Abstract

Open data has the potential to positively transform many aspects of society, but this promise requires investment in data access and use specifically for the public good. As development of technical infrastructure to enable data science proceeds apace, a gap is forming between what is possible technically and what is being achieved in public institutions that serve open data stakeholders. In particular, formal workforce development in data expertise is lagging for public libraries, much as it was for academic libraries when they began assuming responsibility for the curation of research data over a decade ago.

The Open Data for Public Good (ODPG) project, led by Carole Palmer at the Information School at the University of Washington, will make progress on the public library workforce gap through an educational program to prepare both new students and practicing professionals to: curate collections of open data of value to local communities, build infrastructure and preservation environments needed to sustain open data collections, and collaborate with open data providers on advocacy and outreach activities. The program will build on the new Future of Libraries strategic initiative in the Information School, scaling and adapting current data curation and data science activities in collaboration with the DataLAB and Technology & Social Change (TASCHA) group, and the University of Washington eScience Institute. ODPG will engage in collaborative curriculum development and support student field experiences through partnerships with the Seattle Public Library, Washington State Historical Society, Washington State Department of Transportation, and the Washington State Office of Technology. The range of collaborators and partners will assure that ODPG is grounded in current data best practices and demands in the public sector and is responsive to the needs of public libraries.

The three-year project has been designed to serve four stakeholder groups: 1. **Students** in Library & Information Science, through new courses and practical field experiences; 2. **Public Sector Organizations**, through engagement with UW data science experts, mentoring experiences, and student projects; 3. **Practicing Librarians**, through continuing education events and resources; and **LIS Educators**, who will be able to adopt the Data Literacy Framework, curricular materials, and approaches developed by ODPG. The minimum number of students served is estimated at 105 for the Foundations course, with 75 completing the full ODPG course sequence, 40 completing the sequence plus data science electives, and at least 45 participating in Field Experiences. Over the course of the project, ODPG expects to at least double the network of public sector partners and mentors, aiming for 8 participating organizations and 15 mentors, to significantly increase the range of field experiences available to students and benefits back to external institutions. Workshops will serve approximately 60 professionals, with webinars and open educational resources reaching dozens of professionals each year.

A series of formative evaluations will be conducted throughout the project to iteratively improve the course content, fieldwork experiences for both students and mentors, the scope and quality of open educational resources and the overall open data literacy framework. A summative evaluation will be produced at the end of the project, focused on the impact of both the academic and continuing education contributions, including a study of broader workforce needs for public open data expertise. Finally, a strategy will be developed and implemented for long-term sustainability of the program within the Information School.

Open Data for Public Good: Data Literacy Education for Public Information Professionals

1. STATEMENT OF NEED

Open data has the potential to positively transform many aspects of society, from increasing civic engagement in political processes (McClean, 2011), to improving community resilience to a changing climate (Boulton et al., 2011), and even strengthening cultural heritage awareness (Kitchin, 2014). In recent years, there have been many advances in technical infrastructure for discovering, accessing, and using open data (George, 2014; Davies & Bawa, 2012), and libraries are embracing new roles in the collection, storage and preservation of data resources (Shueh, 2015). Library and Information Science (LIS) education programs have made considerable progress in data workforce development, but it has been focused largely on academic libraries (Steinhart & Qin, 2012; Weber et al., 2012; Carlson et al., 2013; Palmer et al., 2014; National Research Council, 2015). A significant gap remains, however, between what is possible technically and what is actually being achieved with data infrastructure and services, especially within public libraries where open data has great potential to serve local communities (Bertot, Butler & Travis, 2014). In short, formal workforce development in data expertise is lagging for public librarians, much as it was for academic librarians as they began to take responsibility for the curation of research data about a decade ago.

We request support for the proposed Open Data for Public Good (ODPG) program through an LB21 project grant for "continuing education and programs to build institutional capacity." Based at the Information School at the University of Washington (UW), ODPG will address both programmatic aims. ODPG will prepare future and current public librarians to curate collections of open data of value to local communities, build the necessary infrastructure and preservation environments to sustain open data collections, and work to make more data open to the public by collaborating with open data providers on advocacy and outreach activities. These roles will be essential to assuring sustainable, equitable access to open data by the public and are a natural extension of the mission of public libraries. The goals of the proposed education project will be achieved by strengthening and leveraging the team's collective capacity in LIS and data science education, resulting in an integrative education framework and curriculum for open data literacy. ODPG's educational activities will be conducted in collaboration with a network of practicing information professionals from the public sector–each of whom represent a municipal or government agency engaged in ongoing open data initiatives and invested in advancing data expertise in the LIS workforce.

Open Data as a Public Good. Librarianship has a long and rich tradition of increasing the value of information resources that act as public goods (Braverman, 1982). These activities include organizing and providing access to government documents, public records, and cultural artifacts. As such, public libraries play a critical role in stewarding and providing access to electronic government information (Bertot et al., 2009); this increasingly includes open transportation, spending, permitting, and cultural heritage data resulting from government transparency initiatives (Davies, 2012). While there is much enthusiasm for the creation of technologies to disseminate open data, relatively less emphasis has been placed on the service models required for meaningful use of these resources, or on how newly designed open data infrastructures will be sustained and preserved over time. Information services and preservation are both areas where LIS can make a strong contribution. For instance, researchers working on personal privacy issues in open data publishing for the city of Seattle note that municipal service agencies lack many of the skills needed to understand retention, disposal, and secure long-term storage of data (Whittington et al., 2015). Smaller municipalities–Waco, TX (Smith, 2015) and Chattanooga, TN (Shueh, 2015)–recognize the potential value of working with public libraries to design and launch new open data initiatives.

LIS approaches to the curation and long-term preservation of open data as a public good are urgently needed as open data initiatives escalate at city, state and federal levels. The current library workforce, however, is lacking the data expertise needed to be strong partners or to lead efforts that respond to the interests and needs of their service communities. The proposed ODPG project will prepare public information professionals to step up to the challenge and make significant contributions to open data initiatives through the following: 1. **Curriculum**

for open data literacy and expertise, built on established, successful advances in data curation and data science education, and available online to reach students regardless of physical location; 2. **Practical field experiences & mentoring**, including mechanisms for bringing practice-based knowledge back into the classroom; 3. **Continuing education and outreach** to advance expertise for practicing librarians to advocate for and implement open data initiatives.

Public Sector Data Literacy and Use. Previous work has considered what it means to establish competency in critically evaluating, using, and managing data as an informational resource (Mandinach & Gummer, 2012; Twidale, Blake & Gant, 2013). LIS data literacy work has addressed, for example, the meaningful interpretation of social science data (Stephenson & Shifter, 2007) and how STEM students manage data in e-Research settings (Qin & D'Ignazio, 2010; Carlson et al, 2011), with numerous efforts focused on subject experts providing data management for research communities. Digital curation curriculum and educational opportunities for professionals also continue to advance, with increasing attention to data resources (e.g., CURATECamp and CurateGear). To date, however, little investment has been made in the specific open data competencies now needed in public libraries and other public sector organizations serving the general public.

We believe an effective way to frame data literacy and technical competencies for the public sphere is to build on current data services within state and federal IT departments, where "data warehousing" approaches are often applied through a three-part ETL model: **extract** data from existing resource pools, **transform** these data for new (often simple) analytical purposes, and then **load** transformed data into appropriate access environments, such as a data portal. This entire process is often referred to as ETL (Kimball, 1998). It is worth noting that organizations applying the ETL model to open city data have created products that require substantial end-user training and have no budget for services or programs needed to assist the public in using these data products (Goerge, 2014). Thus, while ETL can serve as a basic and familiar underpinning for open data work processes, the ODPG framework will extend this base of competencies to ensure that the work of LIS professionals is focused on making open data useful to their broad constituencies and that investments in valuable data resources can be sustained.

Concentrations in the ODPG data literacy framework will cover: **1.** Gathering data based on user community needs, including: eliciting requirements from organizational or community stakeholders; criteria-based collection development plans for acquiring open data. **2.** Presenting data for meaningful reuse, including: transforming open data into meaningful units for consumption (cleaning, restructuring, refining, or renaming files, etc.); standardized metadata descriptions; legal, privacy, and ethical constraints on access; and creating compelling visualizations to accurately summarize data collections or convey specific data narratives. **3.** Sustaining platforms for continued access, including: digital preservation plans for open data repositories; standards for deploying sustainable data curation environments (e.g. compliance with ISO-16919:2014), evaluating solutions from commercial providers to manage large and complex datasets; and promulgating policies for ensuring long-term accessibility of open data resources.

In integrating these areas into the ETL model, our initial data literacy framework will be structured as **GET** (gather, extract, and transform) and **LPS** (load, present, and sustain). The core curriculum will be implemented through two courses: 1. An existing course, 'Fundamentals of Data Curation' will be re-designed and emphasize open data case studies and assignments; 2. A new course on public information services for open data will give students hands-on experience using open-source software to create open data portals (using CKAN) and data visualizations (using R, Python, Javascript/D3/Vega-Lite), as well as designing the policies and service models that ensure long-term preservation of open data resources. Working with the Associate Dean for Academics in the Information School, as we build strong ODPG curriculum in the Masters of Library and Information Science (MLIS) program, we will build bridges to the data science curriculum for cross-fertilization of LIS and data science in ODPG and to attract students to open data librarianship. We will begin by integrating open data topics and applications, informed by LIS principles and practices, in a module on open data curation for a new introductory data science course taught in the UW iSchool.

Practice-based Education. The need to strengthen connections between LIS education and practice, and the value of fieldwork in professional preparation, are well understood (International Federation of Library Associations 2000, Roberts, Madden, & Corrall 2013). To address these concerns, ODPG will bring practice directly into the classroom and send students out into the workplace. Our collaborative practice-based approach will begin with data professionals interacting with students in the core curriculum. Data scientists from partner organizations and UW research centers, including the eScience Institute and the Technology & Social Change Group (TASCHA), will contribute hands-on, small-group instruction in data management and analytics tools to aid student projects. Professional mentors will direct student field experiences providing expert guidance on technical, organizational, and administrative aspects of student projects. The ODPG mentorship model is informed by previous work developing mutually beneficial data curation internships for LIS students and mentors participating in the Data Curation Education in Research Centers (DCERC) project (Kelly et al. 2013, Palmer et al. 2014). In ODPG mentors will have more input into topics and use cases used in the classroom. In addition, students and mentors will produce case studies and document best practices as part of the internship process, to be used in classes and in the development of open educational resources (OERs).

Continuing Professional Development. Educating cohorts of new MLIS graduates as open data experts is essential for establishing a new base of competency and leadership in the field of public librarianship, but it is not sufficient. We will implement a continuing education component, mobilizing students as open data ambassadors to address the data literacy skills gap in the existing public library workforce. Ambassadors will assist the team in developing OERs, such as webinars, self-directed course materials, and best practices guides in curation, customized for non-technical library practitioners. Webinars will be informed by the eScience Institute's success delivering data curriculum to large audiences (see, for example, Howe and Rose 2012). Students who have completed both the ODPG curriculum and fieldwork will employ a social learning approach (Wenger 2000), serving in "train-the-trainer" roles to teach library practitioners to use ODPG data literacy materials locally for outreach and instruction. Ambassadors will work with the ODPG team to deliver a regional workshop, followed by a national workshop at a professional conference (e.g. the Public Library Association annual meeting) for librarians interested in providing local open data instruction for their staff and the public. These activities will form a network of ODPG professionals who can continue to educate peers, develop open data initiatives, and champion data literacy.

A broader outreach effort will be conducted in collaboration with the TASCHA (Technology for Social Change) group and their forthcoming long-term work on the Bill & Melinda Gates Foundation's Global Libraries legacy initiative. Through this collaboration, we will make available and promote OERs to a wider range of public libraries, targeting those with limited resources across the U.S., as well as libraries and information centers in developing countries. Engagement with this array of information service providers will enable us to assess the value and applicability of the OERs for educating public library staff, and to select and further develop these materials to maximize their instructional value.

2. IMPACT

Positioned at the intersection of two new strategic foci at the UW Information School—the Future of Libraries *and* Data for Social Good—ODPG will offer a robust model for innovative LIS data education and raise awareness of our field's growing contribution to social, environmental, and cultural advances in the era of big data. As open data increases in importance for constructive civic engagement, enabling equal and fair cultural opportunities, and sustaining economic vitality, ODPG will prepare active, technically capable, and influential information professionals who can make lasting contributions in their communities and empower others to sustainably curate open data to serve the public good.

The minimum number of students to be served is estimated at 105 for the Foundations course, with 75 completing the full ODPG course sequence, 40 completing the sequence plus data science electives, and at least 45 participating in Field Experiences. As the curriculum will be delivered online, students in the online MLIS program may engage, extending the program's geographic reach. Below we describe an approach to diversity recruitment and inclusion that will assure underrepresented minority students are included in ODPG cohorts.

Over the course of the project, ODPG expects to at least double the network of public sector partners and mentors currently involved, aiming for 8 participating organizations and 15 mentors, to significantly increase the range of field experiences available to students and benefits back to external institutions. Workshops will serve approximately 60 professionals, with webinars and OERs reaching dozens of professionals each year, and actively promoted to small and under-resourced public libraries in the U.S. and developing countries. The broader LIS education community will benefit from the formal curriculum and case studies to be shared through the project website, and evaluation outcomes and implications of the framework will be disseminated formally in journal publications and conference presentations, as well as informally through our project blog. The ODPG partnership approach will also serve as an example for other LIS schools of a strategy for collaborative professional education, well suited for rapidly changing areas in the field where the profession needs direct exposure to best practices and new technologies as they emerge in the workplace.

The design of ODPG, anchored in formal and practical instruction blended with real-world experiences, directly addresses the LB21 programmatic goals of supporting academic programs and professional development related to the National Digital Platform priority, including the identified need to cultivate a digital library workforce and increase hands-on practice in degreed programs.[1] The activities described above are aligned with IMLS strategic goals to prepare the public to be full participants in local communities and global society, promote libraries as community anchors that enhance civic engagement and cultural opportunities, and promote use of technology for discovery of knowledge and cultural heritage.

Primary areas of impact include: **Innovation in curriculum and instruction** that draws on and feeds back into a network of professional partners, reaches students and current practitioners, and graduates diverse cohorts of professionals prepared to lead and mentor colleagues. **Enhanced student learning** based in quality practice-based classroom work, strongly mentored field experiences with carefully designed real-world projects, and engagement with experts in LIS, data science, and public sector information services. **Models for open data products and processes** through exemplar data resources, documentation of best practices, and case studies produced in field experiences, and an adoptable and sustainable partnership approach. **Strengthened workforce** through in influx of public library professionals with data strong expertise and experience, and a network of professionals active in data initiatives for public good.

ODPG will further impact the field by graduating role models who demonstrate the importance and value of technical data competencies in public libraries. In particular, we expect some students in the ODPG program will pursue additional advanced study through the Data Science degree specialization offered by the UW iSchool, which to date has attracted a relatively small number of MLIS students. ODPG students will also have opportunities to become involved in research opportunities in the DataLab and the TASCHA group.

Evaluation. ODPG is aligned with the IMLS agency level goal of "Learning" and the Performance Goal to train and develop library professionals. A combination of formative and summative evaluations will be applied to inform development of the ODPG program and assess progress and outcomes toward these goals, including information on specific IMLS performance measures and the expected data on all ODPG program participants. The ODPG evaluation plan is modeled on a previously developed approach designed for iterative improvement of curriculum and internships (Mayernik, et al., 2015). Students will be surveyed on learning and experiences in courses and field experiences. Continuing education participants will be surveyed on their experience, learning outcomes, and recommendations for improvement. Interviews will be conducted with instructors and mentors to gauge and facilitate progress toward mutually beneficial partnerships.

Course evaluations will be conducted for each course each year to capture both student and instructor perspectives. Evaluation data will support a coordinated course revision process to achieve balance and complementarity across ODPG courses and other Data Science options. In particular, standard UW course evaluation forms will be supplemented with survey questions to assess course content, class experience, and

[1] https://www.imls.gov/sites/default/files/publications/documents/2015imlsfocusndpreport.pdf

learning in relation to specific ODPG goals, and interviews will be conducted with guest lecturers and instructors to capture their experience and recommendations.
Field experience evaluations, from students and mentors, will be collected, analyzed, and shared in group mentor meetings to exchange lessons learned and identify areas for improvement for the next placements. In addition to student surveys to assess field experiences, an online discussion forum for students on learning and challenges will serve as a complementary data source. Mentor interviews will be conducted to collect feedback on project development and student interactions, and recommendations for program improvement.

Continuing education evaluations from participants will be used to assess expectations, learning, and recommendations for improvement. For workshops, paper surveys will be administered in the last session to collect anonymous feedback. For webinars, an invitation to an anonymous online feedback survey will be sent to participants via email.

A broader workforce assessment will be an integral part of the overall evaluation. Interviews will be conducted with employers at partner organizations and those that hire ODPG students, with an extended sample representing employers from a broader base of public libraries and other public sector agencies. This study will provide fuller baseline from which to judge the adequacy of ODPG in preparing graduates in terms of employer expectations and job market trends. It will also be instrumental in designing a sustainable internship model that can be extended to additional partner organizations.

Metrics will be documented throughout the project on the following variables: **students and courses**—number of students who begin and complete specific ODPG courses and the sequence of courses; diversity of students attracted to the program; amount of new course material developed; number and types of capstone and independent study projects completed in addition to primary courses; **field experiences and mentoring**—number of grant funded and partner supported field experiences completed; hours and types of mentoring conducted with students; number and kind of project outcomes—data resources, policies, curation processes; student / mentor collaborative activities; number of partner personnel involved in ODPG and their roles; **outreach and dissemination**—number of educational modules developed; website access analytics; number of workshop participants; number of webinar participants; number of student ODPG dissemination activities; number of team authored papers, reports, and presentations.

A final synthesis of results from all the evaluation components will be completed at the close of the project, including an additional analysis of the institutional resources needed and challenges to sustaining and growing the ODPG partnership network and student field experiences. The sustainability analysis will be essential to continuation of the program, but also of value to other schools aiming to adopt the education model.

3. PROJECT DESIGN

Curriculum for open data literacy and expertise. Beginning in Year 1, the Curriculum Team will concentrate on new content, guided by the GET-LPS framework, to cover collecting, managing, curating, preserving, and providing meaningful access to open data resources for the public good. The focus will be on use of open civic,[2] environmental,[3] and cultural heritage[4] data and also assuring adequate basic introductions to programming and statistical analysis. A sequence of two courses—Fundamentals of Data Curation, redesigned for ODPG, and a new course on Services and Infrastructures for Open Data will be offered and required for students to advance into summer internships. Advising guidelines will be developed to encourage relevant elective courses. The recently introduced Digital Preservation course at the iSchool will an important elective, also revised for ODGP objectives, along will options in Data Science, such as programming, database development, and data analytics. As discussed below, Capstone projects will also be central for students who participate in field experiences at partner institutions or participate in in the DataLab, TASCHA, and eScience Institute initiatives.

[2] e.g., City; County; State; & Country Open Data
[3] e.g., Community Climate Data; Environmental Health ; Agriculture & Local Food Data
[4] e.g., Native American; African American; Latino/a ; and Regional Open Cultural Data

In Years 2-3, modules for the Fundamentals course will be further enriched with practical resources and activities relevant to open data user communities, in areas of selection and appraisal; representation and organization; digital preservation and long-term data management; rights and access conditions; and promotion and sustainability of open data initiatives. The Services course will be enhanced with exemplar case studies developed in collaboration with external project partners. Class topics will include data enrichment and fit-for-purpose curation; data integration and visualization; and interoperability and aggregation, with a focus on practical data repository development using open source software.

Course evaluations will be carried out as described above to enable course content to be iteratively designed and improved each year, and the curriculum as a whole will be monitored to ensure appropriate coverage of technical and service capabilities.

Practical field experiences & mentoring. In Year 1, two options will be developed for students to gain practical experience and work with dedicated professional mentors. **Directed Fieldwork** is an existing iSchool course that allows students to pursue hands-on projects at an organization, directed by a faculty member as a variable credit/no credit independent class. For ODPG, students pursuing this option will undertake open data projects and have additional one-to-one mentoring from a professional at the fieldwork site. **ODPG Internships** will be a new, competitive field experience option for summer placements. A cohort of five high-performing students will be selected and matched with internship projects in Years 1-2. Beginning in Year 2, we aim to increase the number of student field experiences through additional **partner-funded placements,** and work to cultivate **remote fieldwork** options for online students.

Students will work with data mentors at their field site on projects designed to directly benefit local service communities, using open civic, environmental, or cultural data sources. Paid interns will have additional responsibilities to work with their mentors to create professional development resources to be disseminated and used by practicing librarians (e.g. case studies, technical documentation, and best practices guidelines). Directed fieldwork students will be encouraged, but not required, to produce and contribute to these outputs as well. All fieldwork experiences will culminate in formal Capstone projects (https://ischool.uw.edu/capstone), and paid interns will be expected to use their Capstone as the basis for a panel, workshop, or paper presentation describing open data literacy educational activities at a professional venue, such as ALA or SLA.

ODPG external partners identified the following projects as initial fieldwork opportunities within their organizations. In addition, we expect students to assist each site in developing a digital preservation plan.

Seattle Public Library: Branch-directed open data catalog with subsets available for particular neighborhoods; linking Seattle open data with emerging best practices in open data catalog vocabularies; and developing a community outreach service model for open data.

Washington State Historical Society: Linked open data to supplement a Native American digitization project catalog; Women's Consortium Collection migration to an open data portal; and a preservation policy for curated exhibitions past and present.

Washington State Department of Transportation: Creating a compliance strategy for the USDOT Public Access Plan (best practices for persistent identifiers, preservation, etc.); extending existing metadata for WS-DOT open data set catalog; and developing new metrics for open data usage and social impact to justify open data investments.

ODPG Internships will be conducted in the summer, running on site for approximately 6 weeks, with two additional weeks devoted to development of professional development resources. Directed Fieldwork will be allowed in any academic quarter. The core team will work collaboratively with mentors to design appropriate projects. Internship profiling instruments will be used to coordinate matching of student interests to project options and mentors. Project partner Will Saunders, at the State of Washington Office of the Chief Information Officer, will be instrumental in growing the network of fieldwork sites in the region and helping to identify and advise on potential projects. A shared online forum in CANVAS, the distance education teaching platform at UW, will be provided for students to "talk shop" about what they are learning through their field experiences. These discussions will be required for ODPG Interns and used as a virtual meeting space for the cohort,

mentors, and advisor to build community, by sharing achievements and challenges in the workplace, and for ODPG team members and partners to provide input and support for student activities.

The evaluations, outlined above, will be used to iteratively improve field experiences, including online surveys administered to students when they complete fieldwork and analysis of the CANVAS discussion forum content. Interviews will be conducted with mentors following completion of internships covering project development, value of student work, adequacy of student preparation, and the mentoring experience. All three approaches will be used to assess areas for iterative improvement of field experiences.

Continuing education and outreach. Outputs from the ODPG curriculum and field experiences will be mobilized in the creation of educational resources for practicing public librarians and public information stewards. Webinars will begin in Year 1, leveraging the new course content. They will be recorded and made accessible to institutions interested in educating their personnel on open data services. Open educational resources (OERs) will be made available on the web beginning in Year 2 based on the field placement work of interns and mentors in Year 1, with enhancements for online delivery directed by the iSchool's professional Online Learning Administrator. We will also leverage the success of the eScience Institute's online instruction offerings, which have attracted over 200,000 students,[5] as well as their partnership with Software Carpentry[6] for introductory programming, and their Community Data Science events. Example topic areas for ODPG OERs include creating open data catalogs with open-source software, curation workflows for populating a repository with open data, developing open data preservation policies, and the use of open-source visualization tools to analyze and present different thematic collections. A plan for outreach via TASCHA will be developed in Year 1, to leverage synergies with their Gates Foundation Global Libraries legacy partnership activities, with active outreach to under-resourced libraries and global libraries concentrated in Year 3. Outreach will leverage their work on visualization for public library data, conducted in collaboration with Zepheira and Community Attributes.[7]

Materials produced in the first 18 months will be grounding for a community-focused workshop in Year 2 at the Seattle Public Library, informed by SPL's ongoing and well received open data events.[8] A national workshop in Year 3 will draw on material produced throughout the project. The heart of the final report final articulation of the fully tested and iteratively developed ODPG data literacy framework of skill and knowledge.

Project management will be facilitated through an openly accessible team wiki hosted by the University of Washington iSchool IT department, and the use of an open Slack channel for team communication. Software code and example datasets used for classroom instruction will be hosted in an openly accessible GitHub repository. CANVAS—a FERPA-compliant courseware application required by the UW iSchool—will be used as space for coordination with and among students during course instruction. At the conclusion of each project year, we will share results from the evaluations and resulting improvements made for the coming year's activities in an openly accessible technical report.

4. DIVERSITY

ODPG aims to make a significant contribution to sustainable and equal access to open data for the public. This will be accomplished at the broadest scale by educating active and diverse public librarians who are invested in data literacy and the creation of data resources for social justice, equity, and inclusion, all of which are increasingly important, yet historically neglected, in education for librarianship (Jaeger et al., 2015). Diversity goals will be achieved through our collaboration with the Diversity Programs Advisor, a professional position in the Information School. She will play a vital role in attracting diverse students into the program, and identifying opportunities for data projects of interest, through her diversity networks on campus and in the community. Since ODPG will need to recruit students from the population already admitted to the Information School, it

[5] https://www.coursera.org/course/datasci
[6] https://software-carpentry.org/
[7] http://tascha.uw.edu/projects/research-roadmap-for-strengthening-the-library-field/
[8] See, for example, http://events.spl.org/116068151/FromDatatoActionOpenDataandYou

will be difficult to address the national level diversity problems in our field (Gulati, 2010). However, we will be able to attract admitted students to the program through engagement at diversity events and venues to raise awareness of the growing career opportunities for LIS students with data expertise. As ODPG matures, the Diversity Program Advisor expects to promote ODPG with underrepresented minority MLIS students through UW iSchool minority fellowships (the Finley, Alexi-Parks), and align ODPG with the UW iSchool's Minority Ambassador program. She also plans to develop a strategy for leveraging an iSchool matching fund for ALA Spectrum scholars to recruit students interested in ODPG.

The communities that can potentially benefit most from access to open data are also the least prepared segment of the general population to meaningfully use open data (Zuiderwijk et al., 2012). The ODPG project is an investment in closing this institutionalized skill gap. The team aims to dedicate 50% of available internship funds to support engagement of students from underserved populations in the ODPG Summer Internships. Further, we will work with our partners to build a diverse pool of mentors and develop internship projects that respond to the interests of underserved communities. For example, through our collaboration with the Seattle Public Library, we anticipate an ODPG internship project with the Edward S. Curtis collection of photographs from Native American tribes in the Pacific Northwest, in addition to the projects outlined above. The ODPG team will actively seek to support other fieldwork projects focused on access and preservation of cultural open data of interest to underrepresented communities, such as the Women's History Consortium collection option identified by the Washington State Historical Society. In addition, the use of open data for social justice is the proposed theme for the Year 2 community workshop, based on initial discussions with our project partners.

5. PROJECT RESOURCES: PERSONNEL, TIME, BUDGET

The work of the core project team (Palmer, Weber, TBD Ph.D. student) will be organized and monitored through weekly meetings for planning, joint decision-making, and appropriate tracking of project milestones, student progress, deliverables. The core team will collaborate on the evaluation components. Weber, supported by the Ph.D. student who will serve as Project Coordinator, will manage the work of the Curriculum Team (Blumenstock, Howe), the Fieldwork Team (Crandall, Saunders, Oman, Kilmer), and the Outreach Team (Coward, Loter). Del Rosario will work with Palmer and Weber to meet diversity goals. Progress on all activities will be coordinated and communicated to other team members through an openly accessible project Wiki and Slack channel (described previously).

Key Personnel

Carole Palmer, Professor and Associate Dean for Research, will serve as Principal Investigator. She will direct and oversee all project activities, resources, and partnerships. She will contribute to all teams, providing curriculum development expertise and developing and delivering the new Fundamentals course. She is an internationally known expert in data curation and digital collections and has extensive experience leading funded collaborative research, new educational programs, and teaching and mentoring students and staff.

Nicolas Weber, Research Associate, will manage the Curriculum, Outreach, and Sustainability teams and Diversity and Evaluation activities, in collaboration with Palmer. He has directly applicable experience developing data curation and digital preservation curricula for graduate students and continuing education, managing collaborative projects, and formal education in the evaluation methods applied in the project.

PhD Student, TBD, will serve as Project Coordinator. In addition to providing support for the Teams, the Project Coordinator will advise on and manage the fieldwork placements and support students and mentors in development and coordination of fieldwork projects and outputs.

Josh Blumenstock, Assistant Professor and co-Director of the Data Science and Analytics Lab (DataLab), will serve as Data Science and DataLab advisor, providing coordination with the data science curriculum work and DataLab research activities. He is a highly respected and well-known expert in data analytic techniques for understanding poverty and economic development and an accomplished data science educator.

Chris Coward, Principal Research Scientist and Director, Technology and Social Change Group (TASCHA) will serve as the outreach facilitator for broader impacts through TASCHA's work in the U.S. and

developing countries. He has led research and interventions to improve policy and practice in information and communication technologies (ICTs) and international development in more than 50 countries.

Mike Crandall, will provide initial coordination with the Research Roadmap for Strengthening the Library Field project and related partners, and advise on the first phase of sustainability planning for the internship program. Crandall has extensive experience developing and delivering professional education for library and information management professionals as well as designing and leading successful collaborations between the LIS practitioner and educator communities.

Cynthia Del Rosario, Diversity Programs Advisor, will facilitate and implement the project's diversity strategies. She brings extensive knowledge of student profiles and needs, diversity affairs in higher education, and a critical connection to an active network of diversity activities on campus.

Bill Howe, Associate Director, eScience Institute, will serve as a technical advisor and manage the eScience Institute's instructional contributions and field experience placements. He is an active and established figure in scientific databases, data-intensive computing, visualization, and democratization of scientific analysis, and a highly successful educator, known for his achievements with data science MOOCs.

Partners:

Will Saunders, Senior Program Manager, Open Data, State of Washington Office of the Chief Information Officer, will advise on partners and internships, and coordinate access to data, projects, and mentoring.

Jim Loter, Director of Information Technology, Seattle Public Library (SPL), will direct the variety of instructional, fieldwork, and workshop activities planned with SPL and offer data technology advising.

Leni Oman, Washington State Department of Transportation, will advise and provide mentor coordination.

Jennifer Kilmer, Washington State Historical Society, will advise and provide mentor coordination.

Timeline. The proposed start date for the 3-year project is September 1st, 2016, as the best point to launch activities in relation to the academic year at UW. As detailed in the Schedule of Completion, many activities are iterative in nature and extend throughout much of the project. These include Curriculum Development & Revision for the Foundations course (offered in Winter quarter) and the new Services course (offered in Spring Quarter). Electives will be offered and adapted beginning the first Spring quarter. Diversity efforts will begin immediately, with planning for recruitment and activities to build career awareness and specialized advising, and evaluation starting with the first cohort of students. Fieldwork and Internship planning work will start in the first months of Year 1 and intensify as new projects and partners are identified. The Workforce Study will also be launched in Year 1, taking advantage of our partner relationships and participants in webinars, and attendees of the workshops in Years 2-3 to build up a rich base of respondents over the course of the project.

Summer Internships will begin in Year 1, and Directed Fieldwork will be an option for students in any academic quarter throughout Years 2 and 3. OER development and revision is concentrated in Years 2-3, with at least one webinar slated for each year. Evaluation is segmented throughout the project, with analysis of fieldwork beginning in Year 2, and summative evaluation and synthesis concentrated at the end of Year 3. Major events include all-hands meetings to jump-start and wrap-up the project, organized with substantial input and participation from partners. A workshop will be hosted at SPL in the summer of Year 2 and a national workshop conducted in Year 3.

Budget. Total IMLS budget request is $690,858. Direct costs are $507,383, with $183,475 indirect with a 53% federally negotiated indirect cost rate. Breakdown on direct costs: student costs, $268,242; salaries for program development and instruction, $138,296; overall fringe $37,345; workshops $38,500 with associated staff travel of $12K; partner costs $13,500. 1:1 cost share less student support costs has been met with $275,068 committed by the Information School and institutional partners.

6. COMMUNICATIONS PLAN

Communication will focus on promoting and sharing project advances through informal and formal channels. Informal dissemination of progress and outcomes will be shared with the professional community through a project website and weblog. Posts to the ODPG blog will include reports of on-going efforts and topical essays

from students, project partners, and key UW faculty participants. These posts will be aimed at informing practicing librarians, as well as government open data stakeholders. Formal dissemination of findings from use cases, case studies, and evaluations of project impact, and data literacy framework development will be through presentations, research papers, and panels at appropriate conferences (e.g. annual meetings of the Urban Libraries Council, Public Library Association, Association for Library &Information Science Education, ASIS&T, and the iConference). We will also seek out emerging venues where open data scholarship and open data stakeholders are beginning to form a discourse community (e.g., OpenDataCon, the Open Data Research Symposium, OpenSym, the World Wide Web conference). As the project matures, journal papers will be submitted to publications such as Public Library Quarterly, Library Hi Tech, D-Lib, and the Data Science Journal, and the Journal of the American Society for Information Science and Technology. Openly accessible pre-prints of all articles, project reports, and presentations will be archived in the UW institutional repository (ResearchWorks) and made available through the project website. Papers will be collaboratively written with students and external partners. Students will be encouraged to take a leading role in the creation of web-based content as a means to reach out to future students interested in participating in ODPG.

7. SUSTAINABILITY

The ODPG curriculum will be on strong footing from the outset of the project, since it represents two of the Information School's current areas of strategic priority—the Future of Libraries and Data for Social Good—and is consistent with emerging priorities to strengthen both technical competencies and practical experiences of MLIS students. The bridges being built between ODPG and the Data Science specialization in the Information School and the UW eScience Institute will further reinforce the value and strength of the program, allowing it to solidify by the end of the performance period.

Sustaining the network of partners for field experience placements will be facilitated as student projects make contributions of value to the partner organizations and serve as incentives to commit to further mentoring and support for fieldwork. At the same time, recruitment of additional partner organizations will focus on agencies that can commit resources, particularly for paid internships. We already have indications that with advanced planning and budgeting certain agencies will be in a position to dedicate internship stipend fund. Through the evaluation process and work with the Associate Dean for Academics, specific resource requirements will be determined for continuing and expanding Directed Fieldwork and coordinating the overall field experience component of ODPG. Moreover, that analysis will be made available to educators of other MLIS programs interested in fostering a sustainable local internship network.

The principles and practices of ODPG will be propagated through the work of the graduates and professionals completing the curriculum and who become active in data literacy and open data initiatives, especially those that move into instructional and advocacy roles in the profession. They will continue to have access to and be able to utilize the digital content produced by ODPG activities, which will be sustained by project stakeholders as outlined in the digital stewardship supplementary information form. All curriculum materials, including course syllabi, software, data, and visualizations, will be issued open licenses (where appropriate Creative Commons 0 and MIT licenses) to encourage immediate and continued reuse of these resources. Annually, all materials hosted on our project website will be bundled in a compressed archive (.zip file) and deposited in the UW institutional repository. The project website will be hosted and maintained as part of the established DataLab infrastructure as the curriculum transitions into the MLIS degree program web presence as expected after the performance period.

University of Washington

ODPG Schedule of Completion (9/2016-8/2019) Year 1 Year 2 Year 3

Category	*Activity*
Curriculum & Student Recruitment	Curriculum Development & Revision
	Course Design & Revision
	Data Literacy Framework Design & Revision
	Course Delivery
	Diversity Recruitment & Career Activities
Field Experiences	Fieldwork & Internsip Design & Revision
	Summer Internships & Mentoring
	Directed Fieldwork Projects
	Sustainability--Plan & Implementation
	Fieldwork Case Study Development
Continuing Education & Outreach	Workshop Material Development
	Promotional Video w/ Partners
	Workshop Delivery
	Webinar Development & Delivery
	Self-guided OER Curriculum Dev & Revision
	TASCHA Outreach Planning
	TASCHA Outreach Activities
Evaluation	Development of Evaluation Instruments
	Course Evaluations & Revision
	Continuing Education Evaluation & Revision
	Fieldwork / Mentorship Impact Analysis
	Diversity Impact Evaluation
	OER & Webinar Impact
	Workforce Study
Sustainability	Archiving code, data, curriculum
	Preservation plan for hosts
Dissemination	Formal and informal publication
Administration and Coordination	Mgmt. of Project Goals & Budget
	Partner Workplan Development
	All-hands meetings
	Website, Communications Plan Development

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets 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 is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. 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
If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete	Part I
	If your project will create or collect	**Then you should complete**
✔	Digital content	Part II
✔	Software (systems, tools, apps, etc.)	Part III
✔	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (http://us.creativecommons.org) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

Any original digital content produced by project staff at UW will be issued a Creative Commons 0 license, so that these resources can quickly enter the public domain and be used as widely and broadly as possible.

We will obtain IRB approval from the University of Washington for the collection and management of any data that has personally identifiable information. This will include interviews, photographs and / or videos of participants in the ODPG project. We will obtain participant consent before gathering any of this data, and clearly state that participation in this aspect of the ODPG project is voluntary. Where appropriate, we will take measures to restrict data that has personally identifiable information about workshop participants, or students involved in the ODPG project by either redacting this information, or creating non-identifiable pseudonyms for participants.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

The University of Washington will assert no ownership over original content produced as a part of this project. All content will enter the public domain through creative commons or MIT licenses.

Content created by the ODPG project will impose no restrictions on access or use, but we will ask for attribution where appropriate.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

We will collect survey data (including free response data) and interview various stakeholders of the project. (e.g. mentors of fieldwork, students, workshop participants, etc.) Survey and interview data material may contain personally identifiable information. We will obtain IRB approval from the University of Washington before we begin the collection and management of this data. Informed consent will be obtained from all participants.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

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

Software code: Will be developed for the sake of aiding students in classroom or in fieldwork that requires data processing, visualization, or analysis. The primary languages used by our team will be Python, R, and Javascript.

Surveys: Surveys will be designed using Google Forms and administered via the ODPG project web-site. Survey data will be exported to a CSV format and analyzed using the R library.

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

No external vendors will be used in the course of this work.

The ODPG team will receive IT support from the UW iSchool IT department, a staff of full-time professionals that serve the staff, faculty, and students in their software and hardware needs.

For analyzing transcripts, performing statistical analysis or any survey data, or processing any data we will

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

Data: Comma-Separated Value(CSV), Structured Query Language (.sql), Shapefile attribute format (.dbf).
Digital Documents: PDF/A, .Docx
Digital Photographs: JPEG, TIFF – taken at a minimum of 300 DPI
Video: WAV format – recorded with lossless sampling.
Software: R, Python, JS, SQL

B. Digital Workflow and Asset Maintenance/Preservation

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

Software will be managed with a version control system (Git). Digital content (data, documents, photos, videos) will be stored short-term on a secure network managed by the UW iSchool IT department – which conducts nightly backups of all content.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, 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 CFR 200.461).

Software will be hosted in a public repository on Github. Content in this repository will be downloaded, and compressed into an archive (.zip file) for deposit in the University of Washington's institutional repository on an annual basis. The repository will the be versioned so that Version 1, 2 and 3 align with the funded project years of OPDG.

Digital photos, videos, and documents will be stored on a secure UW iSchool server. The server is hosted by the iSchool IT department, and is backed up nightly. At the conclusion of the grant, digital photos,

C. Metadata

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

ODPG will produce administrative and descriptive metadata. Where appropriate we will preserve technical metadata (such as for software dependencies).

Software: All released software will have use the emerging CODE-META JSON-LD standard (https://github.com/codemeta/codemeta/blob/master/codemeta.jsonld_)

Digital Objects (data, documents, photos, and videos): Will use a basic Dublin Core schema (1.1) in XML.

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

Metadata will be created at the time of production or preservation (where appropriate). Metadata records will be preserved alongside digital content that is archived in the University of Washington's institutional repository.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

The University of Washington's institutional repository (ResearchWorks) – is Oan AI-PMH compliant repository based on Fedora Commons software (4.0). All metadata that is archived with ODPG assets here will be harvested through this protocol.

D. Access and Use

D.1 Describe how you will make the digital content 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).

Data and software will be maintained in an openly accessible repository hosted on Github. We will provide links to this archive from our project web-page. At the conclusion of each year of funding (1-3) we will version, and then archive the repository in the University of Washington's institutional repository.

Our project web-site will be a publicly accessible page where we will advertise and promote access to all digital content created by the project.

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

Part III. Projects Creating Software (systems, tools, apps, etc.)

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) this software will serve.

ODPG will not be creating novel software for the use or enhancement of large communities of open data users. The software that we imagine writing will be highly personal to the classroom work that is described in the project narrative - such as scripts to fetch data from API's, cleaning or normalizing open data, visualizing sub-sets of open data, etc.

The primary audience of this software are other ODPG team members, students, and project partners.

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

n/a

B. Technical Information

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

Software will be written in Python, R, and Javascript (e.g. D3.JS).

These languages were chosen because they are openly-accessible to end users, they are regularly used in data science, and they have well supported and active user-communities.

B.2 Describe how the intended software will extend or interoperate with other existing software.

Software will be written in languages (described avove) that are freely accessible, and capable of being used on major operating systems.

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

The software described above will require the following dependencies:

Python - 2.3.0 or higher

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

The software created as a part of ODPG will be documented using plain-text 'readme' files. These files will be updated during development, and will be reviewed by team members before being archived in the UW institutional repository (on an annual basis). All software that is produced for end-users will have descriptive and administrative metadata created in JSON-LD (as described in Part 2).

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

https://github.com/uwescience

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? 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 any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

All software will be given an MIT license. Justification for this choice is found in Section 1.

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

Software will be made available in a publicly / open accessible repository on Github. We will promote and provide links to these resources on the ODPG project web-site.

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository: Github
URL: https://github.com/odpg

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

Please see table attached to end of document. [Page 8]

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?

The ODPG will apply for IRB approval from the University of Washington. The project has not received (nor applied) for approval at this time. The ODPG team will be assisted by the Research Services department of the UW iSchool to obtain IRB approval for the data collection described above, should the project be funded.

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).

Personally identifiable information in data that is collected as a part of the ODPG project will be protected in the following ways:

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.

Informed consent will be obtained from all participants through a signed agreement, or verbal consent. Signed agreements will be stored in a locked drawer in the PI's office. Digital data will be stored in separate files using the coding scheme described above to identify participants. All data containing

5. What 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).

Survey instruments will be developed using Google Forms, and embedded on the project's web-site. Responses will be stored in a CSV format on a password protected server. Survey data will be analyzed using the R platform, and in particular the R library "Survey" for statistical analysis.

6. What documentation (e.g., data documentation, codebooks, etc.) 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?

An interview and survey protocol will be developed by the PI and Post-doc. This protocol will outline procedures for collecting, coding, and performing analysis on both survey and interview data.

The codebook developed as part of the qualitative analysis of the interview data will be developed using a

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

At the conclusion of the grant survey data, transcripts, and codebooks will be bundled into a compressed archive and deposited into the UW institutional repository for long-term storage. Data will be anonymous and reusable by interested parties. We will advertise this data's availability on the project website.

8. Identify where you will be publicly depositing dataset(s):

Name of repository: University of Washington's Research Works Repository
URL: https://digital.lib.washington.edu/researchworks/

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

The plan will be reviewed by the project PI and post-doc on an annual basis. A yearly report produced at the end of Y1,2, and 3 will include details on how and why the data management plan has been modified.

Original Preliminary Proposal

Open Data for Public Good: Data Literacy Education for Public Information Professionals

Project Team. Project Director: Carole Palmer, Information School, University of Washington. Confirmed team members and partners: Information School - Data Science & Analytics Lab (DataLab), Josh Blumenstock (Director) & Nic Weber (Research Associate); Technology & Social Change Group (TASCHA), Mike Crandall (Research Roadmap for Strengthening the Library Field project); UW - eScience Institute, Bill Howe (Associate Director). External - Seattle Public Library (SPL), State of Washington Office of the Chief Information Officer, Washington State Department of Transportation.

Introduction. Open data has the potential to positively transform many aspects of society, from increasing civic engagement in political processes, to improving community resilience to a changing climate, and strengthening local cultural heritage. Advances are being made on technical infrastructure for discovering, accessing, and using open data. In recent years, Library and Information Science education programs have made considerable progress in data workforce development, especially in response to demands for data services in academia. However, there remains a significant gap between what is possible technically, given the current open data infrastructure, and what is being achieved practically in local communities. The proposed Open Data for Public Good (ODPG) program aims to close this gap by preparing public librarians and other information professionals in the public sector to: cultivate data literacy; support open data initiatives; curate collections of open data of value to their communities; and work to make more data open to the public. We request support for this effort through an LB21 project grant for continuing education and programs to build institutional capacity.

Work Plan. The 3-year project will educate LIS students and practitioners in collecting, managing, preserving, and providing meaningful access to open data resources for the public good, concentrating on open **civic, environmental,** and **cultural heritage** data. The program will be grounded in new curriculum and implemented with on-campus and online Master of Library and Information Science (MLIS) students. The model will be scaled to provide additional educational opportunities for current practitioners. The team's collective capacity in data curation, data science, and librarianship will be strengthened through collaboration with professionals in partner organizations, to offer students an integrated academic and practical educational experience.

Phase 1. Curriculum for open data literacy and expertise. A sequence of two courses—Fundamentals of Data Curation, redesigned for ODPG, and a new course on public information services for open data—will cover collection, curation, and repurposing of civic,[1] environmental,[2] and cultural heritage[3] data to address real world problems and interests. Structured as a framework for open data literacy, the curriculum will include modules on selection and aggregation; digital preservation; data enrichment and fit for purpose; long-term management; interoperability; rights and access conditions; and related topics, with an emphasis on practical data repository development with open source software and significant input from partners working at the front lines of the field. The courses will be designed with the additional benefit of providing an onramp for students, including those from less technical backgrounds, who wish to gain expertise in applied data science, through the Data Science degree specialization, and through involvement in DataLab and TASCHA data activities.
Phase 2. Practical internships, mentoring, and capstone. Following these courses, two cohorts of high-performing students will be recruited for internships with partner organizations. Students will work with data mentors in the organization to curate collections designed to directly benefit service communities, using open civic, environmental, or cultural data sources. Interns will also create related resources customized for practitioners, including technical documentation and best practices guidelines. The experience will culminate in formal Capstone projects (https://ischool.uw.edu/capstone), and each cohort will use their capstones as the basis for a panel presentation to promote open data literacy educational activities at ALA or other professional venue.
Phase 3. Extended professional education. ODPG team members, students, and partners will develop instructional modules designed for practitioners in public libraries, NGOs, and government agencies. This material will be built on course materials and internship outputs, and include topics such as creating open-data catalogs with open-source software, curation workflows for populating a repository with open data, and open-

[1] e.g., City; County; State; & Country Open Data
[2] e.g., Community Climate Data; Environmental Health ; Agriculture & Local Food Data
[3] e.g., Native American; African American; Latino/a ; and Regional Open Cultural Data

source visualization tools to analyze and present collections. This content will be presented in a regional workshop held at the Seattle Public Library, and then extended to a national audience in conjunction with a conference such as PLA. A set of free webinars will be developed for two stakeholders: library open data communities, and government agencies retraining their workforce to curate open data.

ODPG will draw upon the highly effective internship and evaluation models developed in the IMLS funded Data Curation Education in Research Centers project (DCERC).[4] Workshops will be informed by SPL's recent open data events.[5] Instruction will leverage the eScience Institute's experience with online courses, which have attracted over 200,000 students,[6] their partnership with Software Carpentry[7] for introductory programming, and Community Data Science events led by a Communications Department affiliate. eScience data scientists will contribute hands-on instruction in data management and analytics tools to aide student projects. We will also extend TASCHA's work with the Gates Foundation on open data practices and visualization for public library data conducted in collaboration with Zepheira and Community Attributes.[8] Additional potential partners include City of Seattle, Washington State Archive, and Departments of Transportation outside Washington for internships with online students.

Relevance. ODPG directly addresses the LB21 programmatic goals of supporting academic programs and professional development related to the National Digital Platform priority, including the identified need to cultivate a digital library workforce and increase hands-on practice in degreed programs.[9] Activities are aligned with IMLS strategic goals to prepare the public to be full participants in local communities and global society, promote libraries as community anchors that enhance civic engagement and cultural opportunities, and promote use of technology for discovery of knowledge and cultural heritage.

Potential Impact. ODPG will further strengthen the unique role that libraries have always held in preserving and providing access to critical information resources. As open data increases in importance for constructive civic engagement, enabling equal and fair cultural opportunities, and sustaining economic vitality, ODPG will prepare a new generation of professionals to make lasting contributions in their communities and to the LIS profession. Through train-the-trainer initiatives, many more individuals will also provide data literacy outreach to their communities and empower more individuals to sustainably curate open data to serve the public good. Moreover, the program is strategically designed to increase placement of technically educated information professionals in public libraries. Finally, positioned at the intersection of two of the Information School's new strategic foci—the Future of Libraries *and* Data for Social Good—ODPG will offer a robust model for innovative LIS data education and raise awareness of our field's growing contribution to social, environmental, and cultural advances.

Projected Performance Goals and Outcomes. New advocates and leaders in promoting open data literacy for public good will include 60+ students who complete the curriculum, with many gaining significant field experience, and more than 100 practitioners participating in continuing education. High quality Open Education Resources (OERs) will be made available to the profession at large, including online modules, recorded webinars, videos, and a rich body of best practices. Formal curriculum and lessons learned will be shared with the broader LIS education community. ODPG will be iteratively evaluated and solidified as an adoptable model and a sustainable long-term educational strategy, including incentive plans for establishing and increasing monetary support for internships by partner organizations. Finally, graduates will begin to propagate data literacy among the public and peers and to apply data expertise in innovate ways in their organizations.

Estimated Budget. Grant budget estimated at $559K. Direct costs at $441K, with 53% federally negotiated indirect cost rate. Breakdown on direct costs: student costs, $217K; salaries for program development and instruction, $113K; overall fringe $31K; workshops $60K with associated staff travel of $12K; partner costs 8K. 1:1 cost share will be achieved with support and resources from key personnel and institutional partners.

---

[4] https://www.youtube.com/watch?v=mbX5bvgT1ME
[5] See, for example, http://events.spl.org/116068151/FromDatatoActionOpenDataandYou
[6] https://www.coursera.org/course/datasci
[7] https://software-carpentry.org/
[8] http://tascha.uw.edu/projects/research-roadmap-for-strengthening-the-library-field/
[9] https://www.imls.gov/sites/default/files/publications/documents/2015imlsfocusndpreport.pdf