INTRODUCTION

The term ‘research data’ refers to quantitative or qualitative information collected by researchers in the course of their work, obtained from experimentation, observation, modelling, through surveys or interviews, or by other means, or created from existing data. Research data provide the necessary information to support or validate the results or conclusions of the investigation. Their
Researchers are often confused or unaware of data life-cycles and the exact definition of data, and also unsure of intellectual property rights.

- Funder policies and European Union norms are the most important motivators for sharing data within food and science researchers.
- Researchers want citation of their data sets and report being far more willing to share if citation is assured.
- Researchers believe that open data must be well formatted, updated, and easy to use and interpret.

To make the research data available, accessible, reusable, assessable, and intelligible (The Royal Society, 2012), we need not only to establish clear policies supporting open science, but also to promote changes in the practices and habits of researchers. New systems for recognition, rewarding, and evaluation of open research practices should also be taken into consideration to change the whole science communication environment. Currently, these recognition systems are based mostly on the number and quality of publications and not on the efforts made to be an open scientist.

Open research data policies are important drivers, but implementation and compliance differ depending on their provenance and disciplinary approach. Institutional and funder policies are considered very important for social sciences and chemical sciences; however, in the field of economics, policies issued by a scientific society are considered more important (Schmidt, Gemeinholzer, & Treloar, 2016).

In Europe, the European Commission and the European Research Council (2019), through the Horizon 2020 Framework Program, have supported open access to scholarly publications and to research data under the principle of ‘as open as possible, as closed as necessary’ (European Commission, 2016). European policy is focused on turning Europe into a place where open science is the new paradigm for conducting research and to making the results of publicly funded research freely available (Guedj & Ramjoué, 2015). The new European Union (EU) Directive 2019/1024 on open data and the reuse of public sector information also includes research data and requires member states to adopt open policies on data obtained through publicly funded research (European Commission, 2019a).

Public and private research funding organizations and research institutions also increasingly require researchers to develop data management plans (see Sherpa Juliet Directory, 2019; SHERPA RoMEO, 2019). A data management plan (DMP) describes the management of the life-cycle of data gathered, processed, and/or created in a project. It is a living document that identifies and describes issues such as the data-gathering process, the metadata standards used in their description, and data preservation, while also reflecting changes or modifications made during the research project. In summary, DMP provides comprehensive information on the data and the context in which they were created. There are numerous resources available to support the creation of a DMP, such as those created by the Digital Curation Center (2019), the Australian Research Data Commons (2018), or the Data Archiving and Networked Services (2019).

Data reuse practices are facilitated by aspects such as efficiency, perceived effectiveness, and the importance of developing the research (Curty, Crowston, Specht, Grant, & Dalton, 2017). However, problems arising from the interoperability of standards and formats, or a lack of understanding of intellectual property issues, may constrain their implementation (Wallis, Rolando, & Borgman, 2013).

Despite the benefits that sharing and reusing data can have for researchers (Costello, 2009; Michener, 2015; Piwowar, 2011; Wiley, 2018), both practices face technical barriers related to the management of data (Michener, 2015) and to social aspects associated with the attitudes and habits of researchers. Genomics, for example, is a field in which technology has overcome the barriers of distance and facilitated data sharing, but there are still other barriers arising from the human factor (resistance, competition, habits) that are slowing down such sharing (Fusi, Manzella, Louafi, & Welch, 2018). In the field of public health, in addition to technical, economic, and motivational difficulties that need structural solutions, there are also political, legal, and ethical considerations that require international consensus to promote fluid data exchange (Van-Panhuys et al., 2014). Fecher, Friesike, Hebing, Linek, and Sauer (2015) surveyed individual academic researchers across all disciplines on their handling of data, their publication practices, and motivators for sharing or withholding research data. The survey found that the main impediments to making data available were the need to publish a paper before sharing the data that supported it, and the efforts required to share them. Eighty percent of respondents cited the fear that others could publish their data before them as a reason for these reservations, evidencing the need for training and information on how to share and reuse data. Fear of losing control over their data was the main reason for the reluctance; however, respondents said they would openly share their data if they had the power to decide how and when their data was reused, and by whom. A big driver for making data available to others was recognition, that is, data citation.

Researchers’ behaviour related to the exchange and use of data also varies between disciplines. According to one of the first studies carried out in this regard (Tenopir et al., 2011), researchers in atmospheric and environmental sciences and ecology expressed a willingness to share (87%) their data. Most authors in other disciplines who participated in the study were also willing to share, but with significant differences among them; in particular, respondents in the fields of computer science and medicine were more reluctant. In biology, data discovery is fundamental and it is also the area where the highest percentage of shared data was reported (Stuart et al., 2018).
Researchers in the areas of engineering and computer science considered it very important to adhere to the FAIR data principles (findable, accessible, interoperable, and reusable; Wilkinson et al., 2016) for the advancement of their discipline, while researchers in earth sciences and economics supported the idea of adhering to the FAIR data principles, but with some reservations (Schmidt et al., 2016). Researchers in social sciences and political science tend to be influenced by their colleagues, so that data sharing is likely to become common practice if their community advocates and practices it (Zenk-Mölten, Akdeniz, Katsanidou, Naßhoven, & Balaban, 2018).

The Spanish National Research Council (2019) (CSIC, for its initials in Spanish) is the largest public institution in Spain dedicated to research and one of the most prominent in the European Commission (2019b). CSIC operates a large number of major research centres where its scientific and technical research activity is generally conducted. Each of these centres is thematically integrated into one of the eight scientific areas of the CSIC, one of which is food science and technology (www.csic.es/en/investigation/institutes-centres-units).

The aim of this study was to examine and analyse the perception and practice of sharing, reusing, and facilitating access to research data in the field of food science and technology among researchers working at CSIC research centres. The study data was collected by means of a focus group discussion and an online survey. This is the first subject-oriented study performed at CSIC, and it could lead to further analysis in other research areas.

METHODS

Focus group

A focus group (FG) was organized to discuss topics related to the research data life-cycle and the practice of creating DMPs. The purpose of the discussion was to gather information about perceptions and awareness of the meaning of data management and data sharing, and attitudes towards research data reuse and sharing. The FG was formed in June 2018, with 7 researchers from IATA-CSIC (Institute of Agrochemistry and Food Technology, 2019) with previous experience in applying for and conducting European research projects.

Online survey

Based on information collected from previous studies (Berghmans et al., 2017; Schmidt et al., 2016; Tenopir et al., 2015; Teperek et al., 2018), a questionnaire was developed with 43 questions organized into seven content blocks (Supporting information, Appendix S1): demographic data; format, volume, and storage of data; funding organization; sharing and reuse of research data; reasons to share open data; obstacles to publishing research data in open access; and citation, licensing, and training.

The initial questionnaire was shared and tested with three researchers who gave us their opinion about the clarity of the questions. After their comments were received, the final version was developed using the Google Forms tool. Email addresses of senior scientists, scientific researchers, research professors, and post-doctoral fellows were retrieved from the web sites of eight CSIC research institutes in food science, nutrition, and agrochemistry. We collected 465 email addresses, but 62 were found to have issues, resulting in a total of 403 researchers to be contacted. Emails were sent individually explaining the aim of the study with a link to the online survey. The questionnaire was available from July to October 2018, during which time three reminders were sent. It was not mandatory to answer all the questions. The survey data were exported to an Excel sheet for descriptive analysis and visual representation. The questions with Likert-type scales (extent of agreement/disagreement) and rating scales (very important/not at all important) with 5 or 4 points, respectively, were coded by assigning a value from the lowest to the highest. They were then grouped into four categories: barriers, motivators, conditions, and quality. A factor analysis was applied to identify which underlying factors were measured by the observed variables within these four categories. Then, a cluster analysis was applied to find, if any, groups of cases that could represent respondents as a whole. Both statistical techniques were performed using the statistics package SPSS v24 (Armonk, NY: IBM). The Kaiser–Meyer–Olkin (KMO) statistic was used to determine how suitable the data were for factor analysis (sampling adequacy), which indicates the proportion of variance in the variables that might be caused by underlying factors (values >0.7 generally indicate that a factor analysis may be useful with the obtained data). Cronbach’s α was used as a measure of reliability, or internal consistency. This metric outlines whether the test designed is accurately measuring the variable of interest.

RESULTS AND DISCUSSION

Focus group outputs

Seven researchers from the IATA-CSIC were part of the focus group. The session lasted approximately 1.5 h and participants discussed two issues raised by the facilitator:

- What do you understand a data life-cycle to be?
- What do you understand a DMP to be? Have you had any experience creating DMPs?

Out of these two topics, the following issues emerged. Each issue is followed by a brief summary based on the group’s comments, and some perceptions from the moderator:

Issue 1: Meaning of the terms ‘data life-cycle’ and ‘research data’

- The participants believed it necessary to clearly define what is meant by ‘data life-cycle’, since it can be approached from two perspectives: the research process that starts in a specific
project and ends when the results are published; or the time during which data are gathered and become ready to use.

Moderator’s perception: Lack of understanding of the concepts.

- It is also necessary to clarify what exactly is meant by ‘data’ in order to define more precisely what their life-cycle is.

Moderator’s perception: Definition of data not clear.

Issue 2: Publication of research data

- The option of publishing raw research data was unknown to participants.

Moderator’s perception: New venues for data publication unknown.

- Doubts were raised about the usefulness of publishing data, considering that all the information has already been used in the published articles.

Moderator’s perception: Lack of culture of acceptance of open raw data before publication.

Issue 3: Data lifespan

- Participants considered it necessary to determine data lifespan to be able to preserve data correctly during their life-cycle. Data are considered to be still useful as long as current technology has not surpassed the method originally used to collect them. Another concern, beyond their obsolescence, was with the validity of data in a different context. There were also participants who felt that data lose their validity once a research project is finished.

Moderator’s perception: The obsolescence of research data is not always clear.

Issue 4: Dissemination, sharing, and publication of data

- The idea of data sharing was not widely accepted because data are created/obtained within the context of a project, which renders them understandable; however, outside that context they could be misinterpreted or misunderstood. In order to avoid this, sharing data requires a previous debugging and description of how the data were collected, treated, formatted, analysed, and the rest, entailing an increased workload.

Moderator’s perception: Fear of misinterpretation hinders data sharing.

- Data are collected/generated for the purpose of publication, but not for sharing and reuse.

Moderator’s perception: Lack of culture of data sharing.

- Participants thought that information is already being shared once the outputs are published, whether through publications, conferences, and the rest.

Moderator’s perception: Publishing is misunderstood as being the same as sharing.

- Data sharing could be more easily assumed by the future generation of researchers, while the current generation continues with the traditional model in which ‘sharing’ is done after publication and not before. There is also a resistance to data sharing because of the hard work involved in obtaining the data, from which others would be able to benefit without any effort.

Moderator’s perception: Reluctance, fear of being scooped/copied.

Issue 5: Data management plans

- Participants said they have never created a DMP associated with their projects. Usually, they design a work plan to achieve their project research aims.

Moderator’s perception: Confusion about the difference between a work plan and a DMP.

- Participants understood a DMP to be a document drawn up at the end of the research, when the outputs are already available, as a final report rather than a living document drafted during the project.

- Participants agreed that DMPs should contain a description of the conditions for obtaining the data, data formats, and data licensing. They also pointed to the importance of following international standards and formats, although there was some concern about the additional workload that this represents.

Moderator’s perception: Standardization/formatting of data means more work for researchers.

Descriptive analysis

Survey-block 1: Demographic data

We sent 403 emails with the link to the online survey and received 101 responses (25%); 57 were from women and 44 from men, in the following age ranges: 26–35 years (3%), 36–45 (21%), 46–55 (50%), 56–65 (24%), and older than 65 (2%). The respondents worked in the following food science and technology disciplines: food safety (11.4%); physical properties of food (4.2%); chemical properties of food (7.2%); biotechnology (14.4%); microbiology (12.9%); nutrition (8.4%); functional foods (17.5%); phytochemistry (5.3%); modelling (2.3%); genomics and proteomics (8%); preservation (7.2%); and sensory analysis (1.1%).

Survey-block 2: Format, volume, and storage of data

Formats of research data used or created by researchers in their most recent project were textual, numerical, and images, while audio–visual formats were almost non-existent (Fig. 1). As was expected given the working areas concerned, the most common types were experimental data (53%), followed by observational or empirical (21%), reference or canonical (12%), derived or compiled (10%), and simulation (4%).

In relation to the volume of data collected as part of the most recent research project, files were mostly small to medium size, although the ‘I don’t know’ response represents nearly 23% of total answers to this question, suggesting that volume does not represent a limitation or they are not aware of it (Fig. 2A,B).
The preferred location used to store data files (Fig. 3) is their personal computer despite the risk this option represents. Worthy of note is the low use of cloud services to store data files. Backups were performed on personal computers (63% performing one backup and 37% performing two), although some respondents said they lost some files during the past year (13%).

On the question of how long data should be stored, the largest proportion of respondents considered that most data files should be stored indefinitely (35%), followed by 5–10 years (26%), 1–5 years (17%), and more than 10 years (16%).

Survey-block 3: Funding source and data ownership
According to the survey data, the biggest source of funding for research projects was the Spanish government (57%), followed by corporate private/commercial organizations based on agreements or R&D contracts with businesses in the sector (21%), regional research funding from projects funded mainly by the European Commission (17%), funding by their own institution (3%), and funded by a charity or other third party (2%).

It should be noted that when respondents were asked about their awareness of funding requirements for FAIR data, the most common response was ‘I don’t know’ (>50%), and when they were asked whether the primary funding agency requires a DMP, 60% said they did not while 38% said they did. This reflects a lack of awareness about how to make data findable, accessible, interoperable, and reusable, and/or about the funders’ requirements for DMPs, despite the fact some respondents were working under the H2020 Open Research Data Pilot. This reveals a lack of awareness of their funder’s requirements for research data.

Intellectual property and data ownership are issues researchers often misunderstand and misinterpret, and many do not even

**FIGURE 1** Formats of research data used or created during the respondent’s last research project.

**FIGURE 2** File size (A) and number of data files (B) produced as part of the researchers’ last research project.

**FIGURE 3** Locations currently used to store your own data (number of responses displayed in the bars).
know how to respond to questions about who owns their data. Indeed, when asked who they thought the research data belonged to before and after publication (Fig. 4), respondents offered very different answers, with 40% more respondents assigning ownership to the publisher after publication. This situation of ignorance about who owns research data is a very widespread problem if we consider the previous literature (Stuart et al., 2018).

Misunderstandings could also arise in relation to what is meant by data when a paper is submitted. Authors might think of data as referring to tables, figures, or images, rather than to the raw data they represent. Authors might not be aware that when a publisher is assigned copyright over a paper, that copyright does not extend to the underlying raw data that support the results or the paper. On the other hand, it was surprising to find that some respondents attributed data ownership to the funder, because, in general, grant agreements do not have clauses regarding ownership but only regarding dissemination and data sharing.

**Survey-block 4: Sharing and reuse of research data**

There is no doubt that open data is important in terms of research progress, science communication (public education and information, supporting applications of science to societal problems), and technology and knowledge transfer, and respondents scored all of these areas highly. This response might be obvious, but what the respondents say is not always what they practice.

This behaviour, whereby data sharing is understood but is not translated into action, has also been described previously (Fecher et al., 2015).

More than 50% of respondents said they shared any or all of the research data used or created as part of their last research project directly with others (i.e. person-to-person); however, only 24% reused data shared directly by others (i.e. person-to-person).

Respondents rated the importance of data sharing before and after publication with other colleagues from different institutions or with the public at large (Fig. 5) considerably higher after publication. The idea that making data publicly available before publication could jeopardize their future outputs could be the one of the main reasons for this.

Nearly 50% of respondents were aware of data repositories, but only 29% had used them to deposit datasets. Respondents were willing to directly share data collected during a research project; however, only 24% said they had used data shared by others. This might be related to a lack of trust in the data of others, or a lack of familiarity with the practice.

When asked what they had done with research data used or collected as part of their last research project, 62% of respondents said they shared them directly with researchers working on the same research project in a research collaboration, 22% shared them directly with project partners, and very few (3%) shared them directly with researchers not working on the same research project whom they did not know personally (see Table 1). This lack of trust and the fear that data may be misinterpreted or misused are attitudes also described in other fields, such as geophysics (Tenopir, Christian, Allard, & Borycz, 2018) and other disciplines (Fecher, Friesike, Hebing, & Linek, 2017).

**FIGURE 4** Who owns research data before and after publication (number of responses displayed in the bars).

**FIGURE 5** Experience sharing the data underlying your paper before and after publication (number of responses displayed in the bars).
Most respondents (69%) believe that a lot of effort is required to make their research data reusable by others (Table 1). This is a significant barrier to the reuse of research data. Familiarity with metadata, archiving venues, rights, and licensing, or with how to create a DMP, is necessary for the production of FAIR data and deserves attention. Recognition of this effort (learning about metadata, archiving venues, IP issues, licensing, etc) could transform this barrier into a motivation (Schmidt et al., 2016).

Respondents were asked to state their agreement or disagreement with different statements regarding the conditions for allowing other researchers to reuse their data (Fig. 6). Acknowledgement, recognition, citation, and providing reprints of articles using the data to the data provider, were the most common statements selected by respondents, which are directly related to collaboration and feedback from data users. Recovery of part of the cost of data acquisition does not seem as important when exchanging data is driven by prestige and reputation rather than money (Fecher et al., 2015, 2017).

Citation in formal venues, certification, compliance with guidelines and standards, and the reputation of the data providers...
are the most important factors related to the use of other researchers’ data. Regarding factors for determining whether to use others’ data (Fig. 7), more than 90% of respondents identified aspects related to the documentation and metadata provided to explain how the data were developed and formatted as important or very important. The provenance and reputation of data providers were also important factors in deciding on their reuse.

Survey-block 5: Reasons for using and sharing open data

The most common situation in which respondents had previously shared their research data (66%) has been as an appendix to a peer-reviewed research publication (journal or book), followed by a standalone peer-reviewed data publication (23%) and archiving in a data repository (17%).

All the reasons for sharing data presented in the survey were identified by most respondents as important or very important, particularly those related to funder policy compliance (although curiously scientific and professional society policies have not scored highly) (see Fig. 8). Conversely, when respondents were asked whether they were familiar with funders’ expectations for FAIR data and funder policy requirements, the response rates were low, 12 and 4%, respectively.

Figure 9 presents the survey data regarding respondents’ agreement or disagreement with certain conditions for reusing and sharing research data. Proper citation when datasets are used, ease of access, and willingness of data sharing across a broad group of researchers were the top three conditions selected by respondents. Researchers would not share all their data in data repositories without restrictions of use; however, paradoxically, they would be willing to reuse others’ data if they were easy to access. Citation of datasets when they are reused was an important issue, as seen previously. In summary, data providers do not want to lose control of who is able to access and reuse their data, but are open to new collaborations as a result of
sharing research data. This same question was asked in a survey distributed to members of the American Geophysical Union (Tenopir et al., 2018); our results are qualitatively in agreement with this previous study, although quantitatively there are some differences in the percentages of respondents who disagreed with most of the statements except for the importance of citation.

The most important attributes of open research data identified by respondents were the following: data should be recent or updated, well documented with quality metadata to facilitate their interpretation, and easily accessible but potentially protected or restricted in relation to their use (Fig. 10).

The most important benefits of publishing research data together with the article were rated as follows: it facilitates reproducibility of research (62.4%); the article is more likely to be cited (54.5%); it complies with funding body mandates (52.5%); there are more possibilities for collaboration (49.5%); it complies with journal or publisher requirements (38.6%); data can be reused (37.6%); the article is more likely to be accepted for publication (32.7%); it encourages other researchers to make their data publicly available (32.7%); and it results in research aggregation (e.g. for meta-analysis, 29.7%). These benefits are directly related to the researchers’ own activities and rewards, rather than to changing the way research data is shared.
Survey-block 6: Barriers

Legal constraints, intellectual property, data rights, potential commercial exploitation, and misuse of data are the main barriers to publishing data as open data (Fig. 11). A lack of trust and misunderstandings about data ownership are the main factors behind these barriers. However, these barriers could be overcome by the establishment of effective data management and the promotion of best practices by policymakers and academic institutions.

Poor quality of data or metadata that could lead to a misinterpretation or misuse of data is also an obstacle to the use of research data. Lack of access to data generated by other researchers was also mentioned as a major impediment to the progress of science (Fig. 12).

Data quality, formats, the time needed to understand how to interpret data, and how to access data are also major obstacles identified by respondents (Fig. 13).

Survey-block 7: Miscellaneous (citation, licence types, and training)

There is a growing need to cite research data that are not included in traditional publishing venues, but that have been shared directly (i.e., person-to-person sharing) or indirectly (downloaded from a repository). However, the culture of sharing, using, and citing research data has not grown as quickly as desired. Standards for citation exist, but awareness of those standards appears to be poor:

- There are no agreed standards for citing research data in my field (62%);
- There are agreed standards for citing research data in my field, and most researchers follow them (21%);
- There are agreed standards for citing research data in my field, but most researchers do not follow them (6%);
- There is no need for standards for the citation of research data (1%).

When asked about the licences that the researchers would use to distribute and share their data, only 8% indicated CC BY; 27% indicated CC BY-NC; 19% identified CC BY-NC-ND, and 39% responded ‘I don’t know’. Lack of knowledge about the meaning of licences might be the reason behind these results, especially because one of the motives for data sharing is the creation of potentially new results/outputs, which means that the use of restrictive licences makes no sense.

More than 70% of respondents recognized the need for training in several areas: copyright and intellectual property (88%); the use of data repositories and open access (84%);
storing and backing up data (81%); sharing data (74%); creating metadata for research data (74%); and ethics and consent for data creation (72%).

In relation to who should provide this training, the most popular response was a professional in data management (38.6%), followed by the research support unit(s) (26%), and the library (16.5%).

**Factor and cluster analyses**

As mentioned in the Methodology section, questions with rating or Likert-type scales in which respondents expressed their preference/importance or agreement/disagreement were subjected to a factor analysis to assess the correlations between variables, reduce their number, and explain the biggest variance with the smallest number of factors. Factor variables were used to group individuals in a cluster analysis to identify similarities among members of the same group.

The variables that have been taken into account responded to four categories according to the information gathered: (1) motivators for using, sharing, and publishing research data; (2) barriers to sharing research data; (3) conditions for accessing and reusing research data; and (4) aspects related to the quality of research data and metadata. The results obtained under these four categories are discussed in the following sections.

**Questions related to motivations for sharing and publishing research data**

The category grouping survey questions associated with the importance of open data and the motivators for sharing research are shown in Table 2.

In the factor analysis, a rotated matrix with the factor loadings was produced. Only factors higher than 0.6 were considered, resulting in a reduction to three factors that accounted for 62.8% of the variance, and a value of 0.748 in the KMO and Bartlett’s test. The importance of data sharing and the motivation for it are represented by these three factors. Cronbach’s α, which measures the degree of reliability, ranged between 0.7 and 0.84, confirming the scale adopted.

Factor 1 grouped items V71, V72, V70, V69, and V64 (Cronbach’s α = 0.743), involving aspects related to philanthropic issues: Dissemination and promotion of research and personal commitment to open data.

Factor 2 grouped items V65, V68, V66, and V67 (Cronbach’s α = 0.854), related to aspects to support open data motivated by institutional, funder, or publisher mandates.
Factor 3 grouped items V17, V18, V16, and V15 (Cronbach’s \( \alpha = 0.743 \)), related to more pragmatic aspects: the importance of data for technology transfer and to support applications of science to societal problems.

The cluster analysis placed the respondents into four groups. Cluster 1 grouped 38 individuals, comprising researchers with high scores for the philanthropic motivators (Factor 1). Cluster 2 grouped 45 individuals who assigned a high mean value to all three factors. Clusters 3 and 4 grouped 7 and 11 individuals with low mean values for Factor 2 and Factor 1, respectively. In summary, more than three-quarters of respondents agreed on the importance of open data to their commitment with society and for compliance with mandates.
Questions related to barriers to sharing research data

The factor analysis of questions classified under the barriers category (Table 3) produced seven factors. The first three factors accounted for 50% of the variance with a KMO and Bartlett’s test score of 0.785. The other factors did not result in a reduction of variables. These factors represent barriers to publishing, accessing, and reusing research data.

Factor 1 grouped items related to legal issues (IP, legal responsibility, misuse).

Factor 2 included aspects associated with understanding the meaning of the data (content and metadata).

Factor 3 encompassed aspects related to access to data and understanding their terms of use, licensing, and citation. Cronbach’s $\alpha$, which measures the degree of reliability, ranged between 0.88 and 0.86, confirming the scale.

---

FIGURE 13 Significant burdens when accessing and reusing data for your work (number of responses displayed in the bars).

TABLE 2 Importance of open data and the motivators to share research.

| Question                                      | Option                          | Item     | Scale          |
|-----------------------------------------------|---------------------------------|----------|----------------|
| OA importance                                 |                                 |          |                |
| In your community, how important is open data to... | Research                        | V15      | Very Important (4) |
|                                               | Informative science (education, information to the general public) | V16      | Important (3)      |
|                                               | Supporting applications of science to societal problems | V17      | Somewhat important (2) |
|                                               | Technology transfer             | V18      | Not all important (1) |
| Motivators to sharing your data               |                                 |          |                |
| Which of the following are important motivators to publish your data as open data? | Requests from other researchers | V64      | Very Important (4) |
|                                               | EU norms                        | V65      | Important (3)      |
|                                               | Organizational/institutional policy | V66     | Somewhat important (2) |
|                                               | Publisher policy                | V67      | Not all important (1) |
|                                               | Funder policy                   | V68      |                 |
|                                               | Scientific/professional society policy | V69  |                 |
|                                               | Dissemination and recognition of your work/research | V70 |                 |
|                                               | To boost research               | V71      |                 |
|                                               | Personal commitment to open data | V72      |                 |

Learned Publishing 2020; 33: 163–179
Learned Publishing © 2020 ALPSP.
www.learned-publishing.org
The cluster analysis performed with the previous three factors produced three groups that accounted for 50% of the variance. Cluster 1 comprised 43 participants who scored all three factors highly, especially Factors 1 and 2, with mean values greater than 4 out of 5 points. Cluster 2 grouped individuals who stressed the importance of metadata quality and the specification of what and how research data were gathered to facilitate their understanding. Cluster 3 comprised 15 individuals who considered legal issues related to data to be very important.

**Questions related to how to access and reuse of research data: Conditions**

The category that groups items of the questions associated with the conditions under which they would share their data or facilitate their access is displayed in Table 4.

The factor analysis produced two factors that accounted for 58% of the variance (KMO and Bartlett’s test = 0.642).

Factor 1 grouped items: V47, V45, V50, V44, and V48 (Cronbach’s $\alpha = 0.825$).

Factor 2 grouped items: V74, V73, V77, V75, and V78 (Cronbach’s $\alpha = 0.789$).

The first factor comprised items associated with permissions and control by the creator for the reuse of datasets. As has been noted in the previous sections, fear of misuse is another reason why data creators may be reluctant to share their data. Factor 2 refers to the ease/willingness with which they could access or deposit datasets in centralized repositories.

The cluster analysis resulted in three clusters of 25, 63, and 11 participants, respectively; Individuals included in cluster 1 seem not to rely on permissions for the use of datasets and they have a moderate opinion regarding facilitating access to their data through centralized repositories. Clusters 2 and 3 exhibit a moderate and similar attitude with respect to Factor 1 and Factor 2.

**Questions related to the quality of datasets**

In this section, issues affecting the quality of the metadata and attributes of the research data (Table 5) are taken into account based on the degree of importance assigned to them by the respondents.

The factor analysis of questions classified in this category (Table 5) produced five factors that accounted for 63% of the variance with a KMO and Bartlett’s test score of 0.741, related to the use of other researchers’ data depending on their characteristics.

Factor 1 grouped items V85, V84, V83, and V86 (Cronbach’s $\alpha = 0.854$), related to the need of a good definition of data and metadata.

Factor 2 grouped V55, V56, and V54 (Cronbach’s $\alpha = 0.827$), associated with the use of recognized standards for data and metadata.

Factor 3 grouped items V53, V51, and V52 (Cronbach’s $\alpha = 0.716$), associated with the reliability of the data providers (reputation and trust).

---

**TABLE 3** Items classified under the category of barriers.

| Question | Option | Item | Scale |
|----------|--------|------|-------|
| What do you see as the major barriers for publishing data as open data? | The potential use and commercial exploitation | V104 | Major barrier (5) |
| | Loss of control over intellectual property | V107 | High barrier (4) |
| | Misinterpretation or misuse of data | V106 | Half barrier (3) |
| | Loss of credit or recognition of original work | V105 | Low barrier (2) |
| | Difficulty of clarifying data rights for work involving multiple inputs or authors | V108 | No barrier (1) |
| | Concerns about legal liability for data or release of data | V103 |
| | Concerns about impact of data release (e.g., on endangered species, cultural artefacts, or vulnerable populations, unwanted disclosure, etc) | V109 |
| | Varying data formats | V124 |
| | Varying standards in how has been gathered | V123 |
| | Varying degrees of data quality in different datasets | V122 |
| | Understanding how to interpret and reuse data | V121 |
| | Accessing information on how to cite the datasets | V119 |
| | Accessing and understanding terms of use/licences | V118 |
| | Needing to register (e.g. in a website) to access data | V117 |
| | Understanding how to access the data | V120 |
Factor 4 grouped items V87, V82, and V81 (Cronbach’s $\alpha = 0.680$), related to unrestricted access to research data.

Factor 5 grouped items V58, V59, and V57 (Cronbach’s $\alpha = 0.703$), related to accreditation and certification of data shared by third parties.

The cluster analysis produced four groups with 4, 50, 46, and 1 individuals, respectively.

If we consider Factors 1 and 4 formed by outliers, we are left with two groups with approximately the same number of people whose attitudes follow the same trend with respect to the five.

### Table 4

| Question                                                                 | Option                                                                 | Item | Scale |
|-------------------------------------------------------------------------|------------------------------------------------------------------------|------|-------|
| Conditions associated with sharing your data                             | Should be co-authorship on publications resulting from use of the data | V39  | Yes (2) |
|                                                                         | Acknowledgement of the data providers in all disseminated work making use of the data | V40  | I am not sure (1) |
|                                                                         | Citation of the data providers in all disseminated work making use of the data | V41  | No (0) |
|                                                                         | Share facilitates the opportunity to collaborate on a project using the data | V42  |       |
|                                                                         | The results based (at least in part) on the data could not be disseminated in any format without the data provider’s approval | V43  |       |
|                                                                         | At least part of the costs of data acquisition, retrieval or provision must be recovered | V44  |       |
|                                                                         | The results based (at least in part) on the data could not be disseminated without the data provider having the opportunity to review the results and make suggestions or comments, but approval not required | V45  |       |
|                                                                         | Reprints of articles that make use of the data must be provided to the data provider | V46  |       |
|                                                                         | The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, and the rest | V47  |       |
|                                                                         | Legal permission for data use is obtained | V48  |       |
|                                                                         | There is a mutual agreement on reciprocal sharing of data | V49  |       |
|                                                                         | The data provider provides and accepts an agreement for its use | V50  |       |
| Motivators to using and sharing data                                      | I would use other researchers’ datasets if their datasets were easily accessible | V73  | Agree strongly (5) |
|                                                                         | I would be willing to place at least SOME of my data into a central data repository with no restrictions | V74  | Agree (4) |
|                                                                         | I would be willing to place ALL of my data into a central data repository with no restrictions | V75  | Neither agree nor disagree (3) |
|                                                                         | I would be more likely to make my data available if I could place conditions on access | V76  | Disagree (2) |
|                                                                         | Integrating data from various sources helps me in my research | V77  | Disagree strongly (1) |
|                                                                         | I would be willing to share data across a broad group of researchers | V78  |       |
|                                                                         | It is important that my data are cited when used by other researchers. | V79  |       |
|                                                                         | It is appropriate to create new datasets from shared data | V80  |       |
dimensions except for some individuals in cluster 3 who rated Factor 2 items (quality of content and metadata) more highly. Almost all respondents agreed on the importance of access, reputation, and certification of data and metadata.

CONCLUSIONS

In the existing literature there are works related to the habits and attitudes of researchers regarding sharing, use, access and reuse of research data, in areas such as psychology, sociology, physics, astronomy, and also from a multidisciplinary perspective. This study focuses on an important field not only from the research point of view, but also in terms of the importance that data could have in other fields like health, nutrition, and consumer attitudes. Reliable data sharing could therefore represent an important contribution to the scientific community and to society in general. Reasons given by researchers for data sharing reveal that many view it as a means of promoting and disseminating research results, although mandates play an important role in their participation. A lack of awareness of legal aspects related to data and a reluctance to share data for fear of losing exclusive control over it are problems that have emerged in various studies, while a lack of knowledge of data management constitutes a barrier to data sharing. All this points to the need for ongoing training and education in the emerging field of data management to make authors aware of the potential of sharing data safely and legally. There is also a need to create institutional policies on research data rather than just on publications, and to ensure their terms are understood, and the work of data managers and librarians should not be based voluntary but supported by academic authorities. Another very important issue to explore is how the work involved in sharing and managing research data is encouraged and rewarded. Some of these aspects have been stated before in other disciplines, which might lead to conclude they are not discipline-specific but something related with intrinsic research habits that need a cultural change to create a real data sharing ecosystem.

This study can serve not only as a snapshot of the attitudes of a particular scientific community but also as a guide for the types of measures needed to accompany the provision of services, training, incentives and recognition for researchers, that could be extrapolated to any other communities.

| Question | Option | Item | Scale |
|----------|--------|------|-------|
| How important were the following factors when deciding to use others' research data that the data | Are from someone I know personally | V51 | Very important (4) |
| | Are from someone I do not know personally but who has a good reputation | V52 | Important (3) |
| | Are from someone at an institute with a good reputation | V53 | Somewhat important (2) |
| | Are from a repository with a good reputation | V54 | Not all important (1) |
| | Comply with guidelines around collection and/or formatting | V55 | |
| | Have good documentation and/or metadata | V56 | |
| | Are certified or accredited by a third party | V57 | |
| | Are commonly used by others | V58 | |
| | Have been cited in a publication in the formal literature (journals, books, etc) | V59 | |
| | Which attributes do you think are most important to open data? | Access and use of data without restrictions | V81 | |
| | | Possibility of reusing data for academic/research purposes | V82 | Very Important (4) |
| | | Well-defined terms of use licence | V83 | Important (3) |
| | | The data are clearly defined in a useful way to interpret them | V84 | Somewhat important (2) |
| | | The metadata is complete, structured according to a standard, and is useful for interpreting the data | V85 | Not all important (1) |
| | | The potential value of the data for easy access and low-cost reuse | V86 | |
| | | That access to the data has no cost | V87 | |
| | | Data are updated | V88 | |
| | | Data are recent | V89 | |
ACKNOWLEDGEMENTS

The authors wish to thank Amparo Tárrega for her valuable advice and help with the statistical analyses. They also thank the Spanish Ministry of Science, Innovation and Universities for granting the national project code RTI2018-094360-B-I00.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:
Data S1. Survey questionnaire

REFERENCES

Australian Research Data Commons. (2018). Search Resources. Retrieved from https://ardc.edu.au/resources/
Berghmans, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., ... Waltman, L. (2017). Open Data: The researcher perspective. Survey and case studies Mendeley Data. v1. https://doi.org/10.17632/bwrmf4bhv1.1#file-79c8917-7e28-4eb7-8bc4-791d96e9743f
Costello, M. J. (2009). Motivating online publication of data. Bioscience, 59(5), 418–427. https://doi.org/10.1525/bio.2009.59.5.9
Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists’ data reuse. PLoS One, 12(12), e0189288. https://doi.org/10.1371/journal.pone.0189288
Data Archiving and Networked Services. (2019). [Web log post]. Retrieved from https://dans.knaw.nl/en/front-page?set_language=en
Digital Curation Center. (2019). Resources for digital Curators. Retrieved from http://www.dcc.ac.uk/resources
European Commission. (2016). Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Retrieved from https://ec.europa.eu/education/about-eu/institutions-bodies/european-commission_es
European Commission. (2019a). Open Access. Retrieved from https://ec.europa.eu/research/openscience/index.cfm?pg=openaccess
European Commission. (2019b). European Research Area. Retrieved from https://ec.europa.eu/info/research-and-innovation/strategy/era_es
European Research Council. (2019). Open Access. Retrieved from https://erc.europa.eu/managing-project/open-access
Fecher, B., Friesike, S., Hebing, M., & Linek, S. (2017). A reputation economy: How individual reward considerations trump systemic arguments for open access to data. Palgrave Communications, 3, 17051. https://doi.org/10.1057/palcomms.2017.51
Fecher, B., Friesike, S., Hebing, M., Linek, S., & Sauermann, A. (2015). A reputation economy: Results from an empirical survey on academic data sharing. DIW Discussion Papers. https://doi.org/10.2139/ssrn.2568693
Fusi, F., Manzella, D., Louafi, S., & Welch, E. (2018). Building global genomics initiatives and enabling data sharing: Insights from multiple case studies. OMICS, 22(4), 237–247. https://doi.org/10.1089/omi.2017.0214
Guedj, D., & Ramjoué, C. (2015). European Commission policy on open-access to scientific publications and research data in Horizon 2020. Biomed Data Journal, 1(1), 11–14. https://doi.org/10.11610/bmdj.01102
Institute of Agrochemistry and Food Technology. (2019). IATA. Retrieved from https://www.ata.csic.es/es
Michener, W. K. (2015). Ecological data sharing. Ecological Informatics, 29, 33–44. https://doi.org/10.1016/j.ecoinf.2015.06.010
Piwowar, H. A. (2011). Who shares? Who doesn’t? Factors associated with openly archiving raw research data. PLoS One, 6(7), e18657. https://doi.org/10.1371/journal.pone.0018657
Schmidt, B., Gemeinholzer, B., & Trelaor, A. (2016). Open data in global environmental research: The Belmont forum’s open data survey. PLoS One, 11(1), e0146695. https://doi.org/10.1371/journal.pone.0146695
SHERPA JULIET Directory. (2019). SHERPA JULIET Directory. Retrieved from https://sherpajuliet.reading.ac.uk/learn/index.php?la=en&fillNum=1&mode=simple
SHERPA RoMEO. (2019). Publisher copyright policies & self-archiving. Retrieved from www.sherpa.ac.uk/romeo/index.php
Spanish National Research Council. (2019). CSIC. Retrieved from https://www.csic.es
Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Luhr, L., & Astell, M. (2018). Whitepaper: Practical challenges for researchers in data sharing. fgshare. https://doi.org/10.6084/m9.figshare.5975011.v1
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. PLoS One, 6(6), e21101. https://doi.org/10.1371/journal.pone.0021101
Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. Earth and Space Science, 5, 891–902. https://doi.org/10.1029/2018EA000461
Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Plesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLoS One, 10(8), e0134826. https://doi.org/10.1371/journal.pone.0134826
Tepeker, M., Krause, J., Lambeng, N., Blumer, E., van Dijck, J., Eggermann, R., ... & Busse-Wicher, M. (2018). Survey: Research data management practice. https://osf.io/54em6/
The Royal Society. (2012). Science as an open enterprise. Retrieved from https://royalsociety.org/topics-policy/projects/science-public-enterprise/
Van-Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., ... & Herbst, A. J., ... & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. BMC Public Health, 14, 1144. https://doi.org/10.1186/1471-2458-14-1144
Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS One, 8(7), e67332. https://doi.org/10.1371/journal.pone.0067332
Wiley, C. (2018). Data sharing and engineering faculty: An analysis of selected publications. Science & Technology Libraries, 37(4), 409–419. https://doi.org/10.1080/0194262X.2018.1516596
Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18
Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naštoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. Journal of Documentation, 74(5), 1053–1073. https://doi.org/10.1108/JD-09-2017-0126