



HAL
open science

The Transformation of the Green Road to Open Access

Joachim Schöpfel, Stéphane Chaudiron, Bernard Jacquemin, Eric Kergosien,
Hélène Prost, Florence Thiault

► **To cite this version:**

Joachim Schöpfel, Stéphane Chaudiron, Bernard Jacquemin, Eric Kergosien, Hélène Prost, et al.. The Transformation of the Green Road to Open Access. Publications, 2023, 11 (2), pp.29. 10.3390/publications11020029 . hal-04097185

HAL Id: hal-04097185

<https://hal.science/hal-04097185>

Submitted on 15 May 2023

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.

The Transformation of the Green Road to Open Access

Joachim Schöpfel ^{1,*} , Stéphane Chaudiron ¹, Bernard Jacquemin ¹, Eric Kergosien ¹, Hélène Prost ² and Florence Thiault ³

¹ ULR 4073—GERiCO—Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication, Univ. Lille, F-59000 Lille, France

² ULR 4073—GERiCO—Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication, CNRS, F-59000 Lille, France

³ UR 4246—PREFICS—Pôle de Recherche Francophonies, Interculturel, Communication, Sociolinguistique, Université Rennes 2, F-35043 Rennes, France

* Correspondence: joachim.schopfel@univ-lille.fr

Abstract: (1) Background: The 2002 Budapest Open Access Initiative recommended the self-archiving of scientific articles in open repositories, which has been described as the “green road” to open access. Twenty years later, only one part of the researchers deposits their publications in open repositories; moreover, one part of the repositories’ content is not based on self-archived deposits but on mediated nonfaculty contributions. The purpose of the paper is to provide more empirical evidence on this situation and to assess the impact on the future of the green road. (2) Methods: We analyzed the contributions to the French national HAL repository from more than 1000 laboratories affiliated with the ten most important French research universities, with a focus on 2020, representing 14,023 contributor accounts and 164,070 deposits. (3) Results: We identified seven different types of contributor accounts, including deposits from nonfaculty staff and import flows from other platforms. Mediated nonfaculty contributions (deposits by libraries, import of bibliographic records, migration from other platforms, etc.) account for at least 48% of the 2020 deposits. We also identified differences between institutions and disciplines. (4) Conclusions: Our empirical results reveal a transformation of open repositories from self-archiving and direct scientific communication towards research information management. Repositories like HAL are somewhere in the middle of the process. The paper describes data quality as the main issue and major challenge of this transformation.

Keywords: open science; open access; open repositories; green road; self-archiving; contributor; research assessment; monitoring



Citation: Schöpfel, J.; Chaudiron, S.; Jacquemin, B.; Kergosien, E.; Prost, H.; Thiault, F. The Transformation of the Green Road to Open Access. *Publications* **2023**, *11*, 29. <https://doi.org/10.3390/publications11020029>

Academic Editors: Francesco De Pretis and Ahmad Yaman Abdin

Received: 14 February 2023

Revised: 4 May 2023

Accepted: 10 May 2023

Published: 15 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

More than twenty years ago, in 2002, the Budapest Declaration advocated open access to scientific information, i.e., the free and unrestricted online availability and re-use of scholarly research [1]. For scientific journals, the Budapest Declaration recommended two complementary strategies, the first of which was self-archiving, or, in other words, the deposit by researchers themselves of their articles in open repositories¹.

In 2004, Stevan Harnad and his colleagues defined this strategy as the “green road to open access”, which meant publishing an article in a traditional journal, followed by self-archiving in an open repository [2]. Harnad also alerted that this strategy should be accompanied by an institutional obligation, i.e., by a policy of mandate on the part of universities, research organizations, and funding agencies.

The prototype and pioneer of the green road strategy is arXiv, a curated, free distribution service and an open-access archive for scholarly articles mainly in the fields of physics, mathematics, and computer science, founded by Paul Ginsparg at Los Alamos in 1991 and now hosted by Cornell². In January 2023, the global Directory of Open Access Repositories OpenDOAR³ contained 6000 repositories, 89% of which were institutional repositories, defined

as digital collections for the management and dissemination of intellectual output created by the institution and its community members, including long-term preservation [3,4].

In France, following the model of arXiv, HAL⁴ was launched in 2001 by the public research organization CNRS⁵ as a multidisciplinary repository for the French research community. HAL is the central “green road” infrastructure of the French Open Science policy [5] and holds, as of January 2023, more than 4.3 million resources, mainly articles but also preprints, conference papers, dissertations, and so on. Its outreach is international, with more than 117 million consultations from all continents in 2022⁶. However, HAL has changed over time, and it is different today from the initial model (i.e., arXiv) in three ways:

- One part only of the deposits is self-archived, while many deposits are carried out by librarians, technical, and administrative staff, via other platforms and databases, and by publishers [6], and aim at a minimum to identify an institution’s production [7].
- Nearly 75% of the resources are non-full text deposits, i.e., metadata with or without an abstract.
- Recently, a government report recommended the use of HAL as a bibliometric tool for the assessment of public research in the social sciences and humanities (SSH) as an alternative to the Web of Science [8].

A couple of studies have provided empirical evidence of this evolution. The analysis of almost 60,000 deposits in the life sciences revealed that 86% had been contributed by institutional, nonfaculty staff [9]. In SSH, this part is about 40%; in law, economics, and management, it is higher than 50% [10]. A recent study on a corpus of 368 journals from five disciplines estimated the share of self-archiving by researchers at 38% [11]. A scientometric analysis of HAL showed that not more than 13% of articles have been self-archived [12]. Only 35% of researchers in SSH self-archive regularly on HAL [13]. Even if in these studies the surveyed behaviors may not be the same (self-archiving of documents and/or self-archiving of records without full text), they all require activity by the authors themselves. Another survey with research laboratories showed that professional follow-up and nonfaculty deposits are significant elements of the institutional support of open access [14]. The longitudinal assessment of HAL metadata quality [15] shows a sharp decrease in 2016, at the same time as a large increase in the number of deposits, which coincides with facilitated and simplified deposit procedures by the HAL team. However, we have been unable to find any other empirical studies that establish such a link between repositories’ functionalities and the quality of deposits.

This progressive transformation, which drives the HAL repository away from the initial green road model based on the principle of self-archiving, is not specifically French and has been observed in other countries for more than fifteen years. Two surveys of institutional repositories from the United Kingdom, Australia, and other countries revealed a low rate of author self-archiving (<40%) and full-text availability [16,17]. Both studies also showed that most documents had been deposited by a librarian or administrative staff. This observation has been confirmed by [18]: “Despite outreach, few faculty self-deposit anywhere (. . .) most (repositories) are being filled by persons at the institution explicitly tasked with doing so rather than eager faculty” and by [19]: “Regardless of (. . .) efforts to disseminate the ideas and the practice of open science, most of the world’s scholars in the early 2020s do not yet publish their works in preprint form and do not self-archive their research articles”. In one case study, many authors who participated heavily in disciplinary repositories did not self-archive their own papers in the institutional repository [20].

A couple of explanations have been given for this unexpected development and slow uptake, such as lack of awareness, low perceived usefulness, and ease of use, but also disciplinary (community) practice, including a competing culture of self-archiving, with significant differences between institutions and departments [21].

In order to cope with this situation, one common recommendation is that librarians should “help faculty archive their research papers (new and old) within the repository, digitizing older papers if necessary” [4], with the purpose of building up a critical mass of content considered to be the most important factor for the development of institutional repositories [22].

In contrast to the original green road approach, a “mediated archive” means that nonfaculty labor fills repositories. This may be less costly and more efficient than self-archiving, especially in the initial phase of a repository [23]. However, such a choice may also have an inadvertently negative effect on outreach and may distance faculty from the idea of self-archiving, as they have no practice doing it [16,18]. Furthermore, to increase the “buy-in” from academic staff, the process of acquiring “research material” (i.e., articles, communications, reports, and so on) could be embedded into the subject liaison role of the academic librarians, in partnership with the faculty, rather than as an entirely separate process [23]. After the launch of the Mediated Deposit Service at Concordia, the number of mediated deposits surpassed but not superseded author self-archiving, with new practices and workflows between library and faculty [24].

Two routines can be observed. First, increased library support; this includes external partnerships with publishers and service providers like DeepGreen, a German infrastructure “that collects journal articles from academic publishers and sends them to authorized libraries for publication in their repositories” [25]. Second, institutional open access policies requiring deposit not only for research dissemination and long-term preservation but also for performance evaluation [26] will, together with mandates from funders, “likely be the only mechanism that will encourage authors to place an open access copy of their work in a repository” [24].

Both processes—library support and institutional mandates—are not opposed but complementary. The rationale behind this development and its result have been described as a transformation of the green road to open access: a functional change of repositories from dissemination of results to assessment of research performance, which on the level of infrastructure means a progressive convergence between repositories and research information management systems [27]. This does not mean that the repositories’ initial purpose is or will be abandoned; also, these processes (and mediated deposits) at least partly aim to increase (and will continue to do so) the availability of scholarly research through incentives and more efficient workflows. However, this development requires reliable information about research, in other words, high-quality and rich metadata about persons, organizations, and so on.

As part of a research project on the open access strategies of more than 1000 French research laboratories, we had the opportunity to assess their contributions to the national HAL repository. The purpose of this assessment is to provide a better understanding of the development of open repositories based on empirical evidence and can be described as follows:

- A description of the development of the contributor accounts;
- A typology of contributor accounts;
- An estimation of the nonfaculty-mediated contribution to HAL;
- An assessment of differences between laboratories;
- An assessment of differences between disciplines;
- The results will be discussed, and recommendations will be made for the further development of repositories and research on open science.

2. Materials and Methods

We assessed the HAL deposits of 1246 laboratories affiliated to the ten most important French research universities (Udice group members⁷), which together represent 33,800 faculty, 24,000 PhD students, and two-thirds of the most cited French publications worldwide (Appendix A). These laboratories cover the whole range of scientific disciplines (Appendix B).

Based on the HAL-specific organizational structure codes of all these laboratories⁸, 1,035,612 deposits have been identified and analyzed. There was no prospective protocol or analysis plan. For the particular purpose of this study, we assessed the information about the contributor, i.e., the entity responsible for the deposit of the resource. The data extraction was carried out via the HAL API in April 2021; the query was built from the

documentation on the HAL platform⁹. The results were verified, checked, and cleaned by three members of the project team.

The resulting spreadsheet contains information about each contributor for each year and each laboratory (=event), together with 213,140 events (lines) with the following data: university, laboratory, research field, research disciplines, contributor, year of deposit, total number of deposits for the given laboratory for this year, total number of deposits for the given contributor for this year. Limiting the analysis to the period 2010–2020, our sample consists of 180,646 events, totaling 1226 laboratories, 39,038 contributor accounts, and 897,097 deposits.

Based on the data for 2020 (14,023 contributor accounts and a cumulated total of 164,070 deposits), we analyzed the names of the contributor accounts in order to select all clearly identifiable non-personal accounts (institutions, functions, tools, etc.). Additionally, we analyzed the most important personal accounts (with the highest number of deposits) and tried to distinguish between authors (researchers) and technical staff (based on information from HAL, social media, and personal websites). In this way, we obtained a sample of 166 accounts, representing 48% of the total annual deposits.

3. Results

3.1. Number of Contributors

Since 2010, the number of contributors for all laboratories has continuously increased from 3787 in 2010 to 14,023 in 2020 (Figure 1).

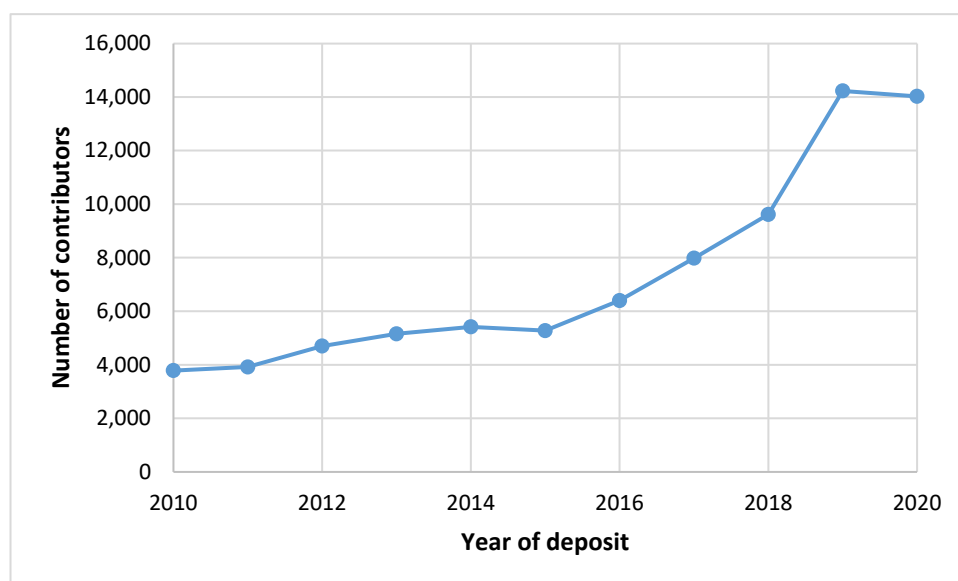


Figure 1. The number of contributors per year of deposit (2010–2020).

We can distinguish three periods:

1. Between 2010 and 2015, there was a slow progression from 3787 to 5274 contributors (+39%);
2. From 2015 to 2018, there was an acceleration of the increase in the number, passing from 5274 to 9615 contributors (+82%), which may be fostered by the improvement (simplification) of the procedures of deposit;
3. From 2018 to 2020, there was a strong and sudden growth from 9615 to 14,023 contributors (+48%), which coincides with the decision of the CNRS to have recourse to HAL for the individual assessments.

This progression is highly correlated with the number of laboratories using HAL and the number of deposits ($r > 0.9$).

The future will show if the stabilization on a plateau of about 14,000 contributors is temporary or definitive. In any case, we are still far from the figure of 33,800 researchers and teachers-researchers of the ten universities of the sample, without counting the PhD students (even if not all of them may have published and/or deposited their output on HAL in 2020).

At the same time, the number of deposits in the laboratories has been multiplied by five, passing from 33,237 deposits in 2010 to 166,939 deposits in 2020¹⁰. This means that the average number of deposits by contributor increased by 36% (from 8.8 to 11.9), while the average number of deposits per laboratory tripled (from 56 to 142), a growth that is probably not due to increased research performance but to increased use of HAL by the laboratories, simplified procedures, support by libraries, and so on.

3.2. Typology of Contributors

The field “contributor” is automatically generated during the deposit from the HAL user account of the person making the deposit; the account is visualized for each deposit as a name or an avatar. Based on this information, a content analysis of the 2020 deposit data with 14,023 contributor accounts reveals seven categories of contributors:

- Authors who self-archive their own publications and/or create metadata (records) of these publications.
- Other researchers who deposit publications for their colleagues working in the same laboratory. One part of the deposits is realized by other researchers than the authors, for instance, PhD students or other early career researchers who are paid for this work by the laboratory or by voluntary researchers in charge of open science and/or the laboratory’s collection on the HAL platform. These contributors may at the same time deposit their own publications.
- Administrative, technical, and library staff of the authors’ laboratory deposit publications for their laboratory (most often metadata without the document).
- Other nonfaculty—often staff from the university library—who deposit publications for several laboratories or for the whole institution (most often metadata without the document).
- Generic contributor accounts corresponding to specific metadata flow from bibliographic databases, reference management software, and catalogs. Some laboratories follow up their scientific production with internal bibliographic databases or other reference management software, and some have created a workflow to ingest the references into the HAL repository with a generic contributor account (avatar).
- Migration flows from other open archives. In the past, the HAL platform has integrated metadata references from other open repositories; this was the case, for instance, when the French National Research Institute for Agriculture, Food, and the Environment (INRAE)¹¹ closed its institutional repository Prodinra and migrated its content to HAL. Some institutional repositories are interconnected with HAL and provide metadata feeds.
- Import flows from other platforms or publishers. A few contributor accounts correspond to workflows from other platforms, like Inspire HEP, the leading information platform for High Energy Physics (HEP), or from publishers who started to feed the HAL platform with their own metadata, like Elsevier. Figure 2 provides an overview of these different categories.

This typology based on the 2020 data describes a quite different landscape than the initial model of open repositories, where all deposits are made by the authors themselves. Only the green part of Figure 2 corresponds to the principle of self-archiving. In fact, faculty do more than self-archiving insofar as they also participate in mediated contributions, along with nonfaculty staff from the research laboratories or from other structures (academic libraries, etc.), and with imports from laboratory-based tools, migration flows, and external platforms like institutional repositories or publishers’ databases. In fact, the reality has changed and is much more heterogeneous.

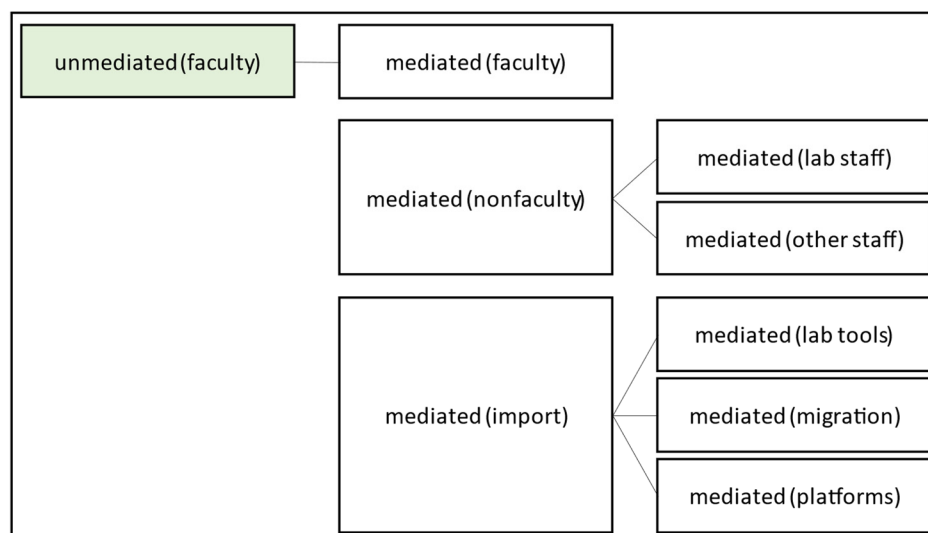


Figure 2. The typology of contributors (2020).

3.3. The Part of Mediated Contributions—Nonfaculty and Import

As we did not collect data for each deposit, it is not possible to match the contributor and author fields of the deposits’ metadata and to produce exact figures on the part of the researchers’ self-archiving. However, it is possible to make a conservative best estimation based on the contributor account data. Here are the results for one year, 2020, with 14,023 contributor accounts and a cumulated total of 164,070 deposits. The curve of the deposits is a Pareto distribution: on the long tail, 20% of publications have been deposited by 83% of contributors, while on the “top of the charts”, 50% of all publications have been deposited by less than 1% of contributors, most of them clearly identifiable as nonfaculty or imports. As Figure 3 shows, many contributor accounts just deposited one, two, or three publications during the whole year 2020.

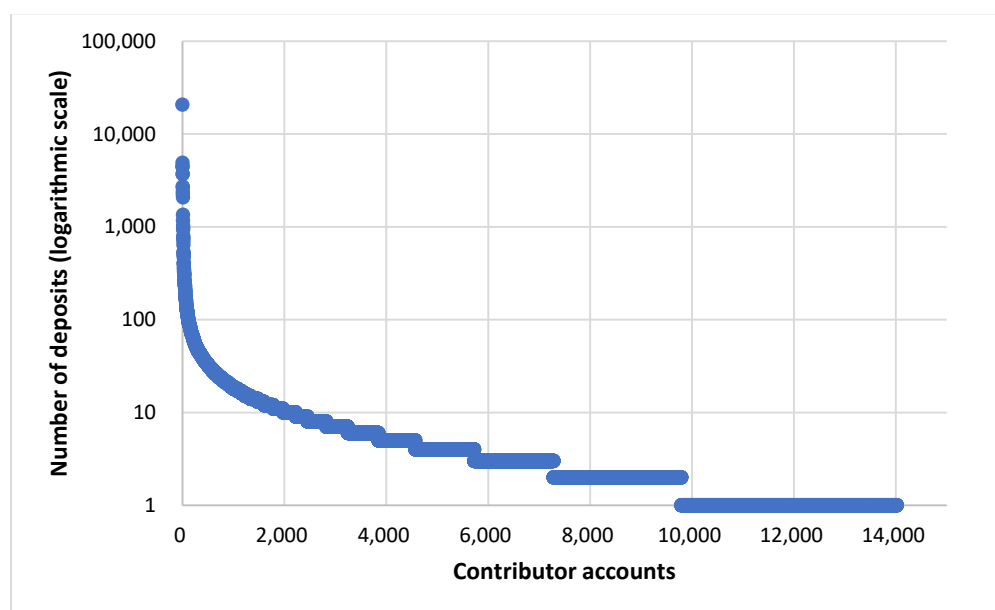


Figure 3. The distribution of deposits per contributor account (2020).

The proportion of mediated contributor accounts, including nonfaculty and import accounts, appears rather low and can be estimated at 1.2% of all accounts. Thus far, this low number of contributors accounts for 78,510 deposits, i.e., 48% of all 165,070 deposits

made in 2020. In other words, nearly half of all publications on HAL match the concept of mediated (nonfaculty, import) contributions (Figure 4).

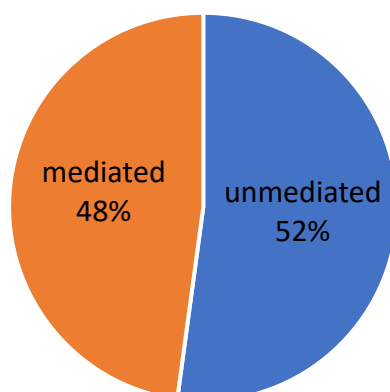


Figure 4. The part of mediated and unmediated deposits (2020).

Input from other platforms accounts for 27% (2% from institutional repositories, 2% from publishers, and the other 23% are migration flows). Input from personal nonfaculty staff accounts for 13%, while the other 8% is input from library, laboratory, or university avatar accounts. Probably some unmediated deposits are in reality mediated faculty deposits (deposits by scientists of other scientists' output), but based on our data, it is impossible to provide a good estimate of this part. Furthermore, it is not possible to make a reliable distinction between laboratory staff and other staff from the academic library or another service, or between laboratory staff and laboratory tools. Finally, it is important to keep in mind that these figures reflect the situation in 2020 and do not represent the whole HAL content.

3.4. Differences between Laboratories

The 2020 sample consists of 1176 research laboratories, ranging from 1 to more than 100 contributors and from 1 to nearly 3000 deposits (median = 67). The correlation between the number of deposits and the number of contributors is 0.54. The higher the number of contributors, the higher the number of deposits in a given laboratory. Thus far, more noteworthy seems the relationship between the number of deposits of the whole laboratory and the number of deposits of the laboratory's first contributor account, i.e., the contributor account with the highest number of deposits for this laboratory; here, the correlation coefficient is 0.86. In other words, while the number of contributors is relevant, the importance of the first contributor is even more relevant for laboratories' total number of deposits on HAL. Figure 5 shows this strong correlation for the whole sample of 1176 research laboratories.

Figure 5 visualizes the large variety between the laboratories—some of them deposited less than 10 publications on HAL in 2020, while others published more than 100 or even more than 1000 items (horizontal axis). Regarding the topic of our paper, the cluster of laboratories in the upper right field of the figure is particularly interesting: these are the laboratories where the first contributor account is “responsible” for a large part of the laboratories' output on HAL, with hundreds of deposits. In other words, this is not self-archiving but a systematic and mediated nonfaculty contribution to the HAL repository.

The differences between universities are less important, except for one (Aix-Marseille), where the proportion of laboratories with mediated contributions seems much lower than expected compared to the other universities. One reason for the special case of Aix-Marseille may be the importance of social sciences and humanities at this university (see below, 3.5). The main reason for the observed (relative) homogeneity between universities is probably that the institutional support and action at the university level are not so important and rather similar, compared to the diversity of tools, resources, and strategies at the level of the research laboratories.

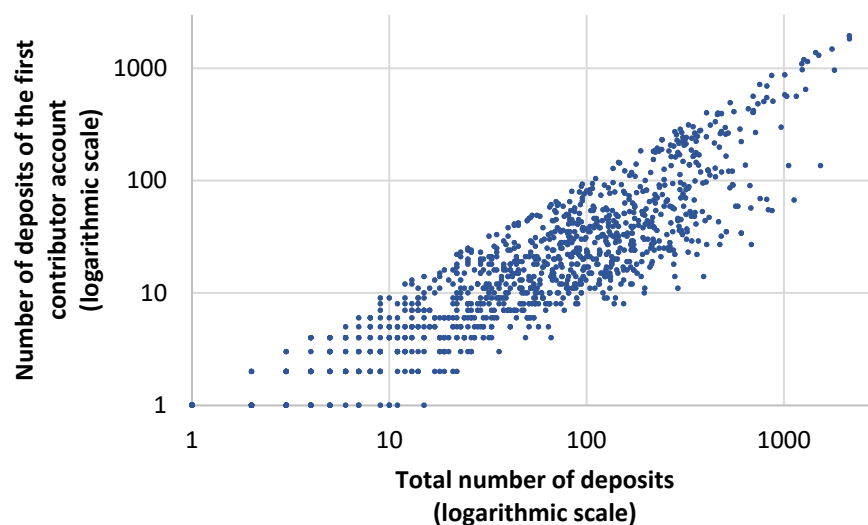


Figure 5. Total number of deposits and the part of the first contributor per laboratory (2020).

3.5. Differences between Research Disciplines

A comparison between research disciplines reveals complementary results. First, in the field of social sciences and humanities, the correlation between the number of deposits and the number of contributors is higher (0.69), while the correlation with the deposits of the first contributor account is lower (0.73). Obviously, for these laboratories, the role of mediated deposits is less important than in other research fields.

Second, the role of the first contributor account seems more important in laboratories in the fields of law, economy, and management, which may be an indicator for a higher degree of mediated contribution here.

Third, the part of mediated contributions is significantly higher for the laboratories in earth sciences, ecology, and agriculture; the main reason is probably the migration of the ProdINRA database to HAL in 2020 (see above, Section 3.2).

4. Discussion and Conclusions

The results of the analysis of the HAL deposit data show an important increase in the overall number of contributor accounts, along with an increased average number of deposits per contributor. Based on a content analysis of the 2020 contributor accounts, seven different contributor categories have been identified. The proportion of mediated contributor accounts (nonfaculty, import) appears rather low and can be estimated at 1.2% of all accounts. Thus far, this low number of contributors accounts for 48% of all deposits made in 2020. Other empirical evidence is presented to illustrate the differences between research laboratories, universities, and disciplines. In particular, the strong relationship between the first contributor account and the total deposits on HAL of a given research laboratory is highlighted.

As mentioned above, the research has two methodological shortfalls. First, we counted events (deposits per laboratory per year). If a deposit (article, communication, etc.) has coauthors from two different laboratories from our sample, it will be counted twice (duplicates), which means that the absolute numbers are overestimated. However, a precise analysis of the 2020 events (166,939) shows that this systematic bias is not important (2869, or 1.7%). Second, we assessed the activity of each contributor's account, but we did not assess each deposit. In other words, we cannot compare the metadata of authors with the contributor account, as performed by [9,10]. Thus far, our results are similar enough to those based on direct matching between the creator (author) and contributor fields to provide complementary, valid evidence to these former studies.

Our empirical evidence reveals on a large-scale level the transformation of the French national HAL infrastructure from an open repository based on the researchers' self-archiving

(like arXiv) into an open platform with publications and metadata (records) from different sources. In our sample of more than 1000 laboratories from the ten most important French research universities, only half of the 2020 deposits are self-archived, while the other half represent mediated, mostly nonfaculty contributions or imports. This mediated contribution requires (and reflects) institutional support and assistance, with three purposes: (still) the development of open access and direct scientific communication by creating content in the repository; the long-term preservation of the resources, as all HAL deposits are backed up in a public dark archive hosted by CINES at Montpellier¹²; and the development of an infrastructure that allows monitoring and assessment of the scientific production of the individual researchers, the laboratories, and the universities. Due to the institutional support and contributions from laboratories and universities, HAL has become a kind of showcase for their scientific production. Moreover, it also provides data for the French Open Science Monitor¹³. This mediated contribution is not temporary in order to create a critical mass during an initial period after the repository's launch, as described by [4,22,23]. Our results show mediated contribution as a significant part of the normal and permanent repository functioning; HAL is somewhere in the middle of the process from an open repository (green road, as recommended by the Budapest Initiative) towards a particular kind of open research information management system.

This transformation is not specific to HAL, and it is not specific to France [27]. Furthermore, our intention is not to say if this is good for open science or not. Instead, we would like to draw attention to one particular but crucial challenge: the impact of this transformation on the importance of data and metadata quality. From the moment the platform performs monitoring and assessment functions, the quality of the data becomes an essential criterion for the quality of the system's functionalities and services and for its acceptance [28,29]. This requires a thorough and continuous assessment of the data quality [30] and specific measures to control and improve the data quality during the whole process and even before (upstream) the data import and creation [31], through FAIRization of the data [32], including a qualified and standardized use of the contributor field and a strict control of the input from other platforms.

Erroneous spelling and homonyms, wrong or missing identifiers, wrong attributions of scientific works, and so on are already serious issues for the findability of resources on open repositories. However, the more repositories and research information management systems converge, the more this will become a crucial problem for repositories because of the potential harmfulness of bad data quality for institutions, projects, and above all, people. Will that transition away from self-archiving be accompanied by a decrease in metadata quality? We do not think so; given the importance of the issues at stake, we would rather expect the opposite: a continuous improvement of metadata quality based on better controls, standardization, and improved curation functionalities, including extensive usage of persistent identifiers.

Attention should be paid to maintaining the initial purpose of repositories, i.e., providing open access to research results. However, at the same time, repository operators and managers must take more care of the curation of metadata quality, which means, above all, the assessment and improvement of the FAIRness of their infrastructure [33].

Beyond the question of data quality, other issues will be raised, such as the development of reliable services and functionalities for data creation and import, data analytics, and relevant reports, or the provision of data for third-party services on top of the repository.

More generally, perhaps we should stop speaking about open repositories in terms of "green roads", as if all repositories followed the same principles and functioned the same way, and instead introduce different types or "colors" of repositories (for instance, see [34]), just as we did for open access journals years ago.

Some of the reviewed empirical studies are not recent, even if they still seem relevant and valuable. However, further research is required to assess this transformation of the green road to open access at the level of the infrastructure (system), the data (content), and the usage by researchers and institutions. We need more evidence about the role and impact

of mediated contributions, especially from new initiatives like the German DeepGreen project or from publishers' platforms and databases, but also from academic institutions and organizations, in order to assess the role of libraries and other staff on the terrain of research. The diversity of contributors and deposit types makes it difficult to understand the process of feeding an open repository and its impact [35]. More evidence is required on the content and quality of the mediated deposits: which is the part of metadata creation (records without full text), which is the part of depositing full text, and do the metadata without full text contain links to the full text on other platforms? What is the impact on metadata quality?

In order to contribute to a better understanding of the transformation and of the laboratories' resources and strategies, we conducted interviews with senior researchers and information professionals of fifty French laboratories, and we will publish the results soon. The results will also be helpful for a better understanding of the contributors' motivations and attitudes, e.g., if they are motivated by a disinterested desire to acquaint colleagues with the latest results, or, on the contrary, if they are primarily subject to administrative requirements.

Author Contributions: Conceptualization, methodology, and validation, J.S., B.J., E.K., S.C., F.T. and H.P.; data curation, B.J. and H.P.; writing—original draft preparation, J.S.; writing—review and editing, J.S., B.J., E.K., S.C., F.T. and H.P.; project administration, J.S. and F.T.; funding acquisition, J.S. and F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the GIS “Réseau Urfist” (project HAL/LO).

Data Availability Statement: The underlying data are available on the French repository Recherche Data Gouv: Schöpfel, J., Jacquemin, B., Kergosien, E., Chaudiron, S., Thiault, F., and Prost, H. (2023). Contributors of HAL Repository, <https://doi.org/10.57745/4UYIYS>, Recherche Data Gouv, V2.

Acknowledgments: The authors acknowledge administrative support from the GERiCO laboratory and helpful advice from the CCSD team at Lyon. We would also like to thank Otmane Azeroual for his review of the final manuscript. Thanks also to four anonymous reviewers for their useful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Research Sample

University	Number of Laboratories
Aix-Marseille	105
Bordeaux	75
Côte d'Azur	42
Grenoble Alpes	99
Lyon-1	143
Strasbourg	76
Paris Cité	220
Paris Saclay	228
Paris Sciences Lettres	123
Sorbonne Université	135
Total	1246

Appendix B. Scientific Fields of the Sample

Scientific Field *	Number of Laboratories
Sciences and technology	445
Medical and life sciences	415
Arts, social sciences, and humanities	301
Law, economics, and management	85
Total	1246

* Main scientific domain, following the French Higher Education typology.

Notes

- 1 The second strategic recommendation of the Budapest Open Access Initiative (BOAI) was the publication of open-access journals.
- 2 arXiv. <https://arxiv.org/> (accessed on 12 May 2023).
- 3 OpenDOAR. <https://v2.sherpa.ac.uk/opensdoar/> (accessed on 12 May 2023).
- 4 Hyper Article en Ligne. <https://hal.science/> (accessed on 12 May 2023).
- 5 Centre national de la recherche scientifique. <https://www.cnrs.fr/en> (accessed on 12 May 2023).
- 6 Activity report 2022. <https://www.ccsd.cnrs.fr/2023/03/ccsd-publication-du-rapport-d-activite-2022/> (accessed on 12 May 2023).
- 7 Udice Group Universities. <https://www.udice.org/about-us/?lang=en> (accessed on 12 May 2023).
- 8 AuréHAL. <https://aurehal.archives-ouvertes.fr/structure/index> (accessed on 12 May 2023).
- 9 HAL API documentation. <https://api.archives-ouvertes.fr/docs/search> (accessed on 12 May 2023).
- 10 NB: These figures are based on the laboratories' deposits and include a small part of duplicates because some deposited publications are affiliated with two or more laboratories (see the discussion).
- 11 Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. <https://www.inrae.fr/> (accessed on 12 May 2023).
- 12 National Computer Center for Higher Education. <https://www.cines.fr/> (accessed on 12 May 2023).
- 13 Baromètre français de la Science Ouverte. <https://barometredelascienceouverte.esr.gouv.fr/> (accessed on 12 May 2023).

References

1. BOAI. *Budapest Open Access Initiative*; Open Society Foundations: Budapest, Hungary, 2002. Available online: <https://www.budapestopenaccessinitiative.org/> (accessed on 12 May 2023).
2. Harnad, S.; Brody, T.; Vallières, F.; Carr, L.; Hitchcock, S.; Gingras, Y.; Oppenheim, C.; Stamerjohanns, H.; Hilf, E.R. The Access/Impact Problem and the Green and Gold Roads to Open Access. *Ser. Rev.* **2004**, *30*, 310–314. [CrossRef]
3. Lynch, C.A. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Portal: Libr. Acad.* **2003**, *3*, 327–336. [CrossRef]
4. Prosser, D.C. The Next Information Revolution—Can Institutional Repositories and Self-Archiving Transform Scholarly Communications? In *Scholarly Publishing in an Electronic Era: International Yearbook of Library and Information Management 2004–2005*; Facet Publishing: London, UK, 2004; pp. 99–117.
5. Berthaud, C.; Charnay, D.; Fargier, N. Diffuser et pérenniser le savoir scientifique: 20 ans d'histoire de HAL. *Hist. Rech. Contemp.* **2021**, *10*. [CrossRef]
6. Charnay, D. Avec HAL, nous voulions créer un arXiv multidisciplinaire. *Hermès* **2019**, *85*, 94. [CrossRef]
7. Magron, A. Utiliser les Archives Ouvertes pour Valoriser ses Travaux L'exemple de HAL-SHS. In *Le doctorat en France: Mode(s) d'emploi*; Schnedecker, C., Aleksandrova, A., Eds.; Peter Lang: Berlin, Germany, 2017; pp. 227–239. Available online: https://hal.archives-ouvertes.fr/sic_01697164 (accessed on 12 May 2023).
8. Thibault, F.; Streliski, S. *Les Indicateurs Bibliométriques pour les SHS—Etat de la Question*; Alliance Athéna: Paris, France, 2022. Available online: <http://www.alliance-athena.fr/rapport-de-lalliance-athena-les-indicateurs-bibliometriques-pour-les-shs-etat-de-la-question/> (accessed on 12 May 2023).
9. Prime-Claverie, C.; Mahé, A. Sites de dépôt en libre accès et formes de médiations: Quelles évolutions? In *La médiation Numérique: Renouveau et Diversification des Pratiques*; Boustany, J., Broudoux, E., Chartron, G., Eds.; De Boeck: Brussels, Belgium, 2013; pp. 125–139. [CrossRef]
10. Mahé, A.; Prime-Claverie, C. Qui dépose quoi sur Hal-SHS? Pratiques de dépôts en libre accès en sciences humaines et sociales. *Rev. Française Sci. L'information Commun.* **2017**, *11*. [CrossRef]
11. Gayoso, E. *La Diffusion sur Hal, Academia et ResearchGate des Articles de Recherche des Revues Françaises de Sciences Humaines et Sociales*; Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation: Paris, France, 2020. Available online: <https://www.enseignementsup-recherche.gouv.fr/cid136723/le-soutien-a-l-edition-scientifique.html> (accessed on 12 May 2023).

12. Larrieu, M.; Pain, D. Ouvrir L'accès aux Dernières Publications de son Etablissement avec HAL—Retour D'expérience sur la mise en Place d'un Chantier D'import Effectué à L'université de Versailles Saint-Quentin-en-Yvelines. *Bulletin des Bibliothèques de France*, 9 September 2021. Available online: https://bbf.enssib.fr/matieres-a-penser/ouvrir-l-acces-aux-dernieres-publications-de-son-etablissement-avec-hal_70089 (accessed on 12 May 2023).
13. Mahé, A.; Prime-Claverie, C. Science ouverte et présence numérique des chercheurs en sciences humaines et sociales Une étude exploratoire à partir de deux plateformes en ligne: HAL-SHS et Hypotheses.org. *Doc. Numéri.* **2017**, *20*, 79–96. [CrossRef]
14. Schöpfel, J.; Kergosien, E.; Prost, H.; Barrié, J. « Pas si simple que ça . . . »: Une enquête sur l'usage de HAL par les unités de recherche des universités IdEx. *I2D—Inf. Données Doc.* **2022**, *2*, 150–183. [CrossRef]
15. Tabariès, A. Vers une métrique pour évaluer les métadonnées de documents scientifiques. *Rev. Française Sci. L'information Commun.* **2022**, *24*. [CrossRef]
16. Xia, J. Assessment of Self-archiving in Institutional Repositories: Across Disciplines. *J. Acad. Librariansh.* **2007**, *33*, 647–654. [CrossRef]
17. Xia, J.; Sun, L. Assessment of Self-Archiving in Institutional Repositories: Depositorship and Full-Text Availability. *Ser. Rev.* **2007**, *33*, 14–21. [CrossRef]
18. Tillman, R.K. Where Are We Now? Survey on Rates of Faculty Self-Deposit in Institutional Repositories. *J. Librariansh. Sch. Commun.* **2017**, *5*. [CrossRef]
19. Ciriminna, R.; Scurria, A.; Gangadhar, S.; Chandha, S.; Pagliaro, M. Reaping the Benefits of Open Science in Scholarly Communication. *Heliyon* **2021**, *7*, e08638. [CrossRef] [PubMed]
20. Xia, J. A Comparison of Subject and Institutional Repositories in Self-archiving Practices. *J. Acad. Librariansh.* **2008**, *34*, 489–495. [CrossRef]
21. Carr, L.; Brody, T. Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. *D-Lib Mag.* **2007**, *13*. [CrossRef]
22. Westell, M. Institutional Repositories: Proposed Indicators of Success. *Libr. Hi Tech* **2006**, *24*, 211–226. [CrossRef]
23. Bevan, S.J. Developing an Institutional Repository: Cranfield QUEprints—A Case Study. *OCLC Syst. Serv. Int. Digit. Libr. Perspect.* **2007**, *23*, 170–182. [CrossRef]
24. Neugebauer, T.; Murray, A. The Critical Role of Institutional Services in Open Access Advocacy. *Int. J. Digit. Curation* **2013**, *8*, 84–106. [CrossRef]
25. Boltze, J.; Höllerl, A.; Kuberek, M.; Lohrum, S.; Pampel, H.; Putnings, M.; Retter, R.; Rusch, B.; Schäffler, H.; Söllner, K. DeepGreen: Eine Infrastruktur für die Open-Access-Transformation. *O-Bib. Das Offene Bibl.* **2022**, *9*, 1–13. [CrossRef]
26. Vincent-Lamarre, P.; Boivin, J.; Gargouri, Y.; Larivière, V.; Harnad, S. Estimating Open Access Mandate Effectiveness: The MELIBEA Score. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 2815–2828. [CrossRef]
27. Schöpfel, J.; Azeroual, O. Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger. In *Future Directions in Digital Information*; Baker, D., Ellis, L., Eds.; Elsevier: Oxford, UK, 2021; pp. 19–37. [CrossRef]
28. Azeroual, O.; Schöpfel, J. Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries. *Publications* **2019**, *7*, 14. [CrossRef]
29. Azeroual, O.; Saake, G.; Abuosba, M.; Schöpfel, J. Data Quality as a Critical Success Factor for User Acceptance of Research Information Systems. *Data* **2020**, *5*, 35. [CrossRef]
30. Azeroual, O.; Saake, G.; Wastl, J. Data Measurement in Research Information Systems: Metrics for the Evaluation of Data Quality. *Scientometrics* **2018**, *115*, 1271–1290. [CrossRef]
31. Azeroual, O.; Saake, G.; Abuosba, M. Data Quality Measures and Data Cleansing for Research Information Systems. *J. Digit. Inf. Manag.* **2018**, *16*, 12–21. Available online: <http://www.dline.info/jdim/v16i12018/> (accessed on 12 May 2023).
32. Azeroual, O.; Schöpfel, J.; Pölonen, J.; Nikiforova, A. Putting FAIR principles in the context of research information: FAIRness for CRIS and CRIS for FAIRness. In Proceedings of the 14th International Conference on Knowledge Management and Information Systems (KMIS2022), Valletta, Malta, 24–26 October 2022; Available online: <https://hal.science/hal-03836525> (accessed on 12 May 2023).
33. d'Aquin, M.; Kirstein, F.; Oliveira, D.; Schimmler, S.; Urbanek, S. FAIREST: A Framework for Assessing Research Repositories. *Data Intell.* **2023**, *5*, 202–241. [CrossRef]
34. Björk, B.-C. Gold, green, and black open access. *Learn. Publ.* **2017**, *30*, 173–175. [CrossRef]
35. Tabariès, A. Étude des notices HAL: Vers des préconisations de dépôt. In Proceedings of the 16ème Journée Jeunes Chercheur-e-s en SIC, Lille, France, 25 May 2022; Available online: <https://hal.science/hal-03690997v1> (accessed on 12 May 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.