DEVELOPMENT AND EVALUATION OF AN ACTIVE LEARNING TOOL FOR DATA DECISION MAKING

Clara Llebot Lorente, Hui Zhang, and Michael Boock*

Three library faculty at Oregon State University (OSU), a research intensive university, created the Data Sharing Wizard (the Wizard), a Scenario-Based, online, interactive learning tool that helps students learn about the sharing of research datasets with a focus on intellectual property and decision making. Using the Wizard, students are able to identify: 1. appropriate documents, laws, and policies pertaining to the sharing and use of research datasets; 2. who has a say in their use; and 3. additional resources that can help answer questions at OSU. In this paper, the authors discuss the design and development of the Wizard, share the results of a usability study, and discuss the lessons learned and how other libraries can adopt and customize the Wizard.

INTRODUCTION

As research data are increasingly gathered and processed digitally, in more easily shareable formats, the goal of sharing and reusing existing data has gained momentum over the last decade, particularly in the sciences. Building on years of work to enhance data sharing in the European Union, the European Commission promotes the use of data as “an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general (European Commission 2017).” The Commission describes data stewardship and sharing—Open Science—as “a move towards better science, to get more value out of our investment in science and to make research more reproducible and transparent (European Commission. Directorate General for Research and Innovation 2016).” Research libraries and librarians, historically interested in the broadest possible access to and distribution of all forms of information and research, work closely with researchers to understand copyright in the context of research data,

*Clara Llebot Lorente, Data Management Specialist, Oregon State University, email: clara.llebot@oregonstate.edu; Hui Zhang, Digital Services Librarian, Oregon State University, email: hui.zhang@oregonstate.edu; Michael Boock, Scholarly Communication Librarian, Oregon State University, email: michael.boock@oregonstate.edu
help make their research available to the world, and help them understand what they can and cannot do with research datasets. Questions from researchers in this sphere are often complex, case-dependent, and involve multiple research partners with varying levels of responsibility and minimal expertise. Student researchers, especially, suffer from a lack of guidance, resources, and norms pertaining to questions of data sharing and reuse (Carlson et al. 2011; Doucette and Fyfe 2013, 168). International students are also affected, as they must adapt to a new language, a new educational system, and different copyright and data sharing norms from those of their home countries. Early career faculty, post-graduate researchers, and research assistants quite often do not have the knowledge to help their students with questions about data ownership, rights, and obligations (Carlson et al. 2011).

To solve this problem, three library faculty at Oregon State University, a research intensive university, created the Data Sharing Wizard (Oregon State University 2020), hereafter the Wizard, a Scenario-Based active learning tool that offers an interactive approach to help researchers learn about copyright for research datasets. Using the Wizard, researchers are able to identify:
1. Appropriate documents, laws, and policies pertaining to the sharing and use of research datasets.
2. Who has a say in their use?
3. Additional resources that can help answer questions.

In this paper, the authors discuss the design and application of a usability study to determine the Wizard’s effectiveness in answering these questions for researchers and suggest how the tool may be adopted by other libraries. The usability study is based on the following assumptions. If the Wizard is effective as a learning tool, users: 1) Will become aware of new concepts related to data use, and 2) Will detect and correct their misconceptions.

**METHODS**

*Design and Development of the Wizard*

The Wizard is designed as a single-page web application (SPA). A SPA has one page that users will visit, with no need for page reloading because all contents are loaded dynamically using JavaScript. Most of the contents (HTML, CSS, and JavaScript) are loaded once; only the request and response data is transmitted afterward for a SPA. The project team developed the Wizard as a SPA because it provides a better user experience with less waiting time and faster page loading than traditional web applications. The SPA design allows the Wizard to be distributed and adopted more broadly because the only requirement to run the Wizard is a modern web browser that supports JavaScript frameworks such as ReactJS.

The Wizard is a software tool developed to support scenario-based learning (SBL), an active learning strategy that requires students to work through interactive questions and answers that are built to reflect real-world and complex problems. SBL is optimal for teaching data sharing practices because it works best when students apply it to decision making in complex situations like selecting the appropriate way to disseminate datasets under different contexts (Clark 2016). The Wizard has three guiding questions:

- What can I do with my data?
- What can others do with my data?
- What can I do with other people’s data?

After choosing one of the guiding questions, another specific question appears, along with several potential answers. The user can choose the answer or the next question based on their particular context. The user’s decision leads to a follow-up question or a final answer. Learners can continue each thread until there are no more questions, they can abort in the middle and change their answer to a previous question, or they can switch to a different guiding question. All the questions and answers are encoded in the JSON (JavaScript Object Notation) format as plain text.
Usability Study

Usability studies are instruments that have the goal of evaluating if a particular tool meets the goals for which it was designed (Krug 2013; Seffah et al. 2006). The project team developed two usability studies using the same instrument to evaluate the Wizard. The first study was conducted in 2018 and the results were presented at the ACRL 2019 conference (Zhang and Llebot 2019). The second study is the subject of this paper and its purpose was to determine whether and how much students learn specific data sharing concepts by using the Wizard. Ten graduate students completed both studies in person and the studies were completed as part of the same 60 minute session. Three students self-identified as international students. The sessions were divided into three parts:

- Part 1 (20 minutes). The participants were provided three scenarios and were asked to answer questions pertaining to each (“Table 1”). The participants answered the questions without using the Wizard or receiving other help. Responses were open-ended.
- Part 2 (20 minutes). After the Wizard was introduced to the participants, they were asked to complete several tasks to determine the functionality of the user interface. Participants completed a questionnaire to provide feedback and suggestions on the ease of use of the tool, not the content. Their answers are not the subject of this paper.
- Part 3 (20 minutes). The participants were asked to answer the same three questions from the same scenarios as in part 1, and were asked to use the Wizard to get the most accurate answers that they could.

The focus of this paper is Parts 1 and 3 of the usability study which use scenarios that align with the Wizard’s three guiding questions. Throughout this paper we refer to the answers that participants gave to Part 1 as “pre-Wizard” answers, and to the answers that participants gave to Part 3 as “post-Wizard” answers.

### TABLE 1

<table>
<thead>
<tr>
<th>Scenario 1. What can I do with my data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne is a grad student in a research lab at Oregon State University. The project that she is working on involves two researchers (the PI and Co-PI of an NSF grant), three grad students (including Anne) and one technician who works part time in the project and takes care of the lab equipment. Anne has been working on the same lab and project for 3 years, but unfortunately some important family issues came up and she has to move out of state. She would like to transfer to another university and continue her PhD there, and has found a program that looks promising. She would like to use the data that she generated during her three years at OSU to continue her dissertation in this other university. She is not sure if she can do that, or who to ask. The field that she works in, genomics, has very consistent standards on sharing data, and she is sure that the data she generated will end up in a public repository at some point. However, the analysis and quality control of the data is a complex process that could very well take months. Even if the process went fast, she is not sure if the plan is to publish the data immediately after the quality control has been finished, or if they are planning on waiting until the research is published. She would like to use it in a month, when she moves. The samples that she has used to generate her data come from human subjects, but Anne is not sure about what this means, exactly. The lab has strict data management rules that apply to all data, from human subjects and from non-human subjects, and she follows them routinely. The only agency funding the work (including Anne’s salary) is NSF.</td>
</tr>
<tr>
<td>What kind of information should Anne try to find to have more information about whether she can take the data with her?</td>
</tr>
</tbody>
</table>
Clara Llebot Lorente, Hui Zhang, and Michael Boock

The usability study was exempt from IRB review. Each participant received a consent form informing them of the study’s purpose, the risks and benefits of participating, the activities that they would be asked to complete, and the payment that they would receive (a $15 gift card). The consent form also included a confidentiality section, in which we told participants that “We may use parts of the anonymous information that you provide in the questionnaire (e.g., quote what you wrote) in reports or articles that we write about the Wizard.”

Data Analysis

We used two types of quantitative analysis—the Relevant Concept Analysis and the Misconception Analysis—to evaluate responses in the usability study. The goal of the Relevant Concept Analysis is to get a general sense about whether participants’ answers were correct or not. We focused on what was right, and when in doubt, we were generous in scoring student answers. The goal of the Misconception Analysis is to look at the details of each answer, and detect any misconceptions present in the answer. When in doubt, we were picky in scoring. In addition, we also interpreted each response from a qualitative perspective.

## TABLE 1

<table>
<thead>
<tr>
<th>Scenarios and questions asked of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario 2. What can others do with my data?</strong></td>
</tr>
<tr>
<td>Maria is a Fulbright student working on her Master’s at OSU. Her research involves collecting population maps that are made publicly available under open licenses by open government institutions. She combines these maps with satellite images to evaluate how population affects land use as seen from satellite images in different countries. She is generating a collection of maps, showing different categories that she created, and she is working very hard on making sure that the maps are both clear and informative. She uses her own computer, where she has installed mostly open-source software. She has an advisor who meets with her once a month to help her with her research. Maria went to a conference and presented a poster about her research. A PhD student from another institution, Jie, approached her and told her that she is working on similar topics and that she would like to use Maria’s data to complement what she is doing. Maria has not published her Master’s thesis or any other scholarly content besides this poster in this conference. She would like to collaborate with other researchers, but is unsure about what will happen to the data if she gives it to Jie. Maria would not like Jie to publish her data before she is done with her thesis, she is not sure if Jie can do that. What control does Maria have over what will happen with the data if she gives it to Jie? What are her options?</td>
</tr>
<tr>
<td><strong>Scenario 3. What can I do with other people’s data?</strong></td>
</tr>
<tr>
<td>Blake is a Ph.D. student in the Fisheries and Wildlife department at OSU studying salmon in the Columbia River, and is interested, among other things, in how the abundances of different species of salmon have changed over time. Blake has been reading literature for the research project, and just found an article that compiles historical data of some of the species of interest. Blake thinks it would be really useful to compare abundance measurements with the data from the published study, and has some statistical tests in mind. Fortunately, the article has a data sharing statement that says that the data is available in a repository. Following the citation, Blake reaches the dataset, and starts working with the data. It looks very promising! Blake wants to include the data in the thesis dissertation, but is unsure about whether it is OK to do so or not. The repository where Blake got the data does not include a clear license for the data, and the documentation of the dataset does have the authors names, but not the author’s contact information. Can Blake use the data for the dissertation? If yes, is there anything that Blake needs to do to use the data ethically? If no, what could Blake do to gain permission to use the data?</td>
</tr>
</tbody>
</table>
The results that are shown in this paper are slightly different than the results shown in the 2021 ACRL presentation. The authors found an error in the tagging of the misconceptions in scenario 1; however, the trends described in both the presentation and the paper remain the same.

**Relevant Concept Analysis**

We developed a rubric with a short list of the possible good answers for each question. Each possible good answer was worth one point. Each scenario was different, so the total number of possible good answers varied. Scenarios 1 and 2 each had nine possible good answers, and scenario 3 had six (Table 2). The three scenarios were broad and written to be similar to real life situations; therefore, different people could answer them successfully from different points of view. For example, scenario 1 could be answered from the point of view of the funder, from the point of view of the university receiving the funds, or from the point of view of the researchers themselves. We were generous with the list of answers that could be considered appropriate. For example, in scenario 3 we included the concept that a data license could be found in different places, and that “Blake” should look for a license thoroughly. It was definitely not an essential concept to answer the question, but it was appropriate and correct, so we included it. A higher number of points per answer meant an answer was more complete, and revealed that the participant had a better understanding of the research data landscape. A lower score did not mean that the answer was wrong, it meant that it was answered from a narrower perspective, or without many details.

**TABLE 2**

<table>
<thead>
<tr>
<th>Rubric scenario 1. What kind of information should Anne try to find to have more information about whether she can take the data with her? Who is the person who can make this decision?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Figure out who is the owner of the data</td>
</tr>
<tr>
<td>2. Anne does not own the data because she is part of a team, and her research requires significant resources from a lab. She cannot make this decision on her own.</td>
</tr>
<tr>
<td>3. Identify the data custodian or decision maker who can make the decision, and who should know all the following. Usually the PI of the project but it may include other Co-PI or researchers in the project.</td>
</tr>
<tr>
<td>4. Anne should figure out which are the plans for data publishing, including looking into the Data Management Plan (it will say when they are planning on making data publicly available)</td>
</tr>
<tr>
<td>6. Funding agencies, and grant requirements</td>
</tr>
<tr>
<td>7. If there are policies that affect data ownership/transfer. Like lab policies, or university policy</td>
</tr>
<tr>
<td>8. If there are other formal agreements, like data sharing agreements or contracts</td>
</tr>
<tr>
<td>9. Anne will need to work with the current university and the future university to be able to take the data with her. Maybe via institutional affiliations, or via a data sharing agreement, or via creating a new collaboration between labs at both universities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rubric scenario 2. What control does Maria have over what will happen with the data if she gives it to Jie? What are her options?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Maria has control over the data before she gives it to Jie. She can give the data with restrictions if she wishes to.</td>
</tr>
<tr>
<td>2. Maria can give the data to Jie right then. It is a risk, because giving the data without any formalised (or semi formalised, like an e-mail) agreement on what Jie can do relies on trust only, Maria loses control.</td>
</tr>
</tbody>
</table>
TABLE 2

Rubric used for the Relevant Concept Analysis

<table>
<thead>
<tr>
<th>Rubric scenario 3</th>
<th>Can Blake use the data for the dissertation? If yes, is there anything that Blake needs to do to use the data ethically? If no, what could Blake do to gain permission to use the data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>If the data is a collection of facts (it seems so, by the description), it is not covered by copyright. Yes, Blake can use the data even if it does not have a license.</td>
</tr>
<tr>
<td>2.</td>
<td>Blake must give attribution if they use the data in their work.</td>
</tr>
<tr>
<td>3.</td>
<td>Contacting the authors may not be strictly necessary (if it is not covered by copyright) but Blake could do it as a courtesy.</td>
</tr>
<tr>
<td>4.</td>
<td>Blake should look for a license thoroughly. The says that the repository does not include a license. Blake should check if there is a license in the documentation, or if the repository has general terms of use.</td>
</tr>
<tr>
<td>5.</td>
<td>If the dataset is covered by copyright (element of creativity), or Blake thinks it may be, then Blake should contact the authors and ask for permission. With the name of the authors Blake can hopefully find the contact information online.</td>
</tr>
<tr>
<td>6.</td>
<td>This scenario describes a dataset published following bad practices: no license and no contact information</td>
</tr>
</tbody>
</table>

Misconception Analysis

For the Misconception Analysis, we scrutinized each response more carefully, and identified misconceptions that the participants may have had. We broke each answer down into more granular concepts, and classified each of these concepts as correct or not correct. For example, a participant answered the Scenario 1 questions in the following way.

Pre-Wizard answer: “Maybe. He needs to reach out to the authors and article publishers first to make sure everything is OK”. Post-Wizard answer: “Yes. Since there is no copyright, the data is in the public domain”.

In our analysis, we identified three concepts that were expressed by the student in the pre-Wizard answer:
- First concept: “Blake” needs to reach out to the authors.
- Second concept: “Blake” needs to reach out to the publishers.
Third concept: “Blake” needs to reach out to make sure that everything is OK.

In the student’s post-Wizard answer, we also identified three concepts that were expressed by the student, but with more specific and correct answers:

- First concept: “Blake” can use the data.
- Second concept: There is no copyright.
- Third concept: The data is in the public domain.

Note that the total number of concepts is variable, so the total number of concepts pre- and post-Wizard are different. In many cases, qualifying whether a concept was correct or not was difficult, because participants tended to write very short sentences that sometimes were ambiguous. For example, the concept “to make sure everything is OK” is unclear. It can be interpreted that the data is reusable, that the data is good quality, that the data is complete, etc. To solve this problem we created “unclear” and “vague” categories. We tagged concepts as “unclear” and “vague” if they were likely to be misconceptions and correct concepts (respectively), but were expressed in a way that gave us doubt. We also created a fifth category for concepts that were correct, but that were not appropriate for the scenario. See Table 3 for a definition of the categories we used to tag concepts.

### TABLE 3

<table>
<thead>
<tr>
<th>Misconception Analysis categories</th>
<th>Misconception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misconception</td>
<td>The concept is not correct, or it is not correct given the scenario.</td>
</tr>
<tr>
<td>Correct Not Appropriate</td>
<td>The concept is correct, but it is not appropriate to the scenario.</td>
</tr>
<tr>
<td>Correct Appropriate</td>
<td>The concept is correct and relevant to the scenario.</td>
</tr>
<tr>
<td>Unclear</td>
<td>Likely a misconception, but it is not clear. It can be interpreted in several ways.</td>
</tr>
<tr>
<td>Vague</td>
<td>Likely correct and appropriate, but it is not clear. It can be interpreted in several ways.</td>
</tr>
</tbody>
</table>

### Qualitative analysis

We did not use any structured methodology for our qualitative analysis. We read the questions and answers, and evaluated (1) if there was consistency between the responses and the results of the quantitative analyses; (2) if there was any trend that had been missed by the previous analyses.

### RESULTS

#### Relevant Concept Analysis

We coded all the responses for the 10 participants and the 3 scenarios. The results of the Relevant Concept Analysis are summarized in Table 4. The average scores of the answers before using the Wizard and after using the Wizard are shown as percentages because the rubrics of each scenario have a different number of concepts. These are the average scores of the answers before using the Wizard and after using the Wizard. Again, the important information here is not the comparison between scenarios, but the comparison of pre- and post-Wizard scores for each scenario.

We observe that the scores post-Wizard are either very similar, or lower than the scores pre-Wizard. We believe that this is because the participants, when asked to answer the same question a second time, instead of rewriting everything again, used what they had learned to complement or refine their previous answers. This hypothesis is supported by the fact that a high proportion of pre-Wizard answers were longer than the post-Wizard counterparts (60%). We also saw that more than 80% of the post-Wizard answers included new concepts. The total number of new concepts explained in each scenario can be seen in the last line of Table 4. As a consequence, the score of the combined pre- and post-Wizard answers was always higher than each of the individual answers.
TABLE 4
Average of scores for each scenario and total, expressed in percentage. Scores were calculated with rubrics of Table 3.

<table>
<thead>
<tr>
<th>Summary of results</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Wizard score %</td>
<td>34</td>
<td>34</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>Post-Wizard score %</td>
<td>27</td>
<td>32</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>Pre and post-Wizard combined score %</td>
<td>46</td>
<td>51</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>New concepts mentioned #</td>
<td>10</td>
<td>15</td>
<td>9</td>
<td>34</td>
</tr>
</tbody>
</table>

Figure 1. Number of tagged concepts for all scenarios, pre- and post-Wizard. Numbers show the number of concepts in each category, and the percentage that it represents over the total.

Misconception Analysis

Figure 1 shows the results of the misconception analysis for all scenarios together. The left column represents the classification of the concepts in the pre-Wizard answers, and the right column represents the post-Wizard answers. Misconceptions are at the bottom. Correct concepts are at the top.

The number of concepts expressed in the pre-Wizard answers was much higher than the number of concepts expressed in the post-Wizard answers. This fit with our theory that participants tended to refine their answers post-Wizard rather than rewrite everything. The proportion of correct concepts expressed is higher in post-Wizard answers (74%) than in pre-Wizard answers (52%). Also, the proportion of misconceptions is lower post-Wizard (9%) than pre-Wizard (20%). The trend of fewer misconceptions and more correct answers is maintained when we take into account the concepts whose correctness we were not sure about; i.e., the unclear and vague concepts.

Responses to each of the scenarios followed the same trends. In Scenario 1 correct concepts were expressed 56% of the time pre-Wizard vs. 94% post-Wizard, and misconceptions were expressed 13% of the time pre-Wizard and 0% post-Wizard. In Scenario 2 the percentage of correct concepts expressed pre- and post-Wizard was similar, 72% and 73%, and the misconceptions expressed by students was 14% pre-Wizard and 10% post-Wizard. In scenario 3 students expressed correct concepts 10%
Development and Evaluation of an Active Learning Tool for Data Decision Making 257

of the time pre-Wizard vs. 58% of the time post-Wizard, and misconceptions were reduced from 55% pre-Wizard to 16% post-Wizard.

We also looked at the difference between the results of international students and those of domestic students. Our sample included 3 international students and 7 domestic students, so the total number of concepts analyzed was larger for domestic students (a total of 69 and 54 pre- and post-Wizard respectively) than it was for international students (19 and 12 concepts in total). The percentages of correct and appropriate concepts were larger for domestic students (54% and 78%) than for international students (47% and 58%), both pre- and post-Wizard. The opposite was true for misconceptions (pre- 21%, post-17% for international students, and pre- 23%, post- 7.4% for domestic students); however, when taking the vague and unclear tags into account, the percentages were very similar. This could be because international students were more confused about the questions, the scenarios, or the concepts. Another possible explanation is that the authors found the written statements from international students harder to understand than the statements from domestic students, and so they tagged them as vague and unclear. Both international and domestic students show the same trends: correct concepts increase and misconceptions decrease after using the Wizard.

Qualitative Analysis

Here we include a few examples of answers that illustrate the observations described in previous sections.

1. First observation: The post-Wizard answers were shorter, and less complete than the first answers.
   
   Sample 1 (scenario 1).
   
   Pre-Wizard: “NSF for funding reasons—they may have ownership over the data and research IRB (institutional review board) since there are human subjects—although IRB rules have transitioned from multi-institutional IRBs to following one IRB (?)
   Anne’s PI/Department Head and/or College since she’s a student”
   Post-Wizard: “colleagues, supervisors, administrators from the institution where you work, and NSF”

   Sample 2 (scenario 2)
   Pre-Wizard: “—Ask to be a co-author on all research that results from this data set if she decides to share—Not share the information — it’s public use data so technically other researchers can put together the data”
   Post-Wizard “She can share or not share the data, or wait to share her data.”

   Sample 3 (scenario 3).
   Pre-Wizard: “Blake is allowed to use the data in his thesis since it was made publicly available in a repository. Blake should do his best to track down the original authors of the study and allow them to comment on the use of their data. At the very least Blake needs to acknowledge and cite where the historical data came from.”
   Post-Wizard: “Blake can use the data in his thesis. He should make sure to cite and acknowledge the authors of the original data set and point out that it was publicly available.”

   Samples 1-3 show answers that were very similar pre and post-Wizard. Understandably, participants did not repeat all the details that they had included in the first answer, but summarized them the second time they were asked.

   
   Sample 4 (scenario 1).
   Pre-Wizard: “She should investigate the funding statement from NSF, because they probably have an in-depth description of whether they own the data, or how it can be used besides this project.”
   Post-Wizard: “Anne should try to determine who the owner and custodian of the data are. She should also check to see if there is a Data Sharing Agreement or similar document that describes how the data can be shared. If there was no such document, then she should reach out to the person who made the data available to her (probably her PI) and simply ask them.”

   Sample 5 (scenario 2).
Pre-Wizard: “she will not has any control”
Post-Wizard: “she can set an formal written agreement with the person about the conditions of use”

**Sample 6 (scenario 3).**
Pre-Wizard: “I think that Blake is able to use the data as long as it is openly shared, although I would be hesitant because of the unclear license for the data. Blake should use the authors’ names to find their contact information and request a clarification of the use of data and where to obtain a clear license for data use.”
Post-Wizard: “If the data set is not protected by a Copyright, regardless of if the license is clear or not, it seems that Blake will be able to use the data as it is publicly owned. Blake would need to make sure he cites the data carefully throughout his work so as not to feign ownership.”

In sample 4 the participant included several new concepts: that it is important to find out who the custodian of the data is, that there may be documents that describe how the data can be shared, and that the principal investigator should be contacted. In sample 5 the participant included the idea that a formalized agreement is an option when considering data sharing. In sample 6 the participant added the concept that the data should be cited.

When reading the answers, we also observed other trends that were not captured by the quantitative analysis:

3. Third observation: The post-Wizard answers were more precise.

Sample 5 above was an example of how before using the Wizard some responses were ambiguous, and after using the Wizard were made more clear. In this case the response was still brief, but they identified one clear option to the scenario, very different from the first answer that suggested that there were no options. Another example of more precise responses was the post-Wizard use of vocabulary that was not present before the Wizard, like copyright, fair use, or public domain. In scenario 3 this was particularly clear. Seven of the post-Wizard answers mentioned “copyright”, and 5 mentioned “public domain”, while none of the pre-Wizard answers mentioned either term. See, for example, sample 6.

4. Fourth observation: Students provided more well-reasoned responses after consulting the Wizard even when the answers were similar.

Quotation 6 above was an example of a participant that gave the same answer (i.e. “Blake” can use the data), but the reason given was more precise.

5. Fifth observation: The tool did not work for everybody.

One of the participants did not find the Wizard helpful for scenario 1, and left the following post-Wizard answer: “I find difficulty to find the answer of this question using this website”. The participant was able to complete the questionnaire for scenario 2 and scenario 3.

After the participants had completed the three parts of the usability study we asked them a final question: “Do you have any comments that you want to share about the content or the usability of the Wizard?” Most responses were positive. For example, one of the participants mentioned: “It seems like a really helpful tool! As someone creating my own data (in this case an original model and simulation results) it is good to know the rights I have to using, sharing, and getting credit for it.” Some of the participants included suggestions for improving the tool: 1. Define the concept of “repository”, 2. add specific people in the list of resources for the university, 3. have a more clear “home” button, and 4. number the questions used for navigating the Wizard.

**DISCUSSION**

The usability study results suggest that the Wizard is an effective tool for students to learn data sharing and copyright concepts. Users of the tool demonstrated an understanding of new concepts related to data use and sharing and were able to use the tool to correct their own misconceptions. The Relevant Concept Analysis revealed that more than 80% of post-Wizard answers included new concepts. With the Misconception Analysis we learned that the proportion of correct concepts students expressed was higher in post-Wizard answers (74%) than in pre-Wizard answers (54%). Also, the proportion of misconceptions expressed by students was lower post-Wizard (6%) than pre-Wizard (22%). This trend was observed for each scenario individually, and for international and domestic students alike.
The usability study we designed has some limitations. Participants tended to provide shorter answers in the second round (i.e., Part 3) compared to their solutions in the first round (i.e., Part 1). Our hypothesis is that most participants did not want to repeat their answers in the second round, and refined their answers rather than repeat everything. However, it is possible that in some cases the participant wanted to substitute their first answer with the second one. Our instrument did not allow us to distinguish between these two cases. The study could be enhanced by providing the participants with their answers from the first part, asking them to modify them as needed.

We believe that the results of this usability study support the use of the Wizard as a learning tool. The study was done in the context of a user working alone, without access to other sources such as a librarian or instructor. However, we think that the Wizard can be useful as a tool in the classroom, with the support of a teacher. For example, students could be asked to use the Wizard before a class or workshop to attempt to answer questions such as those posed in the usability study scenarios as a way to focus the attention of the student on a particular issue. During the class the teacher could answer questions that came up during the use of the Wizard, or do a different activity to center the attention on a concept introduced by the Wizard. The Wizard can also be useful as a stand-alone tool. For example, it could be included in LibGuides or reference materials that students reach on their own, outside of the classroom.

The Wizard is designed so that it can be periodically updated when needed, and we plan to modify the Wizard as necessary. For example, right now the Wizard explains that Oregon State University does not have a research data management policy. If a policy is approved, the answers and information presented by the Wizard will be updated to reflect that change.

From the beginning, the Wizard design emphasized easy adoption and deployment. The Wizard code is available from the GitHub repository (Zhang et al. 2021) under a GNU General Public License v3.0 license. To reuse the Wizard we recommend customizing the database.json JSON template file that has the content of the questions and answers. In particular, the Wizard refers the user to OSU specific resources and offices, which should be adapted to the institution adopting the tool. There are three data types in the JSON template:

- Title type: There is only one title type instance, and its value will be used across the Wizard.
- Question type: This data type includes questions, where they originate, and where they lead.
- Finished Type: This data type is for the last step of a question path.

The second step to reuse is to customize the Wizard application with your own logo, colors, branding, etc. The Wizard application includes two important files:

- App.css: The style sheet.
- App.js: The main program written in the React JavaScript library.

The final step to reuse is to deploy the customized Wizard to production. A system administrator must allocate server space, and the Wizard source code must be “built” locally following the README file instructions before uploading to the server.

ACKNOWLEDGMENTS

The authors wish to thank Sam Young, an Oregon State University computer science student, for his help with the programming of the tool. This work was partially funded by the Oregon State University Technology Commons Fund and by Oregon State University Libraries and Press.

REFERENCES


