

ATTITUDES OF CHARLES UNIVERSITY ACADEMIC STAFF TO DATA SHARING

Adéla Jarolímková

Adela.jarolimkova@ff.cuni.cz

Institute of Information Studies and Librarianship, Faculty of Arts, Charles University

Barbora Drobíková

barbora.drobikova@ff.cuni.cz

Institute of Information Studies and Librarianship, Faculty of Arts, Charles University

Martin Souček

martin.soucek@ff.cuni.cz

Institute of Information Studies and Librarianship, Faculty of Arts, Charles University

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstract

Data management and sharing are an integral part of contemporary research work. At Charles University, we carried out a survey of selected aspects of current data management practices and researchers' attitudes to data management and sharing. In our paper we present a part of its results focused on academic staff and comparison of their answers with the answers of

doctoral students, interdisciplinary comparisons, selected comments and recommendations based on survey results.

Klíčová slova

Research data management, research data sharing, open access, researchers, academic staff

The publication was created as part of the Progres Q15 programme at Charles' University entitled "Life paths, lifestyles and quality of life from the perspective of individual adaptation and relationships between actors and institutions".

Introduction

Work with data is an integral part of contemporary science. However, skills and knowledge that are not currently a common part of higher education are required for the administration, storage and sharing of data. For this reason more and more research is focusing on the level of "data literacy", i.e. a set of skills that make it possible to search for, interpret, critically evaluate, administer and ethically use data (Calzada Prado 2013), the needs that scientists have in the area and proposed education programmes (for example, Carlson 2011, Haendel 2012, Sapp Nelson 2017, etc.). Within the Czech environment, only Pavlásková has dealt with research data in her dissertation (Pavlásková 2016).

For this reason we took the opportunity to take part in an international comparative study of data literacy and the management of research data¹ that was first conducted in France, Turkey and Great Britain and whose initial results were presented at the ECIL conference (Chowdhury 2016).

Within the bounds of this research, data is considered to be any information stored in digital format, including text, numbers, images, video or film, sound, software, algorithms, equations, animations, models, simulations, etc.

Methods

The Czech version of the questionnaire was used in the survey. A description and selected results have already been published (Jarolímková et al. 2017, Drobíková et al. 2017). The questionnaire was sent out by e-mail to all academic workers and doctoral candidates at Charles' University.

¹ Data from all participating countries will be published at a common website. Data from the Czech part of the survey is available as DROBIKOVA, Barbora, JAROLIMKOVA, Adela, and Martin SOUCEK, 2017. Data literacy and research data management survey [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.997844>

In this article we concentrate on the part of the results to concern the attitudes of academic workers to the sharing of research data and their current approach to sharing and which consists of a total of six questions (see Table 1).

<p>Which of the following applies to your research data? (My data is openly available to everyone/ My data is openly available only to my research team/ My data is available openly upon request/ My data has restricted access (e.g. only some parts of the dataset is accessible) My data is not available to anyone else)</p> <p>Do you have any concerns for sharing data with others? (No concerns/ Fear of losing the scientific edge/ Legal and ethical issues/ Misuse of data/ Misinterpretation of data/ Lack of resources (technical, financial, personnel, etc.)/ Lack of appropriate policies and rights protection/Any other)</p> <p>Do you collaborate with other researchers and share data? (No/ Yes, with researchers in the same team/ Yes, with researchers in the same university/ Yes, with researchers in other institutions/ Any other)</p> <p>I am familiar with the open access requirements (Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree)</p> <p>I am comfortable and willing to share my research data with others (Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree)</p> <p>I perceive data ethics could be an issue when research data is shared with others (Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree)</p>
--

Table 1: Questions relating to the sharing of research data

Descriptive statistics were used when interpreting data and Pearson's chi-squared test to search for connections between questions at a reliability level of $\alpha = 0.05$. The answers to questions in which a 5-scale Likert scale was used were merged as follows to ensure clearer arrangement: strongly agree and agree as yes, disagree and strongly disagree as no.

Results

A total of 2,381 responses were obtained, although only 1,434 questionnaires were completed in full. Given that only the section on demographic characteristics was completed in the incomplete questionnaires, these questionnaires were omitted from the analysis. A total of 603 complete questionnaires were obtained from academic workers at Charles' University. To ensure clear arrangement, questionnaires were divided into four basic specialisations - humanities, medicine, natural science and social science (see figure 1). Engineering and agriculture were also represented in 7 questionnaires, but these were omitted due to the low representation of those specialisations.

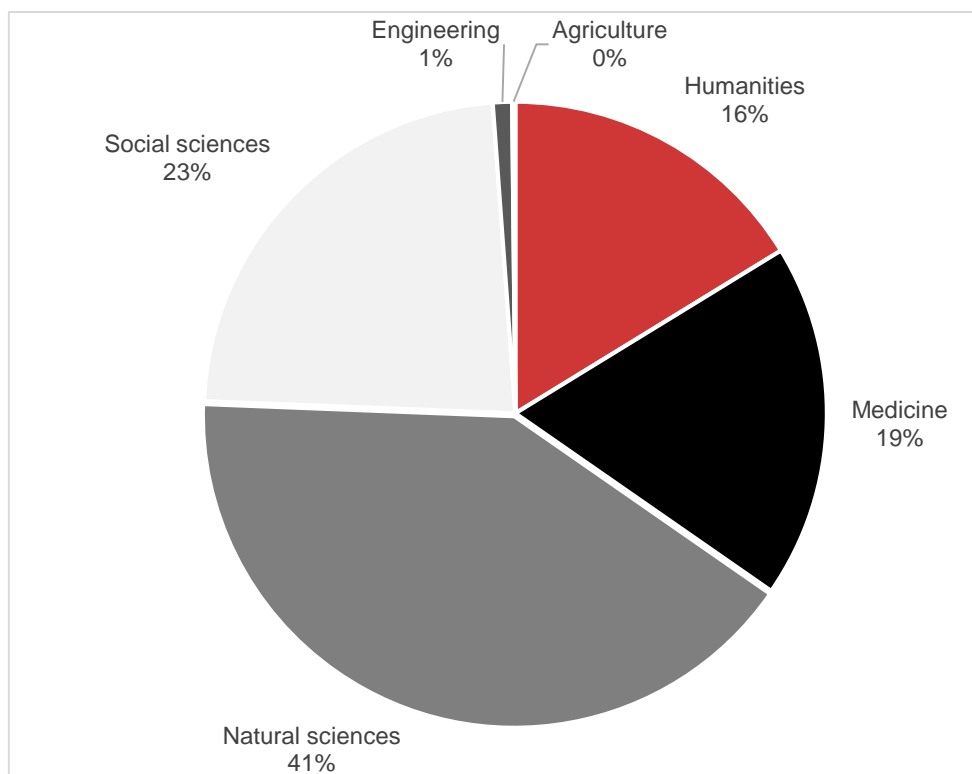


Figure 1. Respondents by specialisation

As far as the age structure of respondents is concerned, most respondents were between 36 and 45, with the category of 65 and over having the lowest representation (Figure 2).

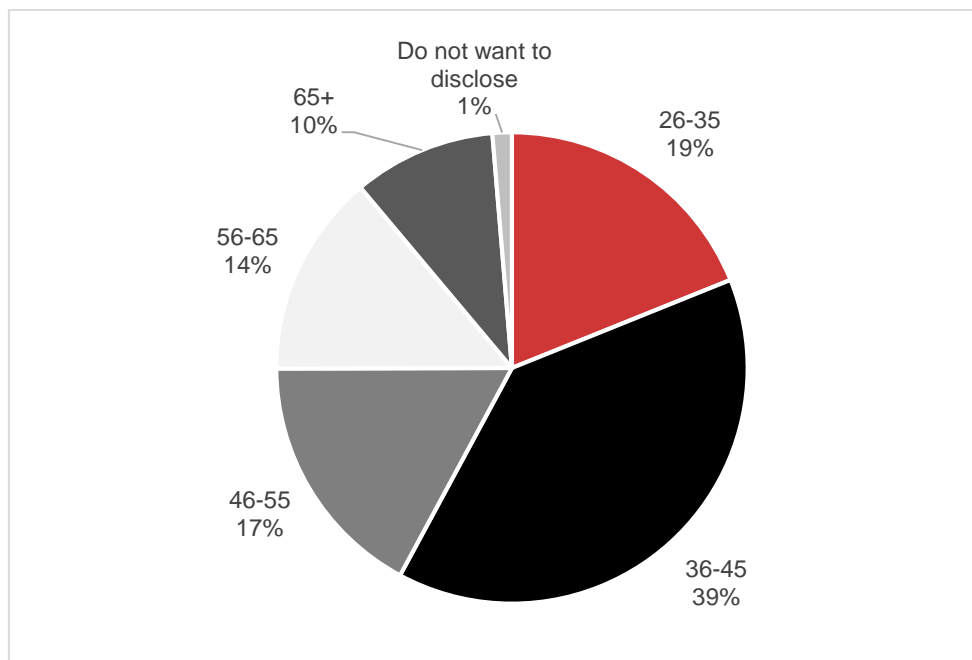


Figure 2. Respondents by age

Some 48 % of respondents from the ranks of academics are familiar with the requirements of open access (n = 596), which is higher than among doctoral candidates, only 36 % of which stated having knowledge of open access (n=826).

Most academics have already shared their data (see Figure 3). Only 13 % (n=596) stated that they did not share data in any way (Figure 3). The highest number of those who did not share came from social sciences (21 % n=140), while the percentage of those not sharing in natural sciences was only 9 % (n=247). Data is most commonly shared within a team. Some 45 % of respondents (n=596) shared their data with scientists from other institutions. In comparison with the results published in Drobíková et al, 2017, academics share their data more frequently than doctoral candidates ($p < 0.001$) and differences were also found between age categories: 8 % in the youngest age category (26-35 years) do not share data, whereas this figure is 25 % of respondents in the oldest age category (65+).

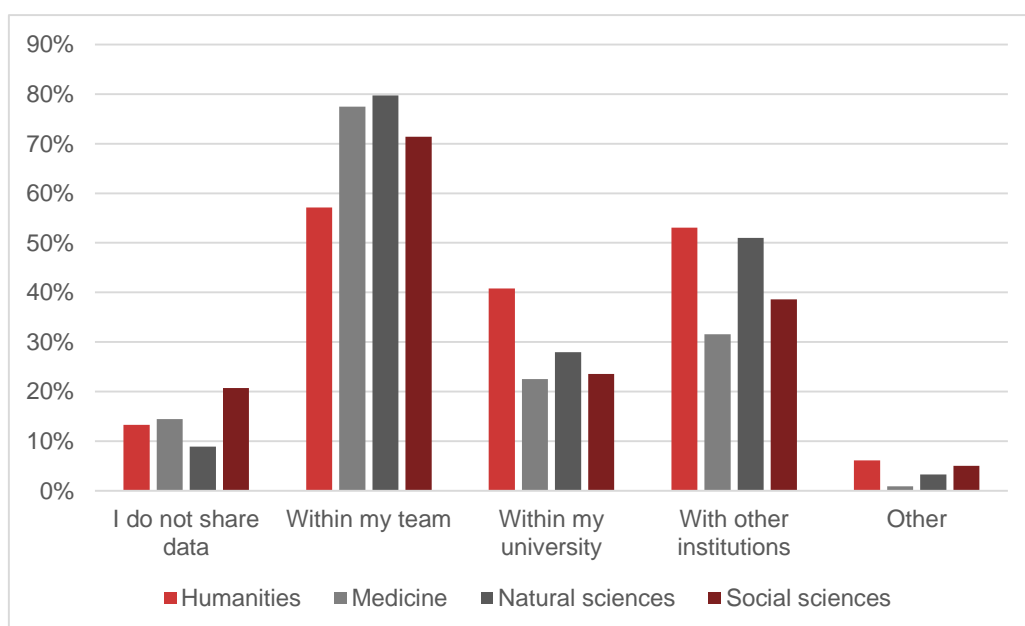


Figure 3. Current practice in data sharing

As was ascertained during the analysis of results for doctoral candidates, however, sharing does not automatically mean open access to data (Drobíková et al., 2017). On the contrary, open access is the least common method of academics sharing their data, in that there are no significant differences between specialisations here (see Figure 4). Those academics that are familiar with the principles of open access provide open access to their data more often than those who are not ($p = 0.002$). Academics also provide their data openly to a greater extent than.

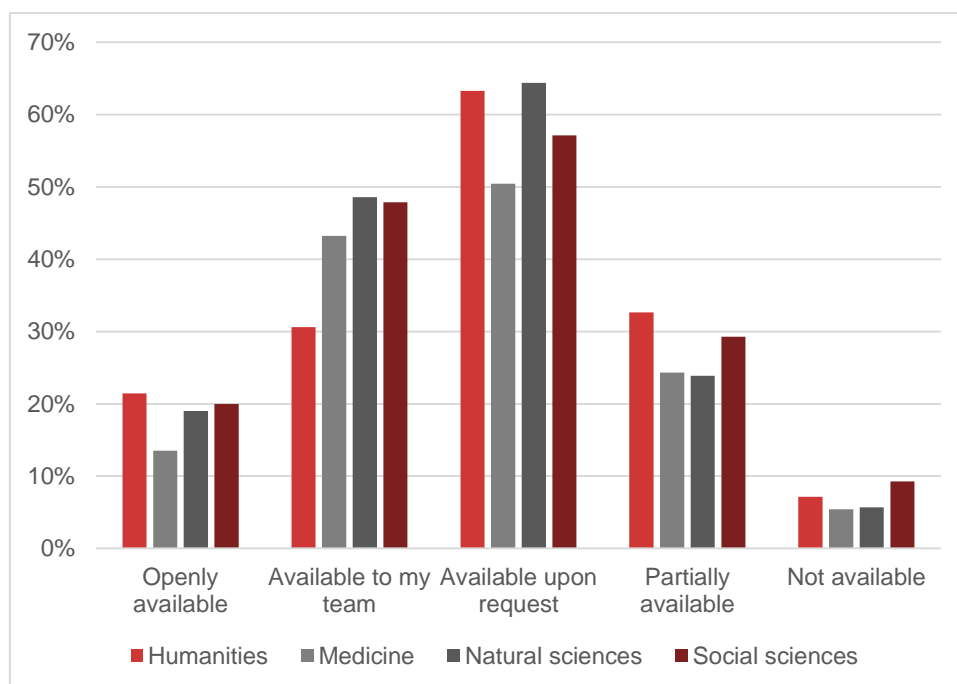


Figure 4. Method of making data available

The answers to the question on fears relating to sharing data brought some interesting results (Figure 5). Respondents were able to choose more than one answer to this question. More than a third of respondents/academics (36 %, n=596) have no fears about sharing data, whereby there is a significant difference between respondents from the humanities and natural sciences, where 41 % (n=98) and 45 % (n=247) respectively had no fears on the one hand and, on the other, medicine with (26 %, n=111) and social sciences (24 %, n=140) on the other. The highest number of academics fear incorrect interpretation of data (35 %, n=596), in that the differences between specialisations are not statistically significant. Thirty-one per cent of respondents fear legal and ethical problems, and there is again a difference here between medicine and social sciences, where respondent fears are more common, and other specialisations. Respondents from medicine also fear the misuse of data more than those from other specialisations (p=0.004). Only a small number of academics fear a lack of resources or the absence of guidelines in the sphere of research data management - on the contrary, it ensued from comments that respondents have greater fears over an excess of guidelines and regulations in this area.

Fears also affect the willingness to share data. Those that have no fears are simultaneously more often willing to share their data (p<0.001). Fears of misuse (p<0.001) and of legal and ethical problems (p=0.003) have a negative influence on the sharing of data.

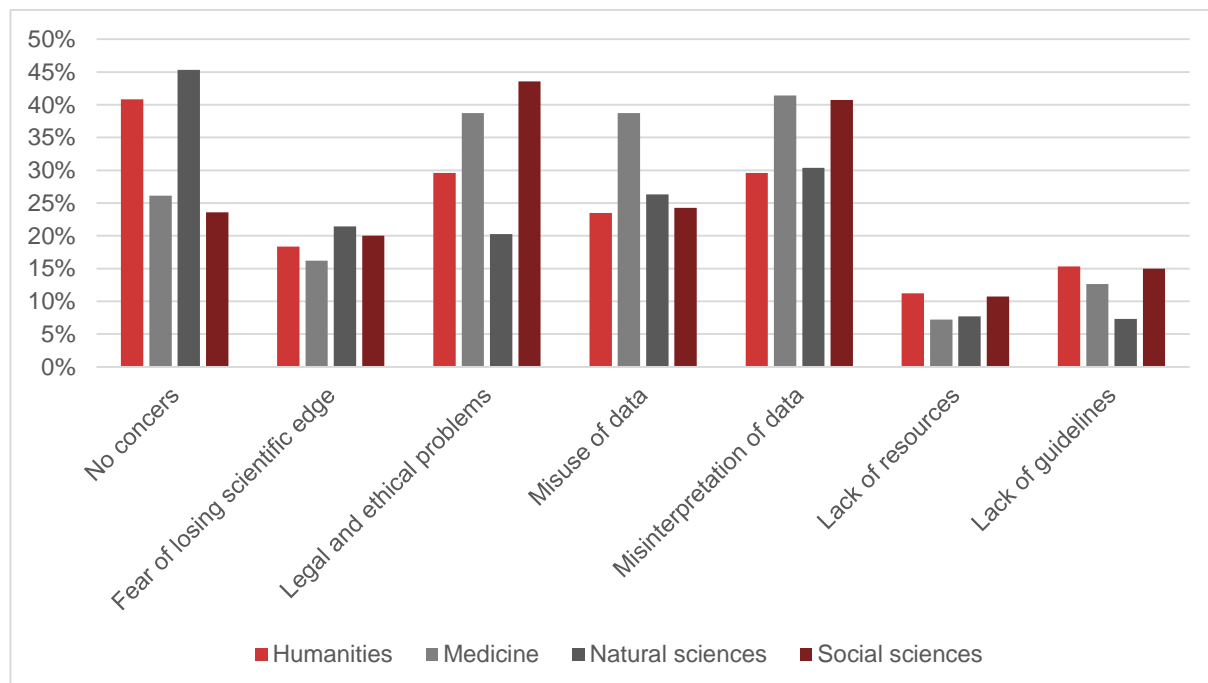


Figure 5: Fears associated with sharing data

A complementary question to the above was whether scientists are willing to share their data (Figure 6). Scientists could answer strongly agree, agree, neither agree nor disagree, disagree or strongly disagree. To ensure greater clarity, we combined the answers strongly agree and agree and the answers disagree and strongly disagree in the graph. A positive response to this question predominated in all areas of science. Scientists are willing to share their data. The answer of "agree" had strongest representation among scientists from the humanities (74 % agree to 12 % disagree, n=98). By contrast, it was lowest among scientists from the sphere of medicine (55 % agree to 23 % disagree, n=111). The attitude of scientists to this issue is consistent with the previous question. Nonetheless, the calculated chi-squared test value of $p=0.13$ does not confirm dependence of membership of a specialisation on willingness to share data.

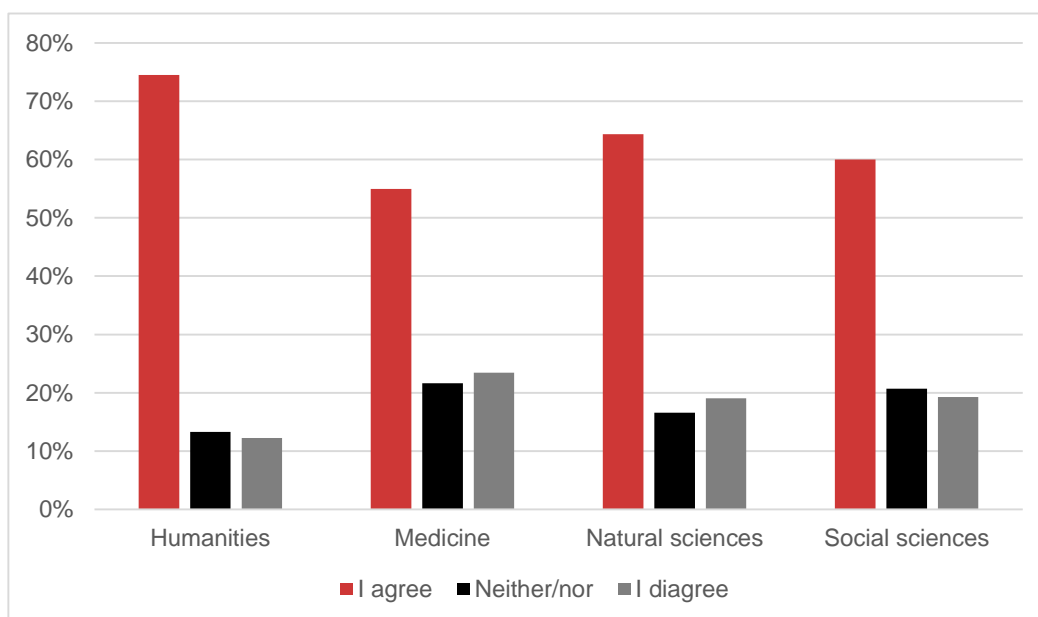


Figure 6. Willingness to share data

The questionnaire also asked the question of whether scientists would consider themselves to be exposed to ethical problems by sharing data (Figure 7). The majority of academics from all four areas of science represented invariably agreed that certain ethical problems could arise. In contrast to other areas of science, more than a fifth academics from the natural sciences (26 %, n=247) think that ethical problems cannot arise, which is a significant difference when compared with the opinions of academics from other areas of science. On the contrary, the vast majority of academics from the spheres of medicine (72 %, n=111) and social sciences (69 %, n=140) think that problems could arise.

The calculated chi-squared test value of $p=0.004$ confirms dependence of opinions in the sphere of ethical problems on scientific specialisation.

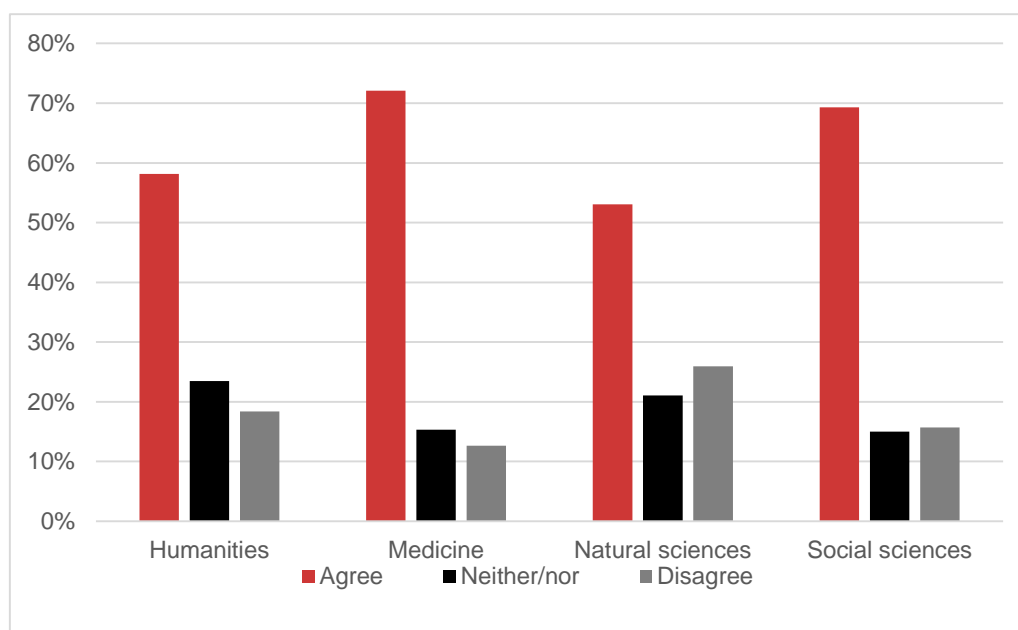


Figure 7. Ethical problems in sharing data

Discussion

The conclusions of the survey brought an initial insight into the issue of research data at Charles' University and it can be said that they do not differ significantly from similar research abroad (for example, Tenopir 2015, Enwald 2017). Academic workers at Charles' University are willing to share their data and most of them do so at present. Their approach is not one of open access, however, a fact that might be influenced by several factors. First in line is that less than half of academics are familiar with the principles of open access. Secondly, there is no simple solution for storing data in the form of an open university repository. Some specialisations have their own area-specific repository (LINDAT/CLARIN, Czech Social Sciences Data Archive), while others are reliant on international services such as Zenodo and Dryad. In respect of the fact that most respondents had not undergone any training in data management, it is understandably more difficult for them to find their way around this area, to choose the right repository and to work with it. It also emerged from the comments to the survey that open access, or indeed data sharing in general, does not make sense in all specialisations.

The willingness to openly share data is also influenced by fears over possible problems, particularly legal and ethical problems, and fears regarding the misuse or incorrect interpretation of data. There are more significant inter-disciplinary differences evident here, clearly arising from the nature of the data in individual specialisations and showing the need to adjust any solution for the administration of data, training and other activities in his area to suit individual areas of science. The data shows differences in the approach to research data between the humanities and natural sciences on the one hand and medicine and social sciences on the other. It ensues from the comments to the questionnaire that respondents also fear an increase in administration associated with data management, which was not an option provided in the questionnaire. Some refer to the technocratisation of research, the obstruction of intellectual activity or even the danger of the idea of open data sharing and there were negative comments regarding experience of the misuse or direct theft of data. As to question of sharing, it is important to respondents whether data is shared before the publication of results in the standard way or after this, something which was not differentiated in the questionnaire. Some comments also refer to the difficulty of creating a central solution with regard to the differences between specialisation, which is confirmed by the results of the questionnaire, although there were comments calling for a central repository.

Conclusions

The important findings of the research are that academics and scientists at Charles' University are willing to share their research data, but that they see a variety of risks associated with sharing, and particularly with open access, and associate a further increase in their administrative load with data management. In order to support open access to data at Charles' University, therefore, it will be necessary to create a secure infrastructure for data sharing that suits the particularities of individual specialisations and to ensure support for data management, for example in academic libraries, so that scientists are not burdened by further administrative duties.

It is also clear that more research is required to deepen our understanding of certain aspects of data sharing, research conducted using quantitative and qualitative methods, and to concentrate mainly on the specifics of individual areas of science.

References

CALZADA PRADO, Javier and Miguel Ángel MARZAL, 2013. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri* [online]. **63**(2), 123-134 [Accessed 9 April 2017]. DOI: 10.1515/libri-2013-0010. ISSN 18658423. Available from: <http://www.degruyter.com/view/j/libr.2013.63.issue-2/libri-2013-0010/libri-2013-0010.xml>

CARLSON, Jacob, Michael FOSMIRE, C.C. MILLER and Megan SAPP NELSON, 2011. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Libraries and the Academy*. **11**(2), 629-657.

DROBÍKOVÁ, Barbora, Adéla JAROLÍMKOVÁ and Martin SOUČEK, 2017. Data literacy of Charles University PhD students : are they prepared for their research careers? In: ŠPIRANEC, Sonja, Serap KURBANOGLU, Joumana BOUSTANY, Esther GRASSIAN, Diane, MIZRACHI, Lorie ROY and Denis KOS. *The Fifth European Conference on Information Literacy (ECIL) : abstracts*. Saint-Malo: Information Literacy Association, p. 41. ISBN 978-2-9561952-0-7.

ENWALD, Heike, Terttu KORTELEINEN and Maija-Leena HUOTARI. Research data management: experiences of scholars in Finland. In: ŠPIRANEC, Sonja, Serap KURBANOGLU, Joumana BOUSTANY, Esther GRASSIAN, Diane MIZRACHI, Lorie ROY and Denis KOS. *The Fifth European Conference on Information Literacy (ECIL) : abstracts*. Saint-Malo: Information Literacy Association, p. 46. ISBN 978-2-9561952-0-7.

HAENDEL, Melissa A., Nicole A. VASILEVSKY and Jacqueline A. WURZ, 2012. Dealing with Data: A Case Study on Information and Data Management Literacy. *Plos Biology* [online]. **10**(5) [Accessed 9 April 2017]. DOI: 10.1371/journal.pbio.1001339.

CHOWDHURY, G., G. WALTON, S. KURBANOGLU, Y. UNAL and J. BOUSTANY, 2016. Information Practices for Sustainability: Information, Data and Environmental Literacy. In: ŠPIRANEC, S., S. KURBANOGLU and H. LANDOVA. *The Fourth European Conference on Information Literacy (ECIL)*. Prague: Association of Libraries of Czech Universities, p. 22. ISBN 978-80-270-0530-7.

JAROLÍMKOVÁ, Adéla, Barbora DROBÍKOVÁ and Martin SOUČEK, 2017. Výzkumná data na Univerzitě Karlově. In: *INFORUM 2017: 23. ročník konference o profesionálních informačních zdrojích, Praha 30.-31. května 2017* [online]. Praha: Albertina icome Praha [Accessed 23 September 2017]. ISSN 1801-2213. Available from: <http://www.inforum.cz/pdf/2017/jarolimkova-adela.pdf>

PAVLASKOVA, Eliska, 2016. *Analýza výzkumných dat na základě fondu disertačních prací Univerzity Karlovy v Praze s ohledem na dlouhodobé uložení digitálních objektů* [online]. Praha [Accessed 12 April 2017]. Available from: <https://is.cuni.cz/webapps/zzp/detail/103851/>. Dizertační práce. Univerzita Karlova. Filozofická fakulta. Vedoucí práce RNDr. Pavel Krbec, Ph.D.

SAPP NELSON, Megan R., 2017. A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs. *Journal of eScience Librarianship* [online]. **6**(1), [Accessed 9 April 2017]. Available from: <https://doi.org/10.7191/jeslib.2017.1096>

TENOPIR, Carol, Elisabeth D. DALTON, Suzie ALLARD a Mike FRAME, 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS One* [online]. [Accessed 3 February 2017]. DOI: 10.1371/journal.pone.0134826. Available from: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>