The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1468-4527.htm

Exploring the social activity of open research data on ResearchGate: implications for the data literacy of researchers

Received 11 May 2021
Revised 20 October 2021
Accepted 30 November 2021

Juliana Elisa Raffaghelli
Faculty of Psychology and Education, Universitat Oberta de Catalunya, Barcelona, Spain, and
Stefania Manca
Institute of Educational Technology, National Research Council, Genoa, Italy

Abstract

Purpose – Although current research has investigated how open research data (ORD) are published, researchers' behaviour of ORD sharing on academic social networks (ASNs) remains insufficiently explored. The purpose of this study is to investigate the connections between ORDs publication and social activity to uncover data literacy gaps.

Design/methodology/approach – This work investigates whether the ORDs publication leads to social activity around the ORDs and their linked published articles to uncover data literacy needs. The social activity was characterised as reads and citations, over the basis of a non-invasive approach supporting this preliminary study. The eventual associations between the social activity and the researchers' profile (scientific domain, gender, region, professional position, reputation) and the quality of the ORD published were investigated to complete this picture. A random sample of ORD items extracted from ResearchGate (752 ORDs) was analysed using quantitative techniques, including descriptive statistics, logistic regression and K-means cluster analysis.

Findings – The results highlight three main phenomena: (1) Globally, there is still an underdeveloped social activity around self-archived ORDs in ResearchGate, in terms of reads and citations, regardless of the published ORDs quality; (2) disentangling the moderating effects over social activity around ORD spots traditional dynamics within the "innovative" practice of engaging with data practices; (3) a somewhat similar situation of ResearchGate as ASN to other data platforms and repositories, in terms of social activity around ORD, was detected.

Research limitations/implications – Although the data were collected within a narrow period, the random data collection ensures a representative picture of researchers' practices.

Practical implications – As per the implications, the study sheds light on data literacy requirements to promote social activity around ORD in the context of open science as a desirable frontier of practice.

Originality/value – Researchers data literacy across digital systems is still little understood. Although there are many policies and technological infrastructure providing support, the researchers do not make an in-depth use of them.

Peer review – The peer-review history for this article is available at: https://publons.com/publon/10.1108/OIR-05-2021-0255.

Keywords Open research data, Academic social networks, Open data use, Open data quality, Researchers data literacy

Paper type Research paper

© Juliana Elisa Raffaghelli and Stefania Manca. *Published by Emerald Publishing Limited*. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode.

This research has been funded by the Project "Professional learning ecologies for Digital Scholarship: Steps for the Modernisation of Higher Education", Spanish Ministry of Economy and Competitiveness, Programme "Ramón y Cajal" RYC-2016-19589.

Online Information Review
Vol. 47 No. 1, 2023
pp. 197-217
Emerald Publishing Limited
1468-4527
DOI 10.1108/OIR-05-2021-0255

Introduction

The most enthusiastic discussions on the availability of data and the feasibility of appropriation by civil society and researchers immediately encountered other factors blocking advanced data practices, crowd science, quality in research and second-hand data usage for industry or research purposes (Molloy, 2011). Research on open data placed in open data repositories has considered several hypotheses in this regard. First, data cultures connected to disciplinary issues, research funding and value given to specific research practices address researchers' attention and practices (Borgman, 2015; Koltay, 2017). Second, any open data quality standard parameters embedded into the digital infrastructures to share data would determine usage and sharing (Berends *et al.*, 2020). In this regard, the FAIR movement has set an agenda pushing the application of quality standard (Association of European Research Libraries, 2017). Third, the methodological difficulties in capturing the social life of open research data (ORD), with some platforms providing more features to study sharing and reusing approaches than others (Quarati and Raffaghelli, 2020).

Moving beyond open data repositories to other types of digital environments promoting scholars' networking and professional learning, research on the researchers' professional practices on social media and the related digital skills must not be left out (Manca and Ranieri, 2017; Raffaghelli, 2017). The literature suggests that scholars have moved to social media from traditional repositories and publication in search of strengthening mutual relationships, facilitating peer collaboration, publishing and sharing research products and discussing research topics in open and public formats (Greenhow *et al.*, 2019; Hildebrandt and Couros, 2016). This is particularly true of increased activity by scholars on academic social networks (ASNs) and ResearchGate (Manca, 2018). Overall, new forms of scholarship aligning with open science ideals have been characterised as open, networked and social (Goodfellow, 2014; Veletsianos, 2013). However, their study moves forward through separate lines of research, between information literacy studies and professional learning in networked and online spaces (Raffaghelli *et al.*, 2016).

In fact, despite the plethora of studies on scholarly practices on social media, no specific research has been conducted to our knowledge on sharing open data usage on ResearchGate. So, it is not clear how researchers engage in such practices as part of their professional learning and identity. Moreover, a preliminary exploration of current practices on data could unravel the existing literacies and spot the skills gaps as a critical piece of open science.

Mind the gap: a way forward to uphold critical data literacy in the data practices of researchers

The tricky situation depicted in the previous section requires scholars to reflect upon data practices from a critical perspective. Emerging forms of research data literacy could be at the cutting edge, aiming at an integrated reflection and action taking in higher education, to provide the necessary support to faculty development (Raffaghelli, 2020; Usova and Laws, 2021). From its inception, the concept of literacy relates to a social activity, namely knowledge that is activated in specific contexts of life or work. This is particularly true when dealing with dynamic social environments like social media (Manca *et al.*, 2021). The visible practices, undertaken by specific groups, show the value given by those social and professional collectives. The hidden or inexistent practices signify both a technical inability and the lack of engagement with a broader view of what developing (a practice) means. In this sense, the open science discussions, including open data infrastructures comprehension, open data production and data sharing and reusage, act as a context of specific professional literacy. The need for knowledge and skills to operate in such contexts is not new. In 2013, Schneider (2013) considered generating a framework to address research data literacy. Some studies also referred to the need for support and coaching by the researchers to develop a more

sophisticated understanding of data platforms and practices, showing basic data usage without technical support from Libraries (Pouchard and Bracke, 2016). Wiorogórska *et al.* (2018) investigated data practices through a quantitative study in Poland led by the Information Literacy Association (InLitAs). The results revealed that a significant number of respondents knew some basic concepts related to research data management (RDM), but they had not used institutional solutions elaborated in their parent institutions. In another EU case study conducted in Slovenia, Vilar and Zabukovec (2019) studied the researchers' information behaviour in all research disciplines concerning selected demographic variables, through an online survey delivered to a random sample central registry of all active researchers. Age and discipline, and in a few cases, gender, were noticeable factors influencing the researchers' information behaviour, including data management, curation and publishing within digital environments. McKiernan *et al.* (2016) studied the literature through 2016 to show the many benefits of sharing data in Applied Sciences, Life Sciences, Maths, Physical Science and Social Sciences, where the advantages are related to the visibility of research relative to citations rates. As a result, the authors pointed out the need to support the researchers on paths to open data practices.

The literature has been concerned not only with detecting the skills gap, but professional development programmes, conducted primarily by university libraries, have taken an active part in developing data literacy amongst researchers. In determining data information literacy needs, Carlson *et al.* (2011) noticed that researchers need to integrate the disposition, management and curation of data and research activities. The authors conducted several interviews to analyse advanced students' performance in Geoinformatics activities within a data information literacy programme. Given the difficulties of finding useful training resources for researchers, Teal *et al.* (2015) developed an intensive two-day introductory workshop on "Data Carpentry", designed to teach basic concepts, skills and tools for working more effectively and reproducibly with data. Raffaghelli (2018) designed some workshops to discuss, reflect on and design open data activities in the specific field of online, networked learning. There is no documentation on whether said activities integrated research on ASNs. ASNs have been primarily considered a space for informal professional learning, which is frequently intuitive and misses the reference to formal, public infrastructures of digital knowledge available for the scholarly work. Researchers move on these platforms, particularly ResearchGate and Academia.edu, using the affordances provided and learning from each other (Kuo *et al.*, 2017; Manca, 2018; Thelwall and Kousha, 2015). However, the literature also portrays the preference for traditional research-related activities to improve reputation, due to incorrect behaviours, lack of quality of the resources shared and gaming within ASNs (Jamali *et al.*, 2016). Overall, it is necessary to understand the extent to which researchers adopt ASNs in appropriate ways, not as a primary space but with the social purpose of sharing and reusing ORDs. The lack of engagement or the erratic behaviour in these contexts would signal the need to develop data literacy as a complex understanding of the open science context, including the appropriate usage of digital infrastructures. Therefore, we purport here that data literacy refers not only to a technical ability but also to strategic, holistic knowledge and the ability to deal with a new context of professional practice, namely, open science. This preliminary picture is also necessary to promote Libraries and Faculty Development services as institutional strategies to promote professional engagement, learning and activism by researchers on digital platforms.

Methods

This preliminary and exploratory study aims to provide evidence on social practices related to open data in the broader professional endeavour of promoting open science. The study was designed to provide an answer to the following research question:

General RQ. What are the characteristics of the social activity related to self-archived open data on ResearchGate as a critical component of promoting open science?

This research question is broad in scope and was, therefore, articulated into three operational questions:

RQ1. What are the characteristics of the social activity related to self-archived open data on ResearchGate when compared to the social activity of the linked published research and in terms of quality?

RQ2. Is there any factor (including researcher profiles, publications social activity or the quality of the ORD) that predicts the social activity of ORD?

RQ3. Do the social practices related to ORD and linked publication show any patterns across specific research groups?

Data collection: instruments and procedures

ResearchGate affordances: ResearchGate is considered as one of the most prominent ASNs (Manca, 2018). Its main affordances encourage researcher visibility and social activity. These affordances include public researcher profiles and pages, access to the researchers' publication through specific links generated by ResearchGate and the possibility of linking supplementary material, such as images, tables or data.

As is characteristic of social and professional network sites, the resources cannot be browsed as in a database but are connected to the researcher's profile. Therefore, the resources selected through an algorithm connected to the researchers' profiling – or the other researchers' profile and reputation – are the "hook" to the curiosity and engagement of others with the information.

Metrics: ResearchGate collects and displays several direct metrics (frequencies, percentages) or metrics built upon layers of data. The metrics are also classified as public (accessed without registering on ResearchGate) or private (only registered users can see them). Public metrics include the number of publications, number of questions and answers, number of research projects opened by the researcher or in which the researcher is engaged, number of reads and number of citations. While the metrics are quite direct, they are always calculated based on activity within ResearchGate, namely: the number of publications uploaded by the researcher or detected by ResearchGate, reads counted as views of a publication summary (such as the title, abstract and list of authors), clicks on a figure or views and downloads of the full texts and citations of articles within the ResearchGate platform.

The second type of metrics includes the ResearchGate Score©, a composite metric showing the researcher's reputation calculated on all research elements, including publications, questions, answers and how other researchers interact with said content, particularly as followers, and through views and citations. ResearchGate metrics are aimed at stimulating social life on the platform. Not only complex metrics such as the RG score but also data views motivate the researcher; for example, to make comparisons with their evolution on the timeline and across the collective of researchers.

Data collection procedure: The use of metrics that are not public would involve requesting research scrapping (Barthel, 2015). Moreover, manual procedures (visiting the researchers' profiles one by one upon agreement) would encompass low feasibility of sampling a sufficient, random number of cases (unless the activity is undertaken via survey or crowdsourced research). As a result, an approach that ensures an initial economy of efforts lies with data-driven procedures through metadata extraction procedures.

Therefore, to obtain fundamental insights on the social activity of a considerable number of users, we analysed two public indicators: the number of online reads and the number of citations associated with each researcher's public profile.

The number of views (as basic interaction with a ResearchGate element) and citations (as an essential reusage parameter) was adopted for the two central ResearchGate elements: namely, the ORD (data item) and the linked publication. The sampling procedure, collecting and transforming data into the final variables, included an initial procedure of web scrapping [1], conducted between November and December 2018, based on a random list of 1,500 objects labelled as data (ORD). The list was applied to search for 1,500 ORDs (sampled on the basis of the links), using the software FMiner (http://www.fminer.com/). After the selection, the linked publications were also searched for automatically. The procedure was repeated to extract the main author's profile (metadata on the institutional affiliation, professional activity/level and RG score). A final data set was assembled with all the information scrapped. After data polishing (removing authors with insufficient metadata, repeated cases, unclear connections between the linked publication and the ORD), 399 items were removed. Finally, 752 cases were considered. At a 95% confidence level and 5% margin of error, the expected sample size is 385 cases. The sample in this study outperformed such values with an margin of error of ±3.57%.

Finally, a set of variables was created through manual analysis by visiting each researcher's profile to ascertain the information retrieved. Variables included gender, scientific domain, geographical region, professional position. The RG score was also rechecked. The metrics, definitions and procedures for data extraction and conversion are synthesised in Table 1. The authors collaborated with two research assistants to classify the ORDs and analyse agreement for the reliability of the creation of variables. On a list of 6% of randomly selected cases, the agreement level was absolute (100%) on gender (including missed or unclear values) and geographical region. However, the professional position required discussion on technical profiles and research practitioner aggregation, which ended up in a 74% agreement. Cohen's kappa coefficient was used to measure researcher agreement: Overall, coded values were 0.66 on the basis of 31 agreements on using a code, seven agreements on not using a code, five disagreements (two using three not using), representing 88% of agreement.

Another variable built was the FAIR quality assessment. One of the authors assessed each of the 752 ORDs, applying the simplified FAIR checklist (https://www.go-fair.org/fair-principles/). If the four FAIR dimensions were fulfilled (an RG data item was findable, accessible, interoperable and reuseable), a score of 4 was assigned. Conversely, one-, two- or three-point scores were assigned when one, two or three of the FAIR criteria were met. A score of 0 meant that none of the FAIR criteria were detected. In this case, the kappa coefficient was applied to 6% of the list above, obtaining a value of 0.30, with 68% agreement.

Data analysis

The analysis encompassed an exploratory approach to data to detect and represent underlying structures in the datasets to be interpreted based on the research questions.

RQ1, the descriptive statistics, including frequencies and percentages, central and dispersion robust measures, were reported to provide an initial synthetic representation that led to insights on the dimensions being studied. The descriptive statistics included univariate and bivariate tables with the overall social activity (reads and citations for ORDs and linked publications) and the social activity characterised by ORD quality and the researcher profile (gender, scientific domain, geographical region, professional position, reputation).

As for RQ2, a relevant issue from the descriptive statistics was the extremely negatively skewed distributions relating to the social activity around ORDs and publications (skewing

Main research Constructs	Variable/s definition	Metric and procedure
Social activity around ResearchGate research items (ORD and linked publications)		
Self-archived ORD	Number of items published on ResearchGate as "data"	• Number of items As published by the researchers at RG
Social activity of ORD and publication	Number of reads Number of citations	• Reads on ORD • Reads on linked publications • Citations on ORD • Citations on linked publications As extracted from RG
Quality of the self-archived ORD	Quality measured from 0 to 4	• Score FAIR Compliance with the four FAIR criteria scale 0 = No compliance 1 = 1 FAIR criterion covered 2 = 2 FAIR criteria covered 3 = 3 FAIR criteria covered 4 = 4 FAIR criteria covered Note: An ORD archived on RG would not be findable. We considered findable an ORD which publishes the link to the self-archived item into a specialised data platform or institutional repository. Only 1 object amongst the 752 satisfied this condition
Researcher profile		
Gender	Nominal classification	• Gender classification using the following values Female, male, not informed/available
Scientific domain	Nominal classification based on a division of scientific domains-8y\g (https://en.wikipedia.org/wiki/Branches_of_science)	• Scientific domain classification using the following values Applied Sciences (medicine and health sciences, engineering and technology), Formal Sciences (computer sciences, mathematics, statistics), Humanities (arts, history, visual arts, philosophy, law) Natural Sciences (biology, chemistry, physics, space sciences, Earth sciences), Social Sciences (educational sciences, psychology, political sciences, business, economics, anthropology, archaeology), not available
Geographical region	Nominal classification based on geographical regions as used at SCOPUS, namely, Tools (https://www.scimagojr.com/countrygraph.php)	• Geographical classification based on the following regions Africa, Asiatic Region Middle-East, Eastern Europe Western Europe, Russia Northern America, Latin America, Pacific Region, not available
Professional position	Nominal classification according to the experience and type of academic activity	• Classification relating the following groups of professional activity Student (post-graduate, PhD), Assistant (Technical, teaching, research); Mid-position, technical (Journalist, Librarian, Technologist, Researcher practitioner)Mid-position, academic (Lecturer, Researcher, Professor), Leader (Coordinator, Manager, director), Retired Scholar
Researcher's reputation	RG score	• Value (continuous scale) As calculated and extracted from the ASN platform

Table 1.
Main research constructs, variables associated, metrics and procedures

for publication citations = 11.33; publication reads = 17.61; ORD citations = 18.30; ORD reads = 25.91. Reference value = 0 for perfectly symmetrical distributions, −1 or +1 for highly skewed distributions). As a result, a non-parametrical correlation (Spearman's rank order correlation) was applied to explore initial relationships. Moreover, the relevant relationships were explored through binary logistic regression. The relevant response variables in this study (reads and citations for ORDs) were recoded as dummy variables (Y/N), taking into consideration a reference value set upon the = 0/>0 reads and citations. As in any regression analysis, the logit model aimed to model potential relationships between explanatory variables and response variables. The aim was to model the response of reading/citing or not reading/citing ORDs, according to the different researchers' characteristics, the quality of the ORDs and the social activity related to the linked publications (explanatory variables).

Finally, for RQ3, an unsupervised *k*-means cluster analysis was performed to observe whether the reads of ORDs and linked publications (as most basic but stable parameters of social activity) generated groups of cases. As expected, the clustering algorithm forms groups (clusters) of observations that should show similar patterns of relationship (in our case, between reading ORDs and reading linked publications). Moreover, to determine each cluster's relevance, the analysis of variance was adopted, computed per variable and its resultant variance table, including the model sum of squares and degrees of freedom as the variance statistics. The other categorical and numerical variables in the study (researcher profile and quality) were adopted to study their behaviour within the clusters on the clusters generated. Thus, the clusters yielded further information on over-usage trends, considering the researcher profiles and ORD quality.

Results

RQ1 – Overall social activity related to self-archived ORDs compared to linked published research and quality

Initially, we researched the distributions related to the social activity (reads and citations) of publications and ORDs. As reported above, these were extremely skewed, with cases deserving the attention of the research community (max number of publication reads = 10,423; max number of ORD reads = 3,438). Most cases of ORDs and linked publications were never read or cited. The medians around 0 (as a robust measure of central tendency) highlighted such phenomena. Table 2 illustrates the social activity related to ORDs and linked publications and the quartiles, mean and standard deviation showing the skewed distribution. A stable relationship between publication/ORD reads and publication/ORD citations can also be obtained.

Combined social activity with the researcher's profile (gender, scientific domain, region, professional position, reputation in terms of RG score dimensions) also showed interesting, specific phenomena within the overall situation (Table 3).

Categories	Min	Q1	PrMedian	All Cases $N = 752$ Q3	Max	Mean	STDV
Publications reads	0	0	0	25	10,423	60.33	465.83
Publication citations	0	0	0	0	106	0.96	6.18
Open data reads	1	2	4	10	3,438	15.46	127.26
Open data citations	0	0	0	0	6	0.02	0.27

Table 2. Social activity around open data and the linked publications

Categories	*N* Total	Publication reads	All Cases *N* = 752 Publication citations	Open data reads	Open data citations
Total all categories	752	45,370	725	11,629	13
Gender					
Female	143	5,192	310	1,439	1
Male	605	40,178	415	10,150	12
NA	4	0	0	40	0
Scientific domain					
Applied Sciences	250	27,984	268	5,638	1
Formal Sciences	77	3,345	39	1,019	12
Humanities	33	1,013	5	410	0
Natural Sciences	290	10,411	329	3,593	0
Social Sciences	102	2,617	84	969	0
Region					
Africa	30	1,672	0	204	0
Asiatic Region	136	4,260	224	4,819	0
Middle-East	39	799	52	503	0
Eastern Europe	49	11,442	20	644	0
Western Europe	293	21,646	259	2,615	7
Russia	16	359	17	417	0
Northern America	107	3,666	126	1,666	6
Latin America	55	1,273	14	514	0
Pacific Region	23	196	4	228	0
NA	4	57	9	19	0
Position					
Student (undergraduate/PhD)	77	1,431	0	675	0
Assistant (technical, teaching, research)	92	7,290	88	4,461	0
Mid-position, technical (Journalist, Librarian, Technologist, Researcher practitioner)	65	1,065	22	319	0
Mid-position, academic (Lecturer, Researcher, Professor)	357	24,907	218	4,254	13
Leader (Coordinator, Manager, director)	73	1,686	75	937	0
Retired Scholar	3	0	48	3	0
Reputation (RG score)					
1–10	239	11,538	273	1966	0
11–20	204	6,137	143	5,829	5
21–30	144	9,150	188	1,293	0
31- . . .	134	16,861	117	2,272	8
NA	31	1,684	4	269	0

Table 3.
Social activity around open data and the linked publications by researchers' identity (gender, scientific domain, region, position, reputation)

First, female researchers were underrepresented in the sample, having the fewest reads ($F = 5{,}192$ publication reads and 1,439 ORD reads compared to $M = 40{,}178$ publication reads and 10,150 ORD reads). Notably, the ORD reads were almost nine times higher for males. The relationship between publication/ORD reads follows the overall pattern detected above. However, regarding citations, the male/female researchers related to publications come closer ($F = 415$/M = 310); this situation is not repeated for the ORD citations ($F = 1$/M = 12).

Regarding Scientific Domain, Applied and Natural Sciences outperform the Formal Sciences, Humanities and Social Sciences. These areas of science established the type of scientific communication based on short articles and citations earlier. The Humanities and Social Sciences have evolved through different forms of research communication (Borgman, 2015). The relationship between publication and ORD reads and citations aligns with the overall situation: Specifically, the more ORDs published in a Scientific Field, the more reads and citations received. Interestingly, most ORD citations come from the field of Formal Sciences (12 out of 13), which include Maths, Statistics and Computer Science. Considering the Open Source movement, we can assume that sharing scripts in programming activity is a more common practice, requiring collaborative literacies, than in other fields (Dabbish *et al.*, 2012).

Regarding the Geographical Regions to which the researchers' institutions belong, most self-archived ORDs fell into the categories of Western Europe (293), North America (107) and the Asian region (136). The relationship between published ORDs and the social activity related to the same ORD and the linked publications is again stable: The more the ORD gets published, the more reads and citations occur, with Western EU and North America showing the highest levels of attention in a typical centre–periphery relationship with knowledge. However, it is interesting to notice some specific cases that could be shaping different cultures of collaboration regarding data. In the Middle East, with a low number of ORD publications (39 out of 752 cases), the ORD is read even as much as the linked publications (Pub Reads = 799; ORD reads = 503) even though there are no citations for the ORD. ORD reads are even higher in the Pacific Region than the linked publication reads (228 compared to 196). As in Western Europe, in Eastern Europe, there is a high concentration of reads on some publications (21,646 reads in the first case and 11,442 in the second case out of 45,313 reads overall). Yet, in the second case, the social activity connected to publication citations (20 for 49 publications) and ORD reads (644 for 49 ORDs published and out of 11,629 reads to all the ORD published) is lower. Moreover, ORD citations are null (0 citations).

The professional position variable shows higher productivity in terms of self-archived ORDs for the academic mid-positions (researchers, lecturers and professors who should have achieved a seniority level) with 357 out of 752 records. They are followed at a distance by the assistant positions (both in research and teaching). The situation is consistent for social activity: Academics in their mid-career positions get more reads on their publications (24,907 out of 45,370) and on their ORDs (4,254 out of 11,629). This is also the case for citations (218 pub citations and 13 ORD citations for 357 publications). More importantly, all ORD citations computed belong to mid-positioned academics. Remarkably, assistants received almost an equal number of ORD reads (4,461 against the 4,254 of mid-position academics) for a relevant fewer number of self-archived ORDs (92 of assistants compared to 357 of mid-position academics).

Finalising the data analysis reported in Table 2, the researchers' reputation in RG score was considered. We discovered that most self-archived ORDs are related to researchers with relatively low reputation (1–10 RG score = 239; 11–20 = 204 out of 752). Nonetheless, when analysing social activity related to ORDs, we observe a slight change in the trend. The highest number of reads for linked publications occurs for the researchers with the highest reputation (31- . . . = 16,861 of 45,730). Also, a relevant number of ORD reads are included in this category (2,272 out of 11,629). However, many linked publication reads are consistent with the lowest reputation (1–10 = 11,538). The highest ORD reads are related to scholars with a relatively low reputation (11–20). The linked published citations and the researchers with a low reputation attract a higher number (273 out of 725 overall citations), followed by scholars with a mid-reputation (RG score 21–30 = 188 of 725). Most ORDs get citations regardless of the reputation (high or low) of the researchers who published them. Even if the ORD citations are negligible, it can be assumed that the researchers focus on the research they know, for

specific purposes (reading and citing despite the reputation). However, the reads and citations on ORDs produced by scholars with higher reputation show that this parameter attracts other scholars' attention.

Moving to the social activity related to ORD quality, results are displayed in Table 3. The social activity in terms of publication citations and open data reads also showed little researcher attention to ORD quality. An overwhelming number of self-archived ORDs were not compliant with the FAIR criteria (562 out of 752) followed by elements compliant with only one criterion (126 of 752). At the same time, only one ORD reaches the top level of quality with four FAIR criteria. Moreover, most citations of linked publications (642 out of 725) and ORD reads (10,201 out of 11,629) were directed to low FAIR scores. However, another unusual pattern is shown in Table 3. A handful of 12 articles linked to published ORDs, compliant with three FAIR criteria, concentrate a very high number of reads (10,750 out of 45,370) and receive 4 of the 13 ORD citations (see Table 4).

In conclusion, Figure 1 represents the relationships between (a) publication reads and ORD reads and citations; (b) ORD reads and ORD citations; (c) these two relationships compared to a third variable, namely, ORD quality. The results are not particularly encouraging and confirm the intuitions emerging from the tables: We observe that highly skewed distributions were most ORDs published, and their linked publications are underseen and underused; and a high concentration of social activity related to specific records. The quality of the ORD is also irrelevant to address the researchers' behaviours.

RQ2 – Factors that predict the social activity of ORDs

The binary logistic regression on the response variable ORD reads and ORD citations did not yield significant models, rather interesting insights. Non-linear relationship, independence of errors and multicollinearity were met as assumptions that support the logistic regression. Therefore, given our study's exploratory nature, we adopted the forced entry method that considered the explanatory variables theoretically identified and showed *t*-values near significance levels.

For the ORD reads, these variables were: quality, RG score and linked publication read. In the latter case, the explanatory variables were transformed into categorical predictors. Annex, published as open data (Raffaghelli and Manca, 2020), respectively, summarise the logit analysis for predicting ORD reads and the predictors for ORD citations. In the case of ORD reads, the high AIC value (−42,348), the non-significant chi-square coefficient (χ 2 = 1, df = 688, $p > 0$) and the negative very high pseudo *R*-squared index (McFadden −2.211247e.03) show a poor model fit and somewhat random behaviour related to ORDs. In any case, the "Quality 3" level and the highest number of reads on the linked publications yield a significant *t*-

Table 4.
Social activity around open data and the linked publications by the quality of open data

Quality of the open data share on RG Compliance with the 4 FAIR criteria scale	*N* Total	Publication reads	All Cases *N* = 752 Publication citations	Open data reads	Open data citations
Total	752	45,370	725	11,629	13
0 = No compliance	562	25,686	642	10,201	9
1 = 1 FAIR criterion covered	126	2,958	63	975	0
2 = 2 FAIR criterion covered	28	412	6	190	0
3 = 3 FAIR criterion covered	12	10,750	5	152	4
4 = ALL FAIR criteria covered	1	43	0	15	0
NA	23	5,521	9	96	0

Figure 1.
Social activity (open data and linked publication reads and citations) combined with open data quality

value regarding the ORD reads, which could point to an association. The more published research linked to an ORD is read, the more the ORD attracts the attention of researchers.

In the ORD citations, the model included the scientific domain, the ORD quality and the ORD reads, converted into categorical variables at two levels (no read = 0 reads, read < 0). While the model got better fit values (AIC -1793), the non-significant chi-square coefficients (χ 2 = 0.99, df = 719, $p > 0$) and the low *R*-squared (McFadden −0.04; Hosmer and Lemeshow 0.11) also showed a poor model fit. However, some interesting relationships appeared. The scientific domain of formal sciences ($p > 0.001$), achievement of at least three FAIR quality criteria ($p > 0.001$) and the presence of ODR reads ($p > 0$) could be associated with ORD citations at significant levels. It can be concluded that while it is not possible to find a model for predicting when ORD citations will occur, there are some specific scientific domains that are moving their social practice towards acknowledging and citing the data of others. As in the case of ODR reads, ODR quality also led to citations.

RQ3 – Social practices related to ORDs and patterns of linked publication across specific groups of researchers

This question was explored through cluster analysis that grouped data points according to ORD reads and linked publication reads, as the most stable parameters of social activity, which proved to be associated. The Silhouette method established three cluster as the optimal number (Figure 2). Figure 3 shows the distribution of the three clusters, where, despite the skewed distribution, it is possible to see that the three groups display diversified patterns. Namely, Cluster 1 relates to self-archived ORDs with linked publications that tend to be read more; Cluster 3 is made up of self-archived ORDs that tend to be read more, with some cases of highly read linked publications; and Cluster 2 shows self-archived ORDs with negligible levels of social activity in terms of reads, both of the ORD and the linked publications. The statistics computed within the cluster (sum of squares = C1 = 119.32; C2 = 163.41; C3 = 104.61) showed a similar distance between cluster centroids. The model explained 72% of the variance (between_Squared distances SS/total_SS parameter).

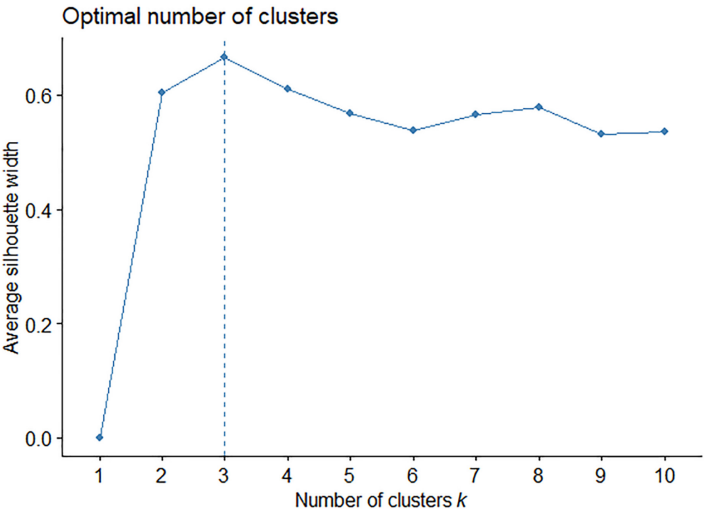

Figure 2.
Silhouette method to determine the number of clusters (three outliers were eliminated)

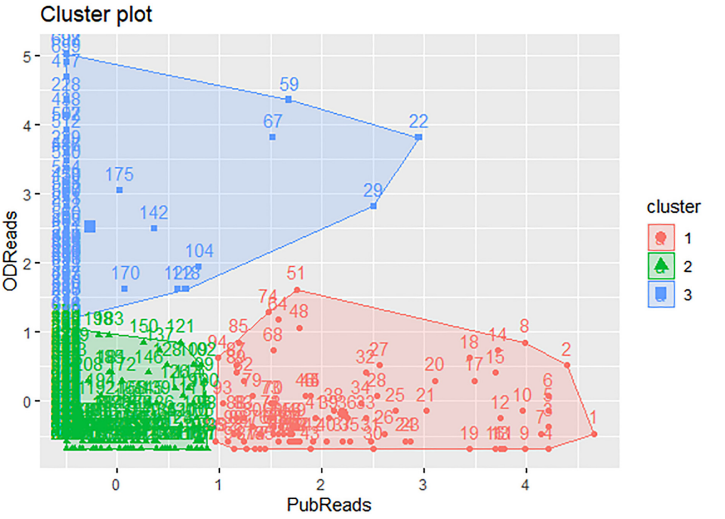

Figure 3.
Cluster plot: Three clusters detected considering the ORD reads (ODReads) and the linked publications reads (PubReads)

Table 5 shows the distribution of self-archived ORDs relating to the profiles of the researchers (gender, scientific domain, region, professional position and reputation) and the quality of the published ORDs per cluster. We can observe that Cluster 2 (negligible social activity related to ORDs and linked publications) is made up of most cases. The profiles of the researchers in this cluster are consistent with the overall situation: More males, coming from Applied and Natural Sciences, mostly located in Western EU, North America and the Asian region and overwhelmingly mid-career academics. However, when analysing reputation and the quality of the published research, we find most self-archived ORDs in Cluster 2 (negligible social activity) published by scholars with very low RG scores. These scholars tend to publish mostly low-quality ORDs (0/1 FAIR criterion covered). Within the second-largest cluster (1), which shows some social activity for linked publications related to the self-archived ORDs,

Table 5. Categories distribution per cluster

Categories	Cluster 1	% of total counts w category (%)	Cluster 2	% of total counts w category (%)	Cluster 3	% of total counts w category (%)	Total	Aggregated % of counts within category (%)
Gender								
Female	15	2.16	101	14.53	18	2.59	134	19.28
Male	77	11.08	432	62.16	52	7.48	561	80.72
Cluster weight over total	92	13.24	533	76.69	70	10.07	695*	100
Scientific domain								
Applied Sciences	22	3.15	186	26.61	25	3.58	233	33.33
Formal Sciences	8	1.14	56	8.01	8	1.14	72	10.30
Humanities	2	0.29	24	3.43	4	0.57	30	4.29
Natural Sciences	46	6.58	195	27.90	26	3.72	267	38.20
Social Sciences	14	2.00	76	10.87	7	1.00	97	13.88
Cluster weight over total	92	13.16	537	76.82	70	10.01	699	100
Region								
Africa	5	0.72	21	3.02	3	0.43	29	4.17
Asiatic Region	23	3.31	90	12.95	15	2.16	128	18.42
Middle-East	3	0.43	26	3.74	7	1.01	36	5.18
Eastern EU	6	0.86	35	5.04	4	0.58	45	6.47
Western EU	38	5.47	213	30.65	22	3.17	273	39.28
Russia	1	0.14	11	1.58	2	0.29	14	2.01
Northern America	8	1.15	77	11.08	11	1.58	96	13.81
Latin America	6	0.86	40	5.76	6	0.86	52	7.48
Pacific Region	1	0.14	21	3.02	0	0.00	22	3.17
Cluster weight over total	91	13.09	534	76.83	70	10.07	695	100

(*continued*)

Table 5.

Categories	Cluster 1	% of total counts w category (%)	Cluster 2	% of total counts w category (%)	Cluster 3	% of total counts w category (%)	Total	Aggregated % of counts within category (%)
Position								
Student (undergraduate/PhD)	10	1.75	49	8.60	4	0.70	63	11.05
Assistant (Technical, teaching, research)	10	1.75	64	11.23	10	1.75	84	14.74
Mid-position, technical (Journalist, Librarian, Technologist, Researcher practitioner)	4	0.70	23	4.04	5	0.88	32	5.61
Mid-position, academic (Lecturer, Researcher, Professor)	48	8.42	245	42.98	35	6.14	328	57.54
Leader (Coordinator, Manager, director)	6	1.05	48	8.42	6	1.05	60	10.53
Retired Scholar	6	1.05	31	5.44	2	0.35	39	6.84
Cluster weight over total	84	14.72	460	80.71	62	10.87	606	100
RG score								
1–10	30	4.43	179	26.44	19	2.81	228	33.68
11–20	25	3.69	145	21.42	18	2.66	188	27.77
21–30	20	2.95	107	15.81	9	1.33	136	20.09
31- . . .	21	3.10	83	12.26	21	3.10	125	18.46
Cluster weight over total	96	14.18	514	75.92	67	9.90	677	100
Quality								
0	73	10.78	393	58.05	53	7.83	519	76.66
1	10	1.48	97	14.33	13	1.92	120	17.73
2	3	0.44	23	3.40	1	0.15	27	3.99
3	2	0.30	7	1.03	1	0.15	10	1.48
4	0	0.00	1	0.15	0	0.00	1	0.15
Cluster weight over total	88	13.00	521	76.96	68	10.04	677	100

Note(s): *The total of cases clustered might vary according to the missed values for each category

the situation also aligns with the global distribution and Cluster 1. But it is also worth noticing that in this cluster, the weight of the Asian region and Western Europe is higher when compared to North America. The RG score is more balanced, with cases with a higher RG score (6.05% of 14.18% of C2 weight for the overall sample). Finally, for Cluster 3 (some social activity related to the ORD), the trends are similar to those described above: more males, scientific domains of Applied Sciences and Natural Sciences, same regional areas (Western EU, the Asian Region and North America), more presence of mid-position academics and low quality of the ORDs published. In this cluster, some interesting trends are related to a better representation of the Middle East (2.16% out of 10.07% compared to 3.17% of the Western EU). It was also more relevant social activity found related to self-archived ORDs published by scholars with a higher RG score (the two levels of 21–30 and 31 or higher RG score contribute with 4.44% out of 9.9% of the overall cluster).

Discussion

In response to the three RQs, we observed that there is still undeveloped social activity related to self-archived ORDs in ResearchGate, in terms of reads, citations and the quality of the published ORDs as their influence to engage with them. We also found that the relevance of the moderating effects on ORDs underpins the assumption of traditional dynamics that is still stuck on "new" data practices. Finally, ResearchGate exhibited a similar situation to other data platforms and repositories.

Our study portrays group characteristics that may support situated values and cultures, preventing or hindering quality data sharing or reuse. Most ORDs were published primarily by males from Western EU, North American and Asian institutions, with a position of academic seniority. They attracted specific citations on the ORDs (as recognition of their work of publishing ORDs) from fields that are already recognised as being dominated by males (Mathematics, Statistics and, particularly, Computer Science). Notwithstanding, female colleagues were underrepresented, but their publications linked to ORDs were cited similarly to male colleagues. Moreover, though the research assistants showed fewer self-archived ORDs than mid-career researchers, the former attracted an equal number compared to the latter. However, the research assistants' self-archived ORDs get much fewer citations. In this regard, even if the ORD citations were negligible, it is related to specific research fields, namely male researchers from Western EU and North America, in their mid-career and of higher reputation. Undoubtedly, centripetal forces attract attention to specific research, which might be linked to several factors.

Our assumptions would be further supported by the sociocritical lens of Bates (2018) who highlights the complex nature of the behaviours of researchers (and other stakeholders) in circulating the ORDs, which entail voluntary or involuntary "data friction". According to Bates, data friction is an emergent effect of the many cultures, jargons and procedures adopted by the researchers in the different disciplines or groups. The friction effect impedes outsider researchers from understanding what the colleagues do with data, making such data unusable. Post-phenomenological approaches might also support data friction in engaging with research objects and representations: It is not the object *per se* that communicates its possible affordances, but the relationship between the prior and present researcher's experience to enable her to engage with it (Trifonas, 2009). Moreover, we might consider Robert Merton's foundational work on science's normative structure (Merton, 1973). There are hidden rules connected to the research cultures across fields that permeate and guide researchers' attention, decision over topics and methodologies. These cultural factors could influence researchers when focussing their attention on the most influential researchers in their fields. The example of the concentration of publication reads in Western and Eastern Europe, with low ORD reads and citations, demonstrates patterns that follow tradition and hierarchies. Such information is confirmed by the prevalence of ORDs published by mid-

career academics compared to all other positions. Contradictory information comes from RG scores, with most published objects from researchers with low RG scores. But the ORDs are consistent with the tradition and/or expertise that get attention (highest RG score in most read and cited ORD). Even if attention also goes to research published by low-reputation scholars, these research items are cited less often.

Another factor is related to expertise and knowledge levels, supported by professional networked learning (Pataraia *et al.*, 2013). Consulting specific resources and recognising where the expertise can be found (as in the case of most ORDs published, read and cited by mid-career academics) is consistent with the idea of intuition and self-determination in searching for the relevant knowledge in one's own field. This is not contradictory to Merton's theory; those that hold power in institutional or professional groups/communities are those whose knowledge is most relevant.

As for the latter factor, there might be values deemed applicable across research disciplines, as Lee *et al.* (2019) expressed. These authors created a model based on 18 factors, amongst which the most important was accessibility, followed by altruism, reciprocity, trust, self-efficacy, reputation and publicity. Nonetheless, data cultures across disciplines are based on the methodological assumption and the research topics' ontological approaches. Borgman's (2015) in-depth qualitative analysis for data practices across disciplines supports this hypothesis, and our work sheds light on the quantitative differences, though not on the motivations. One could consider whether this situation should change: Should all research fields behave similarly and be prolific in opening data? Should Humanities and Social Sciences work on more open data patterns? To a certain extent, as Borgman pointed out, a researcher dealing with unique cultural heritage and/or a social scientist handling sensitive issues would be slower in producing and sharing their data.

In this regard, the values and the ideology of digital and open science could be embraced differently. As Lämmerhirt (2016) and Wouters and Haak (2017) pointed out, data sharing is strongly encouraged by policymakers in some disciplines, such as Physics and Genomics. Still, this concept is far less developed in other fields of research. The recent existence of a field of research could also be considered, as Raffaghelli and Manca (2020) documented for the case of Educational Technology.

Even if our work did not compare the variables explored here within the context of ResearchGate with other contexts of ORD self-archiving like Zenodo, Figshare or institutional repositories, the literature related to these latter cases addresses a similar situation (Quarati and Raffaghelli, 2020). Lack of ORD attention, sharing and reuse is common phenomenon, even if there is an increasing publication trend. Therefore, an infrastructure whose affordances are prepared to support social activity does not encompass specific changes to the researchers' practices, professional cultures and contextual literacies. This aligns with the idea that the technical structure for opening data is embedded in complex sociotechnical ecosystems (Manca, 2018). A clear example is that of ORD quality observed in this study. Most ORDs did not achieve even one FAIR parameter. Within RG, the situation related to the quality of self-archived ORDs and related metadata could be worse than in specialised data repositories due to the lack of specific affordances addressing the appropriate presentation and findability of data sets. However, the very few ORDs of good quality published on ResearchGate deserved attention; thus, when the professional communities engage in specific social behaviour patterns, the participants will adopt the technologies accordingly, rather than the opposite.

The specific situation of groups showing advanced data practices also requires attention, in a changing (and somewhat pressing) situation for digital scholarship to "show up" within social media and ASNs. Weller (2011) highlights the system's pushing effect to adopt digital means supports the idea that many researchers feel obliged to embrace the open practice and "go wild" on some platforms, with somewhat performative practices. The lack of quality of

the ORDs published would go in that direction. Moreover, the openness has been emphasised disregarding the relevance of networking and being digital, another fact supported by the low social activity (no networking) and low quality (low digital abilities to treat the self-archived digital items appropriately). As a result, publishing open data might be a mere performative act. We found that the ORDs published underpin new approaches to data. However, the lack of quality could be an eloquent expression of "no concern/no time to devote" about the life of such objects after being published (open dimension despite the networked and digital). Nonetheless, it could also be the result of behaving as *double gamers* (Costa, 2016). According to Costa, researchers struggle between pursuing the highest values of transparency and public knowledge embedded in the action of publishing open data and the lack of recognition for such an endeavour in most traditional contexts of doing science, where only the final publication supports career advancement.

All in all, there is an emerging scene that hinders scholars' reflection on data practices through more holistic and critical perspectives entailing better quality and reuse. The traditional profiles and social activity (reads and citations) related to the publications underpin the assumption that there is attrition between new professional practices (sharing data in a context of open science) and the consolidated mechanisms of reputation and career advancement. The need for critical research data literacy is part of a culture of pushing for innovations and getting a broader picture of what open science might bring to society in the future, not the present (stuck in the past/tradition). Therefore, better open data quality, replication and second-hand data reusage, as innovative practices of open research, require technical knowledge and require engagement with the policy context, and with strategies to advance the quality and ethics of being an open, networked and social researcher. As an example, we could consider the FAIR data principles. Knowing them helps the researcher publish better quality open data and understand the differences between making data circulate between ASNs or institutional repositories. But knowing the context of generating the FAIR principles might imply attention to familiar patterns and languages, as social knowledge connected to critical data literacy.

Conclusions

This study explored the social activity of researchers related to ORDs, in an attempt to spot areas of conflict relating to making a professional identity as digital scholars in the era of open science. As our study showed, there is still a long way to go for the effective adoption of ASNs to share and reuse ORDs. We considered several hypotheses relating to this phenomenon beyond digital infrastructures and the quality of the digital objects published. In this regard, focussing on practices and the culture supporting them leads to a discussion on sociocultural transformation and development. This element addresses training as a key dimension of institutional cultures, professional practice and critical literacies. In terms of future research, promoting such a critical approach to data practices should be considered. Here, formal training would not be the way to balance a situation where the motivations to publish, read, cite and potentially reuse ORDs could correspond to the researchers' struggle within conflictive institutional and data cultures. To strike a balance between the initial formal learning activities related to institutional repositories and infrastructures for open science and the informal learning occurring in the context of ASNs, engagement and reflective practice within professional learning communities could be explored as a possible way. However, it should also be considered that professional learning requires complex, self-directed pathways including all sorts of engagement with resources, activities and networks to fulfil personal developmental goals into what could be considered an ecology of learning (Sangrá *et al.*, 2019). Once again, the institutional agendas might pressure researchers to focus on specific forms of literacy. As a result, researchers might resist disregarding activities when not rewarded to activism and civil disobedience. It goes without saying that while formal

training imposes an explicit institutional agenda that outlines the types of desired literacies, a critical approach to self-determined data literacy in research might be connected to more informal spaces, particularly activism.

While we found that the factors influencing data practices are relevant, our research could not reach a clear relationship for all the sampled researchers and specific groups. Future research should explore the researchers' (open) data cultures as social contexts of data literacy development, including elaborating open data and publication motivations either in institutional repositories or ASNs. This could be done both by qualitative observational approaches and design-based research on professional learning. In scenarios where the researchers' skills gap predominates, the impact of creating spaces of reflection and informal or non-formal learning amongst researchers could be researched as a source of grounding communication over data to move beyond the sole expression of interest on ORDs. However, in the less optimistic scenarios, where scientific communities' social structures exert power and impose an academic and data culture, such an approach could fail. Other professional learning settings should be explored in tandem with the evolution of policymaking and institutional instruments supporting professional practices.

In our study's practices, we noticed separate worlds between practice and the open science agenda. We purported that the social and cultural implications of being a scholar in the digital age require further understanding of researchers' professional practices regarding social media and their digital skills. We dealt with social activity related to data, which is entangled with many motivations and "know-how" as drivers of informal learning. We purport that the depicted situation points to the need to actively explore the micro-levels of stakeholders' engagement and a more holistic approach to professional learning, to move the agenda of open science and open data forward.

Note

1. This process was undertaken by an external researcher from the company Winged Mercury (http://www.wingedmercury.net/)

References

Association of European Research Libraries (2017), *Implementing FAIR Data Principles: the Role of Libraries*, LIBER, pp. 1-2, doi: 10.1038/sdata.2016.18.

Barthel, M. (2015), *The Challenges of Using Facebook for Research*, Pew Research Center, n.p. available at: https://www.pewresearch.org/fact-tank/2015/03/26/the-challenges-of-using-facebook-for-research/.

Bates, J. (2018), "The politics of data friction", *Journal of Documentation*, Vol. 74 No. 2, pp. 412-429, doi: 10.1108/JD-05-2017-0080.

Berends, J., Carrara, W., Engbers, W. and Vollers, H. (2020), *Reusing Open Data: a Study on Companies Transforming Open Data into Economic and Societal Value*, Publications Office, doi: 10.2830/876679.

Borgman, C.L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, MA.

Carlson, J., Fosmire, M., Miller, C.C. and Nelson, M.S. (2011), "Determining data information literacy needs: a study of students and research faculty", *Portal: Libraries and the Academy*, Vol. 11 No. 2, pp. 629-657, doi: 10.1353/pla.2011.0022.

Costa, C. (2016), "Double gamers: academics between fields", *British Journal of Sociology of Education*, Vol. 37 No. 7, pp. 993-1013, doi: 10.1080/01425692.2014.982861.

Dabbish, L., Stuart, C., Tsay, J. and Herbsleb, J. (2012), "Social coding in GitHub", *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW'12*, ACM Press, New York, USA, p. 1277, doi: 10.1145/2145204.2145396.

Goodfellow, R. (2014), "Scholarly, digital, open: an impossible triangle?", *Research in Learning Technology*, Vol. 21, doi: 10.3402/rlt.v21.21366.

Greenhow, C., Gleason, B. and Staudt Willet, K.B. (2019), "Social scholarship revisited: changing scholarly practices in the age of social media", *British Journal of Educational Technology*, Vol. 50 No. 3, pp. 987-1004, doi: 10.1111/bjet.12772.

Hildebrandt, K. and Couros, A. (2016), "Digital selves, digital scholars: theorising academic identity in online spaces", *Journal of Applied Social Theory*, Vol. 1 No. 1, available at: https://socialtheoryapplied.com/journal/jast/article/view/16.

Jamali, H.R., Nicholas, D. and Herman, E. (2016), "Scholarly reputation in the digital age and the role of emerging platforms and mechanisms", *Research Evaluation*, Vol. 25 No. 1, pp. 37-49, doi: 10.1093/reseval/rvv032.

Koltay, T. (2017), "Data literacy for researchers and data librarians", *Journal of Librarianship and Information Science*, Vol. 49 No. 1, pp. 3-14, doi: 10.1177/0961000615616450.

Kuo, T., Tsai, G.Y., Jim Wu, Y.-C. and Alhalabi, W. (2017), "From sociability to creditability for academics", *Computers in Human Behavior*, Vol. 75, pp. 975-984, doi: 10.1016/J.CHB.2016.07.044.

Lämmerhirt, D. (2016), *Briefing Paper : Disciplinary Differences in Opening Research Data*, Pasteur4Oa, June 2013, pp. 1-8, available at: http://pasteur4oa.eu/sites/pasteur4oa/files/resource/Brief_Disciplinary differences in opening research data APS_MP_FINAL1.pdf.

Lee, J., Oh, S., Dong, H., Wang, F. and Burnett, G. (2019), "Motivations for self-archiving on an academic social networking site: a study on researchgate", *Journal of the Association for Information Science and Technology*, Vol. 70 No. 6, pp. 563-574, doi: 10.1002/asi.24138.

Manca, S. (2018), "Researchgate and academia.edu as networked socio-technical systems for scholarly communication: a literature review", *Research in Learning Technology*, Vol. 26, pp. 1-16, doi: 10.25304/rlt.v26.2008.

Manca, S. and Ranieri, M. (2017), "Exploring digital scholarship. A study on use of social media for scholarly communication among Italian academics", in Esposito, A. (Ed.), *Research 2.0 and the Impact of Digital Technologies on Scholarly Inquiry*, IGI Global, Hershey, PA, pp. 117-142, doi: 10.4018/978-1-5225-0830-4.ch007.

Manca, S., Bocconi, S. and Gleason, B. (2021), "'Think globally, act locally': a glocal approach to the development of social media literacy", *Computers and Education*, Vol. 160, p. 104025, doi: 10.1016/j.compedu.2020.104025.

McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B.A., Ram, K., Soderberg, C.K., Spies, J.R., Thaney, K., Updegrove, A., Woo, K.H. and Yarkoni, T. (2016), *How Open Science Helps Researchers Succeed*, ELife, Cambridge, doi: 10.7554/eLife.16800.

Merton, R.K. (1973), "The normative structure of science", in Merton, R.K. (Ed.), *The Sociology of Science: Theoretical and Empirical Investigations*, University Chicago Press, Chicago, IL.

Molloy, J.C. (2011), "The open knowledge foundation: open data means better science", *PLoS Biology*, Vol. 9 No. 12, doi: 10.1371/journal.pbio.1001195.

Pataraia, N., Margaryan, A., Falconer, I. and Littlejohn, A. (2013), "How and what do academics learn through their personal networks", *Journal of Further and Higher Education*, Vol. 39 No. 3, pp. 336-357, doi: 10.1080/0309877X.2013.831041.

Pouchard, L. and Bracke, M.S. (2016), "An analysis of selected data practices: a case study of the Purdue College of agriculture", *Issues in Science and Technology Librarianship*, Vol. 2016 No. 85, doi: 10.5062/F4057CX4.

Quarati, A. and Raffaghelli, J.E. (2020), "Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case", *Journal of Information Science*, First published on line Oct 4, 2020. doi: 10.1177/0165551520961048.

Raffaghelli, J.E. (2017), "Exploring the (missed) connections between digital scholarship and faculty development: a conceptual analysis", *International Journal of Educational Technology in Higher Education*, Vol. 14 No. 1, p. 20, doi: 10.1186/s41239-017-0058-x.

Raffaghelli, J.E. (2018), *Pathways to Openness in Networked Learning Research - the Case of Open Data*, available at: https://www.networkedlearningconference.org.uk/abstracts/ws_raffaghelli.htm (accessed 30 August 2020).

Raffaghelli, J.E. (2020), "«Datificación» y Educación Superior: hacia la construcción de un marco para la alfabetización en datos del profesorado universitario", *Revista Interamericana de Investigación, Educación y Pedagogía, RIIEP*, Vol. 13 No. 1, pp. 177-205, available at: https://revistas.usantotomas.edu.co/index.php/riiep/article/view/5466.

Raffaghelli, J.E. and Manca, S. (2020), *Dataset Relating the Social Activity of Open Research Data on ResearchGate (Data Set)*, Zenodo, Universitat Oberta de Catalunya, Barcelona.

Raffaghelli, J.E., Cucchiara, S., Manganello, F. and Persico, D. (2016), "Different views on Digital Scholarship: separate worlds or cohesive research field?", *Research in Learning Technology*, Vol. 24, pp. 1-17, doi: 10.3402/rlt.v24.32036.

Sangrá, A., Raffaghelli, J.E. and Guitert-Catasús, M. (2019), "Learning ecologies through a lens: ontological, methodological and applicative issues. A systematic review of the literature", *British Journal of Educational Technology*, Vol. 50 No. 4, pp. 1619-1638, doi: 10.1111/bjet.12795.

Schneider, R. (2013), "Research data literacy", *Communications in Computer and Information Science*, CCIS, Vol. 397, pp. 134-140, Springer Verlag, doi: 10.1007/978-3-319-03919-0_16.

Teal, T.K., Cranston, K.A., Lapp, H., White, E., Wilson, G., Ram, K. and Pawlik, A. (2015), "Data Carpentry: workshops to increase data literacy for researchers", *International Journal of Digital Curation*, Vol. 10 No. 1, pp. 135-143, doi: 10.2218/ijdc.v10i1.351.

Thelwall, M. and Kousha, K. (2015), "ResearchGate: disseminating, communicating, and measuring scholarship?", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 5, pp. 876-889, doi: 10.1002/asi.23236.

Trifonas, P.P. (2009), "Deconstructing research: paradigms lost", *International Journal of Research and Method in Education*, Vol. 32 No. 3, pp. 297-308, doi: 10.1080/17437270903259824.

Usova, T. and Laws, R. (2021), "Teaching a one-credit course on data literacy and data visualisation", *Journal of Information Literacy*, Vol. 15 No. 1, pp. 84-95, doi: 10.11645/15.1.2840.

Veletsianos, G. (2013), "Open practices and identity: evidence from researchers and educators' social media participation", *British Journal of Educational Technology*, Vol. 44 No. 4, pp. 639-651, doi: 10.1111/bjet.12052.

Vilar, P. and Zabukovec, V. (2019), "Research data management and research data literacy in Slovenian science", *Journal of Documentation*, Vol. 75 No. 1, pp. 24-43, doi: 10.1108/JD-03-2018-0042.

Weller, M. (2011), *The Digital Scholar: How Technology is Transforming Scholarly Practice*, Bloomsbury, London.

Wiorogórska, Z., Leśniewski, J. and Rozkosz, E. (2018), "Data literacy and research data management in two top universities in Poland. Raising awareness", *Communications in Computer and Information Science*, Springer, Cham, Vol. 810, pp. 205-214, doi: 10.1007/978-3-319-74334-9_22.

Wouters, P. and Haak, W. (2017), *Open Data: the Researcher Perspective*, Elsevier - Open Science, Leiden, doi: 10.17632/bwrnfb4bvh.1.

Supplementary material

The supplementary material is available online for this article.

About the authors

Juliana Elisa Raffaghelli is a Researcher at the Universitat Oberta de Catalunya (Spain), Faculty of Psychology and Educational Sciences. Her research interests focus on professional development for the

use of technologies in teaching and diversified work contexts, with a strong presence of international / global collaboration; Open Education and Science; critical literacy for the use of technologies, with particular reference to Big and Open Data issues. She has covered roles in research, coordination of international and European projects, learning design and teaching in several universities and research institutions. She did PhD in Education and Cognitive Sciences (University of Venice). Juliana Elisa Raffaghelli is the corresponding author and can be contacted at: jraffaghelli@uoc.edu

Stefania Manca is a Research Director at the Institute of Educational Technology of the National Research Council of Italy. Her research interests include social media and social network sites in formal and informal learning, teacher education, professional development and digital scholarship and student voice-supported participatory practices at school. She is co-editor of the *Italian Journal of Educational Technology* and Editorial board for the internet and higher education.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**