INFORMACIJOS IR ŽINIŲ VALDYMAS

Drop out Factors in Data Literacy and Research Data Management Survey: Experiences from Lithuania and Finland

Jurgita Rudžionienė

Vilnius University, Vilnius, Lithuania
E-mail jurgita.rudzioniene@kf.vu.lt

Vincas Grigas

Vilnius University, Vilnius, Lithuania
E-mail vincas.grigas@leidykla.vu.lt
https://orcid.org/0000-0003-2414-6277

Heidi Enwald

University of Oulu, Oulu, Finland
E-mail heidi.enwald@oulu.fi
https://orcid.org/0000-0003-1953-2157

Terttu Kortelainen

University of Oulu, Oulu, Finland
E-mail terttu.kortelainen@oulu.fi
https://orcid.org/0000-0001-8407-1380

The purpose of this paper is to develop an understanding of factors that affect the respondents to drop out of an already started survey on research data management. We decided to take a questionnaire on data management survey at Vilnius University and Oulu University implemented in 2017 as a case study. The data for the analysis was collected using the questionnaire, which was used in multinational research for Data Literacy and Research Data Management, performed by a group of researchers in more than ten countries, initiated by Serap Kurbanoğlu and Joumana Boustany. This paper describes the analysis of 1 185 survey samples, of which 515 were unfinished and 670 finished in both universities. For the analysis of the data, we used Framework for Web Survey Participation created by Andy Peytchev (2009). The collected data was analyzed using IBM SPSS Statistics ver. 19 with descriptive and inferential statistical tests. The most significant factors on deciding not to finish the survey were the length of the survey, the scientific field, experience, age and the topic of the survey. No statistically significant difference was measured between those who finished the survey and unfinished evaluating the data by gender and job position. An important factor in not finishing the survey was the design of the survey.

Keywords: *survey, research data management, information literacy, Vilnius University (Lithuania), University of Oulu (Finland).*

1 Introduction

Many higher education institutions (HEIs) have recognized the need to develop Research and Data Management (RDM) services and are currently engaged in this activity (in the UK, over 40 universities have been involved in developing RDM services within Managing Research Data (MRD) programs). The increasingly collaborative nature of research is a pressing argument for RDM services. Research data exchange across different platforms is derived from the demand for effective storage, accessing and sharing data securely across multi-institutional research teams.

A significant move toward the management and research of data sharing is implemented at Vilnius University by creating a national, institutional repository for data archiving and preparing legal documents and recommendations for researchers on data management and sharing. Despite the efforts, very few researchers began to archive their data and were eager to prepare a plan for data management and sharing when there are no requirements from research funding organizations or scholarly journals. The Lithuanian research council does not imperatively require that the scholars prepare a data management and sharing plan. As a result, the majority of research data are not archived and cannot be accessed and re-used. Creating a data management and sharing plan would benefit both researchers and universities themselves. It would ensure the possibility of finding and understanding data when there is a need to use it, avoiding unnecessary duplication. For example, when recollecting or reworking data, the data underlying their respective publications are maintained, allowing for the validation of results; data sharing leads to more collaboration and advances in research; research becomes more visible and has greater impact; researchers can cite the data of other scholars, who, in turn, get credited for their studies (Jones, Pryor, Whyte 2013).

In Finland, most universities have published their data management guidelines. It includes the University of Oulu (University of Oulu 2018). Their research data policy highlights the benefits of sharing research data and points out the actions taken at the University of Oulu. Furthermore, the Ministry of Education and Culture promotes research information availability and open science through the Open Science and Research Initiative, which was set out for the years 2014–2017 (Open Science and Research Initiative 2018). The institutions have been evaluated for their openness, a framework for open science has been developed and digital preservation solutions are developed. Also, for example, the Academy of Finland, the biggest scientific funding agency in Finland, requires that Academy-funded projects commit to open access publishing and promotes open access to research data and methods. When applying for funding from the Academy, researchers must include a publication plan and a data management plan in their research plans.

This study is based on the international Data Literacy and Research Data Management survey, performed by a group of researchers in more than ten countries during 2017, initiated by Serap Kurbanoğlu and Joumana Boustany. Vilnius University surveyed Lithuania, and the University of Oulu did the same in Finland. These two universities are similar in size and QS rankings but different regarding their current state of development and policies toward research data management. Scholars at the University of Oulu have broader experiences in the field.

After implementing, the survey researchers found out that almost half of the respondents had started but did not finish the questionnaire. This aspect raised questions about what could be the reasons for not finishing the questionnaire.

The purpose of this paper is to develop an understanding of the factors that had affected the respondents to drop out of the already started survey on research data management.

This paper describes a study that combines the survey data from both countries and presents an analysis of 1 185 survey samples, of which 515 are unfinished and 670 finished. For the analysis of the data, we used the Framework for Web Survey Participation created by Andy Peytchev (2009). The collected data were analyzed using IBM SPSS Statistics ver. 19 with descriptive and inferential statistical tests.

2 Theoretical Background

2.1 Research Data Management as an Important Factor in Research

René Schneider (2018) indicates that data is generally seen as a subset of information in itself (e. g., information), as the content is generally seen as *knowledge communicated*. Also, good data management is a key conduit in ensuring the authenticity, integrity, longevity and utility of datasets and making data easier to use and reuse, which may translate to more collaboration for researchers (Strasser 2015). Data are quite tangible, indicates the author, and eligible to handle, keep, describe, analyze, transform. Data are a vivid structure characterized today as *big data*, *open data*, *linked data*. "Research Data Management in general, or Data Curation in particular, can be seen as minor concerning quantity (as

opposed to big, linked open data), but major, or at least as important, concerning quality, due to the fact that big data, open data and linked data are scientific movements that need proper data to treat, and well-curated data to perform better," indicates René Schneider (2018, p. 140). Sonja Špiranec and Denis Kos note that "openness is one of the key premises of contemporary research, which is reflected in many of its facets and processes, like access or diffusion, collaboration, evaluation etc. Although research and scientific knowledge are open by their very nature and intention, just the last decade brought this intrinsic notion to the fore, mainly owing to the availability of new digital technologies and collaborative tools" (Špiranec & Kos 2018, 148). For real benefits of open science, it is not enough to develop infrastructure; the basic issue is the willingness of researchers to share and publish their research data and build upon open data culture (Špiranec, Kos 2018).

Researchers and academics are important actors and research performers; their research data firstly are in their disposition, and those actors should be motivated and have the necessary competence to process their valuable research data, thus enabling data to be processed, shared and accessed properly. Competences in research data literacy are necessary: identifying, scoping, planning, storing, evaluating, managing and providing your data is imperative (Schneider 2018).

2.2 A Framework for the Web Survey

The survey is a quite popular method to collect data in the social sciences. Online surveying instruments are an effective way to gather information on a variety of aspects in all fields when it is intended to investigate the respondents' opinions on the issue and

enables researchers for a prompt research process as well as the efficient processing of the findings and results. It was found that the average response rate in academic studies was 55.6 with a standard deviation of 19.7 (Baruch 1999).

Researchers are interested mostly in full answers, and only that is the data for them. However, there is another kind of data that could be interesting to analyze – the respondents who dropped out the questionnaire at the very start or middle of the survey. There are various factors that affect the response rates of a web survey, such as content, the presentations of questions, contact delivery modes, the design of invitations, the use of pre-notifiactions, reminders and incentives (Weimiao Fan, Yan 2010). The dropout rate in general invitation web surveys has to be taken into consideration. The proportion of dropout in some cases may reach even 80% (O'Neil, Penrod, Bornstein 2003). On average, the dropout rate could vary from 25.3% to 44.3% (Bosnjak, Tuten 2001).

For classifying the response and nonresponse patterns in web surveys, we chose the Framework for Web Survey Participation created by Andy Peytchev (2009). The framework has three sets of factors:
a) The respondents' characteristics (environment, socio-demographics, survey predispositions, topic involvement, cognitive ability);
b) The survey design (selection and recruiting, topic, sponsor, incentive structure);
c) The page and question characteristics (question content, question type (s), number of questions, real-time validation.

These three factors have an impact on the actions with a survey. The respondents' factors and survey design impact the decision to start and to continue a survey, and they impact how and if the questions are answered. Page and question characteristics have an impact on the decision to continue to answer the questions and how they are answered.

We intend to use the framework to identify the causal mechanisms for specific outcomes and common causal mechanisms.

For naming types of participants, we chose to use response and nonresponse patterns in web surveys created by Michael Bosnjak and Tracy L. Tuten (Bosnjak, Tuten 2001).

They formulated seven processing types: 1) Complete responders; 2) Unit nonresponders; 3) Answering drop-outs; 4) Lurkers; 5) Lurking drop-outs; 6) Item nonresponders; 7) Item non-responding drop-outs.

Having in mind the structure of the survey, we decided to use three of these processing types in an analysis of the results:
1) Complete Responders (answer all questions);
2) Unit nonrespondents (did not participate in the survey), who are either (a) individuals that could have been technically hindered from participating or (b) individuals that have purposefully withdrawn after the welcome screen was displayed but before viewing any questions;
3) Answering Drop-Outs (provide answers to questions displayed but quit before completing the survey).

In our point of view, using all seven types would require to create a very specific questionnaire. Following the advice of the authors, the questionnaire should be with screen-by-screen design, non-restricted question design and each page of the questionnaire should be downloaded from the server. It helps to collect much more data on

the respondents' behavior, but on the other hand, it leads to the creation of a less comfortable questionnaire for the respondents. In our point of view, it is more convenient to have a comfortable questionnaire with the possibility of collecting the necessary data on the respondents' behavior.

2.3 Respondents' Characteristics

Reasons for dropout could be analyzed by such characteristics of the respondents as environment, socio-demographics, survey predispositions, topic involvement and cognitive ability factors.

Education and socioeconomic status have an impact on the response rate. Those of higher status are more prone to finishing a survey (Vincent 1964).

A correlation between the survey topic and the participant's level of interest has a strong impact on finishing the survey (Armstrong, Overton 1977). It is related with the importance of motivation and the ability for processing messages fully (Chaiken 1980; Petty, Cacioppo 1984).

The responses from speedy respondents (those who manage to finish the survey much faster than average) should be analyzed with care. Their answers could be with primary effects when the first response is selected automatically not caring for what is asked (Malhotra 2008; Zhang, Conrad 2014).

An important positive factor for having a higher response rate is the personalization of the invitation (Joinson, Woodley, Reips 2007; Porter, Whitcomb 2003). This has to be done in an elaborate way. Using personalized greetings only has no significant impact on the responses. Also, the message should explain the reasons to why their response is important in a more elaborate way than just a precise introduction (Ding, Poquet, Williams, Nikam, Cox 2018).

A survey has to be mobile friendly. From 15% to 30% of the respondents participate in web surveys using mobile devices such as a tablet or phone (De Bruijne, Wijnant 2014; Lugtig, Toepoel 2016). Nowadays, this proportion is even higher. Some research showed a higher rate of drops outs in the web survey when using a mobile device (Bosnjak, Poggio, Becker, Funke, Wachenfeld, Fischer 2013; Wells, Bailey, Link 2014). We have to have this in mind when deciding on how long a survey should be. It was found that it takes on average 30% longer for mobile device users to finish surveys than for desktop users (Schlosser, Mays 2018).

2.4 Survey Design

Survey design, such as the selection and recruitment of respondents, the survey topic, sponsor and incentive structure, could also be an important factor for not finishing the survey.

The reminding elements in surveys is an important factor for preventing the decline of response rates (Keusch 2015). They do not necessarily have to be wordy (Klofstad, Boulianne, Basson 2008), but they have to be funny. Receiving the humour email remainder could increase the response rate up to 24% (Rath, Williams, Villanti, Green, Mowery, Vallone 2017).

Incentives have a positive impact on the response rate, but this does not necessarily have to be a lottery prize or something similar in value. Tuten, Bosnjak, and Bandilla (1999) found that altruistic motives (contribution to scientific research) stimulate more to finish the survey than a promise to get cash in return as a prize. The same results were proven in another study implemented by O'Neil and Penrod (2001) as well.

Katja Lozar Manfreda et al. (2008) found out that the response rate is closely related to who the sponsors are. Surveys supported by government institutions have a higher response rate than those supported by commercial institutions.

The promise of the survey length has to be precise. Subjective experience during the survey (when one feels that finishing the survey will take longer than promised) increases the dropout rate (Galesic 2006).

Another important factor that influences the response rates is the topic. The topic with high interest to the surveyees helps in motivating to finish the survey (Dillman, Smyth, Christian, Dillman 2009); other studies also proved the topic as the most important factor for the response rates (Edwards, Roberts, Clarke, DiGuiseppi, Pratap, Wentz, Kwan 2002; Manfreda, Berzelak, Vehovar, Bosnjak, Haas 2008).

2.5 Page and Question Characteristics

Page and question characteristics, such as question content, question type, the number of questions and real-time validation have an impact on the response rate.

Surveyees do not like surveys with questions which are wordy, poorly-designed or flawed, with clause sentences, open-ended questions and a long list of response options, because it takes longer to finish the survey (Yan, Tourangeau 2008; Couper, Kreuter 2013; Lenzner, Kaczmirek, Lenzner 2010; Liu, Wronski 2018). Questions that are arranged in tables and in arrangements of questions that are graphically-complex (Frick, Bächtiger, Reips, 2001), inappropriate visual design (Heerwegh, Loosveldt 2002) and some questions in blocks (Galesic, 2006) have a negative impact on response rate.

Asking sociodemographic information at the start of the survey motivates to finish the survey up to 7% more often in comparison with positioned at the end of the questionnaire (Frick, Bächtinger & Reips 1999).

One of the most popular ways to organize the survey is a page-by-page version. One of the negative aspects of this type of survey is a progress indicator, which decreases the participation rates, especially at those cases when the survey is longer (up to 20 minutes) (Matzat, Snijders, van der Horst 2009).

Progress indicators when the survey is long tend to be a reason for higher drop out rates (Matzat, Snijders, van der Horst 2009).

Shorter surveys have higher response rate (Galesic, Bosnjak 2009; Ganassali 2008; Banks, Zeitlyn 2015). Optimal survey length is considered to be up to 13 minutes long (Weimiao Fan, Yan 2010).

Difficult questions should be placed at the end of the survey, especially when surveyees can scroll over the questions. When surveyees found difficult questions at the start of the survey, they concluded that finishing the survey will require too much effort, and this could have potentially influenced a lower response rate (Ganassali 2008).

Interestingly, surveys starting with open-ended questions tend to have a lower response rate in comparison with multiple-choice questions (Liu, Wronski 2018).

Response time could be a measure to evaluate the quality of survey responses. Ting Yan et al. (2015) and Mick P. Couper and Frauke Kreuter (2013) have found out answers alone do not provide enough information about how good or bad their answers are. It is important to measure how long the respondent spends to answer questions. Those who managed to finish much faster than average could be treated as skimmers and their responses as less valid.

Also, respondents need to understand the meaning of the question and determine what the question refers to. Response options are very important, as the respondents have to decide what is the difference between answer options, for example, as "somewhat satisfied" and "not very satisfied." It is very important to propose the questions for respondents seeking to avoid a misunderstanding of the intent of the question (W Fan, Yan 2010; Yan, Ryan, Becker, Smith 2015).

One of the reasons for dropping out could be a form of the survey as well. Katja Lozar Manfreda et al. (2008) analyzed 45 studies examining differences in the response rate between web surveys and other survey modes. They found out that the response rate in the web survey on average is approximately 11% lower than that of other survey modes (Manfreda, Berzelak, Vehovar, Bosnjak, Haas 2008).

3 Method

3.1 Data Collection and Participants

The study was conducted in the form of a survey using Computer-assisted web Interviewing (CAWI), with data being gathered via an online survey tool limesurvey.com. There, we present the basic data of the surveys, which we decided to analyze for drop out factors.

Data were collected between February and March 2017. Respondents to the survey were researchers, doctoral students and emeriti from Vilnius University. By the end of the survey period, data were gathered from 457 respondents from Lithuania, of which had 255 not finished and 202 finished their surveys.

In Finland, the survey was conducted by researchers from the University of Oulu, but the data were also collected by approaching several other universities and research institutions. The survey (in English, Finnish and Swedish) was distributed by utilizing the higher education institutions' and research organisations' personnel involved in the development of the national DMP tool [13]. These persons were contacted via email and asked to distribute the survey in their organizations. The data were collected in June and July 2017. In Finland, the data were collected from 728 respondents, of which 260 had not finished and 468 finished their surveys.

Data analysis from both of the countries was combined, including a total of 1 185 answers. The data were collected in accordance with confidentiality procedures. In Tables 1 and 2, we present the data of those who finished the first part (demographic).

Gender equality is well represented in the survey. In the Lithuanian case: 44.6% of the respondents are males, 53.6 – females,

Table 1. **Respondents' discipline**

Discipline	Number of participants	Percent	Number of participants	Percent
	Lithuania		Finland	
Social Sciences	220	66.3	129	21.1
Science	63	19	404	66
Humanities	47	14.2	72	11.8
Other	2	0.6	4	0.7

Table 2. **Experience in research**

Experience	Number of participants	Percent	Number of participants	Percent
	Lithuania		Finland	
< 5 years	47	14.2	262	42.6
5–10 years	82	24.7	147	23.9
11–15 years	64	19.3	61	9.9
16-20 years	32	9.6	57	9.3
> 20 years	101	30.4	182	13.3
I have never been involved in research	6	1.8	6	1

and 1.8% did not want to disclose or did not answer. In the Finnish case: 38.6% of the respondents are males, 57.6 – females, and 3.1% did not want to disclose or did not answered, and 0.8% selected *other*.

3.2 Measures

The data for the analysis was collected using a survey that was used in multinational research for Data Literacy and Research Data Management, performed by a group of researchers in more than ten countries, initiated by Serap Kurbanoglu and Joumana Boustany. The survey instrument consists of 24 questions arranged into two groups: the awareness of data management issues and demographic information. Mixed type questions (response options according to a five-point Likert scale etc.) were used, and the response options, which ranged between strongly agree/strongly disagree; almost always/never; yes/no/uncertain etc., were presented. The purpose of the survey was to find out the current levels of the awareness and gaps in knowledge of data literacy and management of university community-academic staff and doctoral students.

By analyzing the data from the two surveys, we seek to develop an understanding of the factors that represent the respondents who decided to drop out of already started surveys. Typically, we analyze those who

responded to the survey. Nevertheless, there is another important group of researchers that had showed interest in the survey – they had begun answering it, but after some time decided to drop out. This analysis will help to see the other side of data management and will aid in finding out the factors that influenced some of the respondents to not finish the survey.

3.3. Analytical Approach

For classifying response and nonresponse patterns, we chose the Framework for Web Survey Participation created by Andy Peytchev (2009). We analyzed the data using three sets of factors: a) Respondents' characteristics (environment, socio-demographics, survey predispositions, topic involvement); b) Survey design (topic, sponsor, incentive structure); c) Page and question characteristics (question content, question type (s), number of questions, real-time validation).

For naming the types of participants, we chose to use the response and nonresponse patterns in web surveys created by Michael Bosnjak and Tracy L. Tuten (Bosnjak & Tuten 2001). We decided to use three of these processing types in an analysis of the results: 1) Complete Responders (answered all questions); 2) Unit nonresponders (did not participate in the survey); of these,

there could be either (a) individuals technically hindered from participating or (b) individuals that withdrew on purpose after the welcome screen was displayed but prior to viewing any questions; 3) Answering Drop-Outs (provided answers to questions displayed but quit prior to completing the survey).

SPSS was used to analyze the dataset. A Kruskal Wallis test determined the significant differences in actions with the questionnaire (opened introduction only, finished the first part (demographic only), started the second part but unfinished, and finished the questionnaire) and the demographic factors. Statistical significance was set at the significance level of 0.05. Chi-Square tests for Association at the 0.05 significance level were conducted to find out the correlation between the researchers' disciplines, age, experience, gender, time spent on answering the questionnaire and actions with the questionnaire.

4 Findings

4.1 General Results

More than one quarter (27.35%) of the respondents from Lithuania and 15.52% from Finland opened the introduction only. We excluded these results from further analysis with the SPSS. For further analysis, only those who answered to demographic questions were left. In the Lithuanian case, this constitutes 332 answers, and in Finland's case – 615 answers; in total, the further analysis consisted of 947 answers.

Time spent to answer all questions showed that the questionnaire was more difficult than average. See Table 4.

Surveymonkey.com took a random sample of 100 thousand surveys that were from 1 to 30 questions in length (Chudoba 2018). They analyzed the average time spent to complete the survey with a certain number of questions. Their findings showed that 24 questions-long questionnaires should take

Table 3. **Actions with the questionnaire**

Action	Number of participants	Percent	Number of participants	Percent	Total
	Lithuania		Finland		
Opened introduction only	125	27.35	113	15.52	238
Finished the first part (demographic) only	124	27.13	132	21.5	256
Started the second part (data management), but unfinished the questionnaire	6	1.31	14	2.3	20
Finished the questionnaire	202	44.20	468	76.1	676

Table 4. **Time spent in answering all questions**

Action	Number of participants	Percent	Number of participants	Percent
	Lithuania		Finland	
Less than 5 minutes	2	1	138	22.4
Between 5 and 10 minutes	97	48	247	40.2
Between 10 and 15 minutes	58	28.7	128	20.8
Longer than 15 minutes	45	22.3	0	0

to answer 9 minutes on average. There is a difference between the Lithuanian and Finnish results. In the Lithuanian case, for more than a half of the respondents, it took longer than 10 minutes to answer the questions. The Finnish case is contrary to this – only one-fifth of the respondents stayed doing the survey for longer than 10 minutes.

The topic of a questionnaire itself seems to be one of the reasons to drop out of the survey. At the Lithuanian case, 45.5% of all respondents did not start the second part of the questionnaire at all. In Finland case the number is even higher – 66.2%.

4.2 Correlation between Actions with the Questionnaire and Demographic Factors

A Kruskal-Wallis test showed that in the Lithuanian case, there was a statistically significant difference in actions with a questionnaire between the discipline (p = 0.050) and time spent to answer the questions (p > 0.001). An inspection of group descriptive statistics suggests that scientific representatives were more eager to finish the questionnaire than representatives from the branches of either the social sciences or the humanities. In analyzing those who started the second part but did not finish the questionnaire,

the evident leaders are respondents from the humanities. An inspection of the descriptive statistics of time spent answering the questionnaire clearly showed that almost all who have spent longer than 5 minutes have also finished the questionnaire.

In the Finnish case, there was a statistically significant difference in actions with the questionnaire between age (p = 0.036) and time spent answering the questions (p > 0.001). An inspection of group descriptive statistics suggests that young representatives (age group from 26 to 35) are more eager to finish the questionnaire. Also, they were those who most often finished only the first part and started the second part but unfinished the questionnaire. An inspection of the descriptive statistics of time spent answering the questionnaire clearly showed that almost all who have spent longer than 5 minutes have also finished the questionnaire.

Chi-square test results in the Finland case showed a strong association between the time of involvement in research and time spent on the survey. Descriptive statistics showed that those who have been involved in research from 5 to 15 years were more eager to finish the questionnaire. In the cases of Finland and Lithuania, it is evident that those who have spent longer than 5 minutes have also finished the questionnaire.

Table 5. **Results of the Chi-square tests**

Action with questionnaire vs.	Lithuania	Finland
current primary role	$X^2(6) = 7.751$, p = 0.257	$X^2(9) = 5.147$, p = 0.821
age	$X^2(12) = 17.976$, p = 0.116	$X^2(18) = 23.107$, p = 0.187
discipline	$X^2(6) = 11.658$, p = 0.070	$X^2(12) = 11.050$, p = 0.525
gender	$X^2(4) = 6.489$, p = 0.165	$X^2(9) = 13.788$, p = 0.130
time of involvement in research	$X^2(10) = 9.452$, p = 0.490	$X^2(15) = 53.846$, p = 0.000
less than 5 minutes	$X^2(2) = 292.597$, p = 0.000	$X^2(3) = 575.265$, p = 0.000
between 5 and 10 minutes	$X^2(2) = 79.191$, p = 0.000	$X^2(3) = 109.900$, p = 0.000
between 10 and 15 minutes	$X^2(2) = 40.809$, p = 0.000	$X^2(3) = 45.920$, p = 0.000
longer than 15 minutes	$X^2(2) = 25.531$, p = 0.000	

5 Discussion

An analysis of the unfinished responses is one of the methods for understanding another side of the surveyed community's attitude toward the survey topic. Those who decided not to finish are not interested (or at least less interested) in the topic of the survey. It could be clearly be seen from analyzing the number of participants who finished the first part (demographic) but dropped out soon after starting the second part (question on research data management).

Commercial institutions did not support this survey, so it was not a factor for dropping out. The topic is not very familiar to the scholars both at Vilnius University and the University of Oulu. So not being familiar with the topic could be one of the reasons to drop out. We could support this claim with the data regarding the stage when most of the respondents decided to drop out – after finishing the demographic part and before starting the second part, where questions on research data management were listed. Only a minority of the respondents (who hadn't finished the survey) even began the second part.

The length of the survey could be a reason for dropping out, too. The questionnaire was more than 10 thousand characters long. It required a great effort to finish the questionnaire. As the analysis of time spent finishing the survey showed, the respondents spent more time finishing the questionnaire than is the average time usually spent on finishing surveys of this length. It shows that the questionnaire's questions were either very wordy or difficult to understand (this being an unfamiliar topic). It leads to the conclusion that the survey required a high cognitive effort to be finished. Also, the questionnaire's design (questions in wordy matrix tables) makes the survey friendlier for smartphones.

There were respondents who managed to finish the questionnaire much faster than the average. It leads to the conclusion that they just skimmed over the survey, and that more critical attention should be implemented to these respondents.

The response rate of the web surveys is approximately 11% lower than other forms. We should consider implementing the same survey using the paper and pencil method.

Many answers to the questions were the type of the Likert scale. The reasons to drop out could be an inability to decide how to interpret the difference between the Likert scale's options as well.

In our case, there is another factor, which was not discussed but has a high impact on the success of a survey – that what is done after the survey is delivered to its potential respondents. As we analyzed in the theoretical part, a reminding letter has to be written soon after sending the main invitation. Also, the letter itself has to be humorous and as much personal as possible.

The authors think that there is a need to analyze drop out data from more than two countries to prove the validity of the factors that have an impact on not finishing the survey. It could be valuable information for the improvement of the questionnaire and questionnaires in general. Also, these insights could benefit future studies with any related topic (complex and new), as some adjustments could be made before starting the survey and for making predictions in advance on what could the weakest points be.

There are some limitations to our study. First of all, we have used static questionnaire data and did not manipulate the order of the questions, the length of the survey, which

makes the conclusions on survey lengths and the number of questions not fully proved. On the other hand, we think that we have proved this questionnaire to have been too long using other indirect methods.

What could be done to keep the respondents from dropping out of the survey on data management? In such kind of research where a topic for the surveyees is relatively new (as theoretical and contextual background showed it was), one has to give more attention to capturing the respondent's interest from the start to the end of the survey. The survey itself has to be more enjoyable. The authors of these kinds of surveys have to give infuse their surveys with additional content so as to keep the respondents from dropping out. Also, the survey design has to be suitable for smartphone users.

6 Conclusions

The most significant impact on deciding not to finish the survey was the topic of the survey and the length of the survey (the survey being too long). By analyzing the potential demographic factors behind the reason not to finish the survey, we have found the most important to be the scientific field, experience and age. Humanities are those who were least eager to finish the survey. Researchers at the start of their careers started and finished the survey more often than others. No statistically significant difference was measured between those who finished and not finished the survey when evaluating the data based on gender and job position.

The creators of the questionnaire on research data management should consider shortening the survey, arranging the questions in a less complex form – to give more attention to the terminology used in the survey – preparing different versions of the questionnaire for those who are familiar with some conceptions on research data management and for those who are amateurs on the topic. Also, there should be different survey versions designed for separate research fields.

REFERENCES

ARMSTRONG, J. Scott and OVERTON, Terry S., 1977. Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research* [online]. August 1977. Vol. 14, no. 3, p. 396. [Accessed 24 July 2018]. DOI 10.2307/3150783. Available from: https://www.jstor.org/stable/3150783?origin=crossref

BANKS, Marcus and ZEITLYN, David, 2015. *Visual methods in social research*. 2nd ed. Los Angeles Calif: SAGE. ISBN 9781446269756. 2nd edition. Previous edition: 2001.

BARUCH, Yehuda, 1999. Response Rate in Academic Studies-A Comparative Analysis. *Human Relations*. 1999. DOI 10.1177/001872679905200401.

BOSNJAK, M., POGGIO, T., BECKER, K. R., FUNKE, F., WACHENFELD, A., FISCHER, B., 2013. Online survey participation via mobile devices. In: *The American Association for Public Opinion Research (AAPOR) 68th Annual Conference*. 2013.

BOSNJAK, Michael and TUTEN, Tracy L., 2001. Classifying Response Behaviors in Web-based Surveys. *Journal of Computer-Mediated Communication* [online]. 23 June 2001. Vol. 6, no. 3, p. 0–0. [Accessed 24 July 2018]. DOI 10.1111/j.1083-6101.2001.tb00124.x. Available from: https://academic.oup.com/jcmc/article/4584257

CHAIKEN, Shelly, 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* [online]. 1980. Vol. 39, no. 5, p. 752–766. [Accessed 24 July 2018]. DOI 10.1037/0022-3514.39.5.752. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.39.5.752

CHUDOBA, Brent, 2018. How much time are

respondents willing to spend on your survey? | SurveyMonkey. [online]. 2018. [Accessed 29 July 2018]. Available from: https://www.surveymonkey.com/curiosity/survey_completion_times/

COUPER, Mick P. and KREUTER, Frauke, 2013. Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* [online]. January 2013. Vol. 176, no. 1, p. 271–286. [Accessed 26 July 2018]. DOI 10.1111/j.1467-985X.2012.01041.x. Available from: http://doi.wiley.com/10.1111/j.1467-985X.2012.01041.x

DE BRUIJNE, Marika and WIJNANT, Arnaud, 2014. Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly* [online]. 1 January 2014. Vol. 78, no. 4, p. 951–962. [Accessed 26 July 2018]. DOI 10.1093/poq/nfu046. Available from: https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfu046

DILLMAN, Don A., SMYTH, Jolene D., CHRISTIAN, Leah Melani. and DILLMAN, Don A., 2009. *Internet, mail, and mixed-mode surveys : the tailored design method.* [online]. Wiley & Sons. [Accessed 27 July 2018]. ISBN 9780471698685. Available from: https://www.wiley.com/en-lt/Internet,+Mail,+and+Mixed+Mode+Surveys:+The+Tailored+Design+Method,+3rd+Edition-p-97804716986853rd

DING, Dan, POQUET, Oleksandra, WILLIAMS, Joseph Jay, NIKAM, Radhika and COX, Samuel Rhys, 2018. Increasing Response Rates to Email Surveys in MOOCs. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization – UMAP '18* [online]. New York, New York, USA: ACM Press. 2018. p. 203–206. [Accessed 24 July 2018]. ISBN 9781450357845. Available from: http://dl.acm.org/citation.cfm?doid=3213586.3225247

EDWARDS, Phil, ROBERTS, Ian, CLARKE, Mike, DIGUISEPPI, Carolyn, PRATAP, Sarah, WENTZ, Reinhard and KWAN, Irene, 2002. Increasing response rates to postal questionnaires: systematic review. *BMJ (Clinical research ed.)* [online]. 18 May 2002. Vol. 324, no. 7347, p. 1183. [Accessed 27 July 2018]. DOI 10.1136/BMJ.324.7347.1183. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12016181OBJECTIVE

FAN, W and YAN, Z, 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*. 2010. DOI 10.1016/j.chb.2009.10.015.

FAN, Weimiao and YAN, Zheng, 2010. Factors affecting response rates of the web survey: A system-

atic review. *Computers in Human Behavior* [online]. 1 March 2010. Vol. 26, no. 2, p. 132–139. [Accessed 5 April 2018]. DOI 10.1016/J.CHB.2009.10.015. Available from: https://www.sciencedirect.com/science/article/pii/S0747563209001708

FRICK, A., BÄCHTINGER, M. T., & REIPS, U-D, 1999. Financial incentives, personal information and drop-out rate in online studies. In: *Current Internet science. Trends, techniques, results*.

FRICK, A, BÄCHTIGER, M T and REIPS, U D, 2001. *Dimensions of Internet Science* [online]. Lengerich: Pabst Science Publishers. [Accessed 27 July 2018]. ISBN 3935357524. Available from: https://www.researchgate.net/profile/Andrea_Frick/publication/290392761_Financial_incentives_personal_information_and_drop-out_in_online_studies/links/569e352d08ae950bd7a9479a.pdf%0Ahttp://deusto.academia.edu/UlfDietrichReips/Papers/106583/Financial_incen

GALESIC, M. and BOSNJAK, M., 2009. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* [online]. 1 June 2009. Vol. 73, no. 2, p. 349–360. [Accessed 29 July 2018]. DOI 10.1093/poq/nfp031. Available from: https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfp031

GALESIC, Mirta, 2006. Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics* [online]. 2006. Vol. 22, no. 2, p. 313–328. [Accessed 27 July 2018]. Available from: http://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/dropouts-on-the-web-effects-of-interest-and-burden-experienced-during-an-online-survey.pdf

GANASSALI, Stéphane, 2008. The Influence of the Design of Web Survey Questionnaires on the Quality of Responses. *Survey Research Methods* [online]. 30 March 2008. Vol. 2, no. 1, p. 21–32. [Accessed 29 July 2018]. DOI 10.18148/srm/2008.v2i1.598. Available from: https://ojs.ub.uni-konstanz.de/srm/article/view/598he

HEERWEGH, Dirk and LOOSVELDT, Geert, 2002. An Evaluation of the Effect of Response Formats on Data Quality in Web Surveys. *Social Science Computer Review* [online]. 18 November 2002. Vol. 20, no. 4, p. 471–484. [Accessed 28 July 2018]. DOI 10.1177/089443902237323. Available from: http://journals.sagepub.com/doi/10.1177/089443902237323

JOINSON, Adam N., WOODLEY, Alan and REIPS, Ulf-Dietrich, 2007. Personalization, authenti-

cation and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior* [online]. 1 January 2007. Vol. 23, no. 1, p. 275–285. [Accessed 24 July 2018]. DOI 10.1016/J.CHB.2004.10.012. Available from: https://www.sciencedirect.com/science/article/pii/S0747563204001682

JONES, Sarah, PRYOR, Graham and WHYTE, Angus, 2013. How to Develop Research Data Management Services – a guide for HEIs. *DCC How-to Guides* [online]. 2013. Available from: http://www.dcc.ac.uk/resources/how-guides

KEUSCH, Florian, 2015. Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. *Management Review Quarterly* [online]. 9 June 2015. Vol. 65, no. 3, p. 183–216. [Accessed 27 July 2018]. DOI 10.1007/s11301-014-0111-y. Available from: http://link.springer.com/10.1007/s11301-014-0111-y

KLOFSTAD, Casey A., BOULIANNE, Shelley and BASSON, Danna, 2008. Matching the Message to the Medium. *Social Science Computer Review* [online]. 18 November 2008. Vol. 26, no. 4, p. 498–509. [Accessed 27 July 2018]. DOI 10.1177/0894439308314145. Available from: http://journals.sagepub.com/doi/10.1177/0894439308314145

LENZNER, Timo, KACZMIREK, Lars and LENZNER, Alwine, 2010. Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology* [online]. October 2010. Vol. 24, no. 7, p. 1003–1020. [Accessed 27 July 2018]. DOI 10.1002/acp.1602. Available from: http://doi.wiley.com/10.1002/acp.1602

LIU, Mingnan and WRONSKI, Laura, 2018. Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. *Social Science Computer Review* [online]. 23 February 2018. Vol. 36, no. 1, p. 116–124. [Accessed 29 July 2018]. DOI 10.1177/0894439317695581. Available from: http://journals.sagepub.com/doi/10.1177/0894439317695581A

LUGTIG, Peter and TOEPOEL, Vera, 2016. The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey. *Social Science Computer Review* [online]. 26 February 2016. Vol. 34, no. 1, p. 78–94. [Accessed 26 July 2018]. DOI 10.1177/0894439315574248. Available from: http://journals.sagepub.com/doi/10.1177/0894439315574248

MALHOTRA, Neil, 2008. *Completion Time and Response Order Effects in Web Surveys* [online]. 2008. Oxford University PressAmerican Association for Public Opinion Research. [Accessed 24 July 2018]. Available from: https://www.jstor.org/stable/25548051

MANFREDA, Katja Lozar, BERZELAK, Jernej, VEHOVAR, Vasja, BOSNJAK, Michael and HAAS, Iris, 2008. Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* [online]. 1 January 2008. Vol. 50, no. 1, p. 79–104. [Accessed 26 July 2018]. DOI 10.1177/147078530805000107. Available from: http://journals.sagepub.com/doi/10.1177/147078530805000107

MATZAT, Uwe, SNIJDERS, Chris and VAN DER HORST, Wouter, 2009. Effects of Different Types of Progress Indicators on Drop-Out Rates in Web Surveys. *Social Psychology* [online]. 10 January 2009. Vol. 40, no. 1, p. 43–52. [Accessed 28 July 2018]. DOI 10.1027/1864-9335.40.1.43. Available from: http://econtent.hogrefe.com/doi/abs/10.1027/1864-9335.40.1.43

O'NEIL, Kevin M. and PENROD, Steven D., 2001. Methodological variables in Web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods, Instruments, & Computers* [online]. May 2001. Vol. 33, no. 2, p. 226–233. [Accessed 27 July 2018]. DOI 10.3758/BF03195369. Available from: http://www.springerlink.com/index/10.3758/BF03195369

O'NEIL, Kevin M., PENROD, Steven D. and BORNSTEIN, Brian H., 2003. Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, & Computers* [online]. May 2003. Vol. 35, no. 2, p. 217–226. [Accessed 24 July 2018]. DOI 10.3758/BF03202544. Available from: http://www.springerlink.com/index/10.3758/BF03202544

OPEN SCIENCE AND RESEARCH INITIATIVE, 2018. Open Science and Research. [online]. 2018. Available from: https://openscience.fi/

PETTY, Richard E. and CACIOPPO, John T., 1984. The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology* [online]. 1984. Vol. 46, no. 1, p. 69–81. [Accessed 24 July 2018]. DOI 10.1037/0022-3514.46.1.69. Available from: http://content.apa.org/journals/psp/46/1/69

PEYTCHEV, Andy, 2009. *Survey Breakoff* [online]. 2009. Oxford University PressAmerican Association for Public Opinion Research. [Accessed 24 July 2018]. Available from: https://www.jstor.org/stable/25548063Survey

PORTER, Stephen R. and WHITCOMB, Michael E., 2003. *The Impact of Contact Type on Web Survey Response Rates* [online]. 2003. Oxford University PressAmerican Association for Public Opinion Research. [Accessed 24 July 2018]. Available from: https://www.jstor.org/stable/3521694

RATH, Jessica M., WILLIAMS, Valerie F., VILLANTI, Andrea C., GREEN, Molly P., MOWERY, Paul D. and VALLONE, Donna M., 2017. Boosting Online Response Rates Among Nonresponders. *Social Science Computer Review* [online]. 14 October 2017. Vol. 35, no. 5, p. 619–632. [Accessed 27 July 2018]. DOI 10.1177/0894439316656151. Available from: http://journals.sagepub.com/doi/10.1177/0894439316656151

SCHLOSSER, Stephan and MAYS, Anja, 2018. Mobile and Dirty. *Social Science Computer Review* [online]. 27 April 2018. Vol. 36, no. 2, p. 212–230. [Accessed 26 July 2018]. DOI 10.1177/0894439317698437. Available from: http://journals.sagepub.com/doi/10.1177/0894439317698437

SCHNEIDER, René, 2018. Training Trainers for Research Data Literacy: A Content- and Method-Oriented Approach. In: [online]. Springer, Cham. p. 139–147. [Accessed 26 July 2018]. Available from: http://link.springer.com/10.1007/978-3-319-74334-9_15

ŠPIRANEC, Sonja and KOS, Denis, 2018. Data Literacy and Research Data Management: The Croatian State of Affairs. In: [online]. Springer, Cham. p. 148–157. [Accessed 26 July 2018]. Available from: http://link.springer.com/10.1007/978-3-319-74334-9_16

STRASSER C., 2015. *Research Data Management. A Primer Publication of the National Information Standards Organization* [online]. 2015. Available from: http://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf

TUTEN, Tracy L, BOSNJAK, Michael and BANDILLA, Wolfgang, 1999. Banner-advertised Web surveys. *Marfketing Research* [online]. 1999. Vol. 11, no. 4, p. 16. [Accessed 27 July 2018]. Available from: http://search.proquest.com/openview/d6285b57d5de719d0df82495ab456b84/1?pq-origsite=gscholar&cbl=31079

UNIVERSITY OF OULU, 2018. Research data policy at the University of Oulu. [online]. 2018. Available from: http://www.oulu.fi/university/node/44529

VINCENT, Clark E., 1964. Socioeconomic Status and Familial Variables in Mail Questionnaire Responses. *American Journal of Sociology* [online]. May 1964. Vol. 69, no. 6, p. 647–653. [Accessed 24 July 2018]. DOI 10.1086/223695. Available from: https://www.journals.uchicago.edu/doi/10.1086/223695

WELLS, Tom, BAILEY, Justin T. and LINK, Michael W., 2014. Comparison of Smartphone and Online Computer Survey Administration. *Social Science Computer Review* [online]. 7 April 2014. Vol. 32, no. 2, p. 238–255. [Accessed 26 July 2018]. DOI 10.1177/0894439313505829. Available from: http://journals.sagepub.com/doi/10.1177/0894439313505829

YAN, Ting, RYAN, Lindsay, BECKER, Sandra E and SMITH, Jacqui, 2015. Assessing Quality of Answers to a Global Subjective Well-being Question Through Response Times. *Survey research methods* [online]. 2015. Vol. 9, no. 2, p. 101–109. [Accessed 26 July 2018]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27398099

YAN, Ting and TOURANGEAU, Roger, 2008. Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology* [online]. January 2008. Vol. 22, no. 1, p. 51–68. [Accessed 27 July 2018]. DOI 10.1002/acp.1331. Available from: http://doi.wiley.com/10.1002/acp.1331

ZHANG, Chan and CONRAD, Frederick, 2014. Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods* [online]. 23 July 2014. Vol. 8, no. 2, p. 127–135. [Accessed 24 July 2018]. DOI 10.18148/srm/2014.v8i2.5453. Available from: https://ojs.ub.uni-konstanz.de/srm/article/view/5453

DUOMENŲ RAŠTINGUMO IR MOKSLINIŲ TYRIMŲ DUOMENŲ VALDYMO APKLAUSOSE KLAUSIMYNŲ PILDYMO NUTRAUKIMĄ LEMIANTYS VEIKSNIAI: LIETUVOS IR SUOMIJOS PATIRTIS

Jurgita Rudžionienė, Vincas Grigas, Heidi Enwald, Terttu Kortelainen

S a n t r a u k a

Straipsnio tikslas – nustatyti ir išanalizuoti veiksnius, turinčius įtakos respondentų sprendimui nutraukti klausimyno pildymą. Šiam tikslui įgyvendinti pasirinkta Informacinio raštingumo ir mokslinių duomenų valdymo tyrimo anketa, kuri buvo pateikiama Vilniaus universiteto (Lietuva) ir Oulu universiteto (Suomija) mokslininkams, atliekant mokslinių duomenų valdymo tyrimą abiejuose universitetuose 2017 m. Šis tyrimas buvo tarptautinio (atlikto keliasdešimtyje valstybių) mokslinio tyrimo apie duomenų raštingumą ir mokslinių duomenų valdymą (angl. *Data Literacy and Research Data Management*) dalis. Tyrimo vadovės – Serap Kurbanoğlu ir Joumana Boustany. Šiame straipsnyje analizuojami 1 185 anketų duomenys iš abiejų universitetų, iš kurių 515 – pradėtos ir nebaigtos pildyti anketos, o 670 – baigtos pildyti anketos. Duomenų analizė atlikta taikant *Framework for Web Survey Participation* (Andy Peytchev, 2009). Gautieji duomenys apdoroti taikant *IBM SPSS Statistics* 19 versiją, naudojant aprašomosios ir inferencinės statistikos testus. Tyrimo rezultatai atskleidė, kad svarbiausi veiksniai, nulėmę nebaigtą anketos pildymą, yra anketos apimtis, respondentų atstovaujama mokslo sritis, mokslinio darbo patirtis, amžius ir tyrimo tematika. Einamos pareigos ir respondentų lytis reikšmingos įtakos tam neturėjo. Svarbus veiksnys, lemiantis anketos pildymo neužbaigimą, yra ir klausimyno dizaino bei struktūros sprendimai.

Pagrindiniai žodžiai: apklausa, mokslinių duomenų valdymas, informacinis raštingumas, Vilniaus universitetas (Lietuva), Oulu universitetas (Suomija).