

Exploring Humanities Research Data in Figshare

Samia Takhtoukh
ORCID 0000-0001-8543-5537
GERiiCO Laboratory
University of Lille, France

Abstract

Purpose/Thesis: The purpose of this study is to explore the uptake and use of Figshare by humanities community and to know if opening of the research data is enough to make them visible and findable.

Approach/Methods: The overall research design was to gather a sample of humanities artifacts found on Figshare through its API, and then to analyze such indicators as the number of articles by discipline, the evolution of contribution per year, and the number of views and downloads by affiliation to an institution.

Results and conclusions: The humanities community engagement with Figshare varies depending on the discipline; in addition, even when research data is published, 50% is rarely viewed. This means that making research data open is not enough to make it visible and to ensure their reuse. The final conclusion is that articles affiliated to an institution have more chance to be visible than unaffiliated ones.

Research limitations: The limit of our study is that it can only analyze data present in Figshare and not the data that is missing, so we cannot know the reasons for the small visibility of humanities research data; the study of other data warehouses is necessary to give more explanation, but the results conform to the pattern visible in the data presented in several studies of research management practices.

Originality/Value: The first study that explores the humanities contents the multidisciplinary data repository Figshare.

Keywords

Figshare. Humanities. Open data. Research data.

Received: 14 June 2019. Reviewed: 4 July 2019. Revised: 22 July 2019. Accepted: 12 August 2019.

1. Introduction

Since the sharing of research data is of great scientific, economic and social importance, the movement of Open Access is expanding more and more to research data. Indeed, the sharing of research data has several advantages in terms of efficiency, power and rigor. To take advantage of these benefits, many funding agencies put policies in place to promote the optimal use and reuse of data in which their funds were invested. They encourage good data practices, investing in data infrastructure and raising data awareness (Swan & Brown, 2008). In addition, several formalized data management services and tools (Figshare, Zenodo, Nakala, Dryad) have emerged as a result of funding agency's requirement that the grant applicants explicitly document their methods of storing research data and make them available for future use.

Specific studies of research data management practices were conducted at different universities, in different countries and in different years. They have shown that the management

of research data is still not common among researchers in the humanities (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). A struggle to organize the data in a presentable and useful way, the uncertainty as to the copyright and the licenses, the ignorance of the warehouse to use, the lack of time for the deposit of the data and the costs of sharing the data (Stuart et al., 2018) are the main open data challenges in the social sciences and humanities. With all these obstacles, it seems that the researchers are very far from embodying the ideal of open science in their practice.

But it is also important to conduct studies of general multidisciplinary repositories in order to know if the willingness to make the research data more open depends on the discipline. It is also good to know if opening research data is enough to make it visible and findable and if the constraints of funding agencies and institutions policies have an impact on making research data available to the public.

In this article, we will focus on Figshare, which is a multidisciplinary research data repository. Established in 2011, it allows researchers from all disciplines to upload online any type of electronic information that can be used by others. Technically, Figshare users can make all their search results available in a way that is citable, shareable and discoverable. Data may be shared privately with collaborators or made public in the name of open research, or to comply with the mandates of funders and publishers.

Figshare also offers a Figshare for Institutions service which allows institutions to easily aggregate research at the departmental and institutional levels, automatically providing an institutional repository with reporting functions. They can make use of indicators about all the research they generate, which is not possible with the silo systems that exist in many research organizations today. Figshare for Institutions ensures that their data management requirements are made clear for the researcher.

As Figshare does not target a specific discipline, it allows various types of resources to be uploaded, making it the main universal scientific repository of this type. It will allow us to take account of the disciplines in our comparative study. In addition, Figshare publishes views, shares and downloads counts. We consider these counts informative for digital resources (Konkiel & Scherer, 2013), therefore we will use this alternative online metrics to measure the visibility of data. The following research questions drive this of Figshare: Is there any difference in community engagement depending on the discipline? Does institution affiliation have any impact on the visibility of research data on Figshare?

To answer these questions, we have created a web application to automatically query the Figshare API; the source code is available as open source in GitHub¹.

2. Conceptual background

2.1. *Potential benefits of open research data*

The movement of Open Access is expanding more and more to research data, which is defined as following:

¹ The source code is accessible at: <https://github.com/aallou/figshare-vue-app>

Research data refers to information, in particular facts or numbers collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images (Open Access, n.d.).

The sharing of research data is of great importance; in particular, it helps to accelerate innovation through the exploitation, which can stimulate innovation and increase the collective knowledge. This knowledge will be translated into products and services with significant social, economic and scientific value (Ali-Khan et al., 2018). On the other hand, the sharing of research data guarantees an increase in quality. In fact, openly sharing research data will lead to more controllable and reproducible research, which will make it possible to easily validate the results presented in scientific publications. In addition, making research data open by default will help ensure greater transparency and confidence, and help to prevent massive data loss (Jacobson et al., 2014).

To take advantage of these benefits, many funding agencies put policies in place to promote the optimal use and reuse of data in which funds were invested. For example, in 2013 the European Commission launched a pilot project of open access to data from publicly funded research only (Research Infrastructures, n.d.); the pilot expanded to all disciplines in 2016. All stakeholders involved in scientific research, including people from academia, industry, funding agencies and scientific publishers, have come together to jointly develop and agree on a set of concise and measurable principles known by the acronym FAIR, suggesting that the data should be Findable, Accessible, Interoperable and Reusable. The intention is that these principles will serve as guidelines for those wishing to improve the reuse of research data. It should be noted that the FAIR principles place particular emphasis on improving the ability of machines to search and use data automatically, as well as facilitating their reuse by individuals. Funding agencies increasingly require the research data (and publications) resulting from funded research projects to be published in open access. However, as numerous as the potential opportunities are, there remain obstacles to managing and opening up research data.

2.2. Issues around research data for institutions and researcher

As funding agency requires that the grant applicants explicitly document how research data will be stored and made available for future use, many institutions have adopted models and developed tools to help grant applicants submit competing data management plans (Weber, 2013). Nevertheless, the development of these policies (HLEG, 2010) is not necessarily followed by their implementation in organizations; moreover, studies have shown that there is a gap between funding bodies and best practices for institutions supporting research activities (Weber, 2013). Indeed, the openness of the data represents a profound change, in particular in human and social sciences, and poses several challenges:

To collect, curate, preserve and make available ever-increasing amounts of scientific data, new types of infrastructures will be needed. The potential benefits are enormous but the same is true for the costs. We therefore need to lay the right foundations and the sooner we start the better (HLEG, 2010).

Data sharing can cause practical problems (Borgman, 2012), especially in terms of interoperability, standardization, collaboration, curation, decentralization of content integrity

(Girard, 2017). In addition, researchers who are the major producers of data can be very cautious and their current practices do not necessarily fit into this movement (Prost & Schöpfel, 2015).

The questions for the researcher are not simple either; the openness of the data implies that researchers are particularly concerned about the quality of the data they offer, as well as the clarity of the documentation attached to them. Should the researcher develop new practices and skills to enable others to replicate or reuse her/his data? Is it necessary for him to verify the integrity and interoperability of the data, to decide which data sets will be shared, considering in particular their potential for re-use and the costs generated, respecting the terms of the legal, economic, ethical and technical sharing (Carbou, 2017)?

In addition to depositing a dataset in a repository, the researcher should also provide the appropriate information about the dataset, which is known as metadata. Metadata describes the dataset and makes it possible for others to find, understand, and reuse the data. Besides standard information such as the creator and contributors of the dataset, the title, year of publication, and access rights, it can be necessary to add documentation such as codebooks, lab journals, informed consent forms and used software. There are various metadata standards for different disciplines, according to which the researcher must supply relevant additional information, necessary to make specific datasets comprehensible for the other users. For example, archaeological datasets require metadata about the spatial coverage area, while linguistics datasets require information about the language. Thus, as sharing is not an end in and of itself, the researcher should also develop new practices and skills to make the reusing of her/his data possible.

3. Methodology

The overall research design was to gather a sample of humanities artifacts found on Figshare and then to analyze some indicators like the number of articles by discipline, the evolution of contribution per year, and the number of views and downloads by affiliation to an institution (or lack thereof). In order to answer our research questions, we used the API provided by Figshare², which exposes several services to execute queries on its content. All communications are via https and all data is encoded in JSON format.

3.1. Figshare API

To collect data, it is essential that our first step in exploring this API is to understand how content is structured in Figshare, so that we might pose relevant questions. To accomplish this, we made use of the documentation provided (Figshare, 2018). Any resource uploaded in Figshare is an “article”; the article may be public or private. A private article is inaccessible and is beyond the scope of the study as we are interested in open data only. We found that Figshare users do not declare a specialty domain but are asked to enter one or more categories for each downloaded article. We therefore based our queries on the category to deduce the field of the resource. We used queries applying a filter by “category”, to limit

² <https://api.figshare.com/v2>

ourselves to articles in the humanities. Next, we relied on the “publication date” of the articles to restrict our study to the period between January 2015 and July 2019. We also used the concept of a “group” to differentiate articles deposited by an institution account and those deposited by a researcher without a known institutional membership.

Figshare makes it possible to manage several types of data:

- Book
- Chapter
- Composition
- Conference contribution
- Data management plan
- Dataset
- Educational resource
- Event
- Figure
- Funding
- Journal contribution
- Media
- Monograph
- Online resource
- Peer review
- Performance
- Physical object
- Poster
- Preprint
- Presentation
- Report
- Software
- Standard
- Thesis
- Workflow

Figshare API makes it possible to run out queries by filtering the results by the “type of item”. However, we did not apply a particular filter for the type because we wanted to analyze all types of data.

We encountered difficulties in determining the geographical origin of the data from research conducted in the area of human sciences present in Figshare. The API does not provide this information. To work around this problem we had to recover the ORCID of the author, and then query the ORCID database through its API. Unfortunately, the majority of authors on Figshare do not have expose their ORCID ID. We therefore could not know if this warehouse is used by researchers in some countries more than others; thus our identification of the country of origin of the resource and the design of Figshare and its API was limited.

The set of subcategories of Humanities does not necessarily correspond to disciplines, so we have selected only six, which were the most generic and which corresponded best to universally known disciplines. These were Linguistics, Law, History, Philosophy, Art and Literature.

To measure community engagement, it is possible to use the overall user acceptance taking into account the quantitative (number of users) and qualitative (best practices) aspects (Koureas et al., 2016). The number of users is not significant in the case of Figshare, because an author may not be active (having never deposited resources). In this case her/his resource was filed by a third party, and the author of the resource is a user of Figshare himself. For these reasons, we chose the number of articles per sub-category as an indicator of a community’s overall membership. We began by comparing the number of open access articles in Figshare between January 2015 and May 2019 in the six sub-categories chosen.

The Figshare API also provides access to online metrics, such as the number of views and downloads; we have used these metrics to measure the visibility of articles by affiliation to an institution (or lack thereof).

3.2. *Web application for data processing*

The next step of our study was the creation of a web application that allowed us to automatically execute various queries and to perform the necessary processing to restore the

information that interests us. The creation of such application³ was very beneficial and allowed us to reproduce the results of this research at any time without having to store them in personal databases.

The application retrieved data from Figshare and performed the necessary calculations to generate tables with all expected statistics. The goal of having a reusable application was to eventually expand the scope of this study in the future, and to make any researcher able to reproduce the results at any time. Indeed, our web application allowed restoring the data for a category and a period of the researcher's choosing. For this study, our parameters were the category Humanities and the period between January 1, 2015 and July 1, 2019, but it was possible to change the scope of the study without additional effort to collect and process the data. Note that the application could also download the data generated in CSV format for use by other tools for analysis.

Figshare presents a set of categories on main navigation page⁴ including the category Humanities, but when downloading an article, it allows authors to select additional sub-categories from a longer list. In our case study, we looked at 163 subcategories. The application allows restoring the information for all subcategories of Humanities and groups the results by subcategory; they are then displayed in a web page with a paging system to browse (Fig.1).

By pressing the Details link for a given sub-category, a list of results is displayed. Our web application makes it possible to browse all the articles of a given subcategory and to recover the following information for every article: the total number of articles by category for the time span, the number of articles by item type, the number of articles published each year for each sub-category and the number of articles with group ID and without group ID. The group ID identifier is not empty if the article is affiliated with an institution.

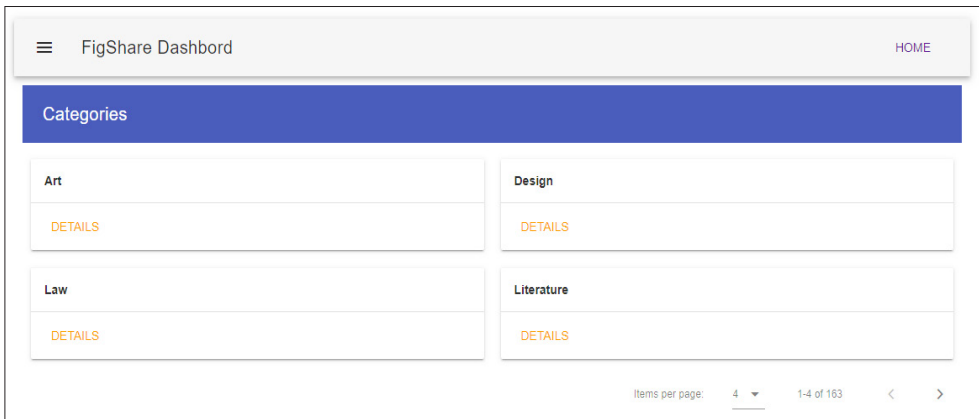


Fig.1. The list of results by subcategory

In addition, we were able to browse all the articles for a given sub-category and retrieve the number of views, downloads and sharing for every article. Although these metrics are

³ The application is accessible at: <https://vigorous-cori-9b6eb9.netlify.com/categories>

⁴ <https://figshare.com/browse>

widely used to indicate the impact of academic articles, they are to be considered with care (Bornmann, 2015).

Altmetrics refers to data sources, tools, and metrics (other than citations) that provide potentially relevant information on the impact of scientific outputs (e.g., the number of times a publication has been tweeted, shared on Facebook, or read in Mendeley). It also allows a broader interpretation of the concept of impact and to more diverse forms of impact analysis (Waltman & Costas, 2014). According to Taylor and Plume

altmetrics hold great promise as a source of data, indicators and insights about online attention, usage and impact of published research outputs (Taylor & Plume, 2014).

Since Figshare tracks the views and downloads statistics for the published research data, these statistics indicators were used in this study as a source for altmetrics to measure the visibility of articles.

The data was imported to Microsoft Excel and prepared for quantitative analysis.

4. Results

4.1. Humanities participation in Figshare by discipline

We've noticed that there is indeed a difference of participation between the sub-categories chosen. In fact, over the same period, the number of articles found in Figshare was 1361 for History, 1342 for Linguistics, 1280 for Law, 681 for Philosophy, 504 for Art and 404 for Literature (Fig. 2).

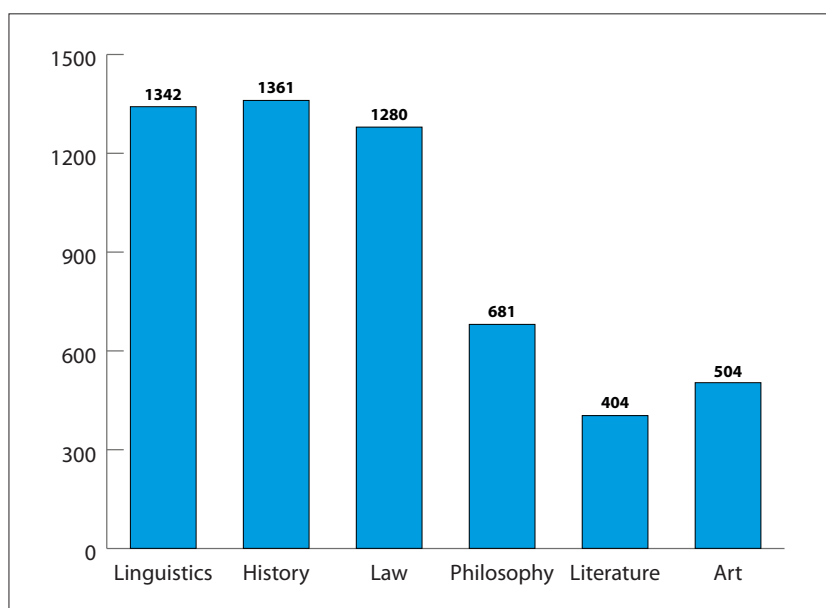


Fig. 2. The distribution of articles per discipline

4.2. Humanities participation evolution by field per year

Next we wanted to show how this participation has evolved between 2015 and 2018. We have excluded the articles from 2019. As the year is not yet finished, the comparison will not be significant, and the indicator will not be relevant.

The participation curve grows at different rates depending on the discipline, for example, the number of articles in Linguistics was three times greater in 2018 than in 2015, while the number of articles in Philosophy was almost the same for four years. A relative increase was visible in 2018 for Law, History and Art (Fig. 3).

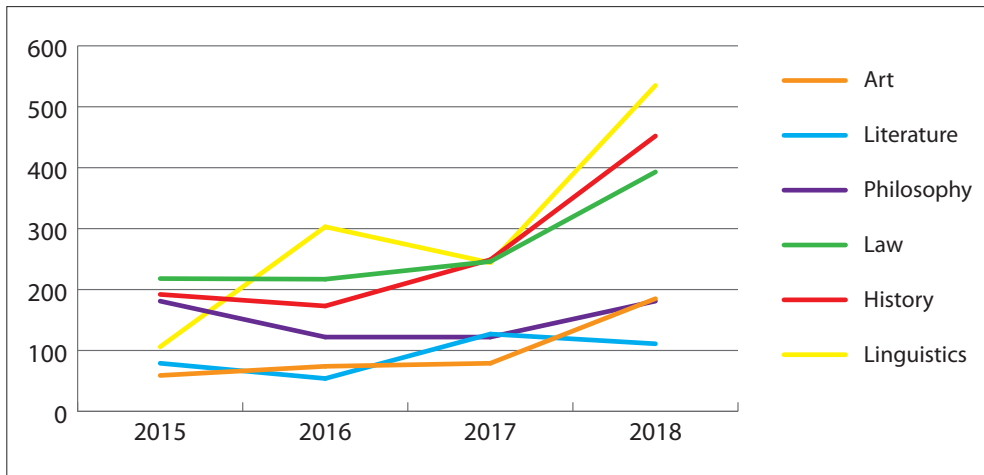


Fig. 3. The growth of the contributions across 4 years

4.3. Affiliated vs non-affiliated datasets

Figshare offers a service for institutions that ensures that their data management requirements are made transparently for the researcher. This feature has attracted some universities' attention which have chosen Figshare for Institutions for management of their research data. For the six chosen fields, the majority of articles were not affiliated to any institution, and only 30% of articles had metadata records about that issue. But there was a small difference between the fields. 60% of articles in Philosophy were affiliated to an institution, followed by 53% for Art. The percentage of affiliated articles for the four other fields was lower: 35% for History, 33% for Literature, 22% for Law and only 12% for Linguistics (Fig. 4).

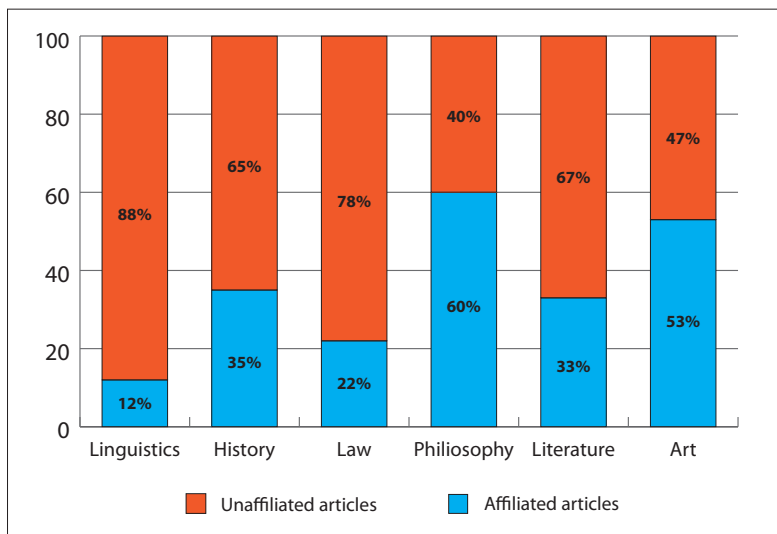


Fig. 4. Articles published with institutional affiliation

4.4. Data types by disciplines

The results show that for the six fields, the most used data types were “Journal contribution”, followed by the “Dataset” type which was in second position for almost all disciplines.

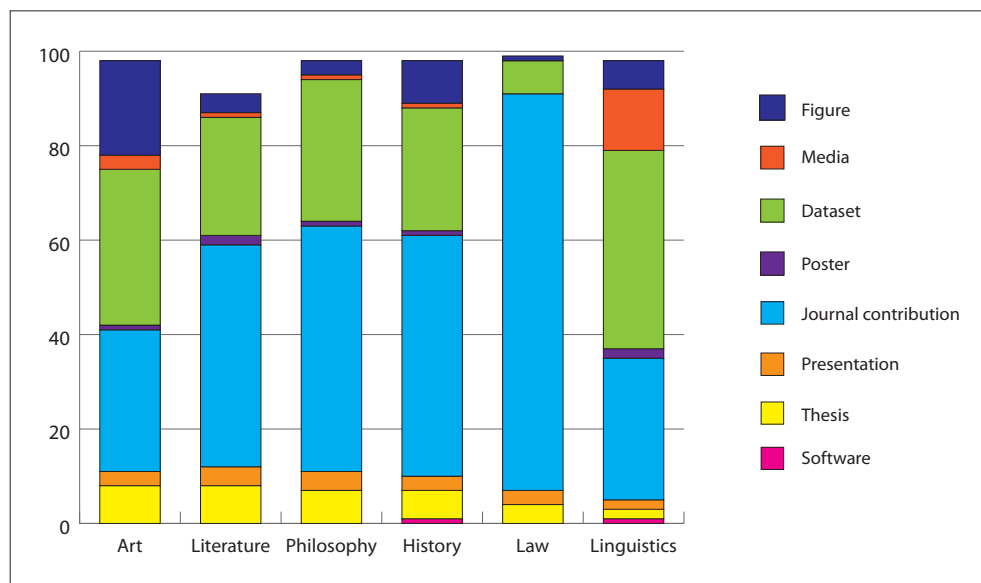


Fig. 5. Data types by disciplines

4.5. Visibility of articles

Figshare provides access to online metrics, such as the number of views that we used for measuring the visibility of open data. The number of views was very uneven. We used the quartiles for each field to calculate the interquartile range, which is a measure of variability around the median. The quartile breaks down the data into quarters so that 25% of the measurements are less than the lower quartile, 50% are less than the mean, and 75% are less than the upper quartile (Tab. 2).

Tab. 2. Quartiles for number of viewed articles

	Art	Law	History	Linguistics	Literature	Philosophy	All fields
Min	0	0	0	0	0	2	0
Lower quartile	15	6	10	13	56.75	14	11
Median	139	15	30	97.5	154	99.5	52
Upper quartile	327	41.5	94.5	252	331.5	245.25	197
Max	7677	13 275	2093	8378	3916	2584	13 275

Overall, 50% of articles has been viewed less than 52 times and 75% less than 197 times. Nevertheless, the variability around the median changes from one discipline to another. For example, 75% of Law and History articles has been viewed less than 95 times. But in the fields of Philosophy, Literature, Linguistics and Art, 50% of articles has been viewed more than 95 times. Literature is the field with the more viewed open research data, with 50% more than 154 times and 25% more than 331 times.

But it seems that some articles from other disciplines have found an audience: the most viewed article is from Law with 13275 views (Ruiz, 2018), followed by a Linguistics article with 8378 (Styles, 2017), then an Art article with 7677 (Snider, 2015).

Tab. 3. Quartiles for number of downloaded articles

	Literature	Art	Law	Linguistics	Philosophy	History	All fields
Lower quartile	4	2	0.25	3	2	1	1
Median	23	2	3	18	15	5	10
Upper quartile	72.5	2	12	65	48.5	13	44.25
Max	388 102	2	5510	40927	11 434	494	388 102

Concerning the number of downloads (Tab 3.), more than 75% of articles has been downloaded less than 45 times. There were no differences concerning the discipline. But some of the uploaded resources have generated much interest (more than 388 000 downloads for a Literature article).

We also compared the mean for the views of articles by publication date, which was very low for both affiliated and unaffiliated articles published after 2018. But the mean view of affiliated articles was generally higher than the mean view of unaffiliated articles for articles published in the same period (Fig. 6).

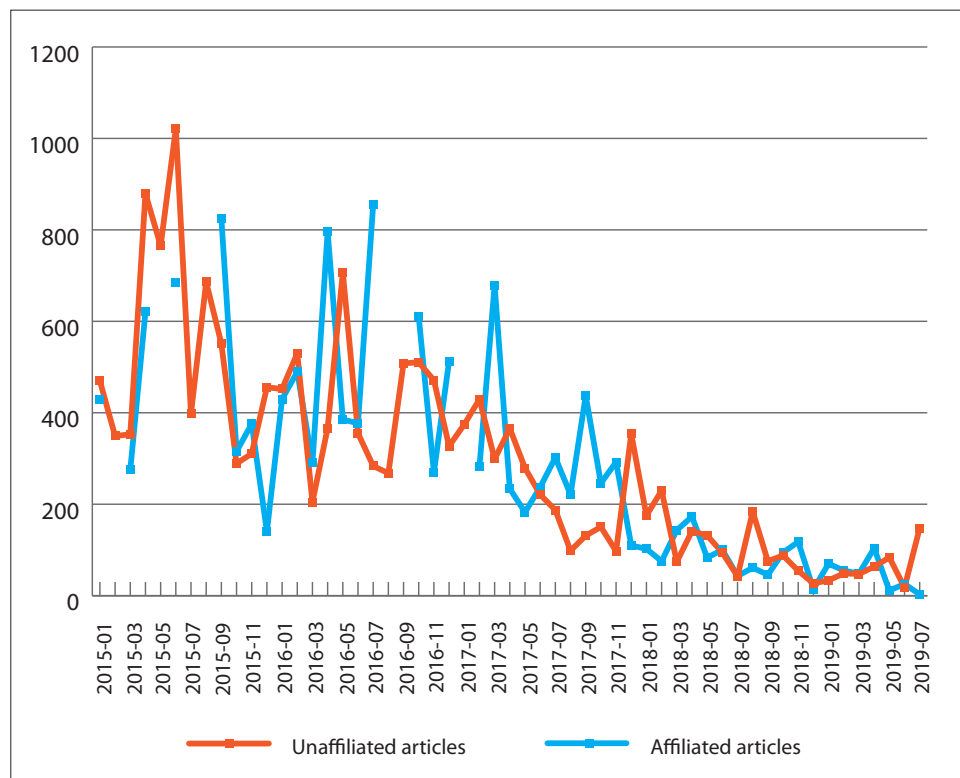


Fig. 6. Mean of views by publication date and by affiliation

5. Interpretation

5.1. Community engagement

The current study found that there is a difference in participation in depositing research data on Figshare across the disciplines in Humanities. The number of articles in Linguistics, Law and History was twice as large as in Philosophy, Art, and Literature. We also noticed that the growth curve increased at different rates according to the disciplines. The disciplines with a higher number of affiliated articles (Philosophy, Art) had more stable evolution curve. However, while an almost identical number of articles had been uploaded on Figshare each year, the disciplines with a higher number of affiliated articles had consistently very low contribution. The contribution in disciplines with a large number of unaffiliated articles (Linguistics, Law) increased quickly from year to year and these were the disciplines with the highest number of articles published. For History and Literature, the distribution of affiliated and unaffiliated articles was identical, but overall, the number of contributions was higher for History, and their respective growth curves changed in very different ways.

The results showed that while the Humanities community engagement for Figshare varied across the disciplines, overall, it remained very weak. Indeed, the total number of datasets

published in the six disciplines was very low (1361 articles for History) for almost five years. The community engagement was not related to affiliation to institution. The presence of humanities data on Figshare was not always framed by institutional membership policies. Humanities researchers were using Figshare to for opening their research data even if the institution they work for is was not affiliated to this data warehouse.

Previous studies have shown that the management of research data is still not common practise among researchers in the humanities (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). In humanities, individual research can rarely generate shareable data, but digitization initiatives create shareable resources such as photographs of works of art, historical documents or cultural artefacts (Gorman, 2007). This explains why History is at the top of the list in terms of the number of articles published on Figshare. Therefore, it may be interesting to compare this results with other data warehouses like Zenodo or Dryad to study if this phenomenon is typical to Figshare or is it common in humanities practices.

Technically, Figshare allows its users to download multiple data formats, so it would seem that it overcame the challenge of managing the heterogeneity of data in human sciences (European Commission, 2015). However, we noticed that the descriptive categories used for the data types are more or less common in the six disciplines. Indeed, the most used item type were “journal contribution”, then “dataset” followed by “figures”, “presentations”, and “media”. Nevertheless, it is interesting to note that journal contribution which is at the top of the list for all the disciplines corresponds to any type of content officially published in a scientific journal, following a peer review process. This implies that Figshare is used to satisfy the requirements of funding agencies or publishers for opening research data. The majority of data published on Figshare was described by users as journal contribution, ignoring the other types even if they are more precise about the real data format.

5.2. Visibility analysis by discipline and by affiliation or not to an institution

More than 75% of articles was viewed less than 197 times and downloaded less than 45 times. This result means that the use of Figshare is far from being enough to insure the visibility and the reuse of data.

Even if some of the uploaded resources have generated much interest, the majority of research data in humanities was not viewed, and even less was downloaded. This calls into question the quality of the metadata. To be reusable, a data must first be findable. A recent study has shown that a large majority of researchers never or rarely use a metadata standards for describing research datasets (Chowdhury et al., 2018).

The program Figshare for Institutions ensures that their data management requirements are met transparently for the researcher. The system assists researchers in data management by providing conservation and metadata options for their research data. The researchers upload files and attach the appropriate metadata. The system can be configured to allow researchers to share files without intervention or route new downloads to an administrator for review before they are posted. Administrative control of file downloads can be done at the level of an institution or group, such as a department (Reed, 2016).

6. Conclusion and perspectives

This study allowed us to highlight that, even if Figshare is part of logic of openness, the design of each tool determines its use, and that the exploitation of data remains closely tied to the warehouse policy and the services it makes available. We encountered limitations with our attempts to determine the country of the researcher or institution responsible for depositing the dataset in terms of the design of Figshare research features and its API. The identification of the resource's country of origin was not possible.

Findings of this study support those of several abovementioned studies. In the arts and humanities, it seems that sharing data globally and reusing it is not yet a common practice among the researchers (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). The humanities community engagement on Figshare remains low. Even when research data was published, 50% is rarely viewed and less downloaded.

The results of our study provided some specific details about the affiliated articles: 70% of humanities research data on Figshare was not affiliated with any institution, but for the 30% of affiliated data, a higher average of visibility was recorded. This study has shown that affiliated articles were more viewed than unaffiliated ones published in the same period.

Thus, it seems that the establishment of the warehouses, is certainly a step in the right direction, but is far from being a sufficient mean to ensure that the potential benefits will be fully exploited. We believe that the supply of these warehouses must become a systematic part of the process of producing scientific knowledge. One way to overcome the challenges we discussed could be the development of an institutional policies that points the recommended warehouse, guarantees copyright, organizes the data, and controls the quality of the metadata. In this study, we acknowledged that data published under the control of a data professional has more chance to be found than data published individually by a researcher.

We think that there is a gap between the one who transfers the content (infrastructure builders) and the one who creates the value (researchers). Defining the institutional policy at the university level more precisely could be a good way of bridging the gap and guarantee the exploitation of this data. The management of research data requires many skills, and the researcher cannot be expected to have them all. Thus, it is important for the researcher to have tools and adequate services, which features are flexible enough to suit her/his needs – an entity able to offer services, support users and meet new needs as they arise, to improve the entire system continuously. Achieving the FAIR goals is not just a matter of a specialist or researcher but rather a shared representation of an ecosystem (Lehmans, 2017). Thus, it is crucial to build on and unleash the potential of collaboration between researchers and specialists, guaranteeing quality, mastering the sharing modalities and embodying the global vision of the ecosystem.

The researcher must be absolutely accompanied and so that she/he is not responsible for the management of the technical and legal problems related to the research data. Specialized entities must deal with these issues in a manner that is transparent to the researcher. We believe that information science professionals have a major role to play at this level. This discipline has investigated some of the most challenging issues related to the use of large-scale information resources, including the organization, access, management, and storage of research products in all their formats and encodings. They could support researchers in order to accomplish this task. Their participation can help increase the efficiency and

ensure the inclusion of default research data management in scientific production. This will ensure both the value of the data and the quality of the metadata.

To conclude, researchers do not get sufficient credit and other rewards for producing and sharing data. It is still the publication in a peer-reviewed journal that matters. Making data sets available should also be rewarded as an important scientific output. Journals adopting a data availability policy and data journals can be an important instrument to change this situation.

References

- Ali-Khan, S. E., Jean, A., MacDonald, E., Gold, E. R. (2018). Defining Success in Open Science. *MNI Open Research*[online], 2, 2, <https://doi.org/10.12688/mniopenres.12780.2>
- Borgman, C. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 6(63), 1059–1078.
- Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., Traweek, S. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. In: G. Buchanan (ed.). *IEEE/ACM Joint Conference on Digital Libraries*[online]. New York: IEEE Press Piscataway, <https://doi.org/10.1109/JCDL.2014.6970177>
- Bornmann, L. (2015). Usefulness of Altmetrics for Measuring the Broader Impact of Research: A Case Study Using Data from PLOS and F1000Prime. *Aslib Journal of Information Management*[online], 67(3), 305–319, <https://doi.org/10.1108/AJIM-09-2014-0115>
- Carbou, G. (2017). Analyser les textes à l'ère des humanités numériques. *Les Cahiers du numérique*, 13(3), 91–114.
- Chowdhury, G., Ünal, Y., Kurbanoglu, S., Boustany, J., Walton, G. (2018, October). *Research Data Management and Data Sharing Behaviour of University Researchers*. [online], [13.08.2019], <http://nrl.northumbria.ac.uk/34963/>
- European Commission. (2015). *Validating the Results of the Public Consultation on Science 2.0: Science in Transition* [online], [13.08.2019], <http://www.eesc.europa.eu/resources/docs/validation-of-the-results-of-the-public-consultation-on-science-20.pdf>
- Figshare (2018). *Figshare API User Documentation* (2018, 7 September) [online], <https://doi.org/10.6084/m9.figshare.4880372.v2>
- Girard, C. (2017). Les mécanismes de centralisation des données de la recherche. Étendre l'accès libre à l'hébergement libre. *Revue française des sciences de l'information et de la communication* [online], (11), <https://doi.org/10.4000/rfsic.3255>
- Gorman, G. E. (2007). The DeLorean or the Mini? *Online Information Review* [online], 6(31), <https://doi.org/10.1108/14684520710841739>
- HLEG (2010, October). High Level Expert Group on Scientific Data. Riding the Wave – How Europe Can Gain from the Rising Tide of Scientific Data [online], [20.05.2019], <http://www.eurosfair.eu/prd/fr/7pc/bibliotheque/consulter.php?id=2043>
- Jacobson, M., Larrousse, N., Massol, M. (2014, 7 July). *La question de l'archivage des données de la recherche en SHS (Sciences Humaines et Sociales)*[online]. Archives et données de la recherche (ICA/SUV 2014), [13.08.2019], <https://halshs.archives-ouvertes.fr/halshs-01025106/document>
- Konkiel, S., Scherer, D. (2013). New Opportunities for Repositories in the Age of Altmetrics. *Bulletin of the American Society for Information Science and Technology* [online], 39(4), 22–26, <https://doi.org/10.1002/bult.2013.1720390408>
- Koureas, D., Arvanitidis, C., Belbin, L. et al. (2016). Community Engagement: The “Last Mile” challenge for European Research e-Infrastructures. *Research Ideas and Outcomes*[online], 2, e9933, <https://doi.org/10.3897/rio.2.e9933>

- Lehmans, A. (2017). Données ouvertes et redéfinition de la culture de l'information dans les organisations. *Communication Organisation*, 51(1), 15–26.
- Neylon, C. (2017, 16 June). *As a researcher... I'm a bit bloody fed up with Data Management* [online]. Science in the Open [28.04.2019], <https://cameronneylon.net/blog/as-a-researcher-im-a-bit-bloody-fed-up-with-data-management/>
- Open Access (n. d.). *Open Access – H2020 Online Manual* [online], [20.07.2019], http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- Prost, H., Schöpfel, J. (2015). *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3* [online], [13.08.2019], <http://hal.univ-lille3.fr/hal-01198379/document>
- Reed, R. B. (2016). Figshare for Institutions. *Journal of the Medical Library Association* [online], 104(4), 376–378, <https://doi.org/10.3163/1536-5050.104.4.031>
- Research Infrastructures (n.d.). *Research Infrastructures, Including e-Infrastructures* [online]. Horizon 2020 sections [28.04.2019], <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/research-infrastructures-including-e-infrastructures>
- Ruiz, R. M. (2018, 28 February). *GDPR Tool for Handling Data-Subject Rights & Requests* [online]. Figshare, <https://doi.org/10.6084/m9.figshare.5853180.v2>
- Schöpfel, J. (2018). *Towards a Data Culture in Social Sciences and Humanities* [online], [13.08.2019], <https://hal.archives-ouvertes.fr/hal-01846849>
- Snider, L. (2015, 21 June). *Geography 101 – Periodos Ges – Hydrographical Iconography in Quire 13 of Beinecke MS408, Folios 75 through 84* [online]. Figshare, <https://doi.org/10.6084/m9.figshare.1309949.v4>
- Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., Astell, M. (2018, 21 March). *Whitepaper: Practical Challenges for Researchers in Data Sharing* [online]. Figshare, <https://doi.org/10.6084/m9.figshare.5975011.v1>
- Styles, S. (2017, 10 July). *The Commonest Speech Sounds: Prevalence Rates for Phonemes of the World* [online]. Figshare, <https://doi.org/10.6084/m9.figshare.5047747.v3>
- Swan, A., Brown, S. (2008). *To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs. A Report Commissioned by the Research Information Network* [online]. University of Southampton Institutional Repository [13.08.2019], <https://eprints.soton.ac.uk/266742/>
- Taylor, M., Plume. (2014). Party Papers or Policy Discussions: An Examination of Highly Shared Papers Using Altmetric Data. *Research Trends* [online], 36. [21.07.2019], <https://www.research-trends.com/issue-36-march-2014/highly-shared-papers/>
- Waltman, L., Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology* [online], 65(3), 433–445, <https://doi.org/10.1002/asi.23040>
- Weber, N. M. (2013). The Relevance of Research Data Sharing and Reuse Studies. *Bulletin of the American Society for Information Science and Technology* [online], 39(6), 23–26, <https://doi.org/10.1002/bult.2013.1720390609>

Eksploracja danych badawczych z zakresu humanistyki w serwisie Figshare

Abstrakt

Cel/Teza: Celem artykułu jest zbadanie obecności danych badawczych z zakresu humanistyki w serwisie Figshare oraz próba odpowiedzi na pytanie, czy samo opublikowanie danych badawczych przekłada się na ich wyszukiwalność.

Koncepcja/Metody badań: Projekt badawczy zakładał zidentyfikowanie zbioru cyfrowych artefaktów z obszaru humanistyki w serwisie Figshare za pośrednictwem jego API oraz analizę ilościową w oparciu o zmienne: liczba artykułów z podziałem na dyscyplinę, zmiany ilościowe w ujęciu chronologicznym, liczba wyświetleń i pobrań artykułów z uwzględnieniem informacji o afiliacji zbioru danych badawczych.

Wyniki i wnioski: Zaangażowanie społeczności humanistów w deponowanie danych badawczych w serwisie Figshare jest różne zależnie od dyscypliny. Nawet jeśli dane badawcze są tam publikowane, to dla 50% z nich odnotowano niską liczbę wyświetleń. Oznacza to, że sama otwartość danych badawczych nie jest wystarczająca dla ich widoczności w sieci i nie zapewnia ponownego użycia. Ostateczny wniosek z badań wskazuje na to, że informacja o afiliacji z daną instytucją badawczą dołączona do artykułów zwiększa szanse na ich widoczność w serwisie.

Ograniczenia badań: Jednym z ograniczeń badań jest fakt, że analizie zostały poddane tylko artykuły opublikowane w serwisie Figshare, co uniemożliwia podjęcie badań nad przyczynami małej obecności danych badawczych z obszaru humanistyki. Istnieje zatem potrzeba przeprowadzenia podobnych badań w odniesieniu do innych platform publikowania danych badawczych. Uzyskane rezultaty są zbliżone z wynikami podobnych badań nad praktykami zarządzania danymi badawczymi.

Oryginalność/Wartość poznawcza: Według wiedzy autorki, są to pierwsze badania nad obecnością danych badawczych z obszaru humanistyki w repozytorium danych badawczych Figshare.

Słowa kluczowe

Dane badawcze. Figshare. Nauki humanistyczne. Otwarte dane.

SAMIA TAKHTOUKH is a doctoral student at GERiiCO Laboratory, University of Lille (France). She holds a Computer Sciences engineer diploma from ENSIAS School in Morocco and a Research Master's degree in Library and Information Science from University of Lille in 2016. Her area of investigation is humanities research infrastructures.

Contact to the author:

takhtoukh.samia7@gmail.com

University of Lille

Laboratoire GERiiCO

B.P. 60149

59 653 Villeneuve d'Ascq Cedex, France