The Transformation of the Green Road to Open Access

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

Posted Date: 16 February 2023

doi: 10.20944/preprints202302.0268.v1

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

Article

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 - GERiiCO - Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication, Univ. Lille, F-59000 Lille, France
[2] ULR 4073 - GERiiCO - 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
* Correspondence: joachim.schopfel@univ-lille.fr

Abstract: (1) Background: The 2002 Budapest Open Access Initiative recommended on self-archiving of scientific articles in open repositories 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 on the French national HAL repository from more than 1,000 laboratories affiliated to the ten most important French research universities, with a focus on 2020, representing 14,023 contributor accounts and 166,939 deposits. (3) Results: We identified seven different types of contributor accounts, including deposits from nonfaculty staff and import flows from other platforms. Mediated nonfaculty contribution accounts for at least 48% of the deposits. We also identified difference 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

1. Introduction

Twenty years ago, the Budapest Declaration advocated open access to scientific information, i.e., the free and unrestricted online availability of research results [1]. For scientific journals the Budapest Declaration recommended two complementary strategies, the first of which was self-archiving, in other words, the deposit by researchers themselves of their articles in open repositories[1].

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 Cornell[2]. The

[1] The second strategic recommendation of the BOAI was the publishing of open access journals ("gold road").

[2] arXiv https://arxiv.org/

global Directory of Open Access Repositories[3] contains in January 2023 6,000 repositories, 89% of which are 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, HAL[4] was launched in 2001 by the public research organization CNRS[5] 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 actually (January 2023) more than 4,3m resources, mainly articles but also preprints, conference papers, dissertations and so on. However, HAL has changed over the time, and it is different today from the initial model 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 aims at a minimum to identify an institutions' production [7].
- Nearly 75% of the resources are non-full text deposits, i.e., metadata with or without abstract.
- Recently a government report recommended the use of HAL as a bibliometric tool for the assessment of public research in 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% articles have been self-archived [12]. Only 35% researchers in SSH deposit regularly on HAL [13]. Another survey with research laboratories showed that professional follow-up and nonfaculty deposits are significant elements of the institutional support of open access [14]. A negative impact of facilitated (and mediated) deposits on the metadata quality (richness, completeness) has been observed by [15].

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 since more than fifteen years. Two surveys of institutional repositories from the UK, Australia and other countries revealed a low rate of author self-archiving (<40%) and of full-text availability [16,17]. Both studies showed also 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 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 deposit 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 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 to build 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

[3] OpenDOAR https://v2.sherpa.ac.uk/opendoar/

[4] HAL https://hal.science/

[5] CNRS https://www.cnrs.fr/en

initial phase of a repository [23]. However, such 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]. Also, to increase the "buy-in" from academic staff, the process of acquiring research material could be embedded into the subject liaison role, 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 authorised libraries for publication in their repositories" [25]. Second, institutional open access policies requiring deposit for performance evaluation [26] which, together with mandates from funders, "will 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, has 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].

As part of a research project on open access strategies of more than 1,000 French research laboratories, we had the opportunity to assess their contributions to the national HAL repository. The purpose of this assessment is 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 part of 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 further development of repositories and research on open science.

2. Materials and Methods

We assessed the HAL deposits of 1,246 laboratories, affiliated to the ten most important French research universities which represent, together, 33,800 faculty, 24,000 PhD students and two-third of the most cited French publications worldwide[6] (Appendix A). These laboratories cover the whole range of scientific disciplines (Appendix B).

Based on the HAL-specific organizational structure codes of all these laboratories[7], 1,035,612 deposits have been identified and analyzed. 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 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 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 1,226 laboratories, 39,038 contributor accounts and 897,097 deposits.

---

[6] See Udice Group Universities https://www.udice.org/about-us/?lang=en

[7] AuréHAL https://aurehal.archives-ouvertes.fr/structure/index

 doi:10.20944/preprints202302.0268.v1

3. Results

3.1. Number of contributors

Since 2010, the number of contributors for all laboratories has continuously increased from 3,787 in 2010 to 14,023 in 2020 (Figure 1).

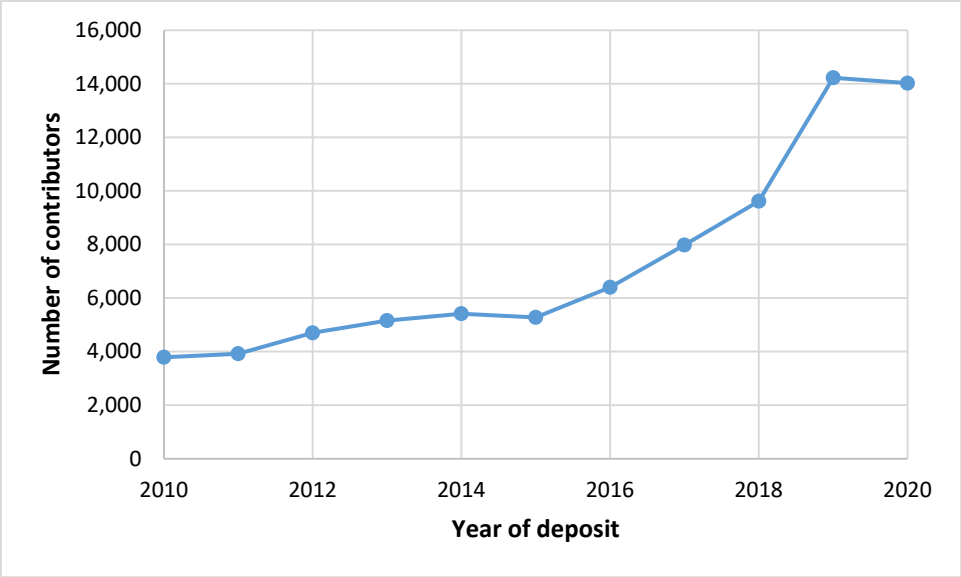

Figure 1. The number of contributors per year of deposit (2010-2020).

We can distinguish three periods: between 2010 and 2015, a slow progression from nearly 4,000 to more than 5,000 contributors (+39%); from 2015 to 2018, an acceleration of the increase of the number, passing from about 5,000 to nearly 10,000 contributors (+82%), which corresponds to the improvement (simplification) of the procedures of deposit; from 2018, a strong and sudden growth from 10,000 to 14,000 contributors (+48%), which corresponds to 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 with the number of deposits (r>.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.

At the same time, the number of deposits has been multiplied by four, passing from 33,237 deposits in 2010 to 166,939 deposits in 2020. This means that the average number of deposits by contributor increased by 36% (from 8.8 to 11.9), while the number of deposits per laboratory tripled (from 56 to 142), a growth which is not due to increased research performance but to increased use of HAL by the laboratories, in order to showcase their scientific production and to facilitate monitoring and research performance assessment.

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 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 by 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.

 doi:10.20944/preprints202302.0268.v1

- Administrative, technical and library staff of the authors' laboratory who 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 flows from bibliographic databases, reference management software and catalogues. Some laboratories follow-up their scientific production with an 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 from time-to-time metadata references from other open repositories; this was the case, for instance, when the French National Research Institute for Agriculture, Food and the Environment [8] 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 (e.g., from HEP Inspire) or publishers (e.g., Elsevier). A few contributor accounts correspond to workflows from other platforms, like Inspire HEP, the leading information platform for High Energy Physics, or from publishers who started to feed the HAL platform with their own metadata. Figure 2 provides an overview of these different categories.

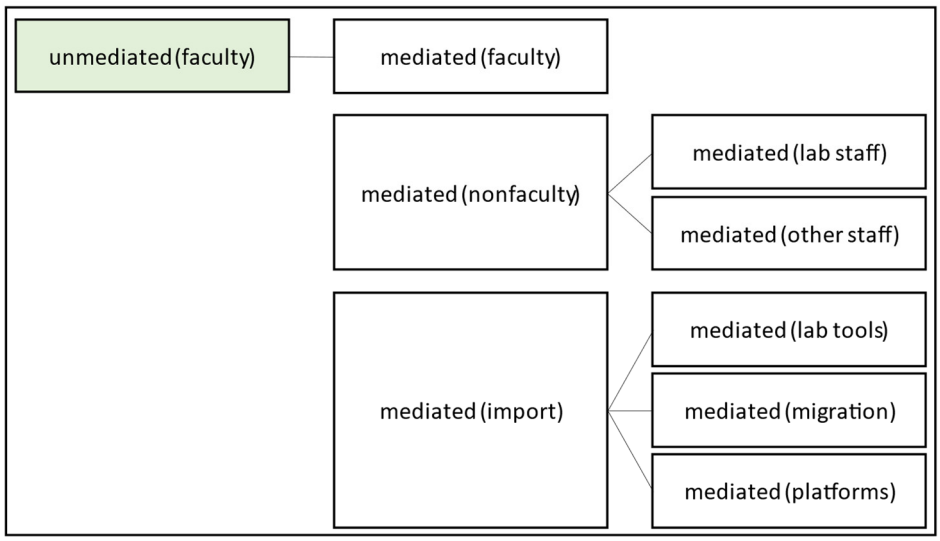

Figure 2. The typology of contributors (2020).

This typology 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 they also participate to mediated contribution, along with nonfaculty staff from the research laboratories or from other structures (academic libraries…) and with import from laboratory-based tools, from migration flows and from external platforms like institutional repositories or publishers' databases. In fact, the reality has changed and is much more heterogeneous.

3.3. The part of mediated, nonfaculty contributions

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 of the part of the researchers' self-archiving. However, it is possible to make an 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

[8] INRAE https://www.inrae.fr/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 February 2023 doi:10.20944/preprints202302.0268.v1

deposits. The curve of the deposits is a Pareto distribution: on the long tail, 20% publications have been deposited by 83% contributors, while on the "top the charts", 50% of all publications have been deposited by less than 1% contributors, most of them clearly identifiable as nonfaculty. As Figure 3 shows, many contributor accounts just deposited one, two or three publications during the whole year 2020.

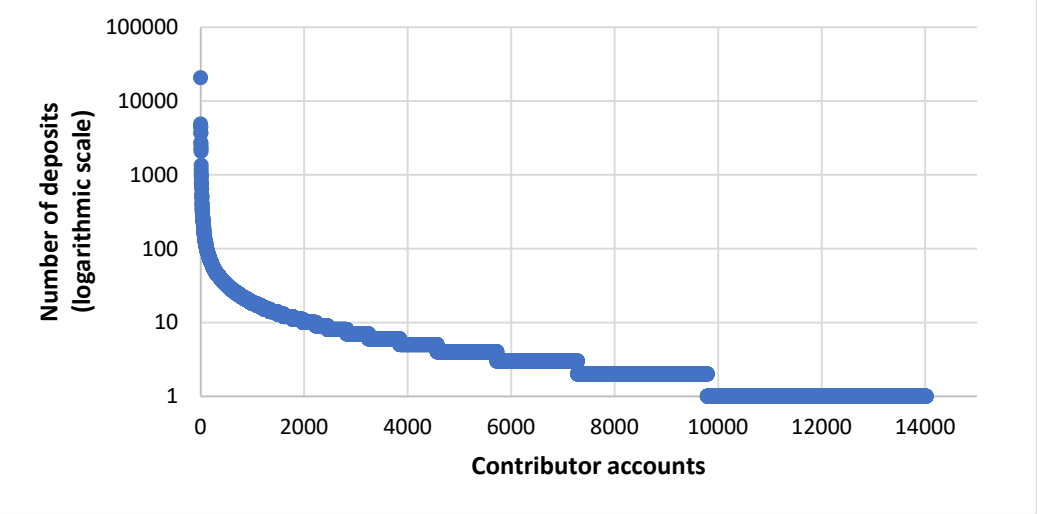

Figure 3. The distribution of deposits per contributor accounts (2020).

The part of nonfaculty contributor accounts appears rather low and can be estimated as 1,2% of all accounts. Yet, this low number of contributors accounts for 48% of all deposits. In other words, nearly half of all publications on HAL match with the concept of mediated, nonfaculty contributions (Figure 4).

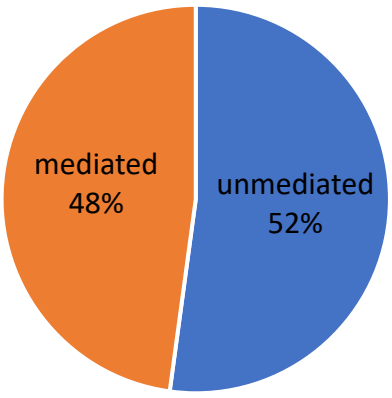

Figure 4. The part of mediated nonfaculty deposits (2020).

Input from other platforms accounts for 27% (2% from institutional repositories, 2% from publishers, the rest are migration flows). Input from personal nonfaculty staff accounts for 13%, while the other 8% is input from library, laboratory, or university avatar accounts.

3.4. Differences between laboratories

The 2020 sample consists of 1,176 research laboratories, ranging from 1 to more than 100 contributors and from 1 to nearly 3,000 deposits (median = 67). The correlation between the number of deposits and the number of contributors is .54. The higher the number of contributors, the higher

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 February 2023 doi:10.20944/preprints202302.0268.v1

the number of deposits of a given laboratory. Yet, more significant is the relationship between the number of deposits of the whole laboratory and the number of deposits of the laboratory's most important contributor account; here, the correlation coefficient is .86. In other words, while the number of contributors is relevant, the importance of the first contributor account is even more relevant for laboratories' total number of deposits on HAL. Figure 5 shows this strong correlation for the whole sample of 1,176 research laboratories.

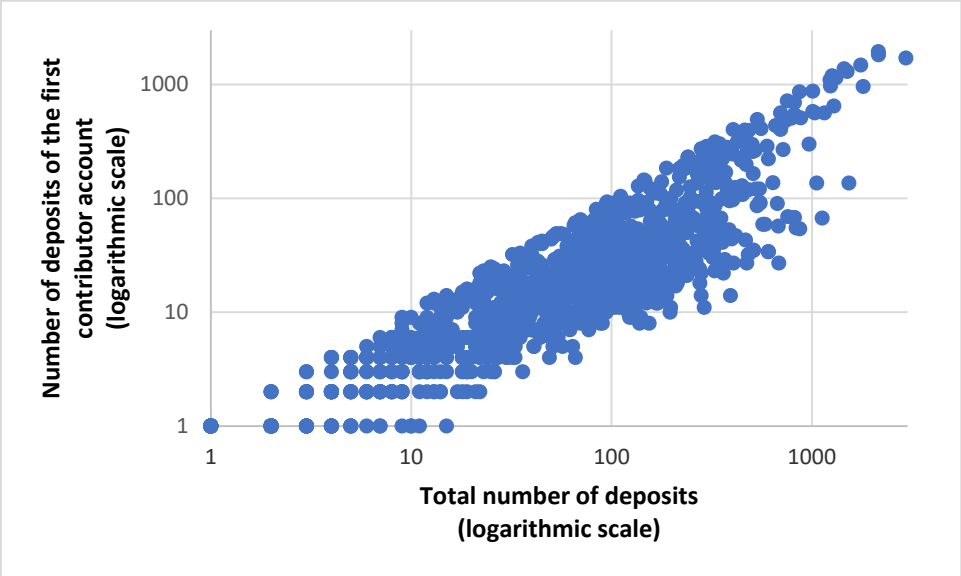

Figure 5. Total number of deposits and the part of the first contributor per laboratory (2020).

Figure 5 visualizes the large variety between the laboratories – some of the them deposited less than 10 publications on HAL in 2020 while others published more than 100 or even more than 1,000 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 and most important 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 systematic and mediated nonfaculty contribution to the HAL repository.

The differences between universities are less important, except for one (Aix-Marseille) where the part of laboratories with mediated contributions seems much lower than expected, compared to the other universities.

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 (.69), while the correlation with the deposits of the first contributor account is lower (.73). Obviously, for these laboratories the part of mediated deposits is less important than in the other research fields.

Second, the role of the first contributor account seems more important in laboratories of the field 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, 3.2).

4. Discussion and conclusion

As mentioned above, the research has two methodological shortfalls. First, we counted events (=deposit per laboratory per year). If a deposit (article, communication…) has coauthors from two

different laboratories from our sample, it will be counted two times 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 (2,869, or 1.7%). Second, we assessed the activity of each contributor accounts but we did not assess each deposit. In other words, we cannot compare the metadata of authors with the contributor account, as done by [9,10]. Yet, our results are similar enough to those based on direct matching between the creator (author) and contributor fields and 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 1,000 laboratories from the ten most important French research universities, only half of the 2020 deposits are self-archived while the other half represents mediated, mostly nonfaculty contribution. 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 back-upped in a public dark archive hosted by CINES at Montpellier[9]; and the development of an infrastructure with allows monitoring and assessment of the scientific production of the individual researchers, the laboratories, and the universities. Because of the institutional support and contribution 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 monitor[10]. 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 an open research information management system.

This transformation is not specific to HAL, and it is not specific for France [27]. Also, 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 the data and metadata quality. From the moment on the platform performs monitoring and assessment functions, the data quality 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] and 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 a serious issue for the findability of resources on open repositories. But the more repositories and research information management systems will 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, persons.

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

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

Further research is required to assess this transformation of the green road to open access, as well on the level of the infrastructures (systems), of the data (content) and of the usage by researchers and institutions. We need more evidence about the role and impact of mediated nonfaculty

[9] National Computer Center for Higher Education https://www.cines.fr/

[10] Baromètre français de la Science Ouverte https://barometredelascienceouverte.esr.gouv.fr/

contribution, 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. 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.

Author Contributions: Conceptualization, methodology, and validation, all; data curation, B.J. and H.P.; writing—original draft preparation, J.S.; writing—review and editing, all; 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 data will be shared on the French repository Recherche Data Gouv.

Acknowledgments: The authors acknowledge administrative support from the GERiiCO 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.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, 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	**1,246**

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	**1,246**

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

References

1. BOAI (2002). *Budapest Open Access Initiative;* Open Society Foundations: Budapest, Hungary, 2002. https://www.budapestopenaccessinitiative.org/
2. Harnad, S.; Brody, T.; Vallières, F.; Carr, L.; Hitchcock, S.; Gingras, Y.; … Hilf, E. R. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review* **2004**, *30*(4), 310–314. https://doi.org/10.1016/j.serrev.2004.09.013
3. Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Portal: Libraries and the Academy* **2003**, *3*(2), 327–336. https://doi.org/10.1353/pla.2003.0039
4. Prosser, D. C. (2004). 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, United Kingdom, 2004; pp. 99–117.

5. Berthaud, C.; Charnay, D.; Fargier, N. (2021). Diffuser et pérenniser le savoir scientifique : 20 ans d'histoire de HAL. *Histoire de La Recherche Contemporaine* **2021**, *10*(2). https://doi.org/10.4000/hrc.6330
6. Charnay, D. (2019). « Avec HAL, nous voulions créer un arXiv multidisciplinaire ». *Hermès* **2019**, *85*(3), 94. https://doi.org/10.3917/herm.085.0094
7. Magron, A. (2017). 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. https://hal.archives-ouvertes.fr/sic_01697164
8. Thibault, F.; Streliski, S. (2022). *Les indicateurs bibliométriques pour les SHS - Etat de la question*; Alliance Athéna: Paris, France, 2022. http://www.alliance-athena.fr/rapport-de-lalliance-athena-les-indicateurs-bibliometriques-pour-les-shs-etat-de-la-question/
9. Prime-Claverie, C.; Mahé, A. (2013). Sites de dépôt en libre accès et formes de médiations : quelles évolutions ? In *La médiation numérique : renouvellement et diversification des pratiques*; Boustany, J., Broudoux, E., Chartron, G., Eds.; De Boeck: Brussels, Belgium, 2013; pp. 125–139. https://doi.org/10.3917/dbu.chron.2013.01.0125
10. Mahé, A.; Prime-Claverie, C. (2017). Qui dépose quoi sur Hal-SHS? Pratiques de dépôts en libre accès en sciences humaines et sociales. *Revue Française des Sciences de l'information et de la Communication* **2017**, *11*. https://doi.org/10.4000/rfsic.3315
11. Gayoso, E. (2020). *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. https://www.enseignementsup-recherche.gouv.fr/cid136723/le-soutien-a-l-edition-scientifique.html
12. Larrieu, M.; Pain, D. (2021). Ouvrir l'accès aux dernières publications de son établissement 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* **2021**, September 9. https://bbf.enssib.fr/matieres-a-penser/ouvrir-l-acces-aux-dernieres-publications-de-son-etablissement-avec-hal_70089
13. Mahé, A.; Prime-Claverie, C. (2017). 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. *Document Numérique*, **2017**, *20*(2–3), 79–96. https://doi.org/10.3166/dn.2017.00010
14. Schöpfel, J.; Kergosien, E.; Prost, H.; Barrié, J. (2022). « Pas si simple que ça… » : Une enquête sur l'usage de HAL par les unités de recherche des universités IdEx. *I2D - Information, Données & Documents* **2022**, *2*(2), 150–183. https://doi.org/10.3917/i2d.222.0150
15. Tabariès, A. (2022). Vers une métrique pour évaluer les métadonnées de documents scientifiques. *Revue Française des Sciences de l'information et de la Communication* **2022**, 24. https://doi.org/10.4000/rfsic.12258
16. Xia, J. (2007). Assessment of Self-archiving in Institutional Repositories: Across Disciplines. *The Journal of Academic Librarianship* **2007**, *33*(6), 647–654. https://doi.org/10.1016/j.acalib.2007.09.020
17. Xia, J.; Sun, L. (2007). Assessment of Self-Archiving in Institutional Repositories: Depositorship and Full-Text Availability. *Serials Review* **2007**, *33*(1), 14–21. https://doi.org/10.1016/j.serrev.2006.12.003
18. Tillman, R. K. (2017). Where Are We Now? Survey on Rates of Faculty Self-Deposit in Institutional Repositories. *Journal of Librarianship and Scholarly Communication* **2017**, *5*(1). https://doi.org/10.7710/2162-3309.2203
19. Ciriminna, R.; Scurria, A.; Gangadhar, S.; Chandha, S.; Pagliaro, M. (2021). Reaping the Benefits of Open Science in Scholarly Communication. *Heliyon* **2021**, *7*(12), e08638. https://doi.org/10.1016/j.heliyon.2021.e08638
20. Xia, J. (2008). A Comparison of Subject and Institutional Repositories in Self-archiving Practices. *The Journal of Academic Librarianship* **2008**, *34*(6), 489–495. https://doi.org/10.1016/j.acalib.2008.09.016
21. Carr, L.; Brody, T. (2007). Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. *D-Lib Magazine* **2007**, *13*(7–8). https://doi.org/10.1045/july2007-carr
22. Westell, M. (2006). Institutional Repositories: Proposed Indicators of Success. *Library Hi Tech* **2006**, *24*(2), 211–226. https://doi.org/10.1108/07378830610669583
23. Bevan, S. J. (2007). Developing an Institutional Repository: Cranfield QUEprints – A Case Study. *OCLC Systems & Services: International Digital Library Perspectives* **2007**, *23*(2), 170–182. https://doi.org/10.1108/10650750710748478
24. Neugebauer, T.; Murray, A. (2013). The Critical Role of Institutional Services in Open Access Advocacy. *International Journal of Digital Curation* **2013**, *8*(1), 84–106. https://doi.org/10.2218/ijdc.v8i1.238
25. Boltze, J.; Höllerl, A.; Kuberek, M.; Lohrum, S.; Pampel, H.; Putnings, M.; … Söllner, K. (2022). DeepGreen: Eine Infrastruktur für die Open-Access-Transformation. *O-Bib. Das Offene Bibliotheksjournal* **2022**, *9*(1), 1–13. https://doi.org/10.5282/o-bib/5764
26. Vincent-Lamarre, P.; Boivin, J. ; Gargouri, Y. ; Larivière, V. ; Harnad, S. (2016). Estimating Open Access Mandate Effectiveness: The MELIBEA Score. *Journal of the Association for Information Science and Technology* **2016**, *67*(11), 2815–2828. https://doi.org/10.1002/asi.23601

27. Schöpfel, J.; Azeroual, O. (2021). 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, United Kingdom, 2021; pp. 19–37. https://doi.org/10.1016/B978-0-12-822144-0.00002-1
28. Azeroual, O.; Schöpfel, J. (2019). Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries. *Publications* **2019**, *7*(1), 14. https://doi.org/10.3390/publications7010014
29. Azeroual, O.; Saake, G.; Abuosba, M.; Schöpfel, J. (2020). Data Quality as a Critical Success Factor for User Acceptance of Research Information Systems. *Data* **2020**, *5*(2), 35. https://doi.org/10.3390/data5020035
30. Azeroual, O.; Saake, G.; Wastl, J. (2018a). Data Measurement in Research Information Systems: Metrics for the Evaluation of Data Quality. *Scientometrics* **2018**, *115*(3), 1271–1290. https://doi.org/10.1007/s11192-018-2735-5
31. Azeroual, O.; Saake, G.; Abuosba, M. (2018b). Data Quality Measures and Data Cleansing for Research Information Systems. *Journal of Digital Information Management* **2018**, *1*, 12–21. http://www.dline.info/jdim/v16i12018/
32. Azeroual, O.; Schöpfel, J.; Pölönen, J.; Nikiforova, A. (2022). Putting FAIR principles in the context of research information: FAIRness for CRIS and CRIS for FAIRness. In 14th International Conference on Knowledge Management and Information Systems (KMIS2022), Valletta, Malta, 24-26 October 2022. https://hal.science/hal-03836525