Challenges Faced By Medical Professionals as Gateway Users: The Case of KnowCOVID-19

Bruce W. Shibo
College of Media & Communication
Texas Tech University
Lubbock, TX, USA
shibobruce.wang@ttu.edu

Roland Oruche
Computer Science
University of Missouri-Columbia
Columbia, MO, USA
rro2q2@mail.missouri.edu

MD Ashraful Goni
College of Media & Communication
Texas Tech University
Lubbock, TX, USA
mgoni@ttu.edu

Xiyao Cheng
College of Engineering
University of Missouri
Columbia, MO, USA
xcheng@mail.missouri.edu

Eric Milman
College of Media & Communication
Texas Tech University
Lubbock, TX, USA
eric.d.milman@ttu.edu

Prasad Calyam
Computer Science
University of Missouri-Columbia
Columbia, MO, USA
calyamp@missouri.edu

Kerk F. Kee
College of Media & Communication
Texas Tech University
Lubbock, TX, USA
kerk.kee@ttu.edu

***Abstract*—Science gateways emerged a couple of decades ago to lower the barriers to computer language proficiency for researchers and educators alike to implement digital tools to further their work. However, gateways remain somewhat esoteric and difficult to use for some potential users. KnowCOVID-19 is a prototype gateway designed to help medical professionals navigate extensive online publications related to COVID-19 research. Via in-depth interviews with 10 medical professionals, we investigated the challenges they faced when using the gateway, namely, slow speed, limited scope, and mixed quality of data. Findings provide a solution for addressing these challenges within a science gateway framework by incorporating a level of evidence pyramid. Lastly, gateway projects could consider conducting similar market research interviews to better understand the work context, including challenges faced by the intended users of specific gateways , and compiling them as user requirements to guide gateway developments.**

Keywords—science gateways, data deluge, KnowCOVID-19 Gateway, in-depth interview.

I. Introduction

According to [1], science gateways aim to increase the discovery and reach of science by fostering connections among highly technical professionals and less technically proficient researchers. It means making big data, high-performance computing (HPC), and cyberinfrastructure (CI) available to the intended users broadly. Solutions for batch and cloud-based systems, management of data solutions, management of workflows, and authorization are some of the building blocks for gateways that make this aim possible [2].

Gateways improve research, cooperation, and data transmission by minimizing barriers to e-science. E-science is the product of the CI community's data-driven research activities, where users need programming expertise and domain knowledge [3]. However, e-science was put to the test by these two important requirements. Therefore, gateways emerged as a solution by their reduction of the need for programming expertise and ease of data integration.

This project was funded by the National Science Foundation under grant numbers NSF-2006816 and NSF-2007100.

The amount of COVID-19 publications became overwhelming and resulted in a 'data-deluge problem' during the pandemic [4]. There is a level evidence-based pyramid that is helpful for clinical researchers and healthcare professionals to find better and more reliable information to combat COVID-19 and the 'data-deluge' problem [5]. Again, if a clinical researcher starts using that evidence-based pyramid then it will take longer time then as usual to solve their problem. To solve this issue, developers of the KnowCOVID-19 gateway integrated this level of evidence pyramid in the gateway to provide high-quality publications in a shorter time [6]. KnowCOVID-19 gateway is the gateway solution for 'data-deluge problem' and Vidura is a chatbot that assists users in navigating and using KnowCOVID-19 Gateway [7].

In this paper, we investigated a new gateway called KnowCOVID-19 (http://54.152.172.214:9300/Covid-19/)for medical professionals (i.e., clinicians, healthcare providers, researchers, etc.) as gateway users. During the pandemic, medical professionals relied on lengthy and laborious manual efforts to find COVID-19-related publications (i.e., journal articles, conference proceedings papers, conference papers, pre-prints, observational studies, opinion pieces, etc.). However, by analyzing 10 qualitative interviews, we explored the difficulties faced by medical professionals in doing literature searches for COVID-19 publications.

II. Related work

A. Brief overview Know COVID-19 Gateway

We implement Vidura Advisor and the Level of Evidence pyramid on the KnowCOVID-19 science gateway, using the Python, AngularJS, and Spring Boot back-end development framework, which is a widely used framework in Java. Spring Boot is an effective alternative framework that is pre-configured with technologies that alleviate the manual efforts of configuration compared to conventional frameworks. We integrate the Level of Evidence pyramid using semi-supervised learning techniques for document classification using Python and ML-based libraries. We collect a set of ground-truth

papers using web scraping tools performed over PubMed search and use this as a basis for classifying papers according to each level of evidence pyramid. Then, we use the CORD-19 dataset [8] for prediction or label generation for the remaining documents. These documents are then organized and populated into a table on the main page of the KnowCOVID-19 science gateway. You can find more details about KnowCOVID-19 science gateway in [4], [6],[7],[8]

The Vidura Advisor is mainly powered by Google Dialog flow, which is a conversational agent service that uses natural language understanding techniques to perform question-answering over user queries. We integrate the Vidura Advisor onto KnowCOVID-19 using the Google Dialog flow API to interact with the user on the science gateway interface. KnowCOVID-19 is the gateway and the focus of this paper. Therefore, the Vidura chatbot is not emphasized.

B. *Level of evidence pyramid* and Chatbot *for a COVID-19 Gateway*

The COVID-19 pandemic posed great challenges for medical professionals during the height of the pandemic. Specifically, they were suddenly thrown into a situation where they must stay current with the best available medical information for finding effective treatments and treating patients while making critical decisions at work when the COVID-19 disease was a rapidly evolving problem and the volume of online publications was overwhelming. However, they can use Google search and level of evidence pyramid separately to find better and more relevant literature, but it would be a very time-consuming and tedious process. Developers of the Know COVID-19 gateway built a solution with filtering domain-specific literature based on the level of evidence pyramid and social filtering to encourage collaboration among medical professionals and researchers to combat COVID-19 [6].

KnowCOVID-19 is a gateway that enables users to sift through voluminous online literature to discover the answers to their search queries based on topics and level of evidence. It has a chatbot (Vidura) built in to help users get through infodemic and low usability, especially for inexperienced users [8]. To address the potential of a chatbot in a gateway, [7] first used the COVID-19 Open Research Dataset (CORD-19) as the database for sorting based on users' topics and levels of evidence in their queries. In addition, the Vidura chatbot was further developed and trained to enhance the search ability for users [7]. In the KnowCOVID-19 gateway level of evidence pyramid helps users to instantly visualize the most relevant literature, and Vidura helps them to search further and navigate properly in the KnowCOVID-19 gateway [7].

Preliminary usability studies indicated that overall, participants had positive impressions of the evidence-based recommender system but indicated that the system needed to be easy to use and navigate [9]. Thus, we investigated deeper through qualitative interviews to gain more insights into the challenges that medical professionals faced when looking for research publications and what they needed to assist them based on the level of evidence metrics. More specifically, we seek to answer the research question (RQ), "*What are the challenges faced by medical professionals as users of a science gateway designed to search and filter medical information related to COVID-19 during the pandemic?*"

III. Interview protocol

Considering that relatively scant prior work has been accomplished at examining the prospect of using the level of evidence pyramid in a science gateway to tackle the problem of data-deluge, we adopt the qualitative interviewing approach [10] to investigate the above-stated RQ. The interview protocol was comprised of 15 open-ended questions. The first 6 questions were devised to elicit participants' routines and experiences of seeking research publications pertinent to their medical fields. Before being interviewed, participants were asked to watch a video about what the KnowCOVID-19 gateway was and how levels of the evidence pyramid assist medical professionals in enhancing their search process. For those who did not watch the video prior to the interview, they had a chance to do so after interview question 6. After participants watched the video, they were inquired about how the KnowCOVID-19 gateway with an evidence-based filtering system could aid them in conducting literature searches and solve the problem of finding the most relevant COVID-19-related literature in the shortest amount of time. The interview protocol was semi-structured, enabling the interviewers to improvise the questions on the spot based on the conversation, and for participants to guide the interviewers along their authentic journeys and experiences. The semi-structured design is optimal for studying emerging topics because the pre-designed questions may entail assumptions that could be shown to be invalid during the interviews, thus necessitating flexible adaptations [10]. Prior to conducting the interviews, we secured IRB approval (TTU IRB #2020-455) on the study, and the interviewers also underwent research ethics training to be IRB certified.

IV. analysis

The crux of our analysis lies in the themes that emerged from our interviews concerning the obstacles that medical professionals encountered while searching for COVID-19 publications andKnowCOVID-19 science gateway. . We employed Otter.ai, an AI transcription service, to produce verbatim transcripts. We subsequently analyzed the interviews by systematically listening to the recorded interviews while examining the transcripts simultaneously to ascertain emerging themes.

Afterwards we applied 'open coding' [11] (a social science data analysis technique) to determine, contrast and progressively refine the themes. Ultimately, we adopted 'axial coding' [11] to correlate themes to each other in meaningful ways. In this analysis, we first addressed the challenges and pinpointed design issues that were specific to gateways. To protect participants' identities, all names used in this paper are pseudonyms. Below is a brief background about each participant. A few participants played several roles and had multiple titles at their institutions. For confidentiality reasons, only two titles most relevant to their excerpts were chosen to introduce the participants.

TABLE I. Participants' names and position(s)

Name (Pseudo)	Selected Position(s)
Steward	Chief Medical Officer/Associate Dean
Simon	Associate Research Professor/Biochemist
Craig	Infectious Disease Medical Doctor
Zavier	Lung Disease Medical Doctor
Jack	Professor of Veterinary Medicine and Surgery
Ivy	Assistant Director of a Health Science Library
Austin	Immunologist
Rachel	Geneticist
Ashley	Head of Veterinary Library
Douglass	Respiratory Physiologist

V. RESULTS

A. Slower Speed versus Desired Speed

The first challenge faced by medical professionals is the speed in getting the answers they are looking to find. This challenge can be summed up by one of our participants, Simon, who stated, "So what I would like to have, is a specific answer rather than, you know, going to hunt for the answers." This is an important challenge to note because, during the pandemic, time was of the essence when it came to treating patients and identifying effective treatment options.

While some may think of Google as a search engine right away, our participants experienced some limitations. For example, Steward reflected that, "Google's pretty fast, but you got to get through a bunch of crap and a bunch of other pointers to the same article." A limitation of Google is its page rank algorithms do not fit what medical professionals are looking for, which is the scope and quality of information, two other challenges that we will elaborate on later. We noted these two challenges to illustrate what Craig noted, "Improvement of the online medical database needs to have some organizational environment and a system to find things... So, everybody can be fast from the get-go."

Moreover, some may think of medical databases besides the general Google search engine. However, Ivy shared her frustration, "There are all kinds of tricks that we can use to increase relevancy, a narrower set of search results. But many of those tricks aren't available to us until the article has been ingested into the National Library of Medicine, and someone has had a chance to use indexing. [Although] indexing is now done somewhat semi-automated, there's very little for us to grab on to other than title and abstract at the very beginning and author keywords." What Ivy shared reveals not only the delay in making medical literature available as soon as possible, but also echoed Craig's point that the current systems lack an effective organizational environment in which users can more quickly find the answers they need beyond titles, abstracts, and keywords.

While the current setup of Google and other medical databases appear to be relying on reasonable logic, Rachel further explained an important work context that gateway developers may find insightful for design decisions. She said "The thing that you have to realize is clinicians are a different group of people than researchers…. A clinician kind of wants to get fast summaries, so they can answer the questions of their patients. They're not going to have the same level of research understanding, because not many clinicians are Ph.Ds." This excerpt suggests that an effective gateway for medical research and literature will need to have features that allow both researchers and clinicians to find the answers with speed, regardless of the differences in formats and details necessary.

B. Limited Scope versus the Range of User Needs

The second challenge discussed by participants is the scope of the publications and datasets. In terms of scope, the participants expressed the necessity to have the latest information available. For example, Craig shared, "You have to be up to date. It is basically a live website meaning you have to update it regularly... It has to be a mechanism to pull the most recent articles of recent information." This expectation to have the most recent publications aligns well with the first challenge of speed.

However, the concept of scope here extends beyond temporal scope to geographical scope. For instance, Douglass added, "There are certain topics that I'm interested in where it seems like a third of the papers come from Korea. I want to make sure I'm getting all those Korean journals, in my search is what it amounts to." What Douglass alluded to is the complexity of COVID-19 being a global problem. Although he works in the US, he needs information and knowledge from a global scope of publications to help him do his work better here.

Besides temporal and geographical scopes, participants also pointed out the need to include original datasets in the gateway. For example, Simon expressed, "More patient data is something that we are lacking. We don't get patient data often. We go to PubMed all the time, and you know how much you can find. Or you go to the New England Journal of Medicine or the Lancet. In those kinds of journals, you can find patient data. So, patient data is a limitation, and we don't have that process (of gaining patient data)". He further explained, "For some journals that do not offer is a creative commons license, you don't see any info [about data]. I guess they can put a link there and say okay, you have to go to such and such site to find it out." However, if the data can be made available with the publications and on a gateway, it would make the information-seeking process more effective for medical professionals. These excerpts above suggest the concept of scope entails temporal, spatial, and data dimensions, which provide important attributes for gateway developers to consider.

C. Mixed Quality versus Ranked Quality

The third challenge raised by participants is that of quality. Given the volume of information that came online, and the speed with which it became available (although not fully captured for effective search by a platform and/or gateway), medical professionals expressed the need to help them assess publication quality. For example, Craig explained, "If I have 10 randomized controlled trials, I will probably put priority to a randomized controlled trial published in the New England Journal of Medicine, compared to a journal that has a very low impact factor." Therefore, a filtering algorithm could be based on the impact factors of journals.

Besides ranking publications in search results based on journal reputation, another approach is to rank publications

based on the levels of evidence. Based on the evidence pyramid [5], medical publications that have the highest level of evidence are systematic reviews/meta-analyses, followed by randomized controlled trials, then non-randomized controlled trials, observational studies with comparison groups, case series/case reports, and the lowest level of evidence is expert opinion. Zavier explained, "That's what we always strive for, is to find articles with the highest level of evidence because you can always find any article out there." Similarly, Douglass shared, "So simply meta-analysis or other systemic reviews, a systematic review of a topic. Getting access to those is a great place to start". Both Zavier and Douglass suggested not to simply take the first few articles from the search results, but to select the ones with the most rigorous methodologies.

For a reader less familiar with the lack of quality control in medical publications, Jack's explanation is helpful. He expressed his concern, "There's a lot of stuff being published right now and when you read it and look at the level of evidence, as well as the quality of the experimental design, it's pretty poor-quality evidence". Steward shared Jack's concern, and thus expressed his enthusiasm for the gateway we studied in this paper, "To me, it (KnowCOVID-19) combines keywords with an objective ranking of the quality of the data. I think those two things alone are a huge boon." The insights for gateway developers to consider is to go beyond making publications, datasets, and even tools available to users, but to find a way to screen and rank them based on what users need to do to do their work. This may appear as the job of the users at first glance, but the insight is that if a gateway can assume this burden instead of leaving it to the users, the gateway will better meet user needs, and be more likely to receive widespread adoption.

VI. OUTLOOK

In this paper, our research team sought to answer the RQ, "*What are the challenges faced by medical professionals as users of a science gateway designed to search and filter medical information related to COVID-19 during the pandemic?"* By systematically analyzing 10 qualitative interviews with medical professionals in various roles across the US, we identified three main challenges: slower speed versus desired speed, limited scope versus the range of user needs, and mixed quality versus ranked quality.

Besides the main challenges identified above, the paper also provides a solution for how a science gateway can overcome those challenges. First, the level of evidence pyramid addresses the first and third challenges of speed and quality, as the gateway narrows search results based on methodological rigor. Second, constant updates on literature databases address the second challenge of scope. Given these, the case of KnowCOVID-19 shows how gateways can be applied to address data-deluge problems by practitioners, who are medical professionals (including clinicians, healthcare providers, etc.) in this case.

Moreover, this paper demonstrates that an important approach to gateway designs is by doing 'market research interviews', to determine if there is a market for a gateway, what the target users need, and what are the gateway implementation expectations or attributes of users for their work. For example, by interviewing medical professionals, we gathered a set of user requirements that the prototype gateway could build towards, to build a useful gateway for the target user community. The user requirements can help outline how the developer team could outline how to get there from a technological perspective. Usability burdens should be removed by design decisions to make users' lives easier, which will also promote widespread adoption. We would not have thought of some of them without conducting the interviews.

While science gateways have much potential to facilitate collaborations between scientists and programmers via easy access to data and research tools, the usability of science gateways remains a challenge [3], [12]. We hope the findings from this paper provide some design and usability foci that gateway developers could take into consideration in their development efforts. Moreover, gateway developers may be interested in conducting a similar interview study to help them (continuously) identify and understand the unique challenges their users face.

REFERENCES

[1] S. Gesing *et al.*, *Science Gateways and AI/ML: How Can Gateway Concepts and Solutions Meet the Needs in Data Science?* IntechOpen, 2023. doi: 10.5772/intechopen.110144.

[2] S. Gesing *et al.*, "A Vision for Science Gateways: Bridging the Gap and Broadening the Outreach," in *Practice and Experience in Advanced Research Computing*, in PEARC '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 1–8. doi: 10.1145/3437359.3465579.

[3] K. F. Kee and A. R. Schrock, "Best social and organizational practices of successful science gateways and cyberinfrastructure projects," *Future Gener. Comput. Syst.*, vol. 94, pp. 795–801, May 2019, doi: 10.1016/j.future.2018.04.063.

[4] K. F. Kee, P. Calyam, and H. Regunath, "The role of Vidura chatbot in the diffusion of knowcovid-19 gateway," *Hum.-Mach. Commun.*, vol. 3, pp. 47–63, Jan. 2021, doi: 10.3316/INFORMIT.099941590265302.

[5] M. H. Murad, N. Asi, M. Alsawas, and F. Alahdab, "New evidence pyramid," *BMJ Evid.-Based Med.*, vol. 21, no. 4, pp. 125–127, Aug. 2016, doi: 10.1136/ebmed-2016-110401.

[6] R. Oruche *et al.*, "Evidence-Based Recommender System for a COVID-19 Publication Analytics Service," *IEEE Access*, vol. 9, pp. 79400–79415, 2021, doi: 10.1109/ACCESS.2021.3083583.

[7] R. Oruche *et al.*, "Measurement of Utility in User Access of COVID-19 Literature via AI-powered Chatbot," in *2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Oct. 2021, pp. 1–13. doi: 10.1109/AIPR52630.2021.9762160.

[8] R. Oruche *et al.*, "Science gateway adoption using plug-in middleware for evidence-based healthcare data management," *Concurr. Comput. Pract. Exp.*, vol. 35, no. 18, p. e7195, 2023, doi: 10.1002/cpe.7195.

[9] M. E. Eren, N. Solovyev, E. Raff, C. Nicholas, and B. Johnson, "COVID-19 Kaggle Literature Organization," in *Proceedings of the ACM Symposium on Document Engineering 2020*, in DocEng '20. New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 1–4. doi: 10.1145/3395027.3419591.

[10] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2017.

[11] Strauss and J. M. Corbin, *Grounded theory in practice*. Sage, 1997.

[12] K. Vekaria *et al.*, "Recommender-as-a-service with chatbot guided domain-science knowledge discovery in a science gateway," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 19, p. e6080, 2021, doi: 10.1002/cpe.6080.