. (2010). *Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement*. Retrieved from https://hal.archives-ouvertes.fr/sic_00497389v1
- Schöpfel, J., & Prost, H. (2015). Degrees of openness: Grey literature in institutional repositories. *The Grey Journal*, 11(2), 103-112. <https://hal.archives-ouvertes.fr/hal-01290339v1>
- Schöpfel, J., Prost, H., Fraisse, 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 novembre 2018, Rabat, Maroc*. Retrieved from <https://hal.archives-ouvertes.fr/hal-01940352>
- 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. Retrieved from https://hal.archives-ouvertes.fr/sic_00379232

¹⁰ Project HAL/LO (*Valorisation sur HAL de la production des laboratoires dans l'environnement de la science ouverte*), duration 18 months, funding by the GIS URFIST.