Library Knowledge Extensions (KNEXT): Data Analytics to Support Innovation Communities

Abstract

Kent State University (KSU-SLIS), in collaboration with the University of Maryland (UMD-CIS), requests $458,319 (in addition to $233,061 in cost share) in grant funding for a three-year project. This project brings advanced data analytics and business intelligence (DA&BI) services to public libraries in order to support small businesses, entrepreneurs, and community advocates. It will be implemented in partnership with local public libraries, small business development centers, economic development organizations, and community advocacy groups. This **Research** project will address the Institute of Museum and Library Services' **National Leadership Grants for National Digital Platform category by addressing the research question**: *How can we integrate advances in big data and data analytics into libraries as essential community anchors in a way that empowers those communities while encouraging economic development and well-being?* This project addresses two strategic goals of the IMLS: to "promote the use of technology to facilitate discovery of knowledge" within local communities through libraries and promote libraries as community anchors "to enhance civic engagement and economic vitality."

As data production grows, the needs of local community actors (e.g., small business, entrepreneurs, and community advocates) for tools and methods to make sense of the complex data landscape increase as well. Developing a community-accessible framework for accessing, learning, and utilizing these tools and methods is an overarching focus of this proposal. In recent years, we have witnessed massive growth in computational methods and discoveries. Advances in information and communication technologies (ICTs) are addressing a wide range of economic and societal challenges; however, we do not yet know *how to successfully and effectively deploy such technologies in community decision making* and *what their role should be in innovation processes.* It is still unclear *how communities can directly benefit from the advances in big data and data analytic technologies*. Further, we are unsure *how such technologies can create an innovation-supporting environment for civic advocates, small businesses, and entrepreneurs that will stimulate growth in recovering communities.*

The KNEXT research project will address these research questions by creating a hub for advanced data analytics and business intelligence (DA&BI) services and integrating it into the existing library systems and economic development networks within two recovering communities in Ohio and Maryland. We believe the best way to address these pressing questions is through making and investigating this hub at the same time. KNEXT will play two roles in this scenario: 1. As an integrated point of access to public open data by aggregating and curating local and public data (e.g., open government data, census data, social media data, etc.); and 2. As the provider of DA&BI services to small businesses, entrepreneurs, civic advocates, and community decision makers.

The KNEXT project will be implemented in three phases (over three years). We will follow a participatory applied research approach to engage with the community and inform our design through their feedback while investigating the overarching research questions of the project. This project will: 1) **Identify the range of DA&BI services** that can best serve the needs of small businesses and community advocates by assessing current services and analyzing unmet needs; 2) **Identify the social and institutional barriers** to integrating advances of DA&BI into the practices of small businesses and community advocates; 3) Further **develop a coalition** of academia, public libraries, small business development centers, small businesses, and community advocates; 4) **Identify sources of useful data** (national, state, county) and consolidate data in a repository for long-term access and delivery; 5) **Create a platform and GUI dashboard** for libraries to provide DA&BI services to the community using machine learning, data mining, web mining, and text analytics; 6) **Train the staff of the partner libraries** to use the analytics dashboard and run DA&BI scenarios; 7) **Evaluate the impact** of the services on community development.

The emergence of low cost data analytics can transform the knowledge services available to communities creating the potential for substantial societal impact by making small businesses more competitive, reducing the risks of entrepreneurial efforts, re-energizing smart cities, and unleashing community innovations. Successful integration of such services into communities helps them to better identify solutions for their ongoing and evolving challenges. KNEXT will be a testbed for evaluating this potential of information technology and libraries for societal change.

Proposal for IMLS National Leadership Grant for Libraries
Library Knowledge Extension (KNEXT):
Data Analytics to Support Innovation Communities

1. Statement of Need

This project brings advanced data analytics and business intelligence (DA&BI) services to public libraries in order to support small businesses, entrepreneurs, and community advocates. It will be implemented in partnership with local public libraries, small business development centers (SBDCs), economic development organizations, and community advocacy groups. This project addresses two strategic goals of the IMLS: to "promote the use of technology to facilitate discovery of knowledge" within local communities through libraries and promote libraries as community anchors "to enhance civic engagement and economic vitality."

The Problem

Data means power—when you have the tools and skills to extract insights from the data (Bertot, Butler, & Travis, 2014; Breeding, 2016; Cervone, 2016; Erdmann, 2014; Louridas & Ebert, 2013; Marshall, Mueck, Shockley, 2015; Mutula, 2016, Soergel, 2015). Previous pilot research has identified a clear need for data that is more meaningful to local decision makers, including neighborhood data and service supply and demand assessments (Bertot, Butler, & Travis 2014). Moreover, the Open Government Partnership has provided federal support for access to big data and open data to generate new and innovative insights (Bertot, Butler, & Travis, 2014; Furlough, 2012).

In economically depressed cities and regions feeling stress from the decline of mass production, small businesses and entrepreneurs are the primary contributors to economic growth (Bertot, Jaeger, Gorham, Taylor, & Lincoln, 2013; Franks & Johns, 2015; Hoppenfield & Malafi, 2015; Leavitt, Hamilton-Pennell, & Fails, 2010; MacDonald, 2010; Martin, 2010; Restivo, 2014; Welch, 2005). However, entrepreneurs need quality information to grow their businesses and, as a result, the local economy. Big data provides insights into customer and market trends that leads to innovation, allowing small businesses and entrepreneurs to make evidence-based decisions (Goyel, 2014; Marshall, Mueck, & Shockley, 2015; Welch, 2005). A small business owner making kitchen cabinets could determine the latest trends in countertop material by analyzing social media posts. An entrepreneur could analyze the pros and cons of competitors' products based on online reviews. Community advocates could view the impact of a new development project on their community through visual analysis of census data on maps. In summary, access to business intelligence can make the difference between success and failure.

However, quality business information and expertise can be prohibitively expensive for the entrepreneur or small business owner to purchase (Ernest, 1993; Leavitt, Hamilton-Pennel, & Fails, 2010; MacDonald, 2010; Welch, 2005). As a result, small business owners cannot compete with major corporations because they cannot afford to access commercial services that provide data-driven insights (Ernest, 1993; Furlough; 2012; Leavitt, Hamilton-Pennell, & Fails, 2010; McDonald, 2010; Welch, 2005). Local level community advocates and decision-makers face similar financial challenges in accessing big data to make informed decisions and better serve their communities.

Libraries serve an invaluable role as community anchors, providing access to information and resources, promoting the use of technology to facilitate the discovery of knowledge, and furthering economic vitality (Scott, 2011; Hill, 2014; Furlough, 2012). With the introduction of the Internet, libraries have expanded their services and remained relevant to the community; the introduction of big

data and open data once again provides such an opportunity to strengthen the role of libraries as community anchors (Breeding, 2016; Scott, 2011; Hill, 2014). Yet big data is not without its own challenges: big data is accumulating at an exponential rate creating the challenges of storage, discovery and access, and management and analysis (Crowe & Candlish, 2013; Mutula, 2016). Fortunately, libraries and librarians are uniquely qualified to address these challenges. Libraries have a long tradition of providing access to knowledge and information services, access to data, and technical support in response to needs of local communities. Clearly, it is a natural progression for libraries, as the bridge between the people and the government and the people and the information, to provide access to big data and open data to the public (Aagaard & Arguello, 2015; Bertot, Butler, & Travis, 2014; Bertot, Jaeger, Gorham, Taylor, & Lincoln, 2013; Cervone, 2016; Crowe & Candlish, 2013; Day, 2002; Erdmann, 2014; Furlough, 2012; Louridas & Ebert, 2013; Marshall, Mueck, & Shockley, 2015; Mutula, 2016; Scott, 2011).

Most existing entrepreneurial outreach programs are focused on providing access to business databases in traditional forms and do not provide advanced analytical or predictive analytic services. Though there have been efforts by information institutions and governments to provide assistance to small businesses and entrepreneurs by providing access to big data (see Environmental Scanning in the supporting documents), there is only one organization, the *Chattanooga Public Library*, that has integrated big data into its entrepreneurial and innovation programs. Most of the current initiatives related to open and local data do not focus on utilizing such information for the advantage of entrepreneurs and small businesses. There is currently no program that uses advanced data analytics such as social media marketing analysis or data mining in relation to entrepreneurial services; even *Chattanooga Public Library* does not offer analytic services, instead they provide the data that a third-party agency then analyzes (Hill, 2011). The KNEXT project would then be the first in its field to utilize big data and analytics to assist entrepreneurs and small businesses in order to promote innovation and economic vitality.

Data Analytics and Its Potentials

Multiple kinds of data analytics could be helpful to local small business owners and community decision makers. Because of the abundance of unstructured data on the Web, the integration of mature and scalable techniques in text mining (e.g., information extraction, topic identification, and opinion mining), social network analysis, and spatial-temporal analysis with existing structured open data or the data which can be extracted from traditional business databases available at libraries can yield enormous value for small businesses and entrepreneurs. However, applying such data analytic services requires skills, infrastructure, and technology know-how (Bertot, Butler, & Travis, 2014; Bertot et al., 2013; Cervone, 2016; Furlough, 2012; Marshall, Mueck, & Shockley, 2015).

Data analytics and business intelligence (DA&BI) refers to technologies which are grounded in data mining and statistical analysis. Five general families of analytics have been discussed in the business literature (Chen, 2012):

- Text analytics (applies to unstructured textual content);
- Web analytics (applies to the content available on the Web in different formats such as HTML, XML, and RSS);
- Network analytics (applies to relational data such as social networking sites);
- Mobile analytics (focuses on the increased use of personalized mobile computing and data generated on such platforms);
- Big Data Analytics (types of analytics that require using solutions which scale well such as MapReduce because of the very large size of data sets).

- Two types of data can be used for analytical purposes:
- Structured data (DBMS; structured data e.g., census data; and structured web-based, e.g., Twitter data) and
- Unstructured (web-based; unstructured user-generated content; social network information; unstructured informal customer opinions).

Vast amounts of business related data about companies, products, and customer information can be gathered from the Web. This data can be used to provide business analytic services to the community. Many social media platforms such as forums, online groups, blogs, social networking sites, photo and video sharing sites have also created an abundance of user-generated content which can be used to collect a large volume of timely feedback and opinions from a diverse population for different types of businesses. Social media analytics presents a unique opportunity for businesses to treat the market as a 'conversation' between businesses and customers and identify trends.

Working with structured data is easier and poses fewer challenges. Usually data mining and statistical analysis techniques can be directly applied to structured data. Unstructured data, however, requires data wrangling processes, to prepare it for analysis. Types of analytics that we can provide to help small businesses, entrepreneurs, and community advocates are subsets of the following general analytics families: association rule mining, database segmentation and clustering, anomaly detection, graph mining, social network analysis, text and web analytics, and sentiment and affect analysis. Benefiting from these general analytical techniques requires data science skills and infrastructure; therefore, in most cases it is impossible for small businesses, entrepreneurs, and community advocates to enjoy such benefits. KNEXT can bridge this gap by making these generic analytical processes accessible and by customizing them for these three groups of patrons to empower them to make better evidence-based decisions. In other words, the service offered by the KNEXT research project will decrease the digital divide which currently favors large corporations, allowing smaller entities to participate in a healthy innovation ecosystem. In such a way, the KNEXT will play the role of "a silent business partner" to small businesses, entrepreneurs, and community advocates (Scott, 2011; Welch, 2005).

As described above DA&BI has the potential to transform communities if it becomes accessible to community members. Following are a set of concrete examples of how different kinds of data analytics could be helpful to local small business owners and community decision makers. For example, KNEXT could identify potential customers in a specific geographic area extracted from census data and visualized on maps to inform a small business owner about its clientele. A business that sells to other businesses (B2B) with 30 clients seeking to grow would need to identify other potential clients. To address this need, KNEXT will design a recommender system using machine learning methods to create a profile of existing clients and identify potential clients matching this profile in possible business databases such as *ReferenceUSA.* Identifying even as few as three additional clients would be a 10% increase, representing tremendous growth. KNEXT could also use social media data to assist entrepreneurs in predicting market trends.

Access to data analytics is important for advocacy groups and community organizations interested in engaging local residents, too. To help community advocates and municipal decision makers understand the impact of a development project, census and local data can be used to visualize the demographics of communities and other socioeconomic variables. Social media data could reveal key influencers in a community, empowering local groups to organize more effective advocacy campaigns. Social media announcements for public meetings could target influential individuals within demographics underrepresented during the public engagement process. In addition, big data could help identify individuals who would benefit from additional social services. For example, data analytics

could reveal individuals susceptible to housing insecurity. Rather than use this data to target predatory loans or other advertisements, as is currently the case, community organizations could connect with people to provide housing education services. Public libraries could serve as accessible platforms to level the data analytics playing field. These are a few examples of the potential of using data analytic and business intelligence services to support innovation and decision making in local communities. The proposed KNEXT applied research project is an attempt to address this need to promote libraries as community anchors "to enhance civic engagement and economic vitality.

The KNEXT applied research project will investigate and assess the DA&BI needs of small business, entrepreneurs, and community advocates and develop a platform to provide such services through public libraries. Its primary goal is to increase access to data, both big data and open data, to promote innovation and economic vitality.

2. Project Design

As data production grows, the needs of local community actors (e.g., small business, entrepreneurs, and community advocates) for tools and methods to make sense of the complex data landscape increase as well. Developing a community-accessible framework for accessing, learning, and utilizing these tools and methods is an overarching focus of this proposal. In recent years, we have witnessed massive growth in computational methods and discoveries. Advances in information and communication technologies (ICTs) are addressing a wide range of economic and societal challenges; however, we do not yet know *how to successfully and effectively deploy such technologies in community decision making* and *what their role should be in innovation processes.* It is still unclear *how communities can directly benefit from the advances in big data and data analytic technologies*. Further, we are unsure *how such technologies can create an innovation-supporting environment for civic advocates, small businesses, and entrepreneurs that will stimulate growth in recovering communities.*

The KNEXT research project will address these research questions by creating a hub for advanced data analytics and business intelligence (DA&BI) services and integrating it into the existing library systems and economic development networks within two recovering communities in Ohio and Maryland. We believe the best way to address these pressing questions is through making and investigating this hub at the same time. KNEXT will play two roles in this scenario:

1. As an integrated point of access to public open data by aggregating and curating local and public data (e.g., open government data, census data, social media data, etc.); and
2. As the provider of DA&BI services to small businesses, entrepreneurs, civic advocates, and community decision makers.

KNEXT provides both a mechanism for investigating the above research questions and an opportunity for developing a tool to address the needs of the community. This research is a collaborative project led by the Kent State University (KSU) School of Information and the University of Maryland College of Information Studies. KSU will be responsible for the overall management and coordination of the project. Dr. Emad Khazraee, of Kent State University, will be the Primary Investigator and lead sociotechnical data scientist on this project. He will work with Karen MacDonald, Business & Entrepreneurship Outreach Librarian at KSU Libraries (senior personnel). Additional contributors will include William Southards, Director of Small Business Development Center at KSU (senior personnel); David Jurca at the Cleveland Urban Design Collaborative (senior personnel); Anastasia Diamond-Ortiz, Director of Lorain County Public Library system; John Skrtic, Director of Public Services at Cleveland Public Library; Linda Hale, Microbusiness Specialist at Akron-Summit County Public Library; Jodie DeLamatre, Adult Services Reference Librarian at Hudson Library & Historical Society; Julie Messing,

Director of LaunchNET incubator; and Kimberly Irvin-Lee, Director of the Minority Business Assistance Centers, Akron.

Dr. Susan Winter will serve as Co-Primary Investigator and is the lead sociotechnical information scientist on the University of Maryland team. She will work with Andrew Fellows (senior personnel) as well as Kathleen Teaze, Director of Prince George's County Memorial Library System and representatives from the Prince George's County Chamber of Commerce, Prince George's County Economic Development Corporation, and the Maryland Small Business Development Center.

This project will follow a truly integrative socio-technical approach by developing the technical component in tandem with a close-reading of social phenomena and community needs. The two dimensions of platform development and community engagement will progress in parallel and provide constant feedback to each other.

The KNEXT research project will:

1. Identify the range of DA&BI services that can best serve the needs of small businesses and community advocates by assessing current services and analyzing unmet needs;
2. Identify the social and institutional barriers to integrating advances of DA&BI into the practices of small businesses and community advocates;
3. Further develop our coalition of academia, public libraries, small business development centers, small businesses, and community advocates;
4. Identify sources of useful data (national, state, county) and consolidate data in a repository for long-term access and delivery;
5. Create a platform and GUI dashboard for libraries to provide DA&BI services to the community using machine learning, data mining, web mining, and text analytics;
6. Train the staff of the partner libraries to use the analytics dashboard and run DA&BI scenarios;
7. Evaluate the impact of the services on community development.

The KNEXT project will be implemented in three phases. We will follow a participatory applied research approach to engage with the community and inform our design through their feedback while investigating the overarching research questions of the project.

Phase One: Needs Assessment and Pre-Development Research (year one):

Phase one of the project consists of three major activities:

1.1 Needs Assessment (Months 1-10)

1.2. Data Source Identification (Months 3-11)

1.3. Readiness Assessment (Months 9-12)

The researchers at both KSU and UMD will identify and assess the needs of small businesses for business information and data analysis. This will be accomplished through engagement efforts with public libraries, Small Business Development Centers, Chambers of Commerce, and economic development agencies. The research team will conduct interviews and focus groups, working closely with the stakeholders to identify and define community members' needs and concerns that can be addressed through access to business intelligence services. Through activity 1.2 the researchers will explore the sources of existing open data (at both the local and national level) that can address the DA&BI needs identified in the activity 1.1. Finally, the research team will conduct a readiness assessment to identify the users' capabilities and knowledge of data analytic in the activity 1.3. See the *Schedule of Completion* page 1 for details of work plan, phase one.

Phase Two: Development and Initial Implementation (year two):

Phase two of the project consists of four main activities:

2.1 Analysis and Development of Analytics Platform (Months 13-24)
2.2 Development of Integrated Data Access Hub (Months 18-24)
2.3 Training the Community Partners (Months 19-21)
2.4 Early Implementation (Months 20-24)

In this phase, researchers will develop DA&BI strategies to meet the community needs identified in phase one. Researchers will define the needed business intelligence products and services and identify the most effective and efficient methods to deliver these products to the local community. The team will turn the identified business intelligence needs of the community into accessible routines available via a GUI dashboard to be used by the librarians in the partner libraries. Moreover, in activity 2.2, the KNEXT project will create data hubs for each community partner. Based on the result of readiness assessment, in activity 2.3, the KNEXT team will provide the required training to community partners in the two research sites. The early implementation on analytics platform will start in month 20 of the project. The goal is to identify the shortcomings and problems and quickly integrate the feedback from the community into a revised scenario. See the *Schedule of Completion* page 2 for details of work plan, phase two.

Phase Three: Implementation and Evaluation (year three):

Phase three of the project consists of three activities:

3.1. Full implementation (Months 25-31)
3.2. Evaluation (Months 31-36)
3.3. Community Building and Sustaining the Project (Months 29-36)

In phase three, the research teams will deploy the full range of DA&BI services developed in phase two. The Research team will conduct follow up interviews and focus groups with community partners to identify areas for improvement. To ensure the sustainability of the KNEXT project after the lifetime of this grant the KNEXT team will introduce the project to the broader community through a series of community building activities. The research team will also partner with community agencies to assess the effectiveness of the programs and evaluate to what extent the project met the performance goal set by IMLS. See the *Schedule of Completion* page 3 for details of work plan, phase three.

Evaluation Plan

The KNEXT team will partner with community agencies to evaluate the effectiveness of the services. Initially three sets of metrics will be used for the assessment of the success of project:

1) ***Application-oriented metrics:***

a. The number of successful fulfilled DA&BI requests;
b. Change in the number and variety of patrons served by that participating libraries (whether libraries could engage with more patrons or new groups of patrons through KNEXT);
c. The effectiveness of KNEXT in enabling the community to make better decisions or innovate based on reported success or failure stories from community;
d. Plans for sustainable implementation of business intelligence service programs.

2). ***Conceptual metrics***:

a. Development of model(s) of DA&BI programs that can be adopted by other communities beyond the two initial sites;
b. The number of barriers to the integration of DA&BI services identified and/or resolved by the end of project.

3) ***IMLS performance measures:***

This proposal aligns with IMLS agency-level goal 2 (community); thus, we will evaluate how this project strengthen libraries as essential partners in addressing the needs of their communities through performance measures provided by IMLS. Survey questionnaire will be used to conduct this part of evaluation.

Resources: Project Personnel and Infrastructure

The estimated total budget is $691,380 over the 3-year period, of which $458,319 is requested from the IMLS and $233,061 will be cost shared. Kent State University is the leading institution, receiving $273,100, and providing $154,780 in cost share. The University of Maryland College Park is the sub-awardee, receiving $185,219 of the grant and providing $78,281 in cost share.

This proposal builds on the tradition of multidisciplinary sociotechnical research in information studies. Our research team brings together expertise from a wide range of disciplines such as data and information science, science and technology studies, organizational studies, urban and regional planning, economic development, and community planning. PI, Dr. Emad Khazraee, is a sociotechnical data scientist experienced in big data analytics and received a Master's degree in architecture focusing on recovering communities. Dr. Susan Winter, Co-PI, was previously a Science Advisor in the Directorate for Social Behavioral and Economic Sciences, studying the social and organizational challenges of data reuse and the co-evolution of ICT and work. Karen MacDonald is Business & Entrepreneurship Outreach Librarian who has formerly done extensive research on the information needs of small businesses and entrepreneurs, examining relationships between information services, innovation, and economic development. Andrew Fellows, formerly the Mayor of College Park, MD, holds joint appointments at the UMD College of Information Studies and the National Center for Smart Growth, specializing in community outreach and engagement. David Jurca is the Associate Director at the Cleveland Urban Design Collaborative (CUDC), where his work integrates professional practice, applied research, and teaching. Responding to dynamic community needs, he advances the CUDC's various advocacy initiatives and spearheads efforts to develop effective techniques for public engagement. W. T. Southards is the director of the Small Business Development Center at KSU-CAEST. Each year the SBDC at KSU-CAEST engages approximately 100 small business clients through one-on-one contact, and more than 100 additional clients through educational events across 22 counties in NE Ohio.

Both the PI (Dr. Emad Khazraee) and Co-PI (Dr. Susan Winter) have offices with contemporary desktops, furniture, phones, and wireless and high-speed Internet access. The laboratory space of the PI is located in the School of Information, Kent State University. The lab space will be used by Ph.D. students (Graduate Assistants), undergraduate students at Kent State University and by the developer, who will be recruited by the project in the second year. PI also has access to the Usability Lab in the School of Information with state-of-the-art technologies — including eye tracking, multiple camera views and detailed analysis software — to assess user behavior and conduct user studies. PI and Co-PI can also use the Digital Laboratory at Kent State with digital library and digital preservation technologies. Both the School of Information at the Kent State University, Kent and the College of Information Studies at the University of Maryland, College Park have instructional spaces with state-of-the-art audiovisual equipment that can host the proposed training workshops in the project. Dr. Khazraee has access to a hybrid cloud computing infrastructure at his research lab at KSU (100 CPU Core, 512 GB RAM, 30 TB of storage) with the capability of extending to Amazon's public cloud. This can be used to host the integrated data hubs for the lifetime of the project and beyond; it can also serve as the host for the backend of the developed platform for DA&BI.

Risks

Four significant, but manageable, risks have been identified for the project:

- Librarians may be poorly prepared to provide data analytics-based information
  - Solution: We will work closely with library staff to provide appropriate training and identify appropriate training that can be integrated into LIS education
- Lack of understanding of the needs of community
  - Solutions: We will engage in participatory design integrating the community into all phases of the project to make sure that the data analytic services actually address the needs of the community
- Reinventing the wheel
  - Solution: We will continually engage in thorough environmental scanning to avoid developing new tools or solutions when there is an existing tool that could be adapted to suit our needs
- Yet another platform abandoned after the grant lifetime
  - Solutions: We will follow open source and crowdsourcing best practices to build a diverse community of practice around the project through strong outreach to the wider community. For example, creating a Python library and an R package makes the tools accessible to a larger audience who can maintain the platform beyond its initial contributors

Communication and dissemination plan

Within the project, we plan to engage in the following communication mechanisms:

- Weekly team meetings between the Ohio and Maryland teams; Monthly progress reports to community partners;
- Project will be kicked-off with a meeting of all stakeholders; Monthly team contact with community partner points of contact; Meetings as requested by community partners;
- Presentations at scholarly conferences and workshops;
- Regular information releases via Kent State and the University of Maryland communications channels and through postings to Listservs, at the beginning of the project, with the release of deliverables, and at the close of each project phase.

We will share our results in scholarly articles and publications with the academic library profession as well as the professional library community. We plan to target business library, small business, and entrepreneurship communities as well as the more general academic library community. Potential conferences and journals include: American Library Association, Association of College and Research Libraries, Association for Information Science and Technology, *Library Quarterly*, *Journal of Association for Information Science and Technology*, *College and Research Libraries*, *Information Systems Research*, and *Library Hi Tech*.

Technical information and documentation of the developed system will be provided via GitHub which incorporates tools for user documentation and access to software code. The developed software will be released under LGPLv3 license, making the source code available to interested parties. The LGPLv3 license allows for the redistribution and use of the software in source or binary form.

3. National Impact

KNEXT will support public library efforts to better accomplish their mission in the 21st century by serving the emerging needs of the public and private sector. The emergence of new forms of data and information production (e.g., big data) is encouraging libraries to redefine and modernize some library services to enhance their relevance in the context of a shifting information and knowledge landscape. As the open data movement gains momentum, additional data resources will continue to be made publicly available to those with the skills and ability to access and analyze them. The library profession has recognized that the availability of big data is a key trend (IFLA Trend Report, 2013). The IFLA report emphasizes that:

> *"vast and expanding data sets acquired by governments and companies through their interactions with Internet users (in conjunction with that generated by scientific research, surveillance and smart object sensors), combined with an accelerated capacity to process and analyse information will expand the possibilities for innovative public/commercial services"* (IFLA, 2013).

The availability of open data enables libraries to enhance the value they provide to their user communities. KNEXT will enable the public to benefit from the availability of massive open data sets by providing the analytic capability to extract insights from data for decision-making. This will move libraries into the locus of community decision-making and strengthen the libraries as essential partners in addressing the needs of their communities. Such a new role for libraries can demonstrate the value of libraries and transform their image among local and national policy and decision makers who support and fund libraries for providing enhanced services to their communities.

The KNEXT project will produce the following outcomes: 1) Needs assessment and readiness assessment; 2) A modular data analytic and business intelligence platform; 3) Training materials and tutorials for deploying and using the platform; 4) Curated set of local open data accessible via an online data portal at each participating county; 5) An evaluation report of the KNEXT implementation and the performance goals of the project.

The outcomes of the KNEXT project have the potential to be widely used by the library community. They will be instrumental in devising a strategy to transfer libraries to be a bridge between the people and the government, and the people and the information, to provide access to big data and open data to the public. KNEXT findings will help us understand community needs for advanced knowledge services, and identify the most urgent data to decision scenarios; therefore, we can focus our attention on identifying types of DA&BI services libraries should provide to best serve the public. The KNEXT project will experiment with different delivery mechanisms and evaluate their success to provide invaluable insights for those investigating the intersections between the digital issues facing libraries and cutting edge work in other fields such as modeling or analytical methods. Finally, KNEXT will create a robust, bottom-up and participatory plan to integrate DA&BI services into local communities' decision making processes. This plan will be applicable to other communities and nationally scalable.

Furthermore, the results of KNEXT readiness assessment will identify the gap in skills of the current library workforce and the type of continuing education required to prepare the LIS community with the skills needed for the 21st century knowledge service. KNEXT will also contribute in capacity building. The project will support multiple graduate students providing them with training in the

sociotechnical perspective, research methods, community engagement, educational techniques, and program evaluation. In addition, local librarians and community members will be trained in finding, curating, and analyzing open data sources. These skills will prepare them to work with the identified constituent groups. Equally important, these new skills in dataset curation and analysis will provide local communities with the infrastructure necessary to extract value from the data generated by local, public, and open data initiatives.

As discussed in the project design, from the initial stages of the project the KNEXT team will work to create a community around the project. This will help to keep the created project platform useful after the lifetime of the grant. For example, the planned Python and R packages based on the KNEXT analytics platform can be continually refreshed by community members in academia, non-profit, or governmental institutions. Because of the modular and open source nature of the developed platform and the strong need for applicable modeling, analytic tools, and civic data initiatives in the library community, the KNEXT platform can be broadly used, adapted, scaled or replicated in libraries and archives across the country by other interested parties.

Beyond the outcomes that directly address the project research questions, the results will also contribute to broader scholarly discussion about innovation and community well-beings by helping us to better:

1) Understand how libraries, in collaboration with universities, can support and facilitate innovation in communities;

2) Understand the benefits and challenges of leveraging existing networks in communities that include libraries, SBDCs, and advocacy groups to foster innovation, civic engagement, and economic vitality in the communities;

3) Identify the social barriers to integrating advances in data analytics into communities to achieve social change.

The emergence of low cost data analytics can transform the knowledge services available to communities creating the potential for substantial societal impact by making small businesses more competitive, reducing the risks of entrepreneurial efforts, re-energizing smart cities, and unleashing community innovations. Successful integration of such services into communities helps them to better identify solutions for their ongoing and evolving challenges. KNEXT will be a testbed for evaluating this potential of information technology and libraries for societal change.

Phase One: Needs Assessment and Pre-Development Research (year one)

Phase One: 2017-2018	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Needs Assessment												
Data Source Identification												
Readiness Assessment												

1.1 Needs Assessment (Start date: October 1, 2017; end date: July 31, 2018)

The researchers at both KSU and UMD will identify and assess the needs of small businesses for business information and data analysis. This will be accomplished through engagement efforts with public libraries, Small Business Development Centers, Chambers of Commerce, and economic development agencies. The research team will conduct interviews and focus groups, working closely with the stakeholders to identify and define community members' needs and concerns that can be addressed through access to business intelligence services. The needs assessment will consider demand for information resources including access to scientific or technical literature, information on the regulatory landscape, assistance with prior art or patent searching, and market assessment. In particular, this research will assess the need for services and tools to assist small business owners with social media marketing and the related data analytics benefiting from big social data. This is an emerging area of business and access to this type of service could be invaluable. By the end of this activity, the research team will have a set of proposed DA&BI scenarios to be implemented in the phase two.

1.2. Data Source Identification (Start date: December 1, 2017; end date: August 31, 2018)

Through this activity, the researchers will explore the sources of existing open data (at both the local and national level) that can address the DA&BI needs identified in the phase 1.1. This search will also investigate any proprietary data sources that may be required (e.g., business databases) and big social data (e.g., Twitter and Tumbler) and determine how they can be accessed through APIs for public library use. The research team at KSU has extensive experience working with big social data. One of the main goals of phase 1.2 is to assess the accessibility of county-level data and to identify points of contact and processing to aggregate the data. This will be an extension of preliminary work already begun by some of the project partners such as the Lorain County Public Library System. During Phase 1.2, the research team will evaluate the existing solutions (such as CKAN or Socrata) for creating an integrated hub for local data and will decide which can best serve the local needs taking into account the nature of the county data, the social and the technical factors. The research team will also identify the transformations that are needed to prepare the open data for analysis to make each source usable (for example, exporting some data formats to commonly used JSON or CSV formats). Finally, we will develop plans to create the required routines and to provide the training for community partners. By the end of this activity we will have an inventory of public open data sources that can address community needs and support economic development.

1.3. Readiness Assessment (Start date: June 1, 2018; end date: September 30, 2018)

As part of phase one, the research team will conduct a readiness assessment to identify the users' needs and knowledge of data analysis. We will investigate the local business and community data analytic capabilities and determine whether they need training on basic data analytics or more advanced topics. We will also assess whether the libraries are ready to deploy the tools and how best to present the outcomes to the community. The research team will conduct interviews and focus groups to identify the required actions to prepare the community to use the KNEXT services. This will define training content, suggest preferred pedagogical techniques, and identify knowledge gaps to be addressed in phase two when we prepare community partners. A major outcome of the readiness assessment will be identification of the types of continuing education interventions required to prepare the LIS community with skills needed for KNEXT.

Phase Two: Development and Initial Implementation (year two)

Phase Two: 2018-2019	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Analysis and Development												
Development of Integrated Data Access Hub												
Training the Community Partners												
Early Implementation												

2.1 Analysis and Development of Analytics Platform (Start date: October 1, 2018; end date: September 30, 2019)

In this phase, researchers will develop DA&BI strategies to meet the community needs identified in phase one. Researchers will define the needed business intelligence products and services and identify the most effective and efficient methods to deliver these products to the local community. Dr. Khazraee will work with a full-time developer and two doctoral students, one at each university, and use advanced computational methods such as machine learning and data mining. Using these methods, along with existing data, the team will turn the identified business intelligence needs of the community into accessible routines available via a GUI dashboard to be used by the librarians in the partner libraries. The design of these routines and the platform will be constantly checked by the community partners to make sure they address the community needs.

2.2 Development of Integrated Data Access Hub (Start date: March 1, 2019; end date: September 30, 2019)

For each county participating in the project, the KNEXT project will create a data hub based on the analysis of existing solutions (e.g., CKAN or Socrata). This Hub will be accompanied by an inventory list of other open data. Collaborators at each site will receive training on how to extract relevant data from each source.

2.3 Training the Community Partners (Start date: April 1, 2019; end date: September 30, 2019)

Based on the result of readiness assessment, the KNEXT team will provide the required training to community partners in the two research sites.

2.4 Early Implementation (Start date: May1, 2019; end date: June 30, 2019)

Following an agile development approach, as the team implements each of the analytic scenarios, they will be made available to community partners who will assist in seeking volunteers for early testing and adoption. The goal is to identify the shortcomings and problems and quickly integrate the feedback from the community into a revised scenario. In addition, one graduate student in each site will work one-to-one with librarians in partner libraries to train them and solicit their feedback to inform the research team and provide guidance for iterative development.

Phase Three: Implementation and Evaluation (year three):

Phase Three: 2019-2020	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Implementation												
Evaluation												
Community Building												

3.1. Full implementation (Start date: October 1, 2019; end date: June 30, 2020)

In phase three, the research teams will deploy the full range of DA&BI services developed in phase two. At this stage only minor improvements and critical bug fixes will be added to the platform. The Research team will conduct follow up interviews and focus groups with community partners to identify areas for improvement and barriers to effective use of DA&BI services for community decision making.

3.2. Evaluation (Start date: April 1, 2020; end date: September 30, 2020)

The KNEXT team will partner with community agencies to evaluate the effectiveness of the services. Initially three sets of metrics will be used for the assessment of the success of project application-oriented metrics, conceptual metrics, and IMLS performance measures as described in the proposal narrative.

3.3. Community Building and Sustaining the Project (Start date: February 1, 2020; end date: August 31, 2020)

To ensure the sustainability of the KNEXT project after the lifetime of this grant. The KNEXT team will introduce the project to the broader community through participation in THAT Camp, Library Hackathons, SBDC National conferences as well as academic conferences, and professional library conferences. While hosting the open source developed code in a public GitHub repository we will work to create an open source governing community around the project following the open source community best practices (Crowston, Wei, Howison, & Wiggins, 2008; Crowston, Howison, & Annabi, 2006; Crowston & Howison, 2005). As part of this attempt, the research team will identify community champions who can advocate for the project, create a blueprint for the long-term use of the platform, and seek stable operational funding.

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions
You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

PART I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

All code created in this project will be released under GNU Lesser General Public License (LGPL) Version 3. LGPLv3 has been widely-used in the open source community. Permissions of this license are conditioned on making available complete source code of licensed works and modifications under the same license or the GNU GPLv3. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights. However, a larger work using the licensed work through interfaces provided by the licensed work may be distributed under different terms and without source code for the larger work. We believe LGPv3 helps the derivative works based on this project also be accessible to public while makes it possible to create a larger work through interfaces be distributed under different terms. Therefore, it can still encourage creating other works based on the product of this grant.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

Kent State University policy and its faculty collective bargaining provisions, faculty members retain copyright to their academic works. Accordingly, digital products developed in this project will be owned in total by the participating authors/developers. As explained about the code will be released under LGPLv3.

The project team will be highly involved in outreach and dissemination of the developed platform as described

in the proposal narrative to engage with interested parties who are willing to use the code. The tools will be shared with potential users via listservs and social media, as well as presentations at academic and professional conferences and publication of articles in journals.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

The product will not involve any privacy concern or copyrighted material.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

We are not creating digital content for this project other than the digital tools and research data described in detail in Parts III and IV. In the project, we will help partner libraries in each county to create a data hub, an integrated point of access to public open data by aggregating and curating local and public data for local county data that they can use for the purposes of data analytics. The quantity of the data is not clear at this stage of analysis but the KNEXT project has capacity of collecting data up to 30 TB. We plan to convert all collected data to commonly used format such as TXT, CSV, JASON, and XML.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

N/A

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

TXT, CSV, JASON, XML, PNG, and ODF.

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

KNEXT will not be responsible for production of data and the data will be collected from the trusted partners such as government agencies. KNEXT will make sure that the data is correctly converted to open formats. Each data conversion process will be audited by KNEXT team to assure the consistency of data conversion.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of

performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

For each partner county, the data will be stored in a cloud-based open source data portal (e.g., CKAN or Socrata). Each portal will be hosted at the could computing infrastructure at the Kent State University for the lifetime of the grant with redundant copies. The data will be backed up daily on Amazon Glacier.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

The metadata that will be produced through this project will consist of description of data sets received from local governmental parties which will be stored using Dublin Core metadata framework. The research team will also use PREMIS data model to support the preservation of digital objects and ensure their long-term usability.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

Since the complete data portal will be backed up, the metadata generated in the system will be backed up as well.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

The open source data portal we plan to use (e.g., CKAN or Socrata) will provide APIs to access the metadata. The metadata then will be accessible via OAI-PHM.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The open government data collected in the county hub will be openly available online. Users should be able to access the system via web browsers. Programmers can access the portals API after receiving an authentication token.

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

N/A

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

The software will be developed in this project will be a web-based application as a dashboard for data analytics. The user can access the frontend via a web browser and the analysis of the content will be done on the containers in the cloud. User can access a list of data analytics routines—identified in the needs assessment activity— in the web-based frontend to upload data, conduct the analysis, and export the results including data visualization. The analytical functions that the software can perform will be special scenarios of the following functionalities:

- Association rule mining
- Anomaly detection
- Graph mining
- Social network analysis
- Text and web analytics
- Sentiment and affect analysis

The scenarios will be tailored for specific needs of small businesses, entrepreneurs, and community advocates. The primary users of the systems will be partners in public libraries who will serve small businesses, entrepreneurs, and community advocates. The sources of data for each scenario will be identified in the first phase of the project, then specific data analysis pipelines will be designed to conduct the required analysis from transforming the raw data to present the output report and visualization. For each data analysis routine, the user (librarian) only needs to enter the required parameters. For example, if the analysis is a Twitter data analysis of brand with sentiment analysis, the librarian needs to enter the keywords or the brand names; then system will walk the librarian through the steps and will send the results in the designated time frame for the analysis.

The Graphic User interface of the platform helps users to pick that analysis scenarios, set the parameters, and input the required data. Then this information will be sent to the backend where each analysis scenario has been set up as a pipeline. The required parameters will be applied to the scenario and the analysis will be done in the cloud. The results then will be sent to the frontend for user. It can include interactive visualizations which allows the user to explore the results.

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

In theory, most of the analysis we describe can be conducted by data science platforms which support data mining and statistical analysis (e.g., KNIME, RapidMiner, Dataiku, Python, and R); however, it requires expertise in data science, statistics, data mining, and machine learning to prepare data and design the analytical

routines, and eventually interpret and present the results. It also requires expertise in working with advanced data analytic environment such as Apache Hadoop or Spark. These are requirements that only major companies or those with access to resources for data science have at their disposal. Most of the public libraries neither are in position to perform such complicated operations nor have the expertise to conduct these tasks. The platform we will create will streamline the process by separating the backend processing from the fronted GUI and it will only focus on the specific limited number of cases identified in the needs assessment activity of the phase one that serves the three patron groups; therefore, we will streamline the process for the end user to simply walk through each step without the need to programming skills or managing the software backend. The output will be tailored for the specific use cases and librarians will be trained to easily interpret them.
The major promise of the KNEXT project is to bring data analytics to the public libraries to better serve the community. Therefore, we need to start this process by removing the existing barriers and make such services accessible. The KNEXT create a first step to bring a limited but powerful set of analytic services to public libraries; however, its open and modular architecture will allow addition of other analytic services to the platform from the community.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

Python and R are the most popular programing languages for data science and most of the data scientist are familiar with R and Python. There are many existing libraries for both languages (such as scikit-learn, matplotlib, and NLTK for Python and Caret, ggplot2, and tm for R). Therefore, we plan to use Python and R to create the data analytic pipelines. Django and Shiny are web application frameworks that combine the computational power of Python and R with the interactivity of the modern web. Through using them we can create interactive web interface for the data analytic platform. Moreover, existence of PySpark and SparkR makes it possible to use Python and R for Big Data Analytics as well.

We also consider publishing the developed analytical pipelines as standalone Python and R packages that individual users who do not need the interactive web interface can used them from within Python and R environment. It helps to increase the use of the developed product and create a community around it.

Our goal is to develop this platform as a cloud application to makes it easier for other users to deploy it at a very low cost. One method to develop reusable applications for the cloud is to use containers such as Docker which isolate applications and make them to run more securely and efficiently. Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: code, runtime, system tools, system libraries – anything you can install on a server. This guarantees that it will always run the same, regardless of the environment it is running in. Thus, using Docker technology makes the application easily reusable for other researchers and users.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

Since we are using the widely-used Python and R programming language and their standard libraries, our platform can be easily extended or modified. The platform will be accessible via an API which makes it easy to programmatically access it or interact with it.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

As mentioned above we will use Docker container technology to wrap up the application in a complete filesystem that contains everything it needs to run: code, runtime, system tools, system libraries. Therefore, those intended to use the platform should be able to deploy it seamlessly in virtual environment or cloud computing environments such Amazon We Services. The whole Docker container will be installed on top of GNU/Linux Operating System. We plan to us Ubuntu Linux as the OS for the project. However, the system can be deployed in other environments as well.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

We will follow an agile software development approach to support iterative, incremental, and evolutionary development, early delivery, and continuous improvement via integrating feedback from the community. We will use the GitHub infrastructure for version control and documentation. Documentation will be a required deliverable for all programmers working on the project. KNEXT internally will use Atlassian's Confluence for documentation, JIRA for task and bug tracking and Github for exposing the code to the public. All documentation will be recorded in Markdown to make it easy to deliver it along with software be accessible via GitHub. KNEXT project will produce detailed documentation for installing and implementing the platform source code. KNEXT developer will review the documentation to ensure that it is understandable to external users. The GitHub Wiki will be used to create user manual, tutorials, FAQs and other required documentation for the users. The documentation will be maintained and updated by the research team during the lifetime of the grant. As described in the proposal the goal is to create an independent community around the project to maintain it beyond the lifetime of the grant.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

N/A

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

The software created in this project will be released under GNU Lesser General Public License (LGPL) Version 3. LGPLv3 has been widely-used in the open source community. Permissions of this license are conditioned on making available complete source code of licensed works and modifications under the same license or the GNU GPLv3. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights.

However, a larger work using the licensed work through interfaces provided by the licensed work may be distributed under different terms and without source code for the larger work.

We believe LGPv3 helps the derivative works based on this project also be accessible to public while makes it possible to create a larger work through interfaces be distributed under different terms. Therefore, it can still encourage creating other works based on the product of this grant.

Kent State University policy and its faculty collective bargaining provisions, faculty members retain copyright to their academic works. Accordingly, digital products developed in this project will be owned in total by the participating authors/developers. As explained about the code will be released under LGPLv3.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

Source code and documentations will be available for public use through a public GitHub repository. The project team will be highly involved in outreach and dissemination of the developed platform as described in the proposal narrative to engage with interested parties who are willing to use the code. The tools will be shared with potential users via listservs and social media, as well as presentations at academic and professional conferences and publication of articles in journals.

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: GitHub

URL: https://github.com/ksusocialcomputing/KNEXT

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

N/A

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

N/A

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

N/A

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans

for preserving the documentation and ensuring that its relationship to the collected data is maintained.

N/A

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

N/A

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

N/A

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

N/A

A.8 Identify where you will deposit the dataset(s):

Name of repository: N/A

URL: N/A

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

N/A