

User Survey Report

CSDA 2022

Michaela Kudrnáčová



Institute of Sociology, Czech Academy of Sciences
Prague, 2022

Brief introduction

The Czech Social Science Data Archive (CSDA) of the Institute of Sociology, Czech Academy of Sciences, is responsible for the professional processing, documentation, and storage of data from social science research projects, linking it with relevant research information and the context of other data and materials, and also supporting secondary analysis in the academia and education. In addition, CSDA organises various activities, such as seminars, conferences and workshops, and participates in numerous research infrastructure projects.

Given our continuous effort to improve our services, we genuinely care about our users' opinions. Therefore, we decided to follow up on a CSDA enquête of 2020 and conduct a similar survey of our users again this year. In addition to the traditional questions on ways of using the data archive, we also inquired in detail about aspects of FAIR data, open science, and data preservation.

Methods

A brief anonymous questionnaire was published online from February 8 to March 31. A total of 3804 registered users of the Nesstar data catalogue were invited. Of them, 80 were excluded for various reasons (undeliverable invitations, job change, own request – no longer using the archive). A total of 281 respondents answered at least one question.

In contrast to the previous survey, the questionnaire was made available in two versions, English as well as Czech. In designing the questionnaire, we chose to focus exclusively on information related to CSDA operation while dropping questions on age, gender and other sociodemographic characteristics.

Main findings

Survey results

- Most users exploit CSDA services under some kind of public organisation (84.7%) such as public educational institutions, academic institutions, or government institutions.
- As for the type of activity, the data archive is predominantly used for study purposes (32%).
- A total of 61.5% of the users surveyed exploit exclusively the CSDA and no other data archives.
 - Those using other archives typically mentioned the German GESIS and the Slovak SASD.
- An overwhelming majority of users (81.8%) have a college degree.
- While the respondents are somewhat familiar with the meaning of FAIR data (43.6%), the term does not mean anything to 43.2%.
- The term Open Science is more widespread among the users, likely due to its explicitness: 46.6% declare knowing the term.
- A total of 56% of the respondents state they produce some data in their study or work, but only 28.4% store the data in a content repository or data archive.
 - OSF is the most frequently used content repository and CSDA is the most frequently used data archive.
- Of those not depositing their data anywhere, 66.3% would be open, under certain conditions, to storing their data in a content repository or data archive in future.

Implications

- The relatively high number of answers to open-ended questions suggests that our users support us and want to contribute to developing the archive, share their ideas with us, and join us for a discussion of data preservation issues.
- Interestingly, only about one-third of the users exploit data archives other than CSDA.
- The findings indicate the need to step up user education on FAIR data but also on Open Science.
- The survey suggests that there is a great potential for obtaining data from our users. Unfounded concerns and a lack of knowledge of existing processes are often among the obstacles to users' openness to storing their data in archives and repositories. The presentation of the data preservation process to users should be improved to assure potential depositors of its simplicity.

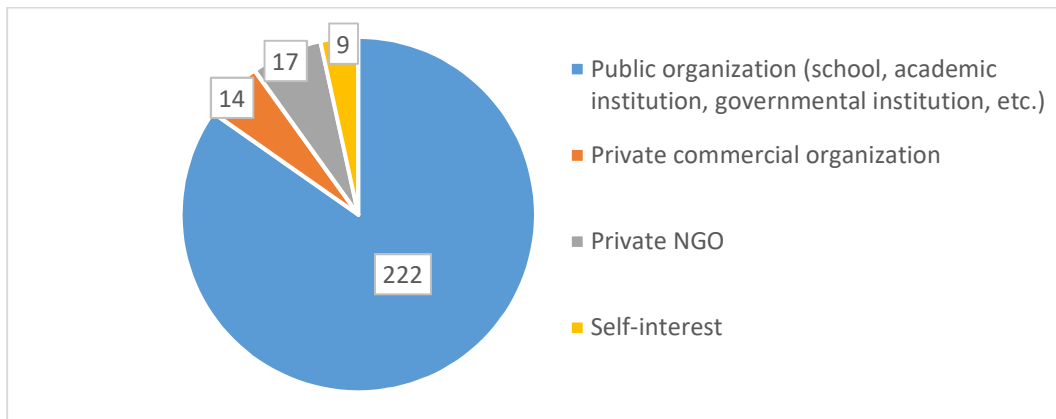
List of Figures

Figure 1: Q1. Under what type of organisation do you currently use our archive?.....	5
Figure 2: Q2. Under what type of activity do you currently use our archive?	5
Figure 3: Q3. Besides CSDA, do you use other data archives to view/download data?	5
Figure 4: Q4. What is the highest level of education that you have attained?.....	6
Figure 5: Q5. Are you familiar with the term FAIR data?	7
Figure 6: Q6. Are you familiar with the term Open Science?	7
Figure 7: Q7. In your study/work, do you produce any data?	7
Figure 8: Q7a. What typically happens to the data when a project/research stage/research assignment is completed?	8
Figure 9: Q7b. Have you ever used a content repository/data archive to store your data?	8
Figure 10: Q7c. What kind of platforms do you typically use?	8
Figure 11: Q8. While you haven't used these platforms yet, would you be open, under certain conditions, to storing your data in a content repository or data archive?.....	9

List of Tables

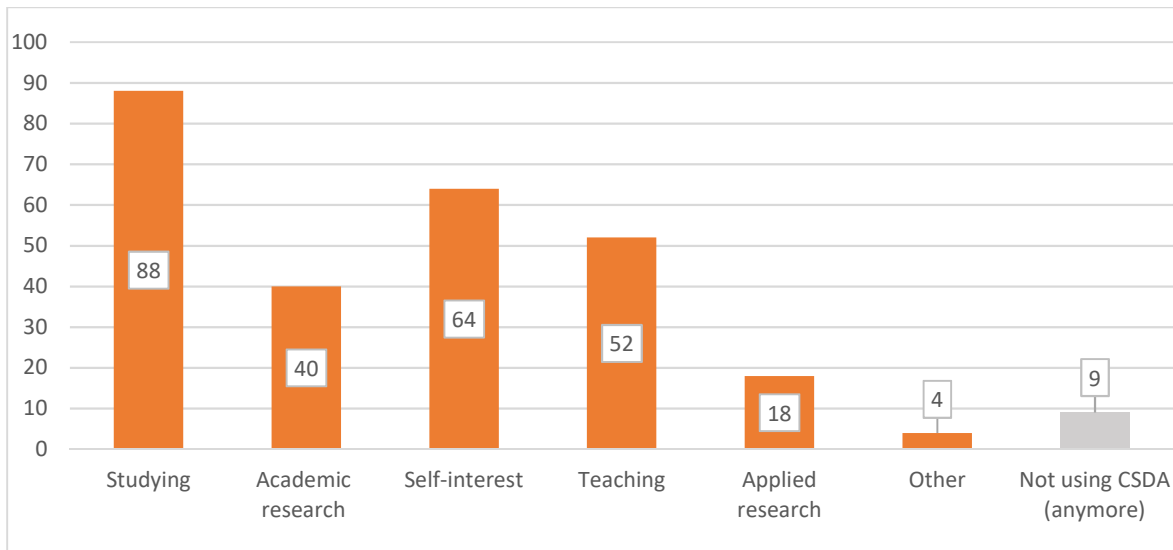
Table 1: Q3a. Which other data archives do you use?	6
Table 2: Q7c1., Q7c2. In which content repository did you store your data?	9
Table 3: Q7c3. In which data archive did you store your data?	9
Table 4: Q8a. Under what conditions would you be open to storing your data in a content repository/data archive?	9
Table 5: Please add a comment if there is anything you wish to share with us on the topic of storing research data.....	10

Figure 1: Q1. What type of organization do you currently use our archive in?1



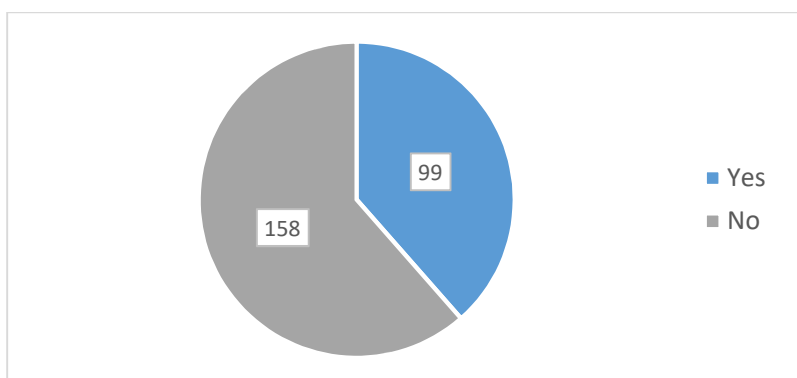
n = 262

Figure 2: Q2. What type of activity do you currently use our archive for?2



n = 272

Figure 3: Q3. Besides CSDA, do you also use other data archives to view/download data?3



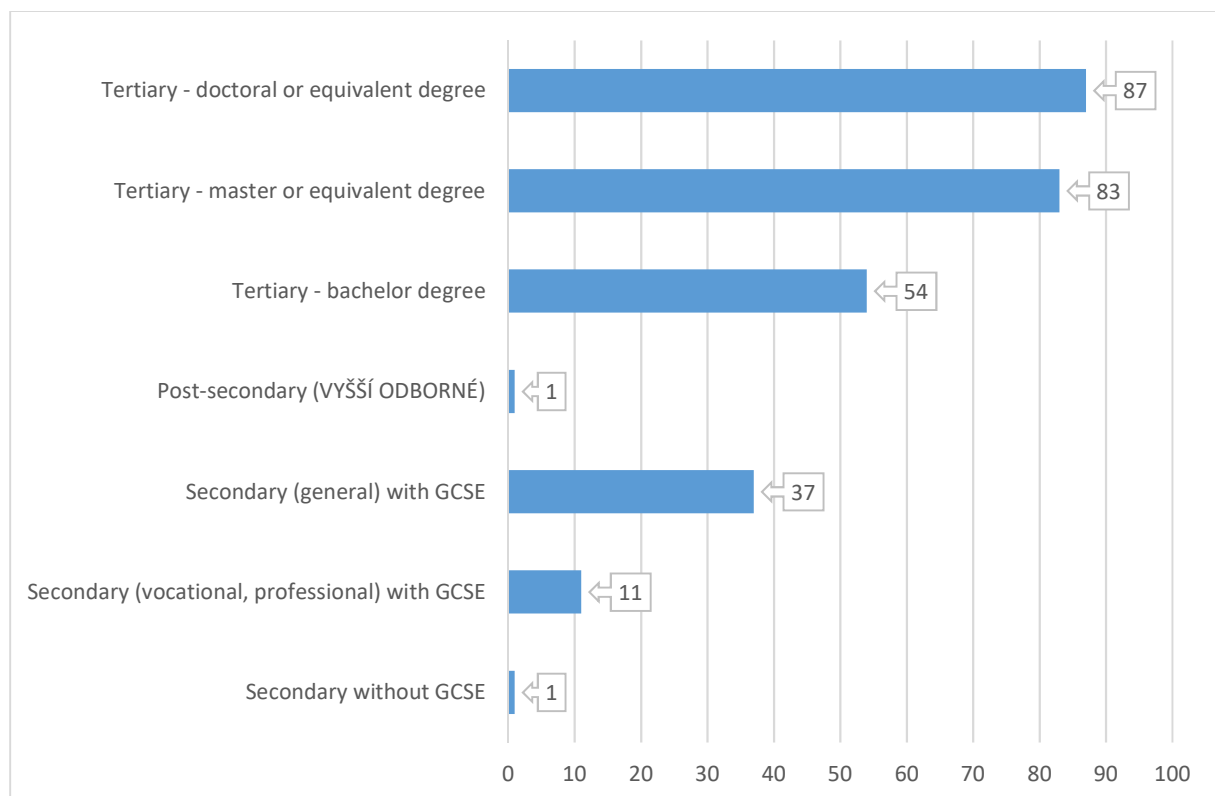
n = 257

Table 1: Q3a. Which others data archives do you use?1

[Other] Which other data archives do you use?	
➤ Germany – GESIS	50
➤ Slovakia – SASD	23
➤ United Kingdom – UKDS	11
➤ Austria – AUSSDA	9
➤ Netherlands – DANS	5
➤ Eurostat	5
➤ Norway – NSD	4
➤ Eurobarometer	4
➤ Italy – UNIDATA	3
➤ France – PROGEDO Research Infrastructure	2
➤ Israel – ISDC	2
➤ Sweden – SND	2
➤ Switzerland – FORS	2
➤ Estonia – ESSDA	1
➤ Finland – FSD	1
➤ Ireland – ISSDA	1
➤ Japan – CSRDA	1
➤ South Korea – KSDC	1
➤ Canada – <odesi>	1
➤ Romania – RODA	1
➤ Serbia – DSC	1
➤ Spain – CIS	1
➤ Other (ESS, ISSP, WVS, special project-based ones, etc.)	13

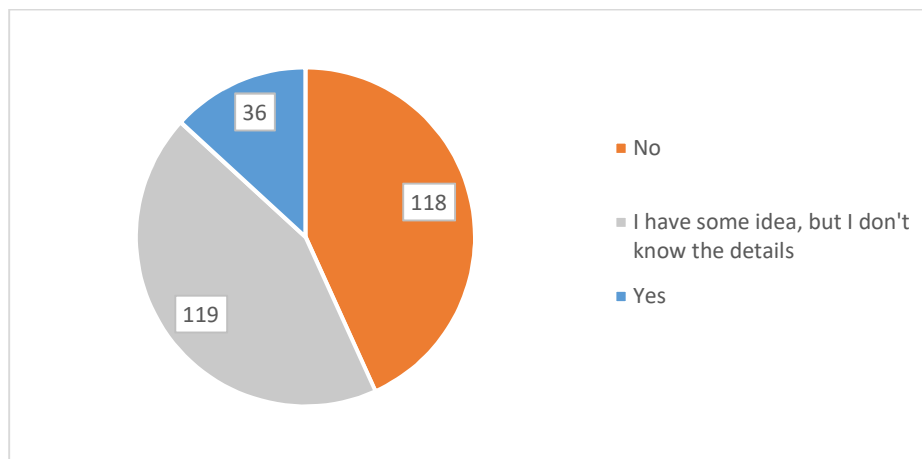
n = 144

Figure 4: Q4. What is your highest completed education?4



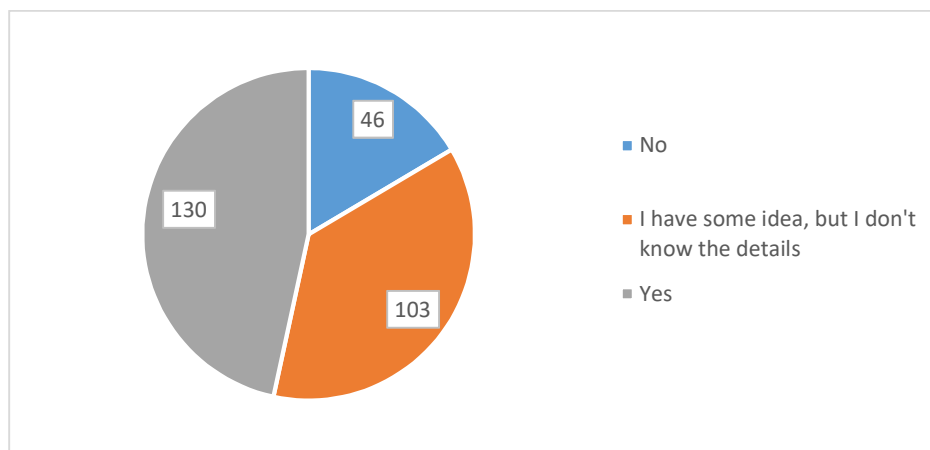
n = 274

Figure 5: Q5. Does the FAIR data mean anything to you?5



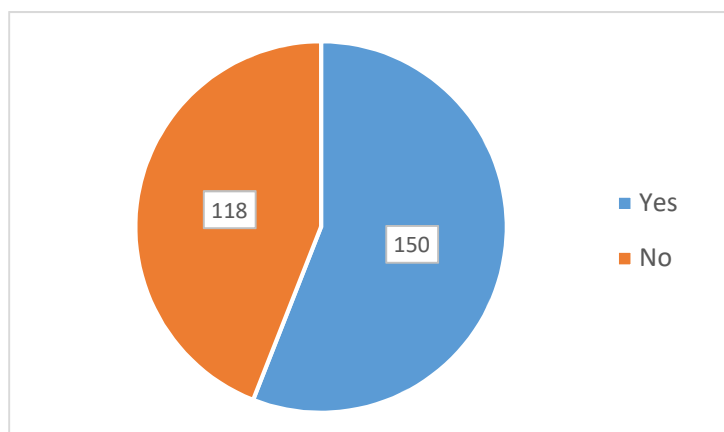
n = 273

Figure 6: Q6. Does the term Open Science mean anything to you?6



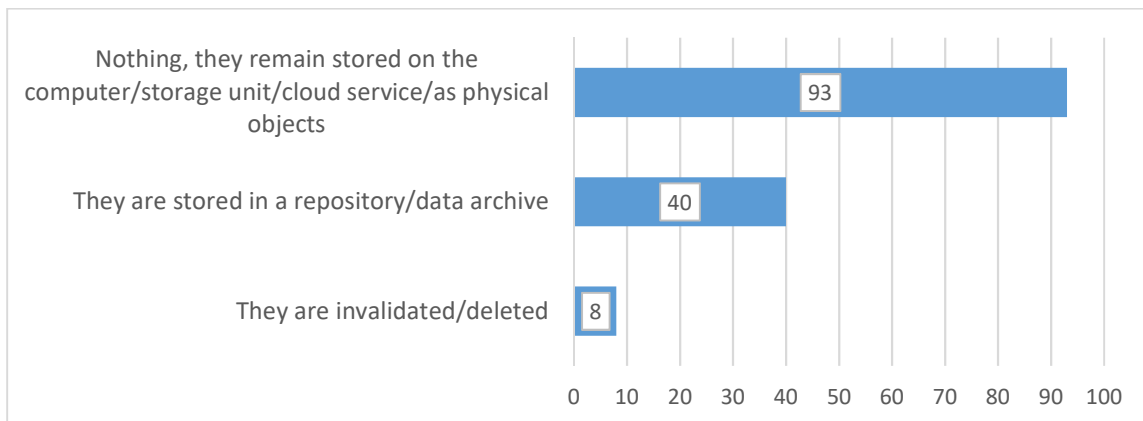
n = 279

Figure 7: Q7. Do you produce any data within your study/work?7



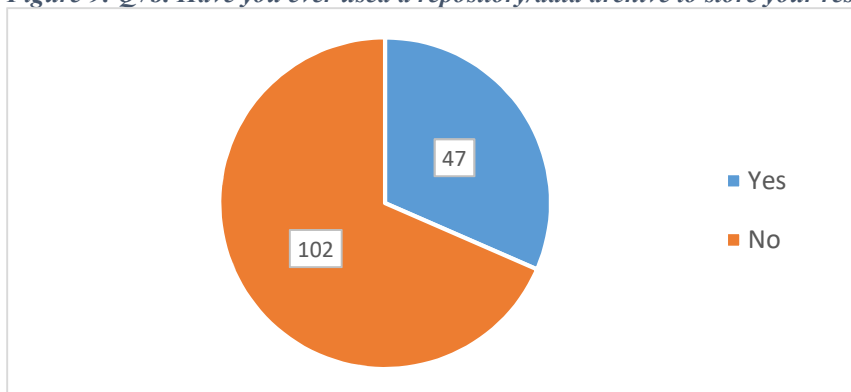
n = 268

Figure 8: Q7a. What usually happens to this data after the end of the project/research phase/research task?



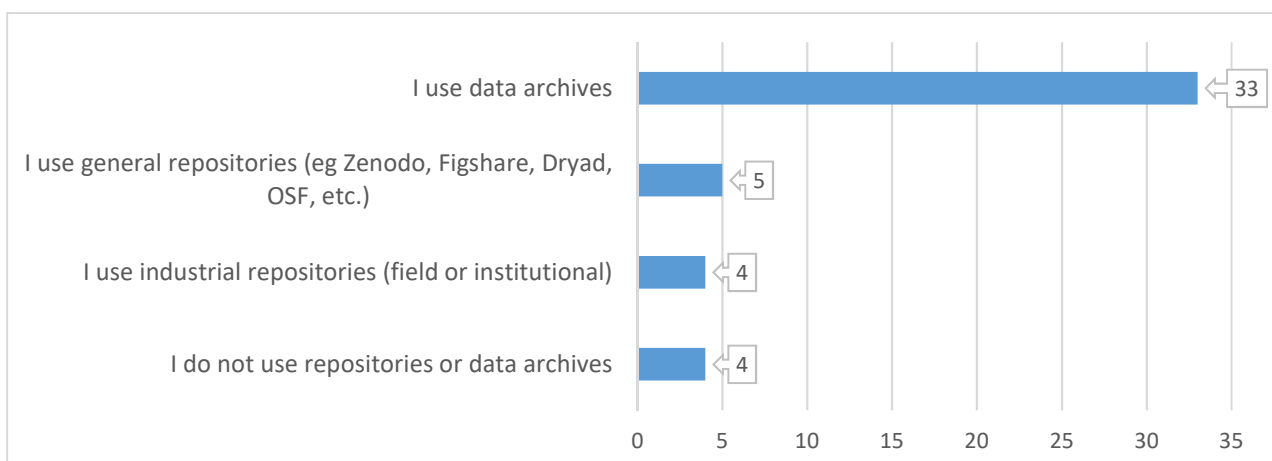
n = 141

Figure 9: Q7b. Have you ever used a repository/data archive to store your research data?



n = 149

Figure 10: Q7c. What platforms do you usually use?



n = 46

Table 2: Q7c1., Q7c2. What repository did you save the data to?2

In which content repository did you store your data?	
➤ OSF	3
➤ Zenodo	1
➤ GitHub	1
➤ Harvard Dataverse	1
➤ Mendeley	1
➤ RepOD	1

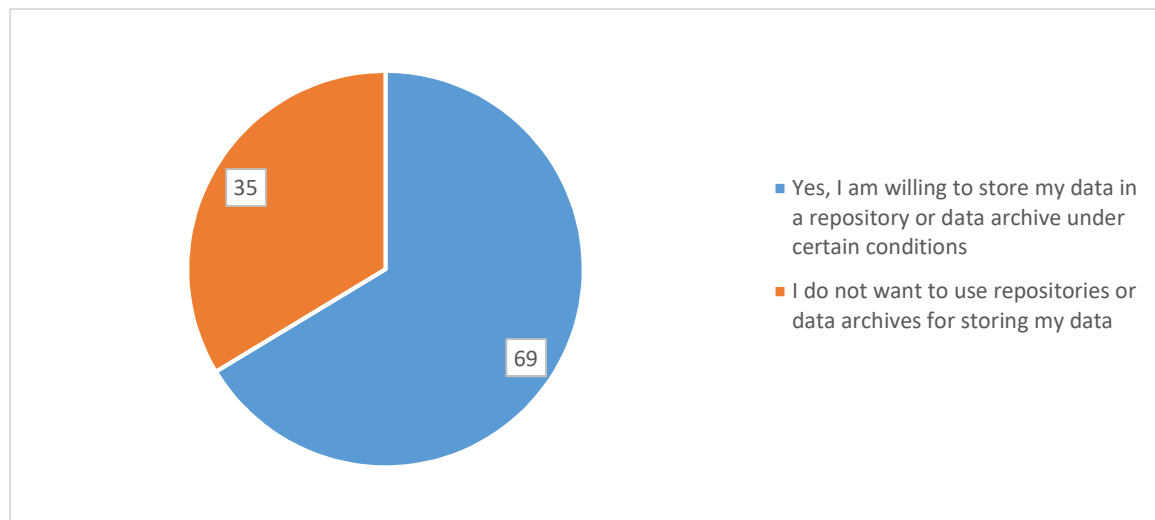
n = 8

Table 3: Q7c3. What repository did you save the data to?3

In which data archive did you store your data?	
➤ Czech Republic – CSDA	28
➤ Slovakia – SASD	1
➤ Project-based archive	1
➤ ICPSR	1
➤ In our own specific way	2

n = 33

Figure 11: Q8. Even if you have not yet used these platforms, would you be willing to store your data in a repository or data archive under certain conditions?10



n = 104

Table 4: Q8a. Under what conditions would you be willing to store your data in a repository/data archive?4

Open-ended answers
➤ <i>After its basic exploitation in publications. Personally, I'd store earlier/immediately, but there are organisational and institutional pressures to use the data before making it available.</i>
➤ <i>When my data have a decent level and are sufficiently representative, not just a "student product"</i>
➤ <i>If the data have been obtained from a publicly funded project. My current data are of internal use to the organisation collecting them.</i>
➤ <i>When I'm confident about the quality/representativeness of the data provided by me; also, ideally, I'd like to know how they were used</i>
➤ <i>It would be easy for me, and other people would have free access to the data</i>
➤ <i>Clear access, anonymity, and ethical handling of the data. Also being able to delete the data provided at any time</i>

➤ <i>It is rather complicated, and despite the amount of unique data we have obtained in several studies, we have not uploaded them in the system due to insufficient capacities. Hence, make it more user-friendly and capacity friendly.</i>
➤ <i>Simplicity</i>
➤ <i>Before launching my study, I need to know how the data is stored so I can collect it in a format other people can exploit</i>
➤ <i>When I know how to do this</i>
➤ <i>Study participants would have to agree to this (as part of their informed consent); but I often store qualitative data</i>
➤ <i>It would have to clearly make sense and, in some cases, our business client would have to agree</i>
➤ <i>I am not sure; I'd have to know more about it</i>
➤ <i>Not for commercial use</i>
➤ <i>After studying the way they are stored</i>
➤ <i>Password-protected</i>
➤ <i>Depending on respondent consent and financial availability of preservation</i>
➤ <i>Storing my data is not a problem if I know that the given data is relevant to the archive and I am told how to prepare it for preservation</i>
➤ <i>If the data was obtained with knowledge of its subsequent preservation in the data archive and if the study/data collection donor okayed that choice</i>
➤ <i>If someone provided me with an infrastructure and easy instructions on how to do this</i>
➤ <i>I definitely wouldn't have a problem if it were user-friendly and my school did not object to providing the data collected to the data archive. As a college student, I realise the importance of data availability.</i>
➤ <i>Yes, if it made sense. We collect logs – long-term storage makes no sense</i>

Table 5: Please add a comment if there is anything you wish to share with us on the topic of storing research data.5

Open-ended answers
➤ <i>It's a great service, thank you for your work</i>
➤ <i>Unfortunately, the data we generate are overwhelmingly paid for by our clients. Given the typical nature of the data, I don't expect our clients to be open to publicly sharing their data like this</i>
➤ <i>CSDA is potentially an important source for our students, especially of high-quality questionnaires and data from publicly funded projects</i>
➤ <i>I would very much appreciate: (a) a less restrictive licence (as it is, even media use is borderline, let alone for one's own project that's possibly funded from contributions), (b) faster publication of recent data so there's always up-to-date data, rather than a kind of data archaeology</i>
➤ <i>Our data is 95% restricted and cannot be shared</i>
➤ <i>I am faced with the dilemma of anonymising qualitative data while preserving its contextual sensitivity, still in the process, no result yet. I would be happy to learn from you.</i>
➤ <i>Perhaps use keywords to ensure easier, less complicated access to international survey data</i>
➤ <i>The last two questions are of no relevance to my situation. Our company has its own data archive in place. I work as a private business researcher; these questions are rather targeted at government, non-profit, or academic sector researchers</i>
➤ <i>There was only one school project collecting data and I would be ashamed to share it. We don't have the skills yet, and sharing would be embarrassing, a shame. I'll share when I have learned how to work with my data :)</i>