FULL PROPOSAL ABSTRACT, NARRATIVE, AND SCHEDULE OF COMPLETION

Abstract

Investigators at the University of Florida (UF) George A. Smathers Libraries in partnership with researchers at OCLC and Rutgers University request $491,882 (with $232,289 in cost share) for the three-year research project titled, *Researching Students' Information Choices: Determining Identity and Judging Credibility in Digital Spaces*. Our project team is composed of academic librarians, research scientists, an educational technology specialist, a school media researcher, and an advisory panel of practitioners. For this research study, we will consider whether students are *format agnostic*. First coined by Abram and Luther (2004), the term refers to students who either cannot or do not identify the *container* (i.e. document type) when making judgments relating to use of digital resources. Several usage studies have reported that students experience trouble distinguishing among different digital resources, such as e-books and e-journals (Croft, 2004; Levine-Clark, 2006; Shelburne, 2009). Soules (2009) goes so far as to say "E-book, e-journal? Users don't care; in fact they never cared, and many only understood book vs. journal in the print world because of the difference in their physical structures. What they want is relevant content" (pg. S4). But how do students distinguish whether relevant digital content is also credible digital content? Given the limited published research on how late primary, secondary, community college, and undergraduate STEM students identify and determine credibility, this project will provide new knowledge to librarians and educators. This information will be promulgated with practitioners for incorporation into information literacy courses, and other teaching and learning environments to aid STEM students in effectively determining credibility in the discovery, access, and use of digital resources.

For this research case study, we propose to study a diverse group of approximately 180 students working in the science, technology, engineering, and mathematics (STEM) disciplines. The students will range in education from primary school to graduate school and be categorized into six groups 1) grades 4-5, 2) grades 6-8, 3) grades 9-12, 4) community college, 5) undergraduate, 6) graduate. Employing a task-based methodology, our case study will observe students' cognition in action. Working with an advisory panel comprised of librarians and STEM instructors who represent our student groups, immersive, STEM-based simulations containing various digital resources will be created with Articulate's Storyline software. The proposed project will take place December 2015 – November 2018. Major activities will include: 1) prescreening STEM student participants using a survey; 2) selecting study participants using survey results to ensure diversity; 3) creating and implementing subject-based simulations using Articulate's Storyline software to devise a controlled lab environment; 4) utilizing a think-aloud protocol to discern ~180 participants' choices, behaviors and rationale during video-recorded sessions; 5) coding (using NVivo for videos) and analyzing qualitative and qualitative datasets; 6) disseminating our findings to the library and educational communities and 7) creating workshops, webinars, and customized exercises to propel information literacy-related instruction for students in K12, community college, and university environments

We will use webinars, workshops, and conferences to engage with the community and share the study results in venues targeted to public, school, and academic library professionals as well as educators. We will gather feedback that will help shape the development of student-centered STEM information literacy instruction through the construction of *real-world exercises* for the library and STEM education community that can accompany information literacy-related instruction. These customizable exercises will be placed in open educational resources (OER) repositories (e.g. OER Commons) for access and reuse.

The proposed research project requests funding for five temporary student employees, software, study participant incentives, data storage, and travel-related expenses for project team conference and workshop presentations.

Researching Students' Information Choices: Determining Identity and Judging Credibility in Digital Spaces

Introduction

Investigators at the University of Florida (UF) George A. Smathers Libraries in partnership with researchers at OCLC and Rutgers University request $491,882 (with $232,289 in cost share) for the three-year research project titled, *Researching Students' Information Choices: Determining Identity and Judging Credibility in Digital Spaces*. Our project team is composed of academic librarians, research scientists, an educational technology specialist, and an advisory panel of practitioners (Supporting Document 1). Based on Round 1 reviewers' feedback, a school media researcher joined the research team as co-PI which accounts for the budget increase in Round 2. For this research project, we will consider whether students are *format agnostic*. First coined by Abram and Luther (2004), the term refers to students who either cannot or do not identify the *container* (i.e. document type) when making judgments relating to use of digital resources. Several usage studies have reported that students experience trouble distinguishing among different digital resources, such as e-books and e-journals (Croft, 2004; Levine-Clark, 2006; Shelburne, 2009). Soules (2009) goes so far as to say "E-book, e-journal? Users don't care; in fact they never cared, and many only understood book vs. journal in the print world because of the difference in their physical structures. What they want is relevant content" (pg. S4). But how do students distinguish whether relevant digital content is also credible digital content? Given the limited published research on how late primary, secondary, community college, and undergraduate STEM students identify and determine credibility, this project will provide new knowledge to librarians and educators. This information will be promulgated with practitioners for incorporation into information literacy courses, and other teaching and learning environments to aid STEM students in effectively determining credibility in the discovery, access, and use of digital resources.

For this research case study, we propose to study a diverse group of approximately 180 students working in the science, technology, engineering, and mathematics (STEM) disciplines. The students will range in education from primary school to graduate school and be categorized into six groups 1) grades 4-5, 2) grades 6-8, 3) grades 9-12, 4) community college, 5) undergraduate, 6) graduate. Employing a task-based methodology, our case study will observe students' cognition in action. Using electronic simulations of digital spaces and think-aloud protocols, the project team will examine "point of selection" behavior in real time (i.e. the moment a user determines a piece of information potentially meets a research need). Working with an advisory panel comprised of librarians and STEM instructors who represent our student groups, immersive, STEM-based simulations containing various digital resources will be created with Articulate's Storyline software. Using Storyline we will create a controlled lab environment where subjects with the same education levels can work through the same education-appropriate exercises and content. A simulated web environment also counteracts problems encountered during live experiments such as websites being unavailable and design changes for digital information resources. Our research project is designed to answer the following research questions:

- Do STEM students differentiate among different types of digital resources during point of selection?
  - What cues from a web search results screen do students use to identify digital resources?
  - How do students' characteristics influence their identification behavior?
- How do STEM students determine the credibility of digital resources?
  - What cues from a web search results screen do students use to judge the credibility of digital resources?
  - How do students' characteristics influence their credibility judgments?

1. Statement of Need

A 2014 study of the "Digital Universe" declared that the digital realm is doubling every two years reaching 44 trillion gigabytes by the year 2020 (Turner, et. al., 2014). Navigating this vast amount of information to find what a seeker needs in order to answer a question, solve a problem, or complete a task,

becomes more challenging as the amount of digital information grows. This is particularly problematic for students who, by necessity, must navigate this sea of information as a critical part of their education. Head (2013) recently described this phenomenon in a study of freshmen as an "information tsunami that engulfed them" (p.28). The issue is further heightened for STEM students, because the majority of STEM information resources are digital. These students are faced with thousands of online materials from which to select when searching any given topic. Likely to become society's future scientists, engineers, and doctors, it is imperative to understand STEM students' behaviors and develop approaches to improving their information literacy. During their schooling, STEM students will compete in science fairs, conduct research projects, write lab reports, conference papers, and journal articles, and present posters and presentations, all of which are intended to prepare them to be top performers in their chosen professions. To be successful students must seek, select, critically analyze, synthesize and reconstitute information already created. An additional component of being information literate is the ethical and appropriate use of found information (including correctly citing the original source) (Julien & Barker, 2009).

Quoting one of his students, Clydesdale (2009) writes "It is imperative that someone studying this generation realize that we have the world at our fingertips — and the world has been at our fingertips for our entire lives. I think this access to information seriously undermines this generation's view of authority, especially traditional scholastic authority". Similarly, the ACRL (Association of College & Research Libraries) Framework for Information Literacy for Higher Education (ACRL, 2015) acknowledges in its authority frame that novices must take responsibility to "critically examine all evidence—be it a short blog post or a peer-reviewed conference proceeding—and to ask relevant questions about origins, context, and suitability for the current information need". The issue of authority in the information age is broad. Our objective is to dissect the variable process of determining credibility, a critical aspect of authority, as it relates to assessing digital information.

A great deal of the library and information science literature on digital information, information-seeking and credibility has examined young adults (high school and college-aged undergraduate and graduate students) seeking, engaging, and using online information. Studies report a variety of criteria used to assess digital resources, some of which are new (i.e. web domain) when compared to the criteria used to assess print resources (Head, 2013; Connaway, Lancios, & Hood, 2013). In the print world, the information's container is readily decipherable and plays a significant role in judging credibility. For example, when presented with a book and a journal issue, a student can use several physical characteristics to identify the container, use it as part of the process for judging credibility, and also attribute it correctly. This identity assessment process appears to change in the digital environment. The identity of the container in digital environments has become less clear given the lack of physicality. As Mirtz points out in her study of first year college students' searching behavior, the "ease with which students can access material is confounded by the absence of physical mediation that used to guide and slow down the process" (2013, p.191). Whether students are abandoning the container as a criterion to judge credibility of digital resources is an open question. Although Williams and Rowlands (2007) write that the "Google Generation are format agnostic and have little interest in the containers (reports, book chapters, encyclopedia entries) that provide the context and wrapping for information 'nuggets'," (p. 20) they also call for more research in the area.

Interestingly, the research also indicates that young adults start developing their information seeking methods before they reach college. This actually begins much earlier during their K-12 education and they retain those often-inefficient techniques and assumptions. For instance, children make few to no attempts to assess credibility, skim information for key words or simply use the title to determine usefulness, find the web much faster than searching in books or other resources, and assume Google results are authoritative, because they are generated by Google (Kafai and Bates 1997, Hirsh 1999, Large and Beheshti 2000, Julien and Barker 2009). Flanagin & Metzger (2010) in a survey of 2,747 children, ages 11 to 18 confirms that children are concerned about the credibility of online information, but 89 percent believe that "some" to "a lot" of it is

believable; and, choosing among several options, they rate the Internet as the most believable information. Most have faith in information found on "Wikipedia", but they also consider an article on the Web site of "Encyclopedia Britannica" to be more believable than the identical article posted on "Wikipedia". Flanagin & Metzger also observed that older kids are more rigorous in their assessment of online information than younger ones; younger children are less analytical and more likely to be fooled. Meyers (2010) notes that further exploration is needed in the area of how young people judge information and more recently, McPherson, Gofine, & Stinson (2014) have stated that children and young people demonstrate a need for more guidance around assessing the trustworthiness of online information.

Additionally, epistemological studies posit that people apply previous experience as a scaffold to understand things in their lives and the scaffold serves as infrastructure for making meaning of information they encounter (Hofer and Pintrich, 2002). For the project, we plan to dig into the fundamentals of a novice's ability to recognize the blog post, the proceeding, book, as well as other containers. We also will examine the less studied group of students in grades 4-12 in addition to community college, undergraduate, and graduate students. In our study we will not only observe behavior during live simulation sessions, but also will administer a prescreen survey to gather data about the students, including their technology and library experiences, to determine whether such experiences are related to particular styles of engagement with digital information.

The Digital Visitors and Residents (V&R) framework will be used for this study (Connaway, Lanclos, and Hood 2013a). The framework is a continuum that represents behavior in the Visitor mode, which is when individuals interact online only when required, and the Resident mode, which exemplifies a more pervasive and energetic online presence. The V&R continuum provides a way to describe a wide range of online engagement modes, helps to identify and describe the behaviors that individuals exhibit while searching online, and is a useful tool for understanding their motivations as they act under varying contexts (Connaway, Lanclos, and White 2014; Connaway, Lanclos, White, Le Cornu, and Hood 2012; Connaway, Lanclos, and White 2012; White and Le Cornu 2011). Prior results from V&R studies indicate students in the Emerging educational stage (last year high school/secondary and first year undergraduate college/university students) tended to consider physical books as the most academically credible source, but often used what they considered less authoritative information from the web, such as Wikipedia, but would not disclose these sources. The findings also suggest that it was important to Emerging students that the digital resources they found were authoritative, specifically that they were identified by trustworthy human sources: people who, in their personal experience, had proven to be knowledgeable, either generally, or specifically about the topic (Connaway, White, Lanclos, and Le Cornu 2012). Moreover, the students often were unaware that many of these sources actually came from the library (Connaway, Lanclos, and Hood 2013b) characterizing them as simply "online" resources.

This project will utilize the V&R framework by asking several of the V&R interview questions within the prescreen survey to help identify how students judge credibility of digital resources in absence of human sources. This will enable comparisons of student responses in the original V&R study as well as expand the types of student groups who participated in the original V&R study to include late primary, secondary, and community college students.

In general, the information-seeking process includes the following steps:

1. **Seek/Discover** – the seeker has an information need, goes online, searches and finds;
2. **Select** – the seeker chooses different information pieces for further review;
3. **Assess** – the seeker evaluates the information and how it meets their need;
4. **Use** – the seeker uses the information they determined is valuable; and,
5. **Cite** – the seeker gives appropriate credit to the creator of the information

During this process the information seeker may categorize the information found by its container (i.e. journal article, book chapter, blog post, or government report), and its quality and value in the context of a works-cited list, which in turn may influence how individuals judge or reassess credibility. The concept of ungluing the

content from the container and being *format agnostic* has only been alluded to in pockets of the library and information science literature over the past decade, most notably in e-book usage studies of university students (Croft, 2004; Levine-Clark, 2006; Shelburne, 2009; Primary Research Group, 2009). Although primarily isolated to the open-ended comments of these studies, this issue is noted and a common theme is revealed when compared to previous studies (Shelburne, 2009). More recently, Swanson and Jagman (2015) state in their book, *Not Just Where to Click: Teaching Students How to Think About Information*, "Students do not recognize the differences between sources and bring assumptions with regards to what research looks like and how it works" (pg.2). In one of the book's chapters, Cole, et al. (2015) study the students emerging behind the millennial generation, whom they dubbed Generation Z, and remark "evolving information environment impacts their information-seeking behaviors and, more importantly, their assumptions about information; however *how* and *in what ways* remain unexamined" (pg. 113). Finding no studies directly related to the subject of distinguishing between different digital containers, Cataldo and Buhler sought opportunities to explore this issue. In a pilot study to assess this issue, Cataldo and Buhler (2012) suggest that, as with Rieh's scholar subjects, labeling and source/brand familiarity (i.e. Google, Wikipedia, Chicago Tribune) play key roles (Rieh, 2002). While the survey had its limitations, this preliminary examination provided some initial insight into students' capability to identify the container and the need to ascertain "point of selection" behavior**.** Our proposed study will employ innovative methods to capture "point of selection" in real time.

2. Impact

What role does being *format agnostic* play in being information literate? How are the two concepts connected? Based on an exhaustive literature review and conversations with other researchers the project team believes the proposed case study will be the first to directly examine this connection, defining what it means to be a *format agnostic* student and what impacts the associated behaviors have on the future of information literacy standards and instruction. By placing STEM students as the central focus of this study and capturing real-time "point of selection" behavior, we can create impactful changes to the discovery and utilization of digital scholarly information, and take strategic action to improve students' digital literacy. This research will provide valuable knowledge for practitioners to understand how learners are making these determinations. The diversity of the student population in Gainesville, Florida, and our effort to ensure this is reflected in our sample strengthens our ability to provide results from our case study that will produce national impact (Supporting Document 2). Additionally, successful implementation and use of Storyline may become adopted as an innovative and interactive means for studying online user behavior in a controlled manner.

Goals and Outcomes:

1. **Goal**: Characterize "point of selection" digital information-seeking behavior in STEM fields; define similarities and differences between educational stages (including late primary, secondary, community college, and university).
    - **Outcome**: Inform the development and increase the effectiveness of both information literacy and classroom instruction that is rooted in the students' experience, given their educational stage and mode of online engagement.
    - **Outcome**: Contribute to conversations regarding standards for digital information engagement (e.g. the ACRL Framework for Information Literacy for Higher Education or the AASL Reading Standards Literacy in Science/Technology Crosswalk).
2. **Goal**: Disseminate results and recommendations with school, public, and academic libraries as well as the STEM instruction community.
    - **Outcome:** Use webinars, workshops, and conferences to share the study results in venues targeted to public, secondary, and academic library professionals and gather feedback that will help shape the development of student-centered STEM information literacy instruction.
        - **Deliverable 1:** Create a minimum of *four face-to-face workshops* to be presented at conference venues such as PLA, AASL, ASCD, ALA, and Educause. Some will be

customized to specific practitioner audiences (both librarians and instructors) while others will be more general in nature to allow for cross-disciplinary discussion and collaboration.
    - **Deliverable 2:** Use feedback from workshops to choose content to incorporate into the design of a minimum of *three webinars* targeted to the study's three main library audiences. Webinars will be hosted via OCLC and recorded to ensure broad geographic participation from the targeted communities and provide asynchronous learning opportunities.
    - **Deliverable 3:** Construct *real-world exercises* for the library and STEM education community that can accompany information literacy-related instruction. These customizable exercises will be placed in open educational resources (OER) repositories (e.g. OER Commons) for access and reuse. They will also be reviewed and updated on a regular basis.
3. **Goal**: Expand and build upon the existing digital V&R framework with new student groups and educational stages (e.g. late elementary, middle school, and community college)
    - **Outcome**: Propose new ways to support STEM students in identifying and determining the credibility of digital resources based upon a theoretical framework that addresses different education stages and modes of online engagement.
4. **Goal**: Demonstrate that simulations are an innovative, useful means to study information-seeking behavior within digital spaces.
    - **Outcome**: Provide evidence of a new methodology for information-seeking behavior research to target audiences via conferences and publications.

3. Project Design

Assessing users' information literacy has been an important topic to libraries and library and information science researchers. In a 2009 study, Walsh identified 91 articles that employed nine different methodologies to gauge information literacy, with the top three being: multiple choice questionnaires (34.1%), analysis of bibliographies (18.7%), and quiz/test (15.4%). From studies employing these methods, we also know students either lack or do not utilize the skills necessary to assess online resources (Fei, Sullivan, & Woodall, 2006; Grimes & Boening, 2001; Head, 2013; Palmer & Tucker, 2004; Taylor, 2012). Yet librarians have little insight into how or why students decide to use different types of digital resources. For instance, in his analysis of student bibliographies, Carlson (2006) writes, "Analyzing bibliographies from student research papers revealed what sources students used to support their research, but it could not determine why they made these choices…" (p. 20).

Related library and information science (LIS) research has examined information-seeking behavior, how users seek and evaluate information and have used think-aloud exercises to collect data. The think-aloud protocol is a widely recognized method to examine the human decision-making process (Wang & Soergel, 1998). This method allows researchers to observe the steps of the cognitive process and makes it possible to ascertain where problems are encountered (Wildemuth, 2009) In 1994, Barry explored user-defined relevance criteria; using a think-aloud exercise with 18 subjects (undergraduate/graduate students and faculty) as they reviewed a printed list of DIALOG search results, and three printed full-text documents that were selected by the participants. In addition, the study asked subjects to highlight cues in the search results and/or documents that influenced their selection process. Subsequent studies targeted graduate students and faculty populations and gathered data from 25 and 15 subjects (Wang & Soergel, 1998; Rieh, 2002). Similar to Barry (1994), Wang and Soergel (1998) used printed DIALOG search results using a read aloud/think aloud exercise where participants read the search results aloud and verbalized their thoughts when making selection decisions. In contrast, Rieh's (2002) study took place in a laboratory setting, where subjects were tasked to find new papers that would be useful for one of their current research projects. Rieh (2002) collected and analyzed data from web search logs and post-search interviews.

The methodology for this proposed research project differs from the existing information literacy and information-seeking studies in several ways. First, we will be studying six diverse groups of students ranging

from primary school to graduate school: 1) grades 4-5, 2) grades 6-8, 3) grades 9-12, 4) community college students, 5) university undergraduate students, 6) university graduate students. Second, our research activities will take place in a simulated environment where students within each group will be presented with the same web search results screen where they will make selections in real time. By controlling this aspect of our study, we will create the framework to allow for determining comparisons within the six student groups. Lastly, our case study builds upon prior research through the use of an innovative real-time method of observation that examines "point of selection" behavior. Specifically we will be able to ascertain *if* a student registers a digital resource's *container*; and if the student uses that determination of the type of *container* (along with other criteria) to judge the credibility of the item. A major objective of using this methodology is to determine whether and how the student groups differ within and across the six educational levels when identifying and evaluating the credibility of digital resources, in order to determine how to effectively intervene at the various educational stages with instruction that cultivates and reinforces information literacy development over time.

3.1 Research Methods

Population and Generalizability

The Gainesville community serves as an ideal venue for this case study, due in part to its large, diverse, and representative student body. Our sample will be drawn from a geographic area that is home to the University of Florida (UF), Santa Fe College (SFC), and 44 schools and centers in the Alachua County Public Schools system, and the 12 branches of the Alachua County Library District. It also boasts the PK Yonge Developmental Research school, a public K-12 school affiliated with UF that requires a population of approximately the same demographics as Florida's student population (Supporting Document 2). SFC has a student population of over 16,000 with associate, bachelors and certificate programs in dozens of STEM fields. UF is comprised of 29% minority students and 12% international students. Additionally, according to 2013 data, the University had approximately 19,000 undergraduates and graduates in STEM majors. Both quantitative and qualitative data will be collected during this study via a web-based survey and during a simulation. Our data collection and analysis procedures are discussed below.

Prescreen Survey

A prescreen survey will be created to collect demographics such as nationality, gender, lifetime exposure to internet access, online engagement mode (i.e. visitor or resident), as well as questions on experiences with citation management tools, library exposure and bibliographic instruction (Supporting Document 3). Our advisory panel will provide an expert review of the survey and we will pilot it on five students from each group using a cognitive walkthrough approach (Groves et al., 2009). Survey participants will be recruited with the assistance of the advisory panel members (Supporting Document 1) who interact with the study's target populations in the schools and public libraries from which we are recruiting. The project's student assistants will facilitate the recruitment process (Supporting Document 4).

The prescreen survey will be administered through a cover page/webpage that describes the simulation study (including age appropriate incentives in the form of Amazon gift cards) and the purpose of the prescreen survey. For students in grades 4-12 parents/guardians will receive this material and a written request seeking permission to have their child participate. The survey will be sent directly to community college, undergraduate, and graduate students with a request for their participation.

From the sample of students who are given permission to participate (grades 4-12) or self-select (students at community colleges and universities), the project team will use the prescreen survey results to select diverse groups of participants for the simulation study. Our objective is to have two groups of participants – one that has a high level of internet access and a high perceived level of online savvy, and another group that has a low level of internet access and a low perceived level of online savvy. The additional demographic variables will be used to recruit a diverse set of participants for each group. Our objective in creating *within group differences* is to determine how students' experiences with technology might influence how they identify and evaluate the credibility of digital resources.

Subject-based Simulation

Subject-based simulations (Supporting Documents 5 and 6) that are appropriate for each educational level will be created using Articulate's *Storyline* software. This approach will create a controlled lab environment for our study. The controlled environment serves two purposes. First, it counteracts challenges resulting from performing experiments in naturalistic settings (i.e. a subject's own search on the open web), such as response time or unavailability of a website, sudden changes in the user interface, subjects' search differences (words and/or typos). By creating a simulation where the search results are the same for every participant allows for the findings within student groups to be more comparable. In addition, the simulation software will collect certain data points automatically that would otherwise need to be captured by the think aloud exercise and then coded (i.e. students can click on the cues that help them identify resources, and the software will capture this rather than having them talk aloud about the cues which will then have to be coded).

A pilot study involving approximately five students from each of the six groups will be conducted to test and refine our tasks, think-aloud exercises, simulation software, and post-simulation questions (same as prescreen survey). We will time the exercise and adjust accordingly to avoid excessively long sessions for participants. We will also pre-test whether it is favorable to complete all the simulations in one session versus two sessions, as well as task comprehension and time needed for these two scenarios.

Following the pilot study, an experimental study involving 20-25 students from each of the six populations will be conducted in three tasks, and depending on the results of the pilot, will occur in either one or two sessions. The sample sizes are based on previous studies that reached the point of saturation (i.e. students begin repeating themes) at approximately 20 participants (Hilligoss & Rieh, 2008; Rieh, 2002; Wang & Soergel, 1998; Barry, 1994). A think-aloud exercise also will be employed to assess subjects' cognitive processes which led to their decisions. All sessions will use a laptop with a mouse and will be recorded using Camtasia's *Relay*.

1. Subjects will be presented with a simulated set of Google search results (these will range from 20-40 items depending on the population). They will be asked to select items from the list that they deem credible. They will highlight what they deem to be the indicators (i.e. criteria) of credibility for each resource (whether within the actual resource or search results) that led them to make their determination. The information container will potentially be one of these criteria. If the container is highlighted, the student will be asked follow-up questions regarding what led them to identifying the container (if it is not obvious). Of the items not selected, they will be asked to describe why they view non-selected items as not credible.
2. Subjects will be presented with a list of 20-40 randomized digital items and asked to sort them by the container type (i.e. book, article, blog post, etc.).
3. In this stage subjects will be presented with a *pre-categorized randomized list of digital items by container type*. They will be asked to look at each bin and categorize each item as credible or not credible. They will highlight what they deem to be the indicators of credibility for each resource that led them to make their determination.

Data Analysis

The data analysis phase will consist of four datasets – the survey, the simulations, the post simulation survey, and the think-aloud exercise with accompanied video capture. Analysis of quantitative data from the study will be conducted using Qualtrics's embedded analysis tools as well as the statistical software, SPSS. Analysis will include standard descriptive statistics. The simulations will be pre-coded to provide a certain numerical code, which will be assigned to the different clicks made by each student during the simulation. This will generate a report to allow for quantitative analysis and to inform the investigators of any trends within and across sessions. A codebook will be developed using the themes that emerge from the video captures of the think-aloud activities during the sessions. The videos will be transcribed using scribie and coded by the research assistants using *NVivo* (Supporting Documents 7 & 8). We will ensure inter-coder reliability using standard practices.

3.2 Timeline (See Schedule of Completion)

Year One (December 2015 – November 2016): Prescreen Participants & Pilot Test Simulations

- Hire student personnel, finalize Advisory Panel, obtain IRB approval for survey and simulations
- Design and construct simulations
- Create think-aloud exercises and code the backend of each simulation
- Administer prescreen survey (using methods listed in Research Methods section)
- Identify and schedule pilot study participants
- Devise preliminary code book for think-aloud exercises
- Attend conferences to discuss methodology and pilot, and gain feedback from the library and information science, and educational communities.
- Begin online simulations and think-aloud exercises

Year Two (December 2016 – November 2017): Data Collection, Data Analysis, & Dissemination of Results

- Conclude online simulations and think-aloud exercises
- Finalize codebook for think-aloud exercises
- Transcribe data
- Begin analysis of quantitative and qualitative datasets
- Attend conferences to discuss preliminary findings and gain feedback from the library and information science, and educational communities.

Year Three (December 2017 – November 2018): Data Analysis & Community Engagement

- Conclude analysis of quantitative and qualitative datasets
- Report findings at conference venues and via peer reviewed publications
- Create a minimum of four workshops and three webinars, based on findings and feedback
- Compare, construct and expand Digital V&R frameworks
- Anonymize data and deposit in appropriate repositories for sharing
- Construct real-world exercises and deposit into IR@UF, open educational resource (OER) and other discipline-specific repositories

3.3 Evaluation Plan

The ultimate success of our research project will be measured by uptake and adoption of the research findings by the practitioners. However, there will be several internal points of evaluation during the project. First, adherence to our timeline will be an important mark to ensure that research and deliverables are on schedule. Second, use of the pilot study to ascertain and make appropriate changes to our simulation and/or think-aloud exercise will ensure that we have been collecting the data needed to answer our research questions. Third, engage our Advisory Panel to support recruitment, ensure that our think-aloud exercises are age appropriate, and that our interpretations of our findings are correct. Finally, we will utilize the practical expertise of our Advisory Panel and feedback from the larger community to create applications that translate our research results into practice. Presentations, peer-reviewed publications, and the creation of applications (e.g. workshops, webinars, and exercises) over the course of this project also will help us to evaluate and improve our work.

4. Project Resources: Personnel, Time, and Budget

4.1 Budget

The IMLS request of $491,822 includes undergraduate and graduate student labor/fringe ($[redacted]) plus tuition ($[redacted]), collaboration/dissemination travel ($[redacted]); simulation, transcription, and coding software ($[redacted]), advisors' honoraria ($[redacted]); storage/participant incentives/publishing fees ($[redacted]), and indirect costs ($[redacted]). UF, OCLC and Rutgers plan to contribute $[redacted] in cost share as described below.

4.2 Key Personnel

Amy Buhler, Principal Investigator, Associate University Librarian (.10 FTE cost share for 3 years, totaling $21,990 state funds) is an Engineering Librarian. Amy has previous experience on two grant-awarded

projects from NIH and more recently as co-PI on the NSF project "Gaming Against Plagiarism." Her primary research interests are in the areas of assessment of information seeking behaviors, library instruction, and marketing of library services. **Project Role:** Buhler will oversee the activities of the project; create survey instrument; provide content for the simulations; create think aloud protocols; conduct think aloud portions of the experiment; analyze resulting coded data; share results and solicit feedback via various professional forums, workshops, and webinars.

Tara Tobin Cataldo, Co-Principal Investigator, Associate University Librarian (.10 FTE cost share for 3 years, totaling $[redacted] state funds): As Biological/Life Sciences Librarian and Collections Coordinator, Cataldo provides collection management, instruction, literature search assistance, and faculty/staff consultations in the life sciences. In addition, she oversees all collaborative collections activities in the sciences, including the management of an over $1.5 million collections budget plus endowments. Her primary research interests are in the areas of subject specialization in academic libraries, information seeking behavior in the sciences, and the use of electronic information resources. **Project Role:** Cataldo will co-manage the activities of the project; supervise assistants, create the survey instruments; provide content for the simulations; create think aloud protocols; conduct think aloud portions of the experiment; code/analyze resulting data; and disseminate results.

Ixchel M. Faniel, Ph.D., Co-Principal Investigator, (.10 FTE cost share for 3 years, totaling $[redacted] private funds) is Research Scientist at OCLC. Her current research interests include examining the reuse of research data in academic communities to identify how contextual information about the data that supports reuse can best be created and preserved. Of particular interested is how scholars evaluate the data when deciding whether to reuse it. She is therefore very knowledgeable of the role trust and credibility play during discovery, access, and evaluation of content. She has received funding from the National Science Foundation, the Department of Defense through Michigan State University, and IMLS. **Project Role:** Faniel will review the data collection instruments; review and provide input on the data analysis plans; and participate in data collection, data analysis, and dissemination of results.

Lynn Silipigni Connaway, Ph.D., Co-Principal Investigator, (.10 FTE cost share for 3 years, totaling $[redacted] private funds), is a Senior Research Scientist at OCLC Research. She has experience in academic, public, and school libraries, as well as library and information science education. Connaway has completed several Jisc-funded and IMLS-funded projects to investigate users' behaviors, including virtual research environments and digital repositories. Connaway currently is studying digital visitors and residents. Prior to joining OCLC Research, she was the Vice-President of Research and Library Systems at NetLibrary (an ebook provider), the director of the Library and Information Services Department at the University of Denver, and on the faculty of the Library and Informational Science program at the University of Missouri, Columbia. **Project Role:** Connaway will review the data collection instruments; review and provide input on the data analysis plans; and participate in data analysis and dissemination of results; provide input on the data analysis plans; and participate in data analysis and dissemination of results.

Joyce Kasman Valenza, Co-Principal Investigator, (IMLS request for summer salary/fringe.09 FTE years 1 and 2, plus .05 FTE year 3, totaling $[redacted]; and .03 FTE for 3 years cost share, totaling $[redacted] state funds) is an Assistant Professor of Teaching, Director of MLIS Program, Rutgers University, has been studying and writing about young people, technology, information seeking, and information fluency for more than twenty years. She has 25 years of instructional experience as a K12 librarian. Joyce has authored books on information skills, developed several video series for Schlessinger Media, and written articles on digital literacy. Joyce is a Milken Educator, a Google Certified Teacher, and a Library of Congress American Memory Fellow. She was selected as one of *Technology and Learning's* 100@30, was honored with the 2011 Edublogs Award for Lifetime Achievement. **Project Role**: Valenza will participate in creation of the prescreen surveys; inform the content of the online simulations; contributing to data analysis; attend and present research findings at conferences engage in dialogue with the practitioner community (particularly school librarians); and contribute to the design and content of instructional outreach targeted to K12 librarians and educators.

Randy Graff, Ph.D., Investigator, (.01 FTE for 3 years cost share, totaling [redacted] state funds) is the Director of Educational Technology at UF Health. Graff has been a teacher for the past 18 years and is experienced with classroom teaching and training in instructional, administrative and assistive technology. He currently teaches faculty and staff at UF in the use of instructional and operational technology. Graff has co-authored inter-professional manuscripts which appear in refereed journals. He has conducted numerous university presentations, taught undergraduate and graduate courses, and designed and taught many technology workshops. He is a two-time award winner of the UF IT Distinguished Leadership Award. **Project Role:** Graff will review data collection instruments; supervise the student constructing the online simulations; and contribute to data analysis.

Rachael Elrod, M.Ed., MSLS, Investigator, (.05 FTE for 3 years cost share, totaling, $[redacted] state funds). As Head of the Education Library at the University of Florida, provides staff supervision, collection management, library instruction, literature search assistance, and faculty/staff consultations in the areas of early childhood education; educational leadership, educational technology; higher education administration; marriage and family counseling; mathematics education; mental health counseling; research, evaluation, and measurement methodology; school counseling and guidance; school psychology; social foundations of education; social studies education; special education; statistics education; and student personnel in higher education. Her primary research interests are in the areas of information literacy and makerspaces in academic libraries. **Project Role:** Elrod will contribute to the survey instrument; provide content for the simulations; contribute to think aloud protocols; and contribute to data analysis.

Fiscal Management Personnel: Project finances will be managed by the Libraries' Fiscal Unit Coordinator, Grace Strawn, who has been with UF since 1994 and coordinates daily management of all Libraries financial and accounting activities for all funding sources (state, government grants, foundation and auxiliary) for managing all grant funds and monitoring/managing budget allocations for all seven Libraries for salary and fringe, OPS and other expense categories; and Robert Sessions, B.S.B.A. in Management from UF, and research administrator for UF's Contracts and Grants Accounting Services assigned to the Libraries for managing the receipt, cost share reporting and sponsor reporting functions related to grants.

5. Communication Plan

Updates on the project and dissemination of findings will occur through several channels including an OCLC Research project website, social media, conferences, publications, workshops and OCLC Research webinars. All team members will take part dissemination, promotion, stakeholder engagement, and outreach. Among the duties of the program assistant will be team communication and promotion including keeping webpages, social media, etc. up to date (Supporting Document 9). The advisory panel also will play an important role in communicating with practitioners in their fields and assisting with translation of research to practice. Communication with our stakeholders on the projects' progress and results will be ongoing and occur via national and regional conferences for libraries in general, higher education, school libraries, public libraries, and education (Supporting Document 10). The ideal forms will be recorded or streamed panels and round tables where communication can be archived and shared to wider audiences. In year three, workshops and webinars will be presented to the various communities to disseminate full findings of the study and translate that research into recommendations for best practices. We will measure professional engagement with audience counts, tracking online communications (listservs, etc.), and monitoring metrics of all publications and deposited applications. As the project team is located in three organizations (UF, OCLC, and Rutgers); regular communication will be vital to the study's implementation and evaluation. Email will be used continuously, with the addition of monthly scheduled teleconference/videoconference calls, and quarterly progress reports. Project management tools will be utilized for best efficiency. The full team will hold a face-to-face meeting in Gainesville each year of the project. All research data will be stored securely and centrally at UF's Research Computing facilities, which will allow all investigators to have easy access. Final anonymous datasets will be placed in an open institutional repository.

ORIGINAL PRELIMINARY PROPOSAL

Researching Students' Information Choices: Determining Identity and Judging Credibility in Digital Spaces

Project:

Investigators at the University of Florida (UF) George A. Smathers Libraries in partnership with researchers at OCLC request $446,247 for the three-year research project titled, *Researching Students' Information Choices: Determining Identity and Judging Credibility in Digital Spaces*. Drawing from prior research we propose students operating in digital spaces (e.g. the open web) are "format agnostic" (Abram and Luther 2004). Usage studies have shown students cannot or do not identify the document type (i.e. container) when viewing digital information resources (Croft 2004; Levine-Clark 2006, Shelburne 2009). Soules (2009) goes as far to say "Users don't care; in fact they never cared, and many only understood book vs. journal in the print world because of the difference in their physical structures. What they want is relevant content" (pg. S4). But how do they know relevant content also is credible content? Using a population of 200 students, this project poses two overarching research questions: 1) *How do students (grades 4-12, community college, and university) working in science, technology, engineering, and mathematics (STEM) disciplines identify types of digital resources?* 2) *How do students determine the credibility of digital resources?*

Through the use of electronic simulations of the digital spaces and think-aloud protocols, the project team will examine "point of selection" behavior in real time (i.e. the moment a user determines a piece of information potentially meets a research need). Working with a project advisory panel of librarians (school, public, college and university) and STEM instructors representing the 3 student populations, immersive, STEM-based simulations with the various digital resources will be created using Articulate's Storyline software. Storyline has the power to create a controlled lab environment where subjects can work through the same age-appropriate exercises (i.e. science projects for 4-12, research papers for college) using the same content. It also counteracts problems encountered during live experiments such as websites being unavailable and design changes for digital information resources.

The Digital Visitors and Residents (V&R) framework will be used for this study (Connaway, Lanclos, and Hood 2013a). When in Visitor mode, individuals interact online only when required, while Resident mode exemplifies a more pervasive and energetic online presence. The V&R continuum provides a way to describe a wide range of online engagement modes, helps to identify and describe the behaviors that individuals exhibit while online, and is a useful tool for understanding their motivations as they act under varying contexts (Connaway, Lanclos, and White 2014; Connaway, Lanclos, White, Le Cornu, and Hood 2012; Connaway, Lanclos, and White 2012; White and Le Cornu 2011). Prior results from V&R studies indicate students in the Emerging educational stage (last year high school/secondary and first year undergraduate college/university students) tended to consider physical books as the most academically credible source, but often used what they considered less authoritative information from the web. The findings also suggest that it was important to Emerging students that the digital resources they found were authoritative, specifically that they were identified by trustworthy human sources: people who, in their personal experience, had proven to be knowledgeable, either generally, or specifically about the topic (Connaway, White, Lanclos, and Le Cornu 2012). Moreover, the students often were unaware that many of these sources actually came from the library (Connaway, Lanclos, and Hood 2013b) characterizing them as simply "online" resources.

This project adds to the V&R work by focusing on how students judge credibility of digital resources in absence of human sources and expanding the student groups to include late primary, secondary, and community college students. Given limited published research on how late primary, secondary, community college, and undergraduate STEM students identify and determine credibility, this project will inform librarians, educators, educational technology developers and providers, and STEM publishers how to develop, teach, market, and provide digital resources for STEM students. This information can be shared in information literacy courses and in other teaching and learning environments to aid STEM students in effectively determining credibility in the discovery, access, and use of digital resources.

Key Personnel: PI, Amy Buhler, Associate Chair & Engineering Librarian, UF; Co-PI, Tara Tobin Cataldo, Collections Coordinator, UF; Co-PI, Ixchel M. Faniel, Ph.D., Associate Research Scientist, OCLC Research; Co-PI, Lynn Silipigni Connaway, Ph.D., Senior Research Scientist, OCLC Research; Investigator, Rachael Elrod Head of Education Library, UF; and Investigator, Randy Graff, Ph.D., Director of Educational Technology, UF.

Goals and Outcomes:

1. Goal: Characterize "point of selection" digital information-seeking behavior in STEM fields; define similarities and differences between educational stages (including late primary, secondary, community college, and university).
    - Outcome: Inform the development and increase the effectiveness of information literacy instruction that is rooted in the students' experience, given their educational stage and mode of online engagement.
2. Goal: Expand and build upon the existing digital V&R framework with new student groups and educational stages (e.g. late elementary, middle school, and community college)
    - Outcome: Propose new ways to support STEM students' in identifying and determining the credibility of digital resources based upon a theoretical framework that addresses different education stages and modes of online engagement.
3. Goal: Provide outreach to public, school, and academic libraries
    - Outcome: Use webinars, workshops, and conferences to share our results in venues targeted to public, secondary, and academic library professionals and gather feedback that will help shape the development of student-centered STEM information literacy instruction.
    - Outcome: Provide real-world exercises to the library community that can accompany STEM information literacy instruction. These customizable exercises will be placed in a repository for access and reuse by the library community. They will also be reviewed and updated on a regular basis.
4. Goal: Validate simulations as an innovative and successful method for studying information-seeking behavior within digital spaces.
    - Outcome: Provide evidence of a new methodology for information-seeking behavior research.

Impact:

By making the student the central focus of this study and capturing real-time "point of selection" behavior when seeking STEM information, this research will provide valuable insights for the practitioner to create impactful changes to the discovery, functionality, and utilization of digital scholarly information, improve student's information literacy, and establish the life-long learning skills tomorrow's professionals need. This research project will:

- Inform the establishment of standards for digital information engagement (e.g. the ACRL Information Literacy Competency Standards for Higher Education Task Force which is charged with updating the 2000 standards).
- Inform new methods for teaching information literacy skills and more targeted bibliographic instruction based on the results of the behaviors demonstrated by STEM students in simulations.
- Provide leverage for those working with writing program instructors by demonstrating evidence of students' natural interactions with and decision-making processes related to digital information
- Inform the development of new library services being established in the digital spaces (e.g. bookless libraries)
- Inform the producers of digital information how to best design resources for the STEM information seeker

Work plan:

In Year 1 the team will hire student assistants; establish the advisory panel; create, administer, and analyze prescreening surveys to identify participants; create, pilot test, and administer the simulations and think aloud protocols. In Year 2 the team will transcribe and analyze the data – surveys, simulation data, videos of think-aloud sessions; submit and present to targeted conferences (PLA, ACRL, ALA, etc.) and run panel sessions at those conferences to gather feedback. In Year 3 the team will hold workshops and webinars to share the findings and recommendations with practitioners from the library community, as well as write articles and white papers.

Budget:

The IMLS request of $446,247 includes temporary student labor/fringe ($[redacted]), collaboration/dissemination travel ($[redacted], simulation software ($[redacted], advisor honoraria/transcription fees ($[redacted]), storage/participant incentives/publishing fees ($[redacted]), tuition reimbursement ($[redacted]), and indirect costs ($[redacted]).Cost share for UF and OCLC personnel effort and additional conference travel will be documented in the round-two budget if invited to submit.