PROJECT DESCRIPTION

Facilitating Museum Evaluation with Real-Time Data Mining

1. PROJECT OVERVIEW AND RATIONALE

This 3-year *Full Scale Project* makes novel use of familiar technology to address the immediate and pressing challenge of affordable, ongoing, large-scale museum evaluation (Adams, 2012; Smithsonian, 2004) while encouraging visitors to explore and engage deeply with museum content.

1.1. Project Goals and Focus

To realize these complementary aims, the project proposes to develop and implement an innovative app for mobile devices, composed of two parts: 1) a front-end "virtual scientist" called *Dr. Universe* (*Dr. U*) for use by museum visitors (and docents) that doubles as an unobtrusive data-gatherer and 2) a back-end analytics portal to be mined by museum staff, evaluators, and researchers (*Figure 1*). The project builds on a pilot version of a simple question-answer app already in beta testing. With a close partnership between universities and museums, we will further develop this app to function as a platform for STEM informal education, research, and cost-effective, data-driven decision making by museum staff.

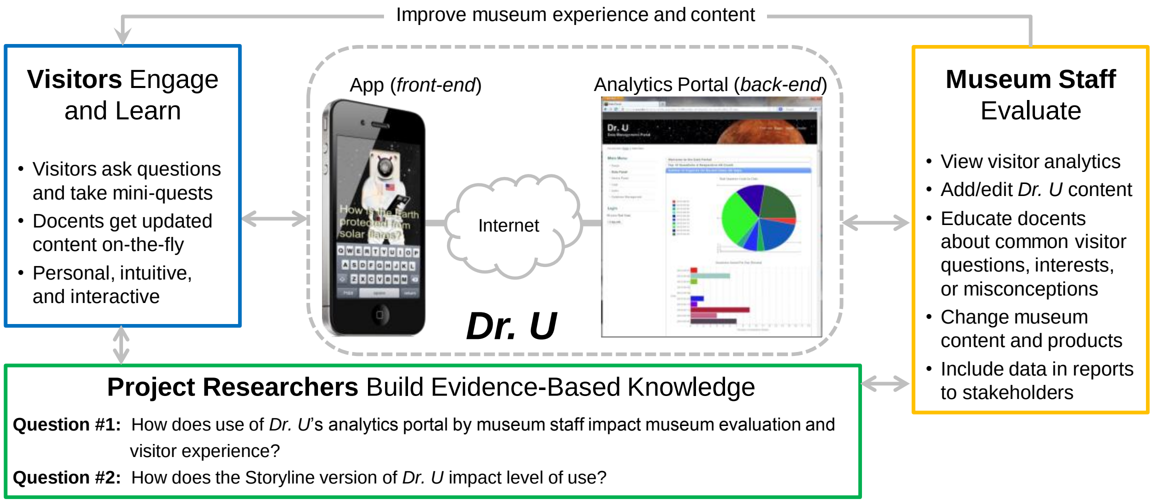

Figure 1. *Project Overview—The Dr. U systems consists of a front-end app and back-end web portal.*

This project aims to exploit the increasing presence and use of mobile devices. Uniquely, the *Dr. U* app will be designed to deeply engage museum visitors (including those who may have been underserved by traditional museum experiences due to prior knowledge or language difficulties), while connecting with an analytic system to make sense, in real time (immediately and continuously), of the large amounts of data produced by visitors' use of the app. The analytic system will help museum staff access and interpret ongoing evaluation data, regardless of experience or museum resources, informing the practice of professionals at the front lines of informal STEM education in diverse communities. The intent is *not* to replace the expertise and services of a professional evaluator, but to enable museums to engage in daily, practical assessment when the services of an evaluator are not available or attainable.

To achieve these goals, we will 1) modify the beta *Dr. U* app to include situated mini-quests featuring simple storylines. We will investigate the power of this enhanced app to bolster sustained engagement for visitors diverse in terms of age, race, language, and literacy; 2) develop a set of novel, open-source analytic tools that make the gathering and interpretation of contextual information from visitors' app use accessible to museum staff and educators, building their capacity for using data in their day-to-day work; and 3) provide the complete system as an open-source research platform to allow others to probe questions about informal learning and motivation, effective application of large datasets for museum evaluation, and ways to encourage and understand use of mobile virtual experiences. This project will complete all four components of "design and development" research specified by the *Common Guidelines for Education Research and Development* (2013) and complies with the justification and evidence guidelines recommended therein. The close relationship between the project's goals, research questions, evaluation questions, outcomes, and impacts is outlined in detail in *Figure 2* (*Section 3*).

With the aid of our partners, the *Arizona Science Center* (ASC) and *Arizona Museum of Natural History* (AZMNH) in greater Phoenix, the project will serve three audiences:

Museum Visitors**:** Both museum partners serve large and diverse populations. The ASC serves 500,000 guests per year, including nearly 140,000 students. The AZMNH has served over 1.8 million visitors since 2000, including thousands of students from hundreds of schools annually. *Dr. U* will enhance the museum experience for visitors of all ages and science knowledge/comfort by 1) answering visitor questions in an understandable, unintimidating way (that is safer and more focused than a general search via commercial systems such as Apple's *Siri*, IBM's *Watson*, the *Wolfram Alpha* engine, etc.) and 2) increasing their engagement with the exhibits. *Dr. U*'s Q&A database will be written at an 8th grade level and the app will support speech/audible input-output to encourage use by visitors with lower literacy. The app will be offered in English and Spanish to be inclusive of the 20% of the populations served by our partner museums who speak Spanish as their primary language (U.S. Census Bureau, 2012).

Museum Staff**:** The back-end database and analytic tools will provide museum staff with continuous, accessible, real-time evaluation data that may be used to make just-in-time tweaks to exhibit content (during formative periods), to modify exhibit pathways to reflect user preferences or facilitate learning, or to change flexible materials and resources (such as multimedia, lecture topics, docent training or locations, and experience carts) as current events or visitors' needs and interests change. The questions and types of data used in traditional museum evaluations (from time spent at an exhibit to what visitors found most interesting), are extensively documented in the literature (e.g., guides such as Diamond et al. 2009; journals such as *Visitors Studies, Journal of Museum Education;* websites such as *Assessment Tools in Informal Science*), and provide the baseline of analytic tools that will be included in the *Dr. U* back-end for use by our museum partners. However, the questions traditionally asked are often constrained by cumbersome data-collection techniques or limited resources, so we aim to capitalize on the ease and flexibility of app-based data collection to develop new analytics as well, guided by our museum partners and advisory panel members who will provide critical ideas and feedback on the tools throughout the project (see *Section 4*). Additionally, *Dr. U* may be used as a resource by human docents or to augment offerings in lower-resourced museums that are not able to have extensive human docent presence.

Although the "hardscape" of museum exhibits (such as physical exhibits, artifacts or permanent interpretive text) may be difficult to change, the content of a holistic exhibit experience includes many flexible aspects that may be adjusted, including docents (their location throughout the museum and their content-specific training, as well as the content of their experience carts), multimedia aspects (content on computer kiosks, television monitors, or other recordings), public event opportunities (lectures, specially-themed family days, etc.), posters and other more temporary signage and text, and the museum's website. In discussion with our museum partners, they are very enthusiastic about access to a continuous window into the minds of their visitors in order to make changes to existing products and services and uncover new ideas for opportunities to engage and serve their visitors. For examples of the analytic outputs and what actionable information they can reveal, please see the "Data Mining" section of this response above.

Research Professionals**:** The system will serve as a platform to allow project researchers (and others with interests such as parent-child interactions, effects of current events on engagement, etc.; see *Section 2.4*) to probe interesting questions that relate to visitor learning and motivation, the effective application of "big data" for museum evaluation, and ways to encourage and understand use of mobile experiences.

The **intellectual merit** of this project is that it will advance the fields of visitor studies and museum evaluation (including providing tools to support practitioners), informal science learning, and situated engagement. This is outlined in *Section 1.2*. **Broader impacts** are addressed within *Section 2.4.*

"Only a handful of institutions—13 percent—have someone on staff that spends all or part of his or her time on evaluation"
Smithsonian, 2004

1.2. Prior Research, Project Innovation and Advancement of Knowledge

This project's research pertains to: museum evaluation issues, data mining of large datasets, engagement, mobile device use, and situated learning. Here we highlight key literature that informs this proposal and describe how this project will advance knowledge in these areas as well as advance museum evaluation practice and the capacity of museum staff to use data in their day-to-day work.

Museum Evaluation Issues: The Smithsonian (2004) found in a national study that although almost three-quarters of museums said they were emphasizing evaluation more than ever, "only a handful of institutions—13 percent—have someone on staff that spends all or part of his or her time on evaluation of exhibits or programs" (p. 15). Museum evaluation in most institutions is tacked on to the responsibilities of other staff. As Adams (2012) points out in her museum review, "While the 'want to' in the practitioner's desire to conduct evaluation is much stronger than ever before, the 'can do' is a different matter" (p. 28). While many museums have adopted mobile apps, "improving the ability to measure impact using new digital technologies is a largely unmet but critical need" (Johnson et al. 2011, p. 6). Though the need is clear, with museum staff stretched thin it is crucial to put data in their hands in a quick, reliable, cost-effective way. The *Dr. U* system will do so by unobtrusively collecting and presenting data from visitors engaged with museums' content.

Real-Time Data Mining: Interest in data mining of online user actions has increased as portals such as Google, Facebook, Twitter, etc. have learned to capitalize financially on the information social networks collect via users. Log files have been used as research tools in e-learning, virtual environments, and computerized agent research for years (e.g. Cockburn & McKenzie, 2001; Ingram & Northcote, 1999). The proliferation of smartphones (Pew, 2013) allows informal learning settings to leverage data mining practices to better understand their visitors and experiences as log files from mobile apps can provide snapshots and longitudinal information on visitors' thinking, use, and engagement with exhibits. A flexible, affordable app providing such data could be useful to museums of all sizes (science museums range from ~3,500 to 700,000 ft^2 [ASTC, 2011, p. 2]) and resources, with or without staff evaluators.

Dr. U will generate such data via log files containing records of visitor interactions and inquiries (and the path and evolution of inquiries); the novel application of existing techniques for analysis of "big data" from other fields can shed light on visitor interests, understandings, and misconceptions. By representing log files as vectors of word frequency in user questions to *Dr. U*, we can employ latent semantic analysis (Deerwester et al., 1990) and closely related principal and independent component analyses to classify sets of questions into more sophisticated broader topics. This technique has been applied successfully to analysis of internet chat room logs (Kolenda et al., 2001) and for counterterrorism threat detection in instant messaging conversations (Bengel et al., 2004), and is currently in use by one Co-PI in the NSF-funded SAVE Science project. We will advance basic topic detection by employing temporal clustering analysis, wherein log files are divided into short-duration segments and analyzed jointly to track changes in questions and topics with time; analyze length of question-answer interactive sessions, types and complexity of entered questions, and patterns and repetitive interactions; explore additional statistics such as distribution functions of derived data parameters and more advanced pattern recognition algorithms to identify common behaviors both between and within individual user sets, including threshold clustering functions (Lee & Spergel, 2010) and scientific matched filter techniques (Allen et al., 2012, Malloy & Lidz, 2013), such as those developed for radio telescope data analysis by PI J. Bowman (Bowman et al., 2013). As useful metrics are identified, they will be added to live analytic tools available to the museums.

Engagement in Informal Environments: Informal learning is characterized as motivating, spontaneous, learner-centered, open-ended, and non-assessed (Wellington, 1990; Ramsey-Gassert, 1997; Eshach, 2007; Vadeboncouer, 2006). The characteristics of informal learning are considered to be factors that contribute to learning and interest in science by offering a new way to experience science, different from traditional classroom learning. Abraham-Silver (2006) explains that students in informal learning environments show increased attention and enthusiasm, as well as a willingness to observe, question, and discuss objects around them. Thus, informal learning environments may serve to engage those who do not have a strong interest in science. A primary evaluation question museums ask is whether exhibits encourage deep exploration while exposing visitors to up-to-date content, as visitors often "flit from exhibit to exhibit" (Bickmore et al., 2011, p. 55), not engaging extensively with the content. The personalized content afforded by mobile technologies may offer a solution. There is an inherent tension, however, between allowing visitors to free-flow through the museum and directing them to deeper interactions. Decades of findings on how visitors learn in museums indicate that "one should expect learning to be highly personal and strongly influenced by an individual's past knowledge, previous experiences, and personal interests… [and] by an individual's desire to choose and control his/her own learning" (Falk, 2007, p. 6). The storylines in *Dr. U* will be designed accordingly, and its impact on engagement assessed.

Situated Learning: Situated learning theorists posit that all learning is situated in specific contexts, and that the contexts in which people learn can help or hinder their learning (Lave & Wenger, 1992; Brown et al., 1989). A good match between the thing being learned and the context in which it is learned will help promote learning. Thus, a possible solution to the tension between free-flow and directed pathways through a museum is to provide mini-quests, featuring simple situated storylines, tied to interactions with Dr. U. These mini-quests can, we propose, bolster engagement without interfering with the open-ended, visitor-driven museum experience. Visitor use of the mini-quests will also generate data for the evaluation analytics component of our research. Successful creation of interactive storylines that promote engagement with exhibits and *Dr. U* requires a theory-based, systematic approach to design to create exhibit integration opportunities that produce meaningful evidentiary data for more powerful evaluation. We will make use of our *Global Evidence Channels* (*GEC*) framework (Nelson et al., 2012) in designing the mini-quests. The *GEC*s track visitor interactions with *Dr. U* at a given exhibit in three categories: Location/Movement, Object Interaction, and Communication Activities. The combined *GEC* data that *Dr. U* will gather will provide information on visitor engagement and learning as they move through an exhibit, allowing for more rapid formative evaluation and richer summative evaluation over time.

Smartphone facts:

- ✓ 56% of all U.S. adults own a smartphone
- ✓ 80% of 18-34 year-olds own a smartphone
- ✓ Black and Hispanic American ownership rates are ***above*** the national average

Pew Res. Center, 2013

Mobile Device Use: As of 2013, 56% of **all** U.S. adults owned a smartphone (91% of all adults own some mobile phone) (Pew, 2013). The prevalence is even higher among young adults aged 18-24 (79%) and 25 to 34 (81%). Defying common stereotypes, Pew found 64% of African Americans and 60% of Hispanic Americans own smartphones, compared to 53% of White Americans, and that even among those with household incomes under $30,000, 43% own a smartphone. These data suggest that there will be relatively equal barriers (or lack thereof) to ease of use and adoption of the *Dr. U* app by museum visitors. Exploiting the trend in ownership, many museums are adopting mobile experiences (AAM, 2011). The *New Media Consortium Horizon Report* found that "Mobile Apps are the most relevant features of mobiles for museums right now" (Johnson et al., 2011, p.7). Tallon (2012) surveyed hundreds of museums, finding that the top reasons to enact mobile experiences are to provide additional information, engage visitors, and offer interactive experiences.

Implementing a mobile app comes with inherent challenges: Tallon's list includes producing content, keeping experiences up to date, and encouraging users to use the mobile experience—three things that we will address actively with this project. Falk and Dierking (2008) point out, "it needs to be demonstrated that these new technologies enhance the visitor experience better than competing technologies and in ways that are cost effective" (p. 28). The National Research Council (2009) acknowledges the potential of mobile devices in museums, but cautions that "further research is needed on the potential for the devices to support learning” (p. 270). This project will contribute to such knowledge-building.

STEM Content Areas: The STEM content area for this project is Earth and space science. The project’s role is not to be the primary content provider, but to enhance the ability of the museum exhibits to deliver Earth and space science content by providing additional, detailed information about those topics (including the latest science results) through the *Dr. U* system (see *Section 2*). This system could be used for any content area; Earth and space science was selected because: 1) The PI and one Co-PI have expertise in the field and positions within ASU’s School of Earth and Space Exploration and 2) Museum partners indicated this content is of particular interest to their visitors and one they would like to enhance. The specific themes in *Dr. U* will be tailored to the exhibit(s) selected for study at both partner museums.

1.3. Results from Prior NSF Support and Related Research

Dr. U builds upon findings from previous research with the *Dr. C* agent and *SAVE Science*. (Prior support for J. Bowman’s astrophysics research is summarized in *Section 4.1*.)

Dr. C: Developed with in-kind support from NASA by C. Bowman, two APA scientist mentors (*Dr. C-1 & -2*) were designed to provide career or career-and-psychosocial (interpersonal-focused) mentoring (Kram, 1985) in a simple, low-bandwidth online system created for use in schools. Middle school students (n=532) were randomly assigned to two versions of *Dr. C* or to a control during a three-week NASA science inquiry curriculum. Log files showed students accessed the *Dr. C*s on an “as needed” basis with peaks that corresponded to introductory, information-gathering, and synthesis and reporting curricular segments (Bowman, 2011). Findings (Bowman, 2008) indicated that, on average, students gained in content knowledge but there was no statistically significant difference between the three conditions. Interviews with a sub-sample (n=70), however, suggested students believed the agents facilitated learning, eased the workload, provided trusted information, and were enjoyable to use.

SAVE Science: The on-going *SAVE Science* project (DRL-0822308, 9/1/08-8/31/14, $2,772,482), conducted by Nelson, uses assessment modules embedded in a virtual environment. Log files are used to understand students’ problem-solving behaviors, allowing assessment of students’ problem-solving processes and their solutions. The project created four assessment and two introductory modules and a dashboard for teachers and researchers providing access to raw and aggregated student data. 2000 students and 15 teachers at 12 schools have participated thus far. **Broader impacts:** The project has helped improve teacher practices by providing them with detailed data to inform differentiated instruction. Two journal articles (and three in preparation), seven proceedings, and a dozen presentations have resulted (e.g. Ketelhut, 2010; Ketelhut et al., 2009; Majerich et al., 2011; Majerich et al., 2010; Schifter et al., 2012). **Intellectual merit:** Results indicate the assessment lends new insights into student understanding and misconceptions not captured by traditional tests (Ketelhut et al., 2013; Nelson et al., 2010); the assessments can be designed to lower extraneous cognitive load through visual cues ($p<0.05$; n=193; Nelson et al., 2009); students are engaged and interested in taking the test (Ketelhut et al., 2013).

2. RESEARCH DESIGN AND METHODOLOGY

We propose a 3-year effort to investigate *Dr. U* as a museum evaluation data-gatherer and a computerized virtual scientist with whom visitors interact. We plan a series of multi-day museum implementation studies in which we will provide familiar mobile devices (iPad Mini) with *Dr. U* to museum visitors in order to conduct the proposed research activities. The final product will be usable for visitors on their own smartphones or with the research set of devices we will leave with our museum partners following the study. The project is guided by the research (*Section 2.2*) and evaluation questions (*Section 3*).

2.1. Project Deliverables and Products

The *Dr. U* platform consists of: 1) the front-end virtual scientist that runs on a mobile device and interacts with museum visitors, and 2) the back-end central server database and website portal that collects data from use of the app and provides access to the data and associated analytic plots, tools, and outputs to museum staff and project investigators.

Mobile App (Front-End)**:** The app will be implemented in iOS for Apple devices (iPad, iPhone, iPod) and feature an interface in the form of an animated astronaut, *Dr. U*, floating against a star field. *Dr. U's* appearance will be informed by design guidelines from APA research literature (e.g., Bailenson et al., 2006; Baylor & Kim, 2009; D'Mello et al., 2008; Lusk & Atkinson, 2007; Mayer & Moreno, 2002; McQuiggan et al., 2008; Woo, 2009) and will have simplified, neutral facial features that are partially obscured by the iconic gold-coated helmet and a body encased in a spacesuit. The neutral design of *Dr. U* is chosen considering the conflicting evidence in existing research on the effects of APA qualities like gender, race, age, and realism (e.g. Baylor & Plant, 2005; Wang et al., 2007; Yee et al., 2007). *Dr. U* will exhibit moderate, coordinated levels of simple facial expression and gesture (e.g., smiling, blinking, waving, and "texting" an answer). Visitors will interact with *Dr. U* through text and speech interfaces in both English and Spanish. Spell correction will be utilized on typed input. The app will comply with guidelines for public audience materials (plainlanguage.gov; usability.gov/guidelines/index.html).

Alpert (2013) points out that when creating materials for museums one "must design… offerings to attract and to hold audience attention" (p. 17) but must also be user-driven. With this in mind, engagement with *Dr. U* must allow the visitor to drive the experience pathway. Two challenges to the effective use of the *Dr. U* app as a tool for sustained engagement in museums are 1) the decontextualized nature of the app and 2) the need for of an engaging set of activities used in conjunction with *Dr. U* to increase the likelihood of voluntary user uptake of the app and to foster continued use throughout a visit. The beta version of *Dr. U* offers a simple "question-answer" mode: a museum visitor types a question into *Dr. U*'s text box, *Dr. U* "reads" the question and "texts" back an answer. To bolster user uptake, engagement, and learning with *Dr. U*, we will extend the beta version with the "Storyline" version explained below.

We will investigate the use of the Unity game development tool to create Dr. U. Unity allows for rapid porting of mobile apps across platforms. We will focus first on the iOS application, and second on the Android platform. Should time and resources permit, we will also port to Windows Phone devices. Prof. Nelson has extensive experience developing Unity applications across multiple platforms for two previous NSF-funded projects.

Storyline mini-quests (Front-End): We will design a "storyline" mini-quest approach to encourage museum-goers to engage with the museum exhibit and interact with *Dr. U* in order to generate the data needed for analytics and evaluation. The **Storyline** approach will encourage use of the question/answer capabilities of *Dr. U*. In this version of the app, *Dr. U* will be floating in space with a spaceship visible in the distance. When visitors first open the app, they will be introduced to a narrative in which *Dr. U* asks them for help getting back to the ship. *Dr. U* will state that every time the visitor asks a question, *Dr. U*'s jetpack will fire one impulse, bringing *Dr. U* closer to the safety of the ship. If a visitor's question-asking gets *Dr. U* back to the spaceship, *Dr. U* will then request help in getting to a nearby satellite to investigate it for repairs. There will be many additional space objects that *Dr. U* can ask for aid in visiting if necessitated by extensive question-asking by the user (so that there is not an unintended, system-imposed limit on asking questions). Based on a decade of research into use of gamelike features for improving student science learning (Clark, Nelson, Sengupta, & D'Angelo, 2009), we hypothesize that the storyline motivators will promote use of the app and help accumulate evaluative data for the museum. Further, we hypothesize based on situated theory that they will enhance the visitor's engagement with the exhibit. As part of the study, we will investigate the comparative power of this approach against the simple Q&A app to provide information to museum staff and researchers on visitors' needs, likes, and learning.

The Simple Storyline version of Dr. U will be designed to encourage use of the question/answer capabilities of Dr. U. When visitors first open this version of the app, they will be introduced to a narrative in which Dr. U asks them for help in completing a series of short, simple mini-quests. To keep the museum visitor focus centered on the exhibit itself, while encouraging greater use of the Q&A system in Dr. U, the mini-quests will specifically be designed to follow a casual game ethos: simple to use, easy to master, quick to complete, and engaging.

As they start playing the first mini-quest, visitors will see a “Space Map” showing Dr. U floating far out in space. On the opposite side of the map will be Dr. U’s spaceship. The overall goal of the quest is to help Dr. U get back to the safety of the ship. Between Dr. U and the ship, there will be a number of space objects that can be reached to replenish oxygen, water, and other supplies used to keep Dr. U healthy on the way to the spaceship. For example, near to Dr U’s initial location will be an abandoned spacelab. Beyond that there will be several telecommunications satellites, the Hubble Space Telescope, and other human-made objects. The app will tell visitors that every time they ask a question, Dr. U’s jetpack will fire one impulse, bringing Dr. U closer to the next space object and toward the ultimate destination…the ship. Reaching each space object will replenish Dr. U’s supplies for the next leg of the journey. In addition, visitors will ‘level up’ in expertise from “Space Newbie” through to “Master Explorer”, with intermediate levels of expertise incremented upon reaching each space object (and, consequently, asking more questions of the Dr. U app to bolster the volume of evaluation data).

There will be multiple routes available to move Dr. U to the safety of the spaceship. The ‘Safety First’ route will let visitors move Dr. U from space object to space object with relative ease by asking an increasing, preset number of questions between each. The ‘Daring Doc.’ route will let visitors take a more direct path to the spaceship, but require more questions be asked to move from Dr. U’s given location to more distant space objects along the way. A ‘Go for Broke’ route will allow visitors to aim directly for the spaceship from any given location, but require them to ask a long series of questions all at once in order to reach safety. Should players ask too few questions in a given attempt to move between space objects, Dr. U will retreat back to the most recent space object.

In the initial version we will hard code in the numbers of required questions and space objects; following initial testing we will develop the system to self-adjust based on the number of exhibit items a museum has, and/or the number of topics for which a museum enters questions and answers into the backend system. So, for example, if a given exhibit has 10 items to see, the app would generate 10 space objects on the map, including the spaceship.

In our experience, it is not possible to completely isolate the development of the evaluation tool from the user experience. In order to get the volume of data needed to conduct meaningful analytics for use by museum staff in evaluating their offerings, it is necessary to make the app something that improves visitors' experiences and encourages them to use it. We are determined to design an experience that bolsters the museum experience, promotes greater learning about museum content, and facilitates rapid museum evaluation. Dr. C. Bowman and Prof. Nelson will bring to bear more than a decade of experience in designing virtual agent and game-based learning environments to do both.

The theoretical rationale for the Simple Storyline version of Dr. U is rooted primarily in a situated learning perspective. As we write in the proposal, situated learning theorists posit that all learning is situated in specific contexts, and that the contexts in which people learn can help or hinder their learning. A good match between the thing being learned and the context in which it is learned will help promote learning. In our case, we propose that a good match between context and use can also bolster sustained engagement with an activity (asking questions of Dr. U). The more visitors use the Dr. U application, the richer the data that can be provided to the museum. We hypothesize (and will investigate) that instantiation of situated learning principles via a simple storyline scenario featuring short mini-quests will improve sustained usage of the application.

It can be argued that the Simple Storyline version of Dr. U is ‘doubly situated’. On one level, visitors will be using Dr. U in the context of exploring a museum exhibit. Dr. U will be there with the museum visitors to provide immediate answers to questions that arise in situ. This is in contrast to a decontextualized FAQ Q&A application that museum visitors might have access to online, away from the museum. By being in the museum with visitors, Dr. U shares the context and experience with them in real-time: a prototypical situated experience. One additional benefit of this approach is that visitor focus remains firmly on the exhibit, with Dr. U available in the background at any time to answer questions.

The storyline version of Dr. U adds a second level of 'situatedness' through the use of its basic narrative and associated mini-quests. By providing visitors with an engaging, game-like framework for asking questions of Dr. U, we provide an additional context for use that we posit will have appeal beyond the foundational context of asking questions as part of the exhibit exploration experience. However, use of storyline with Dr. U poses some design and implementation challenges. The richer and more complex the storyline, the greater the likelihood that visitors will shift their focus from the museum exhibit toward the Dr. U app. Our approach is to keep the storyline simple and the quests short. The mini-quests will situate visitors in an engaging scenario to bolster engagement with the app, but will be designed to be completed in the background, with each action in the mini-quest consisting of simply asking a question of Dr. U.

Analytics and Management Web Portal (Back-End)**:** The back-end of the platform is a password-protected web portal hosted on a central server. The portal provides a single access point for researchers and museum staff to explore the information obtained from user interactions with the mobile app, as well as to modify the content of the question/answer database. The mobile app is able to share its logs and ancillary information with the central server and receive content database updates. Fewer than half of museums targeting mobile experiences currently offer free public WiFi networks throughout their galleries (Tallon, 2012) so the *Dr. U* app will operate without requiring a dedicated network connection. Each mobile instance will maintain its own copy of the question/answer database and temporary logs. It will attempt to contact the central server periodically and make all pending data transfers when a successful connection is present (e.g., at night when the devices are able to connect to a museum's private staff network). Logs will contain entered questions and retrieved answers along with time-stamps, available position information from iBeacon, device ID, mini-quest interactions, and other ancillary data for each entry. On the server, all log information is stored in an SQL-compatible database. It is processed and made accessible through the web portal where information and statistics are compiled into high-quality user-friendly charts and data visualizations using the freely-licensed Google Chart API (or open-source equivalent) and the Joomla content management system used by millions of websites. Simple input forms on the web portal allow museum staff to update *Dr. U*'s question/answer database. All changes to the database will be logged by user and time-stamp to facilitate engagement research.

The Dr. U log files will contain records of visitor interactions and inquiries (and the path and evolution of inquiries). Using analytic techniques developed for use with "big data," we will be able to rapidly analyze all logged data in an automated, real-time way. As log files are transferred and stored on the backend server, individual user interactions will be analyzed, as well as the aggregate of many users. The automated analysis will occur at least once every 24 hours (or as frequently as needed by our museum partners) through use of scheduled "cron" jobs on the server.

Our data mining analysis begins with the automated generation of simple statistics related to the core visitor experience. We will generate statistics for total length of question-answer interactive sessions with Dr. U, number of interactions per visitor, and the types and complexity of entered questions (using word length and number of words). The Dr. U analysis web portal will present these statistics and derived parameters (e.g. distribution functions such as histograms of interaction times, a "Top 10" list of popular questions, etc.) to museum staff in order to provide a foundation for understanding the visitor experience. We anticipate that inspection of the complexity of entered questions will provide museum staff with an independent assessment of the communication proficiency of visitors that may lead to future exhibit materials and/or docent interactions better targeted to the particular populations served by each museum.

The more powerful, second tier of data mining will employ latent semantic analysis (Deerwester et al., 1990) and principal and independent component analyses, in which we perform a SVD of the matrix of word-question combinations in our knowledge database to reduce the dimensionality from the initial 10,000+ questions to a subset of primary singular vector modes (of order 100). We will then compare the primary singular vector modes to user-entered word vectors (similarity to each mode will be assessed through a dot-product operation). This technique has been applied successfully to analysis of internet chat room logs (Kolenda et al., 2001) and instant messaging conversations (Bengel et al., 2004), and is currently in use by one Co-PI in the NSF-funded SAVE Science project. Using the comparison to the lower-dimensional space resulting from the SVD, we expect to be able to:

1. Identify the primary topics of interest of a given user more accurately than through simple keyword matching.

2. Identify non-obvious cross-cutting interests of users by classifying sets of questions into more sophisticated broader topics. For example, visitors may ask large volumes of questions associated primarily with human exploration of space. In such a situation, our SVD analysis could reveal a secondary significant theme focused on biological factors. Knowing this, the museum staff could invite a topical speaker or give docents a set of prompts about the sub-topic to use to engage visitors on the floor.

3. Track the evolution of a user's interests with time as they explore an exhibit or multiple exhibits. We will advance basic topic detection by employing temporal clustering analysis. In this, log files are divided into short-duration segments and analyzed jointly to track changes in questions and topics with time. Specifically, we will break each user's interaction log into temporal bins and perform the comparison for the word vectors derived from each bin independently to identify primary singular vector modes (interests) associated with each bin. Our analytics portal will provide data views to allow staff and researchers to identify patterns in the temporal analysis (e.g. do visitors spend more time thinking about one topic or another, and in what order to do they engage in topic themes throughout the museum experience). We will further investigate the extent to which the temporal clustering of interests correlates with visitor location in the museum, under the hypothesis that it serves as a proxy indicating the location of the visitor associated with specific exhibits or content within exhibits. Through this data mining analysis, we expect to be able to inform museum staff of the level of visitor engagement with particular exhibit content.

4. Track the broad evolution of all visitors' interests with time by analyzing the aggregate log data in temporal bins. This may identify large shifts in interest in response to e.g. popular news coverage of a NASA mission or of a natural disaster.

The third tier of proposed data mining analysis will implement algorithms to identify patterns in the visitor populations. Here, we will take the derived parameters from the first two tiers of analysis as inputs and employ pattern recognition to identify common behaviors both between and within individual user sets. One algorithm we plan to investigate is threshold clustering functions (Lee & Spergel, 2010). Use of clustering functions may allow us to identify natural divisions in the parameter space (e.g. to be able to separate visitors that ask many, complicated questions from those that did not engage significantly with Dr. U) in order to study specific categories of visitors in more detail. For example, we may be able to identify visitors that enter the museum with significant background knowledge of space science and then assess how their experience with exhibits differs from that of visitors who likely have limited background knowledge.

As our investigation of pattern recognition progresses, we expect to accumulate a library of template visitor patterns. This library will be used in the analytics portal automated data analysis to identify and separate logs into distinct populations so that museum staff will have access to statistics for each identified visitor population.

If unique log patterns are identified through this analysis, we will integrate matched filter techniques (Allen et al., 2012, Malloy & Lidz, 2013) into the automated analysis in order to extract all occurrences of the behavior for detailed study.

As useful metrics are identified, they will be added to live analytic tools available to the museums.

Although these techniques may seem to diverge from standard museum evaluation data analysis, they actually align with them closely—but on a much larger scale. For example, museum evaluators conducting exit interviews or surveys might ask visitors specific factual questions about exhibit content, or more general questions such as, "What did you find most interesting about…?" Data mining of Dr. U's log files can provide insights into both the specific (such as misconceptions revealed by the questions asked) and the general (inferring visitor curiosity and interest through topic frequency for individuals and participants overall). An additional strength of collecting visitor data via the Dr. U app is that all visitor interactions with the app will be unobtrusively captured, not just the potentially filtered responses that a visitor might provide to an interviewer or the self-edited comments made by a visitor wearing a microphone during their visit (e.g., Allen, 2002). The app's analytics should yield a richer picture of visitor understandings, interests, confusions, misconceptions, and even frustrations—both "in the moment" and across their visit. Combining data from hundreds or thousands of visitors will allow museum staff to target their efforts on changes or enhancements that will be the most impactful in their setting.

Analyses of these large-scale results will be informed by project research data (including demographic information) and potentially by docent and evaluator observations. Together, these may show patterns in the types of questions asked (or not asked) by certain groups, which could suggest changes to docent staffing, explanatory materials, or exhibit routing to be more understandable, less intimidating, or more culturally sensitive.

Heritage and preliminary testing: The *Dr. U* system builds on direct heritage in algorithms and software design from the *Dr. C* precursor. First implemented as a virtual science mentor for a middle school science inquiry study, *Dr. C* now serves as the primary public FAQ system for NASA's highly-visited Mars Exploration Program site. Preliminary development and beta testing of a basic English-language version of the *Dr. U* app and analytics portal has been conducted. A team of undergraduate students at ASU from the schools of Earth and Space Exploration, computer science, art, and engineering developed the beta system in 2012, including a 10,000+ question/answer database on the Earth, Solar System, and Universe from NASA's public-domain information. Basic functionality of the system has been demonstrated in tests at several museum spaces on the ASU campus. The proposed system stemming from this beta will be built at ASU.

2.2. Measures of Outcomes

We will invite museum visitors to participate in our project and randomly assign them to one of three groups: no *Dr. U,* Q&A-only *Dr. U,* and Storyline *Dr. U*. All participants will be given a pre- and post-survey (details below) including demographic questions. See subsequent section for analytic process.

Research Question #1 How does use of the analytics web portal by museum staff impact museum evaluation and visitor experience? The technical aspect of this research question will be investigated by J. Bowman. Outcomes related to the technical data mining component are summarized above. Enhancements to analytics and the web portal component will be guided by extensive existing museum evaluation literature and the wishes and requirements of the museum partners, and will be informed by best practices from the robust field of data-driven decision making in formal education (e.g., Mandinach et al., 2006). The analytics research will center on developing automated statistical tests and metrics that provide insight into usage patterns and characterize users in ways relevant to museum evaluation. The elements of this question focused on use and utility of this system for museum staff will be investigated by C. Bowman and are informed by the data sources outlined in *Tables 1* and *2*. The Evaluation Plan is designed to provide external evaluation of the project and simultaneously to contribute to the knowledge-building efforts. The categories that relate to the outcomes for museum staff are centered on understanding/knowledge as well as skills (defined by Friedman, 2008). The measures include interviews and surveys to probe how museum staff perceive and use the analytic tools and web portal to understand their visitors, alter their museum content, and inform their evaluation efforts and reporting. Secondary analysis of the database to see when, in what ways, and with what frequency the museum staff update the database content or access and manipulate the analytic tools will also inform these questions.

Research Question #2: How does the Storyline version impact level of use of *Dr. U*? Lead by Nelson, overall usage level of *Dr. U* will be tallied as a count of questions asked of *Dr. U* by each visitor as well as number of quests undertaken. Sustained usage will be measured by recording the frequency of questions asked from the time of the first to the last access of the *Dr. U* app by a visitor in a single museum visit. We will match questions asked to specific items in the exhibit, examining question patterns for variations in usage between the two *Dr. U* versions and against the *Global Evidence Channels*. Although not a primary research question, changes in visitor interest following use of *Dr. U* may be investigated in concert with related evaluation questions (see Section 3).

2.3. Procedure for Evidence-Based Knowledge-Building and Broadening Participation

Target Audience and Sample: Museum visitors will be offered the opportunity to use *Dr. U* during test implementations. Our museum partners serve large and diverse urban populations, so participants are anticipated to reflect the diverse demographics of the museums' locations. Small-scale tests will be conducted in the first year. Following this, we will conduct two implementation tests in each partner museum each year. Members of the research team will attend each implementation test to work with the museums. In each test, we will lend 25 mobile devices to museum visitors over the course of 3-5 days. Assuming each mobile device is given to a minimum of four visitors (or visitor groups) per day, the sample on any given day of implementation testing will be approximately 100 users. Assuming average 3-day implementation tests (24 days of testing in total), we forecast a minimum total sample of roughly 2,400 study participants, with the potential for many more. Our museum partners (as indicated in their letters of commitment) plan to engage in this project at all levels of their organizations, from the directors of education to the docents and teen volunteers. These professionals and volunteers represent our second target audience as partners in exploring the evaluation capacity of the system. They will work closely with the research team (described in *Section 4*), providing expertise and substantive feedback.

All visitor groups entering the museum will be offered the opportunity to use Dr. U when devices are available for lending (if visitors express interest when all the devices are out on loan, they will be invited to "play" with the booth-based device either at that moment or on their way out). Note that only one device per group will be lent. While this is experimentally messy, it is the reality of doing research in a museum setting. The data collected will be analogous to traditional museum evaluation methods which may ask questions of an individual visitor, but cannot "control" for that visitor's responses or experiences being mediated by interactions within their group during the visit. On the post-survey, the primary visitors (who took the pre-survey) will be asked to quantify and describe their impressions of their group's dynamics and use of the device.

We believe that the inclusion of a control group is appropriate. The control group is not selected randomly from all museum-goers. Instead, the control group, as well as the two treatment groups, will be selected, at random, from a group of participants who have already: 1) agreed to participate in the study, 2) signed the informed consent, and 3) participated in the pre-survey (see page 8 of the project description). Of those who have already gone through those three stages, some will receive a mobile device with the Q&A version of Dr. U, some will receive a device with the Q&A + Storyline version, and some will not be given a device, but will be told they may try one out at the end when they return to take the post-survey.

Regarding the sample, our museum partners serve large and diverse urban populations, so participants in the study are anticipated to reflect the diverse demographics of the museums' locations. Since visitors to a science center or natural history museum (free-choice learning environments) are already somewhat self-selected to be interested in the topics presented in the museum, which intersect with science and technology, we expect most visitors entering the museum to be interested in participating with Dr. U and

for there to be at least one person in their visitor group familiar with smartphones. As Pew, 2013, reported, as of 2013 56% of all U.S. adults owned a smartphone. The prevalence is even higher among young adults aged 18-24 (79%) and 25 to 34 (81%). Defying common stereotypes, Pew found 64% of African Americans and 60% of Hispanic Americans own smartphones, and even among those with household incomes under $30,000, 43% own a smartphone. While we do not anticipate that the self-selected sample will skew in a specific demographic direction, if we notice specific trends in demographics (reported on our pre-survey) that do not reflect the particular museum's overall visitor demographics, we may engage in purposeful selection (Maxwell, 2005). Both the pre- and post-survey will include closed- and open-ended questions that are informed by our research questions and also the needs and interests of our museum partners. There are extensive guidelines for crafting both types of questions (everything from word order, to scale length, to question order, to techniques for minimizing error and encouraging item response) as well as demographic questions. In this, we will mainly look to three primary resources: Dillman et al., 2009; Fowler, 2009; Krosnick & Presser, 2010. The first two, as well as Converse & Presser, 1986, also provide guidance for testing our instruments, which we will follow before using the instruments for data collection. Our external evaluator will design the external evaluation instruments per her extensive experience and will also provide expert feedback to us on our pre- and post-surveys. We will use quantitative (data mining, multiple regression) and qualitative methods of data analysis to explore constructs related to our research questions. Data sources include participant demographics, survey and interview data, observations and log file data. We will look for patterns and themes, developed first within-case and then across-cases and visitor qualities, and build matrices (Miles & Huberman, 1994). We will cross-check codes to attend to inter-rater reliability and alternative interpretations.
Given Pew's (2013) research on the ownership of smartphones, we expect most museum visitors to be users of the app (Pew found that as of 2013 56% of all U.S. adults owned a smartphone. The prevalence is even higher among young adults aged 18-24 [79%] and 25 to 34 [81%]. Defying common stereotypes, Pew found 64% of African Americans and 60% of Hispanic Americans own smartphones, and even among those with household incomes under $30,000, 43% own a smartphone). Visitors to a science center or natural history museum (free-choice learning environments) are already somewhat self-selected to be interested in the topics presented in the museum, which intersect with science and technology. Therefore, during the study, we expect most visitors entering the museum (all of whom will be offered the opportunity to participate when devices are available) to be interested in participating. During the study's implementation tests, visitors will be limited by the number of devices available. In the final year, however, the app will be available for download on a visitor's own device, which will extend participation greatly. To extend the opportunity to those without smartphones, a version of the app will be made available to our museum partners for inclusion on a kiosk (and they will keep a set of the study devices to lend as they wish). While we do not anticipate that the self-selected sample will skew in a specific demographic direction, if we notice specific trends in demographics (reported on our pre-survey) that do not reflect the particular museum's visitor demographics, we may engage in purposeful selection (Maxwell, 2005).

Broadening Participation: Our museum partners take seriously their commitment to broadening participation in STEM learning for the diverse communities they inhabit. This project is closely aligned with their efforts in both direct and indirect ways. Though counterintuitive, it is the *in*direct mechanism (providing evaluation data) that is most likely to broaden participation, as museum staff use the evaluation data provided by the data portal and analytics to uncover and act on visitor interests, misconceptions, and concerns for a range of visitor types. For example, project research data (which collects demographic information) along with docent observations may show patterns in the types of questions asked (or *not* asked) by certain groups, which could suggest changes to docent staffing, explanatory materials, or exhibit routing to be more understandable, less intimidating, or more culturally sensitive. Additionally, it is possible that certain groups of visitors who might be less likely to ask questions of staff (such as visitors with lower content knowledge or lower language skills) would be more likely to inquire within the nonjudgmental online space (Ketelhut et al., 2005). In a more obvious way, the design of the app itself aims to appeal to all audiences. Functionally, the spelling-correction feature of the software and the fact that visitors may speak rather than type their question (using the same native capabilities in iOS employed by *Siri*, effective in even high-noise settings) is inclusive of younger visitors or those with lower literacy.

Logistical Considerations: The project mobile devices will be provided for each test and reused from one implementation to another, allowing us to make the devices available to exhibit visitors who would like to use *Dr. U* regardless of whether or not they have a device of their own. The user interface and all associated text will be made available in English and Spanish. Project team members will provide usage instruction (in both languages) to those without prior experience with mobile devices. For decades, museums have been lending auditory devices and other materials to visitors and have developed their own mechanisms for loan and retrieval (e.g., Laursen, 2013; Ucko & Ellenbogen, 2008). We will use each museum's method of choice and provide the logistical/human support for lending and use of the *Dr. U*-enabled devices. Every effort will be made not to inconvenience the museum staff or visitors in any way.

Implementation Procedure: As visitors enter the museum, they will be asked if they want to participate in the *Dr. U* research project (with a signature required on an IRB-approved consent form for participation). Following approval, participants will be asked to take a short survey (marked with a unique identifier that associates it with the mobile device being lent) on demographics and science interest (data from which will inform the research and the external evaluation). Using the museum's normal procedure for delivering electronic aids, the mobile devices with *Dr. U* installed will be handed out to a randomly selected subset of participants, with the rest serving as a control group. Of those receiving the mobile devices, a random half will receive the Q&A version of *Dr. U* and half will receive the Q&A + Storyline version (the control group participants will not be given a device, but will be told they may try one out at the end when they return to take the post-survey). The research team will be available to explain use of the mobile device. Treatment participants will be told that *Dr. U* is designed to complement a specific exhibit (the actual exhibit name will be substituted). All participants will be left to explore the museum and targeted exhibit on their own with their *Dr. U* interactions recorded in the log files. The research team (and external evaluator, as appropriate) will take observational notes of participants' behavior and interactions with *Dr. U*. As participants leave, they will return to the mobile devices (if in a treatment group) and be given the post survey on interest and their experience (marked with a unique identifier associating it with the returned device or the control). Randomly selected participants will be asked to participate in a short interview. All participants, including control subjects, will be given a fun sticker or temporary tattoo that has a code allowing them to access *Dr. U* on a website from a non-museum computer to provide an exploratory indication of interest. All participants will receive a small reward.

Data Analysis: We will use quantitative (data mining, multiple regression) and qualitative methods to explore the research questions (see Section 3 for additional detail). Data sources include participant demographics, survey and interview data, observations and log file data. We will look for patterns and themes, developed first within-case and then across-cases, and build matrices (Miles & Huberman, 1994). We will cross-check codes to attend to inter-rater reliability and alternative interpretations. Qualitative data will complement quantitative findings. As an example, to investigate the impact of *Dr. U*'s treatments on participants' use of *Dr. U.*, participants will be randomly assigned to one of two treatments, represented by a dichotomous predictor: *DRUSTORY*=1 if assigned to storyline *Dr. U*; *DRUSTORY*=0 if assigned to Q&A *Dr. U*. To assess the impact of treatment on use of *Dr. U*, we will regress *DRUUSE*, defined as the total number of user-initiated interactions with *Dr. U*, on treatment (*DRUSTORY)*, controlling for demographics (*AGE*, *RACE*, *ENGLISH*, *FEMALE*) and prior science interest (*INTEREST*), as collected from the pre-survey. The hypothesized multiple regression model is below:

$$DRUUSE_i = ß_0 + ß_1 DRUSTORY_{ij} + ß_2 INTEREST_{ij} + ß_3 AGE_{ij} + ß_4 RACE_{ij} + ß_5 ENGLISH_{ij} + ß_6 FEMALE_{ij} + \varepsilon_{ij} + u_{ij}$$

where $ß_0$ is the intercept parameter; $ß_1$ represents the effect on total usage due to using *Dr. U* Q&A or Storyline; $ß_2$ - $ß_6$ represent the effects of the control variables, and $\varepsilon_{ij} + u_{ij}$ represent random effects. We will investigate expected interactions between control and predictor variables in explaining *DRUUSE*. Power and effect size calculations indicate we should be able to detect even small effects with high power across sites. If exploratory analyses of the pre-post survey on interest suggest potential differences, a similar model may be employed with change in interest (*CHGINT*) as the outcome variable and a series of dichotomous predictors indicating assignment to one of the two treatments or the control.

2.4. Intended Impacts

The logic model in the External Evaluation Plan (Section 3) summarizes the overarching strategic impacts of the project and shows their alignment to the project's activities, outputs, and outcomes, while *Table 2* outlines the public (visitor) and professional (museum staff and researcher) audience impacts.

Broader Impacts: The *Dr. U* app and its editable server database and analytic tools will be documented and disseminated through multiple outlets and made freely available to any informal education venue that wishes to use or adapt it to their context and content (*Section 5* below). This will provide museum staff with a cost-effective, manageable way to collect evaluation data that may be unattainable otherwise. Another impact is the development of a framework for designing situated, quest-based learning experiences utilizing mobile agents for use in informal settings and helping a diverse population of museum visitors better engage with exhibit content. By enhancing the visitor experience and improving museum access to data for evaluation, *Dr. U* will have direct and indirect impact on all types of museum visitors. To aid researchers investigating evaluation in informal settings, app source code will be released as open-source at the conclusion of the project. The project also will have impact outside the education field. Identifying users' geographic locations from web-post content and related tags is an active area of research for online companies like Twitter, Google, Foursquare, and Facebook (Ahmedx et al., 2013; Cho et al., 2011; Cheng et al., 2011; Eisenstein et al., 2010; Mei et al., 2006). *Dr. U* will enable novel investigation of spatiotemporal clustering of questions in log files that may extend existing efforts and lead to new techniques to identify mobile user locations within well-described indoor environments.

3. EVALUATION

The project aims to develop an app to serve three goals and audiences (visitors, museum staff, and researchers), so the evaluation goals and questions follow suit. The goals flow chart (*Figure 2*), logic model (*Figure 3*), questions and analytics chart (*Table 1*), and impacts and indicators chart (*Table 2*) show the relationships between the project goals, research and evaluation questions, and the audiences and their outputs, outcomes, and impacts (Friedman, 2008). The evaluation is designed to provide external evaluation and contribute to project knowledge-building through shared data collection.

The external evaluation consists of a needs assessment and implementation evaluation in Year 1, formative evaluation in year 2, and a summative evaluation in year 3. These will be conducted by the external evaluator, Jodie Hoffman, M.A., who spent over a decade as an educational evaluator at WestEd. She will be aided by Dr. C. Bowman, who will assist with the development of measures and the collection of data during the field site implementations. C. Bowman has training and experience in education evaluation and in the creation and implementation of animated agents and is local to the sites. Dr. C. Bowman will be the primary liaison with the external evaluator.

The needs assessment and formative evaluation will include critical review of the app, mini-quests, and the analytic system in light of design theory, museum evaluation literature, and museum visitor and staff needs (via site visits, surveys, and interviews with partner museum staff and visitors). The results will help the team balance research goals with critical considerations of usability, usefulness, and use. Hoffman will determine whether project activities are on schedule and aligned with the research plan, will confirm compliance with Alpert's (2013) "principles of good stewardship" for work between museums and researchers (pp. 41-45), and will provide recommendations in annual reports. All appropriate human-subjects protective measures will be followed. Approval of research and evaluation activities by the Institutional Review Boards on Human Subjects at ASU will be acquired before study funding.

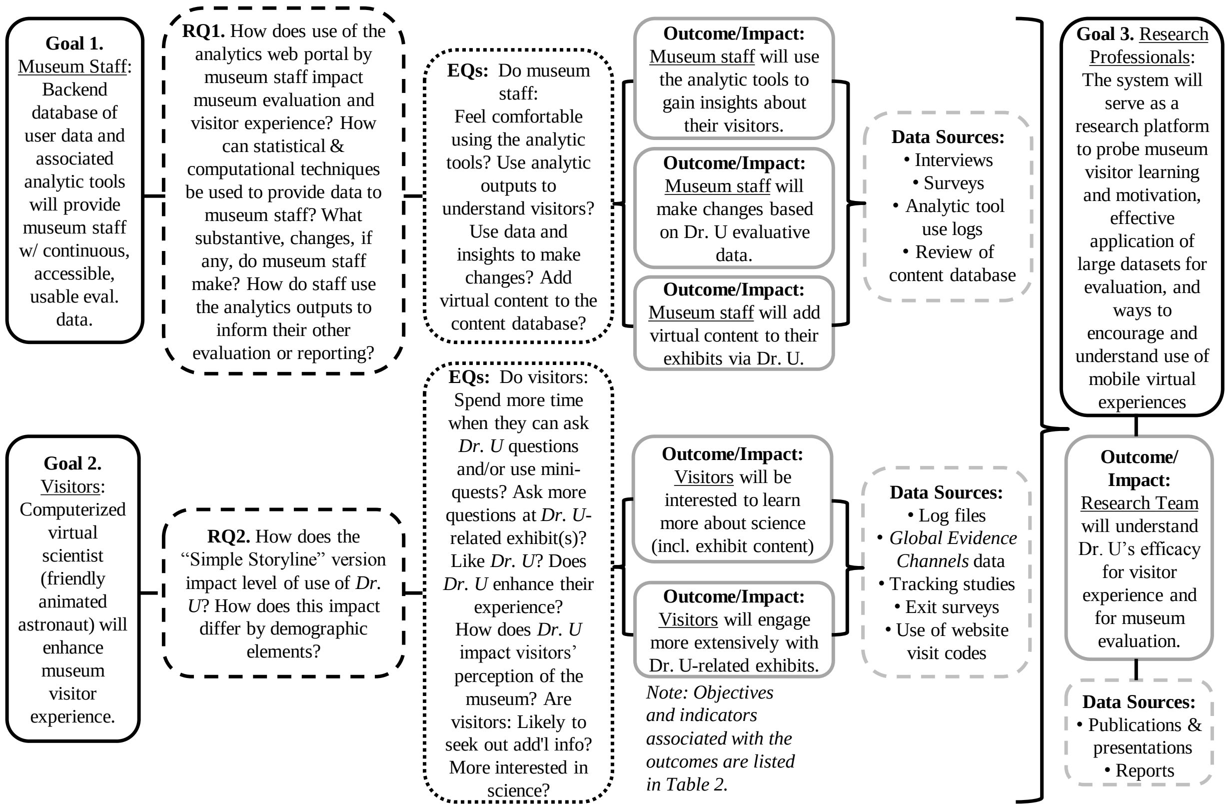

Figure 2. *Goals flow chart*

Timeline: Year 1: Hoffman will design a needs assessment guided by best practices from museum evaluation resources (e.g., Diamond et al., 2009; *informalscience.org*) to establish a baseline of the current evaluation capacity of the partner museums, their overarching evaluation goals, and what evaluative data and tools would be of most use. She also will conduct an implementation evaluation "describing and documenting" project activities, assuring that "essential elements are in place and operating" (NSF, 2010, pp. 8-9) and the project is adhering to "principles of good stewardship" (Alpert, 2013, p. 41) for collaboration with museums. At the end of Year 1 she will provide a report charting progress toward goals. Year 2: The questions that guide the formative evaluation align to the outcomes/ impacts identified in our Logic Model (*Figure 2*) and impacts/indicators chart (*Table 2*). Hoffman will provide ongoing feedback to the team about whether project activities are on schedule and aligned with the research plan and about the health of the museum-research collaboration. She will provide a formal annual report with findings and recommendations. Year 3: The summative evaluation will assess whether the implementation and research activities met the goals, served the audiences, and answered the research questions of the project. A final report will be provided to the research team and to practitioners and researchers through *informalscience.org*.

Methods and Procedures: We will develop survey, interview, and tracking instruments as appropriate following best practices (e.g., Dillman et al., 2009; Fowler, 2009; Krosnick & Presser, 2010). Surveys will be tested, refined, retested, and finalized following standard procedures (Converse & Presser, 1986; Dillman et al., 2009). They will be administered on paper or digitally, depending on the audience and setting. Interview questions will be drafted, reviewed by project researchers and museum partners, and revised. The questions will be designed following established interview question development guidelines (e.g., Kvale, 1996; Maxwell, 2005; Maykut & Morehouse, 1994; Strauss & Corbin, 1998). Interviews will be conducted by phone or in person. Tracking studies will be conducted by C. Bowman with the aid of project undergraduate students, under direction of Hoffman. Depending on the implementation size, all or a subset of visitors will be included in the sample for data collection. If it is a subset, they will be selected randomly or, as appropriate, using purposeful selection (Maxwell, 2005).

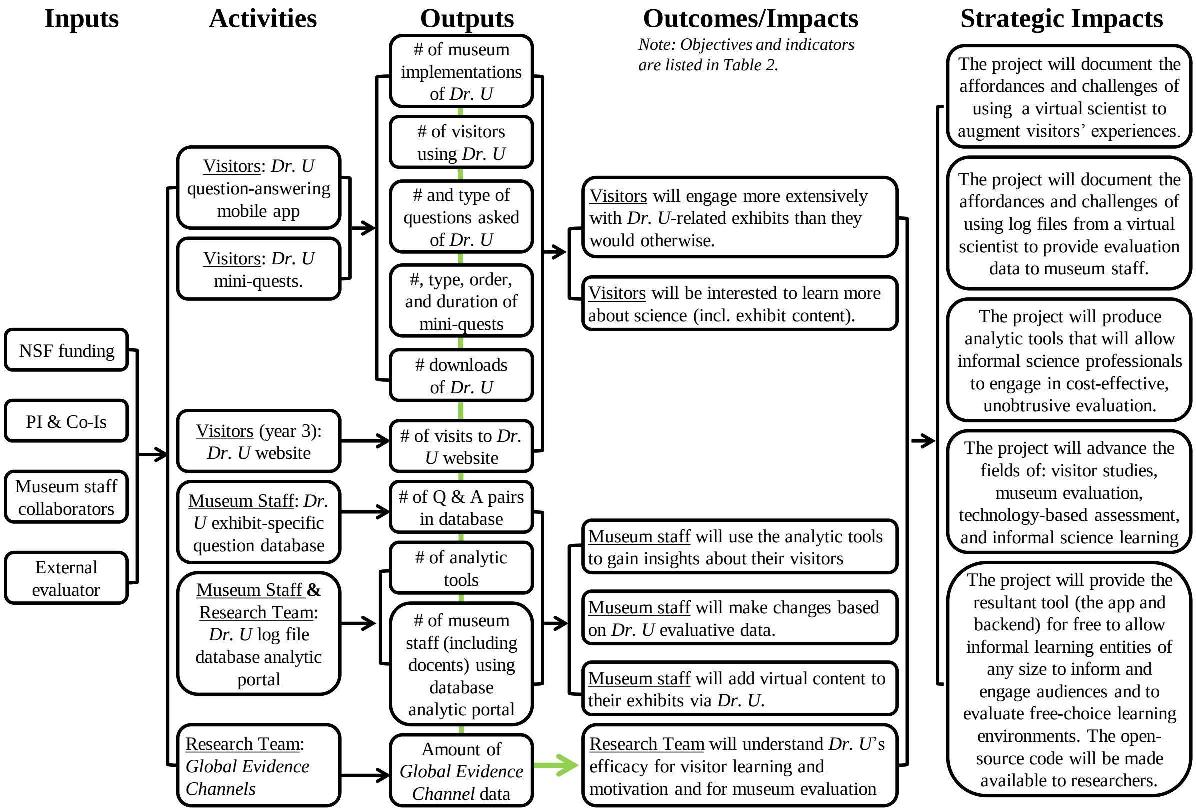

Figure 3. *Logic model*

<u>Data Analysis</u>: Quantitative survey and tracking data will be analyzed within and across implementations. Survey data will be analyzed by individual item and theme. In smaller implementations, the sample sizes will preclude statistical inference (Fowler, 2009; Huck, 2000), though appropriate descriptive statistics, along with insights from qualitative data, will inform and guide improvements to the project and suggest areas to probe through subsequent data gathering. When constructive, the results from the regression analyses by the research team will inform the evaluation. Qualitative data from the surveys (short-answer questions) and interviews will be analyzed following established techniques (e.g Miles & Huberman, 1994; Strauss & Corbin, 1998), including verbatim transcription, open coding, identification of emergent themes, development of analytic questions, and analysis of response matrices to identify similarities, differences, and discrepant data. The evaluation will also make use of demographic and other data collected by the research team (e.g., *pre-post surveys)*, when such data inform the evaluation questions.

Table 1. Evaluation questions, data sources, and analytic plans

Project Activities	Evaluation Questions	Data Sources	Analytic Plan
YEAR 1 • Test beta *Dr. U* • Design storyline • Design GEC data points • Develop and test analytics framework • Planning mtgs w/ museum partners, advisory board	What do the museums need and want from *Dr. U* as a resource for visitors? What do the museums need and want from *Dr. U* as a source of evaluation data? What back-end analytic tools might best meet the evaluation needs of the museums?	- Published materials for building evaluation capacity among museum professionals - Museum partner needs assessment survey	Review existing museum evaluation capacity-building materials. Quantitative and qualitative analysis of needs assessment survey results.
	Is the design of *Dr. U* and the storyline mini-quests aligned with the results of the needs assessment?	- Draft versions of storyline mini-quests	Review mini-quests in beta; compare mini-quests with needs assessment results.
	Is the design of the analytics framework & collection system aligned with needs assessment results?	- Beta of analytics framework & collection system	Review analytics framework and collection system in development. Compare with needs assessment.
	Is the *Dr. U* project on track to meet its Year 1 deliverables?	- *Dr. U* project implementation checklist	Comparison of project accomplishments with implementation checklist.
YEAR 2 • Mods of *Dr. U* & mini-quests • Develop/integrate GEC & analytics • Pilots & full-scale implementations of revised *Dr. U* • Pilots & full-scale implementations of revised data collection system, analytics display • Meetings with museums, advisors • Data analysis	Do visitors spend more time at related exhibits when they are able to access *Dr. U* and/or use mini-quests?	- Tracking studies - Exit surveys	Analyze tracking data for use/time w/ *Dr. U* & exhibits; exit surveys of self-report of time spent.
	Do visitors ask more questions at *Dr. U*- related exhibit(s) than otherwise?	- Tracking studies - Exit surveys	Analyze tracking data; surveys of visitor self report about question-asking behavior.
	Do visitors like *Dr. U*? Does use of *Dr. U* enhance perceptions of the exhibit experience? How does *Dr. U* impact visitors' perception of the museum?	- Exit surveys	Analyze exit surveys on visitors' perceptions of *Dr. U*; effects of *Dr. U* on visitors' experience with the exhibit and perception of the museum.
	Visitors likely to seek add'l info about *Dr. U* exhibits?	- Exit surveys; website (yr 3)	Analyze and compare treatment/control responses.
	Are visitors more interested in science?	- pre-post survey	Pre-post analysis of surveys
	Do staff feel comfortable using analytic tools?	- Interview, Surveys, logs	Analyze interviews, surveys, and tool logs.
	Do museum staff use analytic outputs to understand their visitors?	- Interviews - Surveys	Analyze interview, survey responses for staff use of outputs to understand visitors.
	Do museum staff use data and insights to make changes to materials and resources?	- Interviews - Surveys	Analyze interview, survey responses for staff use of outputs to make changes.
	Do museum staff use the Q&A input forms to add their own virtual content to the content database?	- Interviews & Surveys - Content database	Analyze interview, survey responses & database re: staff comfort with/use of content database.
YEAR 3 • Full-scale imple-mentations, data analysis, dissem.	Summative evaluation questions will be guided by the previous formative evaluation and by the project's overarching goals and research questions.	- Formative evaluation - Research results - Products - Add'l instruments as needed	Full scale review & qualitative/quantitative analysis of early evaluation & research results; review of resultant products & dissemination channels. Add'l data collection/analysis as needed.

Table 2. Impacts and indicators (E/I: Engagement/interest; K/A/U: Knowledge, Awareness, Understanding; S: Skills [Friedman, 2008])

Outcome/ Impact	**Audience Objective**	**Indicator/Evidence**
Visitors will engage extensively with *Dr. U*-related exhibits (E/I)	Visitors will spend more time at related exhibits when they are able to ask *Dr. U* questions and/or participate in mini-quests.	Tracking studies will determine whether visitors with *Dr. U* spend more time at related exhibits than those without, and the effect of mini-quests. Exit surveys of visitors with *Dr. U* will allow self-report on where they spent the most time.
	Visitors will ask more questions at *Dr. U*-related exhibit(s) than otherwise	Exit surveys report whether visitors asked more questions at related exhibits than they would otherwise. Obs. will note use/other behaviors (e.g. talking to docents)
	Use of *Dr. U* will enhance visitor perceptions of the exhibit experience.	Exit surveys given to visitors with *Dr. U* will allow them to report its effect on their experience with the exhibit.
Visitors will be interested to learn about science, exhibit content (E/I)	Visitors will report that they are likely to seek out additional information about the *Dr. U*-themed exhibit.	Through exit surveys, visitors with and without *Dr. U* will report on interest in learning more about exhibit content. Treatment and control group responses will be compared, with attention paid to emergent demographic patterns.
	Visitors are more interested in science.	Pre-post comparisons of interest-related questions will indicate increased visitor interest in science following use of *Dr. U*.
Museum staff will use analytic tools to gain insights (S; K/A/U)	Museum staff will feel comfortable using the database analytic tools.	Through interviews and surveys, museum staff will report understanding and comfort using the analytic tools. Analytic tool use logs will show extent of use.
	Museum staff will use the analytic tools to understand visitors	Through interviews and surveys, museum staff will report use of the analytic tools and provide examples of insights they gained from the outputs of the tools.
Museum staff will make data-driven changes (S)	Museum staff will use data and insights from analytic tools to make changes to materials and resources	Through interviews and surveys, museum staff report they used the evaluative data to make specific, substantive changes to materials and resources. They provide examples of changes they would not have known to make w/o the *Dr. U* data.
Museum staff will add content via *Dr. U* (S)	Museum staff will use input forms of Q&A database to add content specific to their context, exhibits, or visitor needs.	Through interviews and surveys, museum staff will report being comfortable adding content via database input forms and will report using the forms to augment their exhibits with virtual content. Content database will indicate staff additions.
Researchers understand *Dr. U*'s efficacy (K/A/U)	Outputs contribute to understanding of *Dr. U* for visitor experience & museum evaluation	Research and evaluation reports, published journal articles, and conference presentations will provide evidence of this outcome.

4. PROJECT MANAGEMENT

4.1. Research Team

Prof. Judd D. Bowman (PI) will oversee all aspects of the project, including adherence to the schedule and budget. He will be responsible for delivering the Dr. U app and analytics portal, including all software and art design and implementation, and leading activities to investigate Research Question #1. He is an assistant professor in ASU's School of Earth and Space Exploration and a former software developer (resulting in two NASA Software Release Awards) and architected algorithms for the Dr. C precursor. He has a long-standing interest in questions that require analysis of large datasets. He is PI of two NSF collaborative research grants: AST-1109257 to identify optimal statistical techniques to analyze new astrophysical observations, and AST-1109865 to develop techniques to improve processing of large systems of radio astronomy antennas. He is PI of the *EDGES* astrophysics experiment (AST-0905990, 9/1/09-8/31/13, $392,446; AST-1207761, 6/1/12-5/31/15, $842,117). **Intellectual merits:** *EDGES* placed unique limits on the cosmological epoch of reionization (Bowman & Rogers, *Nature* 2010), extending the findings of preliminary investigations (Bowman, et al. 2008, Rogers & Bowman 2008), documented by publicly available memoranda http://loco.lab.asu.edu). **Broader impacts:** EDGES characterized radio interference at areas in the U.S. and Australia (Bowman & Rogers, 2010) and led to a new method for calibrating radio receivers (Rogers & Bowman, 2012).

Prof. Brian C. Nelson (Co-PI) will be responsible for conducting activities to address Research Question #2. He is an associate professor in the School of Computing, Informatics, and Decision Systems Engineering at ASU. Dr. Nelson's research focuses on the theory, design, and implementation of educational immersive virtual environments. He has published and presented extensively on the viability of virtual environments for situated inquiry learning and assessment, including *Design for Learning in Virtual Worlds* (Routledge). Nelson is Co-PI on the SAVE Science project, Co-PI on the recent NSF study Scaffolding Understanding through Redesigning Games for Education (SURGE), investigating the use of an online physics game, and Co-PI on the MacArthur Foundation 21st Century Assessment project.

Dr. Catherine Bowman (Co-PI), will assist with development of *Dr. U* as it relates to museum evaluation. She will support the external evaluator in developing measures and collecting data and will contribute to Research Question #1. She is the evaluator for NASA's Mars Public Engagement Program (Raytheon contract, 0.75 FTE) and an ASU affiliate. She developed the precursor *Dr. C* system to explore expert-novice mentoring. She has 15 years of experience with NASA-specific STEM education efforts.

4.2. Larger Project Team

Ms. Kathy Eastman (AZMNH) and Ms. Sari Custer (ASC) will act as leads at the partner sites to coordinate museum involvement with the project. Additional museum staff, docents, and volunteers will interact with the project under their guidance. Partner museum staff have participated in the design of this proposal; they will work with the team, providing expertise on museum education, operations, visitors, and the evaluation tools and data they require. Funds are included to offset museum partner expenses.

Ms. Jodie Hoffman (Consultant) will act as the external evaluator. She holds an M.A. in education policy and evaluation from Stanford University. She has over a decade of experience as an educational evaluator at WestEd, focusing on research study design, instrument development, data collection and analysis, project direction and coordination, and proposal and report writing.

Mrs. Margaret Hufford (Educational Coordinator) will provide logistical support for the project, assisting with the development, distribution, and implementation of the project and materials and coordinating the undergraduate team members. Hufford has a decade of experience in K-12 professional development and education and public outreach, and a four-year affiliation with NASA's Space Grant. One **part-time staff programmer** and one **undergraduate programmer** will develop the mobile app and web portal. **Two undergraduate** majors in Earth & Space Exploration and graphic design will provide content and artwork. **One native Spanish-speaking undergraduate** translate all content.

To provide external input and oversight for the project, we will convene twice annually an Advisory Board (see Supplementary letters) consisting of **Prof. Chris Dede (Harvard; virtual environments)**, **Prof. Shari Tishman (Harvard; learning in museums**), **Dr. Jake Hartman (Google; big data analytics)**, and the two museum liaisons. The Board will provide critical external review of the project's progress and outcomes and offer counsel in their areas of disciplinary expertise relevant to the project.

4.3. Schedule and Work Plan

We will take a design-based research approach to this project that focuses on cycles of design, implementation, analysis, and redesign based on implementation results (Brown, 1992; Collins, 1992). Year 1: Basic application usability testing; back-end development of the full framework for collection and display of analytics; and design of the storyline-based mini-quests and Global Evidence Channels (GEC). Year 2: Development of the storyline activities and GECs; tests for usability of new agent features; tests and refinement of the framework for collection and display of analytics; iterative changes to *Dr. U* (front- and back-ends) based on feedback and full-scale tests (multi-day implementations at the partner museums) of usability, portal evaluation, and data analysis. Year 3: Refinements to the front- and back-ends; final full-scale implementations with full data collection, analysis, dissemination, documentation.

Year 1	• *Dr. U* (existing version) usability studies in AZMNH and ASC • Design of *Dr. U* situated storylines with AZMNH and ASC • Design of Global Evidence Channel data points for storylines • Develop and test analytics framework and collection system • Planning meetings with all museum partners and Advisory Board: roadmap for collaboration
Year 2	• Development and integration of Global Evidence Channel system with analytics system • Pilot of revised *Dr. U* with storyline: AZMNH and ASC • Pilot of revised data collection framework and analytics display system with all museums • Evaluate program and make changes as needed for more effective implementations • Design/Development meetings with AZMNH and ASC and Advisory Board • Conduct a preliminary analysis of data; disseminate preliminary results • Iterative refinement pilots of *Dr. U*, storyline, and analytics system with AZNHM and ASC • Evaluate program and make changes as needed for more effective implementations • Conduct a preliminary analysis of data; disseminate preliminary results
Year 3	• Design/Development meetings with all partner museums and Advisory Board • Full-scale implementations at all partner museums • Cumulative data analysis and evaluation of research questions • Dissemination of *Dr. U* to museums and public download on App Stores • Dissemination of research and evaluation findings, frameworks, *Dr. U* software and code.

We believe the proposed project is within the scope of the resources requested. The fact that both study sites are in the same place (the greater Phoenix area—both within ten miles of ASU and within 20 miles of one another) greatly improves the project's efficiency. The additional statistical power for the study provided by the sample sizes afforded by two distinct sites is valuable, as is the access to additional museum staff expertise. Our two partner museums serve similar but distinct groups (one is a larger institution based in downtown Phoenix; one is smaller, based in Mesa). Another factor that contributes to the efficiency of this project is that it is not starting from scratch but from a version 1.0 app that was developed by undergraduates funded by local funds, so the technical components of this project are being modified and enhanced, not designed from scratch. Additionally, the workload is equally distributed, by expertise, among the team.

The programming of the app itself (the computerized astronaut, its functions, and its analytic backend) are common to both sites (we will investigate the use of the Unity game development tool to create Dr. U. Unity allows for rapid porting of mobile apps across platforms). The programming costs outlined in the budget are based on Profs. J. Bowman and Nelson's extensive personal programming experience and experience in hiring professional programmers for projects.

What is customizable to each museum context is the question-and-answer database itself. A simple web-based interface allows the project undergraduates in charge of content creation, as well as authorize museum staff themselves, to add and edit questions and answers whenever appropriate--in the normal course of building the database (which already includes over 10,000 space science-related questions), in response to user feedback and use of the app (unveiling their interests and misconceptions), or in preparation for a museum event or in response to a current event. There is no additional programming required to enter new content into the Dr. U system.

The Dr. U mobile app is designed to contact the central server at least once every 24 hours (when in use) in order to download content updates. The content database of 10,000 question/answer pairs for the existing prototype Dr. U system was researched and entered primarily by one ASU undergrad student over approximately 100 hours. That corresponds to entering approximately 1 question/answer pair per minute - including the time required to research the information from a reputable source (such as NASA) and write content. This level of efficiency is achieved in part because any one answer can be assigned to multiple questions so it is not necessary to retype an answer for every question.

To support museum staff in implementing Dr. U, making use of the Dr. U analytic portal, and interpreting data derived from museum visitor interactions, we will make use of a modified framework for professional development (PD) created for the ongoing SAVE Science Project (DRL-0822308, 9/1/08-8/31/14, Prof. Nelson) by Catherine Schifter (2008).

The framework uses an approach to PD that is a combination of processes to facilitate participants' (museum staff, docents, and volunteers) understanding and use of digital technologies and related data visualizations. The Dr. U project will extend the SAVE Science approach to help museum staff establish control over use of the Dr. U app and analytic portal according to their local context and museum and visitor needs, and guide the museum staff's use of data derived from Dr. U to support evaluation efforts or operational decisions related to the museum visitor experience.

Schifter's PD model includes several related and fundamental factors designed for integration of innovative technologies by formal educators that will be adapted for use with museum staff as follows:

(1) Training. Training time will be reserved in Year 1 for museum personnel to a) practice (with both the app and the early analytics portal), b) develop interest and knowledge related to the range of evaluative information Dr. U might be used to reveal, and c) assess the usefulness of the Dr. U application and analytics portal in controlled trials before 'going live' with museum visitors. This will be informed by the needs assessment carried out by our external evaluator in Year 1, which will establish a baseline of the current evaluation capacity of the partner museums, their overarching evaluation goals, and what evaluative data and tools would be of most use.

(2) Workshops. Informed by our external evaluator's Year 1 needs assessment and ongoing formative evaluation, we will host annual workshops for personnel engaged in the project at each museum partner site (ASC and AZMNH) that provide an overview of data mining and web analytics in the context of the Dr. U project, introduce personnel to the analytics portal and demonstrate the key information presented in the portal, and engage the museum personnel and research staff in a discussion of the information they would like to see included in the portal analytics and how to effectively translate the analytics into actionable information for each museum.

(3) Technical support. Effective and ongoing post-training technical support in the museums and through online channels (e.g. embedded tips in the Dr. U analytics portal and a user forum for museum personnel hosted on the portal) will be provided to support those museum personnel who are not technology experts so that any technology hurdles do not prevent appropriate application of the Dr. U app and analytics portal.

(4) Learning resources. In coordination with our museum contacts, we will develop online tutorials and fact sheets to aid museum personnel as desired.

(5) Communication. Ongoing communication and social support will be provided through regular

team teleconferences (including the research team and designated museum contacts) and in-person visits of the research team to the museum partners and of museum personnel to the ASU campus (which our primary contacts already do several times per year) in order for each group to understand and participate in the respective activities in the context of each environment. These factors are important for any PD when technology is involved and will be made even more relevant to the museum context by our adherence to Alpert's (2013) "principles of good stewardship" for work between museums and researchers, particularly "cultivat[ing] cross-cultural insight and understanding" (p. 43) to bridge the cultures related to professional learning within the museums and at a research university, and "collaboration on educational content" (p. 44), in this case, the professional development content and procedures created to support the museum staff learning.

Our progress in these efforts will be monitored and assessed by our external evaluator on an ongoing basis, particularly as they relate to the first two "knowledge/understanding/awareness" and "skills" outcomes (Friedman, 2008) for museum staff outlined in our logic model (Figure 3 of the proposal) and our impacts and indicators chart (Table 2 of our proposal). In helping museum staff in understanding and making use of the analytics portal data, we will draw upon the lessons learned in the SAVE Science project. For example, using Schifter's model, SAVE Science researchers implemented a data-driven decision-making PD workshop for the participating teachers in 2012. During this session, teachers used their student performance data from an online project dashboard developed for SAVE Science to improve their instruction and identify their students' misconceptions. Informed by the 6-step conceptual framework proposed by Light, Wexler and Heinze (2005), which outlines how teachers can transform data into knowledge, and based on the SAVE Science teacher dashboard, the Dr. U analytic portal will: store raw data, organize the data into easy to use formats, and summarize the data from each data set. This will allow museum staff to focus on engaging in generating practical knowledge from data, and similarly allow Dr. U researchers to focus their PD efforts on these aspects of knowledge generation and use.

We note that initial discussions between the research team and museum representatives in preparation of this proposed project have already established that our museum partners have a strong understanding of the potential roles and benefits of the Dr. U system within their institutions. We anticipate that the continued PD planned during the project will build on this foundation to effectively engage the museum personnel, including docents, volunteers, and staff, who will interact with Dr. U and lead to a well-integrated research and evaluation system.

5. DISSEMINATION

At the end of the project, both partner museums may retain a subset of the mobile devices for their ongoing use and museum visitors will be able to download the app onto their personal devices (through online app stores). As a novel dissemination approach we will embed information about the study in the final database of the app so that ongoing users may discover and learn about the project and its results. The software code for *Dr. U* system will be released as open-source. The final integrated database of content questions and verified answers in English and Spanish languages will be released for download.

The findings and outcomes of the research activities and design of the system will be documented and disseminated through multiple outlets, including as reports submitted to academic journals targeting museum professionals (e.g., *Visitor Studies*), those targeting researchers (e.g. *Journal of Science Education and Technology*) and through websites and conferences relevant to practitioners (e.g., CAISE's http://informalscience.org/ and ASTC). Collected logs of user data, as appropriate, will be made available to legitimate researchers upon request. Results and insights into log data processing applicable to other fields will be documented through submissions to suitable journals, such as IEEE and ACM publications.

We intend to reach out to practitioners through relevant websites and conferences such as CAISE's http://informalscience.org/ and the Association of Science-Technology Centers (ASTC) annual conference (included in our travel budget). We will work with our museum partners to identify additional practitioner-focused venues, web sites, social networks and other dissemination platforms to rapidly inform museums across the country about the findings from the Dr. U project. These same dissemination channels will be utilized to provide access to the final Dr. U system.