Breaking Down Barriers: How Conversational Agents Facilitate Open Science and Data Sharing

Milad Mirbabaie
University of Bamberg, Germany
milad.mirbabaie@uni-bamberg.de

Jonas Rieskamp
Paderborn University, Germany
jonas.rieskamp@uni-paderborn.de

Lennart Hofeditz
University of Potsdam, Germany
lennart.hofeditz@uni-potsdam.de

Stefan Stieglitz
University of Potsdam, Germany
stefan.stieglitz@uni-potsdam.de

Abstract

Many researchers hesitate to provide full access to their datasets due to a lack of knowledge about research data management (RDM) tools and perceived fears, such as losing the value of one's own data. Existing tools and approaches often do not take into account these fears and missing knowledge. In this study, we examined how conversational agents (CAs) can provide a natural way of guidance through RDM processes and nudge researchers towards more data sharing. This work offers an online experiment in which researchers interacted with a CA on a self-developed RDM platform and a survey on participants' data sharing behavior. Our findings indicate that the presence of a guiding and enlightening CA on an RDM platform has a constructive influence on both the intention to share data and the actual behavior of data sharing. Notably, individual factors do not appear to impede or hinder this effect.

Keywords: Open science, research data management, digital nudging, conversational agents.

1. Introduction

Pre-registered reports, data exchange and the transparency of how results were achieved has become an important factor for the success of research processes (Doyle & Luczak-Roesch, 2020). According to FAIR data principles, data should be findable, accessible, interoperable, and reusable (Stieglitz et al., 2020). In this way, studies can be replicated and the quality of science can improve (Heise & Pearce, 2020). Researchers often perceive barriers associated with data sharing and the organization and preparation of data is often time-consuming and costly (Hofeditz et al., 2020; Tenopir et al., 2011). Research often involves decentralized data management and no explicitly defined research workflows (Ince et al., 2020).

Publications often lack transparency, as full data sets are not always published (Nosek et al., 2015) or studies were not pre-registered (Doyle & Luczak-Roesch, 2020). Overall researchers are willing to share their data when the perceived benefits not only outweigh the disadvantages of data sharing, but also when additional uncertainty factors (such as the fear of losing one's unique value, a fear of data misuse, and the fear of competition, pseudonymization requirements imposed by ethics boards), can be mitigated (Stieglitz et al., 2020). When researchers do not share their data, this can negatively impact knowledge sharing, transparency, and reproducibility of research findings (Perrier et al., 2017). Recent research emphasizes that an RDM platform should prevent exploitative use (Sivizaca Conde et al., 2023).

For an increasing number of journals, fundings, and researchers, free sharing of scientific knowledge to the public – which is also known as open science (OS) practice – is becoming a basic requirement (Doyle, 2021; Link et al., 2017). OS approaches embedded in information technology (IT) can be both a driver for and a result of innovation (Schlagwein et al., 2017). However, these technologies tend not to consider that researchers often have very inadequate knowledge about research data management (RDM) and OS and that, in addition, various fears discourage researchers from sharing their data.

As one type of innovative technologies conversational agents (CAs) can serve as a natural way to guide people through processes and answer questions (Hofeditz et al., 2022; Stieglitz et al., 2022). CAs can be represented as chatbots or virtual agents (Quarteroni, 2018) and assist users in several tasks such as work related organizational tasks (Mirbabaie et al. 2021). The CAs' potential has also been considered in OS practice (e.g., to provide relevant metadata for researchers) (Langer et al., 2022). Tools that aim to manage research

URI: https://hdl.handle.net/10125/106457
978-0-9981331-7-1

data tend to be difficult to use for researchers who are less technically proficient (Hofeditz et al., 2020). CAs can not only make technically complex processes easier through natural language interaction but are also able to address individual researcher needs (e.g., lack of self-discipline).

One approach to improve the perceived value of sharing research data could be the application of digital nudges which are information and design elements (Weinmann et al., 2016). Nudging can influence the behavior of people in a predictable way while preserving freedom of choice (Thaler & Sunstein, 2008) and can also be used to guide people in a certain direction. CAs not only offer the possibility of a natural interaction (Rietz et al., 2019) but also provide a way to implement nudges (Bawa et al., 2020). We therefore derived the following research question (RQ):

RQ: *How can digital nudges in a CA-based conversation improve researchers' data sharing behavior?*

To answer the research questions, we conducted an online experiment with 60 researchers from several disciplines. The participants needed to complete an OS task following a typical research workflow in which they interacted with a CA. The CA used different textual nudges in the dialog with the researchers while providing guidance for the OS task. The participants were asked to answer questionnaires on their data sharing intentions before and after the task.

We aimed to provide knowledge on how the intended use of RDM tools during the research workflow can be positively influenced by using guiding CAs that apply textual nudging elements.

2. Literature background

2.1 Open science practices and data sharing

Data exchange, which is necessary for scholarly research activities, has been based on long-established, trusted sources and channels (Wilms et al., 2018). However, due to digitalization, additional new channels and platforms have emerged such as commercial and institutional cloud services or open access and OS data repositories (e.g., Zenodo or Open Science Framework) (Hofeditz et al., 2020). In an increasingly digital environment, many factors influence how academic researchers decide what to read, what to cite, where to publish their work and where to share their research data (Stieglitz et al., 2020). In certain fields such as science and engineering, research becomes more data-intensive, making data sharing and reusing a critical topic of scholarly communication (Theobold et al., 2020). By reusing and sharing larger amounts of various scientific data, researchers can provide meaningful insights that would never be possible without interdisciplinary exchange of knowledge (Betz et al., 2019). Overall, the digital advancement of scholarly communications has allowed researchers to share their expertise, discoveries, and raw data sets in less limited ways than in the past and holds potential for great scientific breakthroughs. But despite the possibilities, data sharing and OS practices are not as established as it could be among researchers (Stieglitz et al., 2020).

Although the idea of OS is promising and can enhance scientific discovery (Stieglitz et al., 2020), the involved barriers such as fears have not been well understood. To better understand these barriers, researchers' attitudes toward sharing their data need to be considered (Kim & Zhang, 2015). Researchers often decide not to share their datasets before and after a publication as it requires additional effort on top of their actual research to organize and prepare their data. Perceived career benefits and risks, perceived effort, and perceived availability of data repositories have significant impacts on researchers attitudes toward data sharing (Kim & Zhang, 2015). On the one hand, studies found that researchers are reluctant to share their data because of perceived risks, including concerns about losing publication opportunities and the exclusive rights to their data (Kim, 2012; Stieglitz et al., 2020). On the other hand, researchers seem to increasingly share their data when they perceive more career benefits than risks (Stieglitz et al., 2020). We therefore hypothesized:

H1: Perceived career benefits positively affect the attitude towards data sharing behavior.

H2: Perceived risks negatively affect the attitude towards data sharing behavior.

2.2 Guiding users through conversational agents

There are different options to provide support to a user by innovative technologies. CAs can support users in several tasks by providing a natural way of interaction (Diederich et al., 2019). The function as an assistant has already been investigated in many contexts. For example, a CA can be used to support students in campus life (Dibitonto et al., 2018). Feine et al. (2019) define CAs as "a software-based system designed to interact with humans using natural language." (p. 6). CAs have also been tested in higher education to support students in providing information on attendance at

lectures, grades, and financial records (Heryandi, 2020). They can also be used in collaborative learning to guide students (Kumar & Rosé, 2011). Therefore, we assume that CAs could be used in academia to assist researchers as they provide a high level of user satisfaction and user experience (Diederich et al., 2019).

There are several ways to influence and guide researchers towards a certain desired behavior. For example, social cues of a CA can have a significant influence on the user's behavior towards the CA. Social cues can be defined as "a cue of a CA that triggers a social reaction of the user towards the CA" (e.g., emotional or behavioral reaction) (Feine et al., 2019). Due to these social reactions through the application of social cues we assume the following:

H3: Applying a CA with social cues positively affects the intention to use RDM tools.

Furthermore, previous studies suggested considering individual factors such as technology affinity (i.e., "tendency to actively engage in intensive technology interaction"; Franke et al., 2019, p. 456) in the context of CAs (Sohn, 2019) and the intention to use CAs (Lee et al., 2020). We therefore assume the following:

H4: Technology affinity mediates the influence of a CA with social cues on intention to use RDM tools.

Our goal is not only to guide researchers through CAs, but also to mitigate uncertainty factors. This requires a suitable strategy that does not take away users' freedom by being too restrictive (which would be contrary to the OS idea). Thaler and Sunstein (2008) introduced the concept of nudging people to make better decisions that benefits themselves or a society. One of these decisions can be the sharing of research data which can benefit both research and society. The concept of nudging aims to make marginal changes in the choice environment to guide people in a specific direction by using or overcoming psychological effects (Mirsch et al., 2017). Nudging keeps every choice option when changing the choice environment, meaning that no option can simply be excluded (Thaler & Sunstein, 2008).

The concept of nudging originally takes place in a non-digital environment but can be transferred to a digital environment. "Digital nudging is the use of user-interface design elements to guide people's behavior in digital choice environments" (Weinmann et al., 2016). Kissmer et al. (2018) stated that digital nudging can positively influence adoption of, and therefore collaboration through, new technologies. As one of those new technologies, CAs can provide guidance and information though a user friendly way of communication (Diederich et al., 2019). Previous research already showed that CAs can use digital nudging (Ali Mehenni et al., 2021). We therefore derived the following hypotheses:

H5: Applying digital nudges on an RDM platform via CAs positively impact the intention to use RDM tools.

H6: Applying digital nudges on an RDM platform via CAs positively impact attitudes toward data sharing behaviors.

3. Research design

To determine how CAs can be use digital nudging that affects the intention to use RDM tools and their data sharing behavior, we conducted an online experiment. As participants, we recruited experienced researchers from various disciplines and career levels via email and social media. The study was aimed at people who work in research at e.g., a university (PhD candidates, post-docs, and professors) and are therefore familiar with the activities during a research project along a research workflow.

The online experiment can be divided into three parts. First, a questionnaire consisting of demographic data and questions about research were requested. Following, questions about data sharing attitude, data sharing behavior, career cost and career benefit were asked (Kim & Zhang, 2015). Responses to all questionnaires were recorded using a 7-point Likert scale. The previous questionnaires were followed by a task for the researchers in the second part of the study.

Participants should gather their own experience in using a self-developed RDM platform during the simulated execution of a research project. To make this possible, we specified a typical RDM workflow which the participants should follow on the platform. We used the literature review by Ince et al. (2020) as orientation. In addition, we studied various research processes in different scientific fields and derived a general workflow that is divided into four steps. Accordingly, the research process starts with "information literacy", where the relevant information needs to be found and evaluated by looking at different sources. This is followed by "information management", which focuses on the ordering and sorting of the identified literature. The ordered information is then transformed into knowledge in "knowledge management". This was done, for example, by making sense or performing analysis, either alone or in collaboration. Finally, "scholarly communications" focuses on writing,

reviewing, and publishing or sharing papers and other scientific outputs. Based on this workflow, we developed a platform prototype called science base (Figure 1). We used bubble, a service for the creation of web applications, to develop the platform[1]. On this platform, users could perform tasked related to all steps previously described.

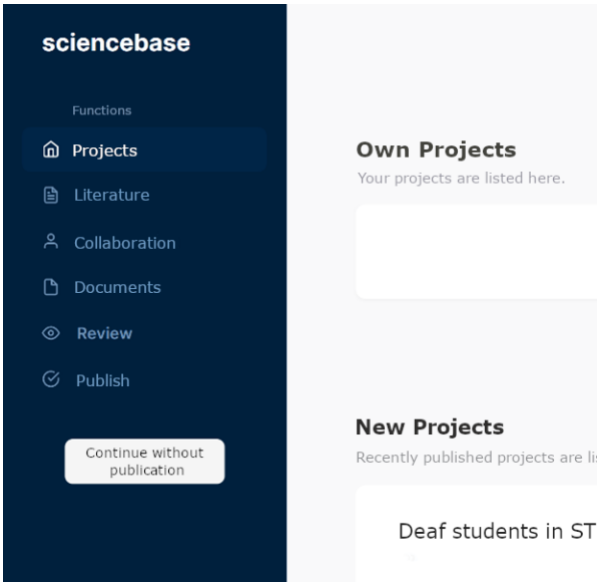

Figure 1. The developed sciencebase platform showing the nudge condition.

To guide researchers through the platform, we developed two CAs using Google dialogflow[2]. The CAs provided basic information about the tasks according to the workflow and the structure of the platform (Figure 2). To support social interaction with the CAs, we implemented social cues according to Feine et al. (2019). The CAs could respond to small talk, send emoticons, or provide self-disclosure on certain topics.

On the platform, participants were asked to conduct a literature review as this is relevant to most disciplines. Hence, the data to share includes bibliographic information about the articles and the created structure. The task was only roughly explained to encourage the CA's assistance to complete the task. At this point, we divided the researchers into two random, equally distributed groups. One group interacted with a CA which tried to actively guide and nudge the participants towards advanced data sharing and OS practices, and one group completed the task with the help of a very simple chatbot that just provided answers to functions of the platforms.

To realize a CA that can nudge researchers to share research data, we considered the target audience and the OS practices: In fact, the nudges are intended to benefit researchers and are designed to encourage OS practices such as data sharing. Consequently, nudging has the potential to simplify the use of an RDM tool and thus support researchers in their work. We integrated six text-based nudges based on the previously gathered knowledge into the vocabulary of the CA. First, the 'informing' nudge presented the benefits of an RDM tool while asking the user to create a project in the tool (Meske et al., 2019). Second, the 'loss aversion' nudge presents disadvantages if a user does not share the literature, he/she found via the search on our platform (Wijland et al., 2016). While the users were supposed to work collaboratively and invite collaborators to his project, we presented the third 'social influence' nudge. Social influence is the process by which external inputs lead individuals to change their attitudes and behaviors (Malhotra & Galletta, 1999). We pointed out how important it is to collaborate and that collaboration is a commonly accepted success factor among researchers (Kroll et al., 2019; Schneider et al., 2019). Fourth, we divided the nudge 'pre-commitment' into two parts. First, the users were asked if there is a general interest in a review when they created a project. This should lead the users to be more likely to submit their work to a review process later in the project. As the user confirms his/her general interest in a review, this laid the foundation for the actual review process. We assumed that it would contradict the theory of cognitive dissonance if a user did not want to have their work reviewed, even though they had initially shown interest (Luoto et al., 2014). Another nudge which is frequently used in the literature is 'default setting' that utilizes the status quo bias (Kahneman et al., 1991). In the last step of the workflow, the user should be encouraged to share all publication-relevant data when publishing his/her work. For this purpose, the CA pointed out that all publication-relevant data are shared by default. The user had to actively deselect this function. In addition, the option was preselected in the tool. How this nudge was realized through the CA is visualized in Figure 2. A default setting was also used when requesting a review.

During the simulation of the research project, variables were recorded, which contained information about, for example, the publishing of the paper or other research data. After completing the tasks on the platform, we guided the researchers into the third part of

[1] https://bubble.io/

[2] https://cloud.google.com/dialogflow

the study, where further questionnaires should be completed. The questions referred to the attitude towards and perception of the CA, the attitude towards data sharing, the technology affinity (Franke et al., 2019) as well as the perception of nudging elements. We covered the attitude toward and perception of the CA by questions from Malhotra and Galletta (1999), supplemented by a question on general perception of a CA by Novielli et al. (2010).

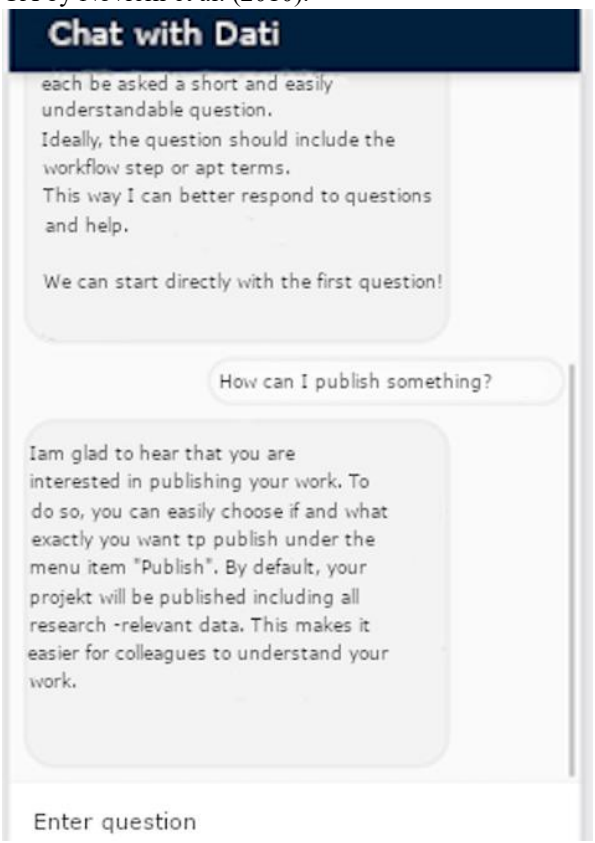

Figure 2. Guiding CA showing the application of the nudge 'Default Setting'

Data sharing attitude was again queried accordingly by Kim and Zhang (2015), to ensure that a comparison of the data before and after the task in part two of the study was possible. The question on technology affinity was asked according to Edison and Geissler (2003). The group with the nudging elements in the task was additionally offered a picklist with the individual nudging elements and their occurrence on the platform. The awareness of the individual nudges was measured here, so the participants should reflect on which nudging elements they perceived. According to this information, a modified question on usefulness from the technology acceptance model (TAM) (Malhotra & Galletta, 1999) was asked for each remembered nudging element, which was intended to measure the improved use of the platform due to the nudging elements. We adapted the TAM related questions with regard to the intention to use to use the RDM platform (Malhotra & Galletta, 1999). After the exclusion of 7 respondents who did not answer the attention checks correctly, we had a final sample size of $N = 60$ from different career levels. Of the researchers, 29 were male, 30 were female and 1 was diverse. Most of the participants were between 25 and 35 years old. The professions that the participants occupy show the reason for the tendency toward young age ($M = 30.2$). Research associates (82 %) were the main participants in the experiment. The remaining participants consisted of 3% professors, 8% postdocs or senior lecturers and 7% of other types of researchers. The researchers were from 9 main disciplines. Nevertheless, 3 areas emerge for which our experiment seemed to be of particularly high interest: Psychology (30 %), Computer Science (28 %), and Human Computer Interaction (17 %). We published our dataset via OSF[3].

4. Findings

We used a paired t-test to investigate the impact of the CAs on researchers' attitude towards data sharing indicating that the difference in mean attitudes toward data sharing before the intervention ($M = 5.4$; $SD = 1.11$) and attitudes toward data sharing after the intervention ($M = 5.64$; $SD = 1.13$) was significant ($t\ (59) = –2.34$, $p = 0.023$). Further, we calculated a linear regression to investigate a possible reason for the improved attitude. In this case, whether the participants were in the group with or without guiding and nudging was the predictor (i.e., we introduced a dummy variable) of how the data sharing behavior of the researchers would turn out. The results show that the condition nudging or no nudging has an impact on how positive the attitude towards data sharing is, the regression model is significant ($F\ (1, 58) = 10.8$, $p = .002$). If switching from the nudging group to the no nudging group, attitudes toward data sharing would decrease by 0.89 units on the 7-point Likert scale ($\beta = –.89$, $t\ (58) = –3.28$, $p = .002$). 14.2% of the variance in attitude toward data sharing is explained by the presence of nudging elements ($R^2_{adjusted} = .1429$). According to Cohen (1988) this is a large effect ($f^2=0.86$). Corresponding to the participants' statements about their attitude towards data sharing, their behavior in sharing the data during the intervention also emerged.

[3]https://osf.io/d5jzv/?view_only=e3a540fec52947db8c8d14060c9e281d

This behavior shown on the platform was recorded as an additional variable. In the group with guidance and nudging, 14 participants shared all data and only 3 participants shared only the manuscript, while in the group without nudging elements, 9 participants shared all data and 8 participants shared only the manuscript. The remaining people in the respective groups did not publish anything.

Furthermore, we wanted to investigate to what extent a CA that provides guidance and that aims to nudge researchers influences the intention to use RDM tools. Again, we calculated a linear regression. The nudging vs. no nudging condition as a dummy variable represented the predictor for the intention to use data management tools. However, no significant influence could be shown here: $R^2_{adjusted} = .01$, $F(1,58) = .62$, $p = .436$. The descriptive results suggest minimal differences in mean values between the nudged and non-nudged groups regarding their intention to use RDM tools. Finally, the individual nudging elements were examined more closely. The nudging elements were analyzed descriptively and ranked according to their average usefulness, which can be seen in Table 1 (the two further nudges were not considered useful by the participants). Overall, on average, an improvement in RDM could be perceived from the four nudging elements ($M = 4.22$).

To identify influencing factors on data sharing behavior and the attitude towards RDM, we calculated additional regressions. First linear regressions were calculated, looking at two different predictors of the attitude towards data sharing. Both career benefits ($R^2_{adjusted} = .002$, $F(1,58) = 1.17$, $p = .283$) and career risks ($R^2_{adjusted} = .055$, $F(1,58) = 4.45$, $p = .039$) showed no significant effect on the attitude towards data sharing.

Table 1. Ranking of the nudging elements

	n	M	SD
1. Default Setting	10	5.25	0.69
2. Social Influence	5	4.57	1.53
3. Informing	11	4.56	1.26
4. Pre-Commitment	15	4.47	1.65

Furthermore, a mediation was calculated between the perception of the CA and the intention to use the RDM tool with the affinity for technology as a mediator. The mediation by technology affinity did not show significant results. Nevertheless, there is a direct significant effect of the perception of the CA on the intention to use the RDM tool (Figure 3).

Table 2. Summary of the hypotheses

H1: career benefits ↑ data sharing	Not supported
H2: risks ↓ data sharing	Not supported
H3: CA ↑ use RDM tools	Supported
H4: tech. affinity mediates H3	Not supported
H5: Nudges ↑ use RDM tools	Not supported
H6: Nudges ↑ data sharing	Supported

As indicated in the mediation, the perception of the CA as a predictor of intention to use the RDM platform led to a significant regression model with a $p < .001$ ($F(1,58) = 19.3$, $p < .001$). Accordingly, a one unit increase

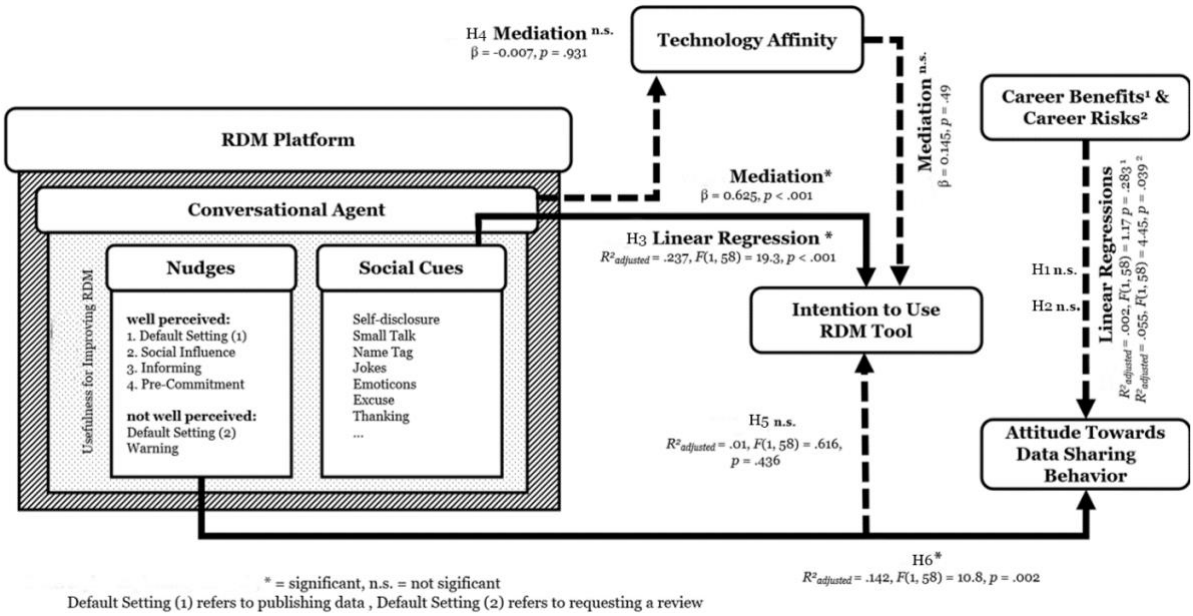

Figure 3. Digital Nudging and Conversational Agents (DNA) research model

in positive perception of the CA would also increase the intention to use the RDM platform by 0.63 units (β = -.626, t (58) = 4.40, p = .002). 23.7% of the variance of the intention to use the RDM platform can be explained by the perception of the CA ($R^2_{adjusted}$ = .237). According to (Cohen, 1988) there is a large effect here with an f^2 = .557. Our findings are summarized in Table 2.

5. Discussion

Our findings indicate that a guiding CA that uses text-based nudging influences data sharing attitudes of researchers and leads to an increased willingness to share research data. Moreover, a CAs offer the possibility of a user friendly and satisfying way to guide and nudge researchers toward more data sharing and an increased intention toward RDM. Previous findings already indicated that a CA could be suitable for integrating nudging elements into a platform (Ali Mehenni et al., 2021). Additionally, our analysis highlights that the perception of the CA positively affects the intention to use the RDM platform, which confirms other studies that found an influence of CAs and the promotion of positive behavioral changes and increased participation engagement (Scholten et al., 2017; Wang et al., 2018). Thus, nudging through a CA can be used to mitigate the problem of reluctance to share academic data and positively influence researchers consequently strengthening the OS movement to improve interdisciplinary research. However, of course within the boundaries of ethical and legal limitations (e.g., GDPR) (Tenopir et al., 2011).

We could not confirm an impact of perceived career risks and benefits of researchers on data sharing behavior and data sharing attitudes. This leads to the assumption that career risks and benefits associated with research data sharing might become less important for scholars (except if they are bound by legal restriction). According to Kim and Zhang (2015) attitudinal beliefs such as perceived career benefits and risks should have significant impact on researchers' attitudes regarding data sharing. Our findings could indicate an attitude shift in researchers' perceptions towards OS, as the acceptance and reputation has grown over the last years. A more recent study of social scientists found that hopes concerning the benefits of OS significantly exceeded concerns (Abele-Brehm et al., 2019). This, in combination with our findings, gives reason to believe that today there are generally more hopes associated with OS and data sharing, which overshadow individual beliefs regarding one's career.

Despite the great potential, RDM tools are not widely used yet and there is lack of perceived availability and standardized platforms (Kim & Zhang, 2015). This could also be a reason why not much is explored in this field yet, especially regarding the combination of CAs and digital nudging (Ince et al., 2020). Our results add to existing design principles (cf. Sivizaca Conde et al., 2023) by showing how such a platform could be designed and how CAs can be used to assist researchers on such platforms. We further showed how a CA can influence researchers' intention to use RDM tools.

Some research has shown that the default setting can be used as an effective nudge (Meske et al., 2019; Schirrmacher et al., 2019), other work suggests that the default setting could backfire, as some users have concerns regarding their privacy (Schneider et al., 2019). We could show that some nudges are more useful than others, regarding influencing researchers' data sharing behavior within an RDM tool. Signifying that indeed the default setting to share data was the most useful nudge on our platform, confirming the already existing work on positive influence of this nudge.

Although previous research suggested to examine the impact of individual factors such as the influence of technology affinity on the intention to use a CA (Sohn, 2019), our findings did not support this assumption. We found that technology affinity did not mediate the effect of our CA on the intention to use our RDM platform. Thus, we conclude that CAs can be used independently of individual technology expertise to positively influence researchers' attitudes toward data sharing.

To provide an overview of our research design and to support understanding of the results, we present our Digital Nudging and Conversational Agents (DNA) research model in Figure 3. We mapped the structure of each element from the online experiment. Our findings strengthen the suggestion that a CA is capable of nudging users (Ali Mehenni et al., 2021), especially by applying the default setting and social influence nudges towards more data sharing and a better intention towards OS and RDM.

6. Contributions to research and practice

With this work, we contribute to IS research by providing a way of practicing OS in an empirical study to encourage data sharing and a positive intention towards RDM among researchers. We have gained initial knowledge on the application of CAs that guide and nudge researchers to influence their behavior

without violating the freedom of choice. Our DNA model outlines the relationships among constructs of CAs and nudging in the OS context and integrates empirical evidence regarding RDM use and data sharing. It offers an approach for both researchers and platform developers for integrating digital nudges into the dialog vocabulary of a CA to facilitate RDM. We therefore provided knowledge on the understanding of how a platform that aims to benefit OS practices can be designed or even complemented by a CA. By this, hindering factors for sharing research data such as individual fears or uncertainty factors can be mitigated.

Practicing OS affects almost all research disciplines and can contribute not only to an increase in the quantity of data sets, but also to an increase in quality through more interdisciplinary exchanges. With our work, we bridged different IS research topics such as designing and examining collaboration systems and technologies (such as digital RDM platforms), collaboration with CAs, and research on behavior intention and digital nudging.

Our findings suggest that individual factors such as fears and uncertainty in the context of data sharing – which have been shown in previous studies (e.g., Hofeditz et al., 2020; Stieglitz et al., 2020) – seem not to impact the increased data sharing behavior and intention towards RDM in our study. This may indicate that the use of CAs that guide researchers through platforms for managing and publishing research data (such as Zenodo or OSF) is not only suitable for people with a high affinity for technology, as it is the case on many RDM platforms and tools. The use of a user-friendly platform and a guiding CA that gently nudges researchers into living OS practices thus holds great potential to stimulate interdisciplinary exchange between researchers.

We could also show that a CA with social cues can positively affect the intention to use RDM platforms. Moreover, our research allowed us to gain initial insights into the effectiveness and integration of nudges as part of a CA's vocabulary as a user friendly and satisfied way of user interaction in academia. For this, the default setting and social influence seems to be the most promising digital nudges. We highly recommend implementing these nudges in real world scenarios on RDM platforms.

Our findings also bring further implications for practice. CAs are often being used as a simple tool to answer users' question on a website (Cui et al., 2017). We showed how they can provide more active guidance and texted-based digital nudging through a CA. This type of CAs could be applied on already running RDM platforms such as Zenodo or OSF or university cloud storages such as ownCloud or Nextcloud. The integration of a guiding CA with nudges could potentially lead to an increased acceptance and use of RDM tools and thus in the long term, the willingness to share data could be increased. To summarize, our research and findings support the OS movement and will hopefully contribute to increase research quality.

7. Limitations and future research

Our research comes with some limitations. First, one of the most important challenges was the small sample size. Our requirement to recruit only experienced researchers and cover many different disciplines limited our sample. In addition, the participation took some time and could not be completed in less than 30 minutes. We recommend that future research further examine the effects of CAs and nudging on RDM as an important OS practice. It should also be mentioned that there might be interaction effects between the different nudges. In a future study, the study should be conducted with more participants and a CA that only uses one specific nudge. Another limitation is that the CA and platform are only a prototype. Future research should test the effectiveness of a guiding CA that uses text-based nudging in a real-world environment on an already established RDM tool. In this line, researchers can examine how nudges lead to an increased sharing intention and why individuals decide for or against sharing. Ultimately, for science to benefit from open data, research can study the quality of data shared. Furthermore, possible cultural differences are to be mentioned as a limitation. As this was one of the first studies in this research area, we did not include cultural differences in the research. In a future study, the influences of different cultures or countries should be considered. In addition, since RDM tools are not universally adopted, a possible information overload should be considered. Further research needs to test our findings in practice by examining more advanced RDM platforms and CAs with a higher level of sophistication. Moreover, it would be beneficial to develop trainings for researchers to learn about OS practices (e.g., data sharing, crafting metadata).

8. Conclusion

In this study, our findings suggest that CAs can guide and nudge researchers on an RDM platform

towards more data sharing, a more positive attitude towards data sharing and an intention to use RDM tools to practice OS. We suggest that a CA that uses the text-based nudges default setting, social influence, informing and pre-commitment can lead to a more positive attitude towards data sharing and RDM in general. Guiding a CA that nudges researchers differed from a CA that was limited to general assistance on an RDM platform regarding the impact on researchers' data sharing behavior. Using a CA with social cues led to an increase on the intention to use an RDM platform. Individual factors such as technology affinity did not affect this intention. Therefore, we provide an approach to overcome challenges for practicing OS such as researchers' ignorance, uncertainty, and fears.

9. References

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes Toward Open Science and Public Data Sharing: A Survey among Members of the German Psychological Society. In *Social Psychology*.

Ali Mehenni, H., Kobylyanskaya, S., Vasilescu, I., & Devillers, L. (2021). Nudges with Conversational Agents and Social Robots: A First Experiment with Children at a Primary School. In L. F. D'Haro, Z. Callejas, & S. Nakamura (Eds.), *Conversational Dialogue System for the Next Decade* (pp. 257–270). Springer Nature.

Bawa, A., Khadpe, P., Joshi, P., Bali, K., & Choudhury, M. (2020). Do Multilingual Users Prefer Chat-bots that Code-mix? Let's Nudge and Find Out! *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1–23.

Betz, U. A. K., Betz, F., Kim, R., Monks, B., & Phillips, F. (2019). Surveying the future of science, technology and business – A 35 year perspective. *Technological Forecasting and Social Change*.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A Customer Service Chatbot for E-commerce Websites. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-System Demonstrations*, 97–102.

Dibitonto, M., Leszczynska, K., Tazzi, F., & Medaglia, C. M. (2018). Chatbot in a Campus Environment: Design of LiSA, a Virtual Assistant to Help Students in Their University Life. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10903 LNCS* (pp. 103–116). Springer International Publishing.

Diederich, S., Brendel, A. B., Lichtenberg, S., & Kolbe, L. (2019). Design for fast request fulfillment or natural interaction? Insights from an experiment with a conversational agent. *European Conference on Information Systems*, 1–17.

Doyle, C. (2021). How can I share my work? A review of the open access policies of IS journals. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 660–669.

Doyle, C., & Luczak-Roesch, M. (2020). This paper is an artefact: On open science practices in design science research using registered reports. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 5026–5035.

Edison, S. W., & Geissler, G. L. (2003). Measuring attitudes towards general technology: Antecedents, hypotheses and scale development. *Journal of Targeting, Measurement and Analysis for Marketing*.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human Computer Studies*, *132*(September 2018), 138–161.

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467.

Heise, C., & Pearce, J. M. (2020). From Open Access to Open Science: The Path From Scientific Reality to Open Scientific Communication. *SAGE Open*, *10*(2), 1–14.

Heryandi, A. (2020). Developing Chatbot for Academic Record Monitoring in Higher Education Institution. *IOP Conference Series: Materials Science and Engineering*, *879*(1).

Hofeditz, L., Mirbabaie, M., Erle, L., Knoßalla, E., & Timm, L. (2022). Automating Crisis Communication in Public Institutions – Towards Ethical Conversational Agents That Support Trust Management. *Wirtschaftsinformatik 2022 Proceedings*, *3*.

Hofeditz, L., Ross, B., Wilms, K., Rother, M., Rehwald, S., Brenger, B., López, A., Vogl, R., & Rudolph, D. (2020). How to design a research data management platform? technical, organizational and individual perspectives and their relations. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12185 LNCS*.

Ince, S., Hoadley, C., & Kirschner, P. A. (2020). Research workflow skills for education doctoral students and postdocs: A qualitative study. *Journal of Academic Librarianship*, *46*(5), 102172.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives*, *5*(1), 193–206.

Kim, Y. (2012). Factors influencing STEM researchers' data sharing behaviors. *Proceedings of the ASIST Annual Meeting*.

Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library and Information Science Research*, *37*(3), 189–200.

Kissmer, T., Potthoff, T., & Stieglitz, S. (2018). Enterprise Digital Nudging: Between adoption gain and unintended rejection. *Proceedings of the Twenty-*

Fourth Americas Conference on Information Systems, 1–5.

Kroll, T., Paukstadt, U., Kreidermann, K., & Mirbabaie, M. (2019). Nudging People to Save Energy in Smart Homes With Social Norms and Self-Commitment. *Proceedings of the 27th European Conference on Information Systems*, 1–10.

Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, *4*(1), 21–34.

Langer, A., Schmolke, L., & Gaedke, M. (2022). BIRD : Using Conversational User Interfaces to Provide Relevant Metadata for Interdisciplinary Research Data Publishing. In *E-Science-Tage 2021. Share Your Research Data.* (pp. 326–332).

Link, G. J. P., Lumbard, K., Conboy, K., Feldman, M., Feller, J., George, J., Germonprez, M., Goggins, S., Jeske, D., Kiely, G., Schuster, K., & Willis, M. (2017). Contemporary issues of open data in information systems research: Considerations and recommendations. *Communications of the Association for Information Systems*, *41*(1), 587–610.

Luoto, J., Levine, D., Albert, J., & Luby, S. (2014). Nudging to use: Achieving safe water behaviors in Kenya and Bangladesh. *Journal of Development Economics*, *110*, 13–21.

Malhotra, Y., & Galletta, D. F. (1999). Extending the Technology Acceptance Model to account for social influence: Theoretical bases and empirical validation. *Proceedings of the Hawaii International Conference on System Sciences*, *00*(c), 5.

Meske, C., Amojo, I., Poncette, A.-S., & Balzer, F. (2019). The Potential Role of Digital Nudging in the Digital Transformation of the Healthcare Industry. In A. Marcus & W. Wang (Eds.), *Design, User Experience, and Usability. Application Domains* (pp. 323–336). Springer International Publishing.

Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., & Frick, N. R. J. (2021). Understanding Collaboration with Virtual Assistants – The Role of Social Identity and the Extended Self. *Business and Information Systems Engineering*, *63*(1), 21–37.

Mirsch, T., Lehrer, C., & Jung, R. (2017). Digital Nudging: Altering User Behavior in Digital Environments. *Proceedings of 13th International Conference on Wirtschaftsinformatik*, 634–648.

Nosek, B. A., Alter, G., Banks, G. C., et al. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Novielli, N., de Rosis, F., & Mazzotta, I. (2010). User attitude towards an embodied conversational agent: Effects of the interaction mode. *Journal of Pragmatics*, *42*(9), 2385–2397.

Quarteroni, S. (2018). Natural Language Processing for Industry: ELCA's experience. *Informatik-Spektrum*, *41*(2), 105–112.

Rietz, T., Benke, I., & Maedche, A. (2019). *The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance*.

Schirrmacher, N.-B., Ondrus, J., Ter Chian Tan, F., Ai-Chi Loh, Y., & Hardoon, D. R. (2019). *Overcoming Status Quo Bias : Nudging in a Government-Led Digital Transformation Initiative*. *August 2017*, 1–8.

Schlagwein, D., Conboy, K., Feller, J., Leimeister, J. M., & Morgan, L. (2017). "openness" with and without Information Technology: A framework and a brief history. *Journal of Information Technology*, *32*(4), 297–305.

Schneider, D., Klumpe, J., Adam, M., & Benlian, A. (2019). Nudging users into digital service solutions. *Electronic Markets*.

Scholten, M. R., Kelders, S. M., & Van Gemert-Pijnen, J. E. W. C. (2017). Self-guided Web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of Medical Internet Research*, *19*(11), 1–19.

Sivizaca Conde, D., Rößler, D., Kliewer, N., & Stegemann, L. (2023). Enabling Data-Driven Mobility Research: Design Principles and Design Features for an open Platform approach. *Proceedings of the 56th Hawaii International Conference on System Sciences*, 532–542.

Stieglitz, S., Hofeditz, L., Brünker, F., Ehnis, C., Mirbabaie, M., & Ross, B. (2022). Design principles for conversational agents to support Emergency Management Agencies. *International Journal of Information Management*, *63*, 102469.

Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., Lopez, A., & Rehwald, S. (2020). When are researchers willing to share their data ? – Impacts of values and uncertainty on open data in academia. *PLoS ONE*, *15*(7), 1–20.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6), e21101.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Theobold, A. S., Hancock, S. A., & Mannheimer, S. (2020). Designing Data Science Workshops for Data-Intensive Environmental Science Research. *Journal of Statistics Education*, *December*, 1–31.

Wang, H., Zhang, Q., Ip, M., & Fai Lau, J. T. (2018). Social Media-based Conversational Agents for Health Management and Interventions. *Computer*, *51*(8), 26–33.

Weinmann, M., Schneider, C., & vom Brocke, J. (2016). Digital Nudging. *Business and Information Systems Engineering*, *58*(6), 433–436.

Wijland, R., Hansen, P., & Gardezi, F. (2016). Mobile nudging: Youth engagement with banking apps. *Journal of Financial Services Marketing*, *21*(1), 51–63.

Wilms, K., Brenger, B., Lopez, A., & Rehwald, S. (2018). Open Data in Higher Education – What Prevents Researchers from Sharing Research Data? *39th International Conference on Information Systems (ICIS) 2018*, 1–9.