A National Forum on Web Privacy and Web Analytics

Abstract

The Montana State University (MSU) Library seeks $100,000 funding from the Institute of Museum and Library Services (IMLS) through its National Forum Grant program under the National Digital Platform project category to hold a meeting around the topics of web privacy and web analytics involving librarians, library administrators, systems developers, product designers, user experience designers, lawyers, technologists, and privacy researchers. A gathering of these stakeholders is needed in order to build consensus and develop an actionable agenda for implementing values-driven web analytics across a nation-wide spectrum of libraries: public, academic, community college, tribal, school, and special.

Measuring visits and visitors to a website through third-party analytics tracking is common practice by website owners. Google Analytics—a leading third-party analytics tracker—is widely implemented on library websites and well represented in the library literature, yet the related factor of web privacy is underrepresented in our discourse and practice around web analytics. Google Analytics and its related services such as AdSense and Tag Manager present a potential threat to user privacy and to the library profession's long-held value of intellectual freedom. Websites that implement Google Analytics and other Google tracking services are participating in the extensive network of third-party trackers that are capable of sharing user data across sites. In many cases, the user has not given fully informed consent for this data sharing, nor does the website owner fully understand the capabilities and consequences of web analytics and other third-party trackers.

We propose hosting a *National Forum on Web Privacy and Web Analytic*s to address the challenges of user privacy on the web. The National Forum Grant will support the creation of a community-driven strategic direction to be developed over a one-year period, from May 2018 to April 2019. The activities of the Forum will be framed around three supporting goals: 1. Activating and focusing our professional discourse around privacy and analytics; 2. Improving our field's understanding of third-party web tracking and web analytics; and 3. Developing practical approaches and policy guidelines for implementing privacy-focused analytics within the contemporary landscape of the web. The realization of these goals will ultimately produce three key outputs that will offer immediate practical and scholarly advances for privacy and analytics in libraries: (1) a scholarly white paper that synthesizes forum activities and articulates the direction for a national privacy-focused values-based agenda for web analytics, (2) a practice-based action handbook that provides background, resources, and best practices to guide libraries in responsibly implementing web analytics, and (3) conference presentations and peer-reviewed publications that will further disseminate forum activities and generate engagement and buy-in around forum topics.

In short, a community-fueled effort focused around the Forum and its follow-up activities will enable our profession to take important strides toward a better analytics practice that protects our users' privacy from unwanted third-party tracking and targeting. In convening experts from across disciplines and professions, we can break down silos, learn new perspectives, focus our privacy discussions around analytics and third-party tracking, and produce a roadmap for enhancing our analytics practice in support of privacy.

National Forum on Web Privacy and Web Analytics

The Montana State University (MSU) Library seeks $100,000 funding from the Institute of Museum and Library Services (IMLS) through its National Forum Grant program under the National Digital Platform project category to hold a meeting around the topics of web privacy and web analytics involving librarians, library administrators, systems developers, product designers, user experience designers, lawyers, technologists, and privacy researchers. A gathering of these stakeholders is needed in order to build consensus and develop an actionable agenda for implementing values-driven web analytics across a nation-wide spectrum of libraries: public, academic, community college, tribal, school, and special.

---

1. Statement of National Need

1.1 Summary

Libraries need web analytics. The usage measurement statistics generated from web tracking software help tell a story of value and impact for library stakeholders, including funding agencies, university administrators, and community boards. Analytics also serve a crucial role for tuning, tailoring, and improving services to better help library users. Google Analytics—the powerful, comprehensive, and easy-to-install software package—has emerged as a leading choice for web analytics in libraries. From the perspective of monetary cost and implementation, Google Analytics is free to use and easy to install. These pragmatic factors amplify its widespread use. From the perspective of user privacy, however, Google Analytics is anything but free to use. As a third-party web tracker, Google Analytics participates in a vast network of third-party tracking on the web that exposes user data to countless additional trackers, often without the fully informed consent of the user or the library.

Libraries understand that users need privacy. Libraries have historically offered safe spaces of intellectual freedom that are underpinned by a commitment to privacy in the pursuit of information, yet the widespread implementation of commercial analytics packages such as Google Analytics on library websites may conflict with the library profession's long-held values of privacy and intellectual freedom. This potential conflict is often not addressed in publications and presentations on Google Analytics and its related services such as Google Tag Manager. Indeed, the MSU Library's research into privacy and web analytics (as a part of IMLS grant LG-06-14-0090-14, "Measuring Up: Assessing Accuracy of Reported Use and Impact of Digital Repositories") has revealed that the library community appears to have an imperfect or incomplete understanding of web privacy, web analytics, and third-party web tracking. The proposed Forum is therefore driven by the underlying need to enhance our profession's understanding and practice of web analytics with a view towards our unique commitment to privacy.

1.2 Web Analytics and Third-party Tracking : A Brief Introduction

The practice of third-party tracking on websites is widespread (Narayanan and Reisman 2017), and has only increased in prevalence, variety, and complexity over time (Lerner et al. 2016; Englehardt and Narayanan 2016). One of the most common trackers found on the Web is produced by the Google Analytics web service, which is used to measure and analyze the visitation to a website (Lerner et al. 2016; Schelter and Kunegis 2016). In exchange for this easy-to-implement and free-to-use analytics service, websites execute Google Analytics JavaScript code and pass user visit data to Google through browser cookies set by Google Analytics (Krishnamurthy and Wills 2009). Such data is considered to be "leaked" if the user is unaware of its collection and does not consent to the data being shared with additional third parties (Sar and Al-Saggaf 2012). An

1

analysis of 1 million websites found that nearly 9 in 10 websites leak user data to third parties without the user's knowledge (Libert 2015).

The Google Analytics tracker is not designed to leak user data across sites on its own—but when a website has implemented both Google AdSense (Google's popular cross-site advertising service) and Google Analytics, the unique identifiers from each service can be linked in a way that allows Google to merge user profiles and track users (Roesner et al. 2012). Data leakage via Google Analytics can also occur when websites activate the additional Google tracking service known as Tag Manager, which allows for cross-site tracking and targeted advertising (Bashir et al. 2016). Under these expanded tracking conditions, third-party trackers can match user behavior data with user profiles, thereby allowing users to be tracked and targeted across the web (Olejnik et al. 2012; Kalavri et al. 2016).

This kind of data leakage and user profiling via web tracking represents a privacy issue for users because of a lack of transparency and the lack of opportunity of users to consent to the sharing of their tracked behavior. The following example illustrates this case: a user logs into Gmail through a web browser on a public computer terminal at a library that uses Google Analytics and Google Tag Manager. This user then clicks through a TurboTax email and, later in the same session, searches for tax relief resources through the library website. Because Google can track this user's behavior across multiple sites, Google's advertising network can then target the user with ads for credit cards or personal loan services. This style of tracking is pervasive; Google was shown to be capable of tracking users across nearly 80% of sites through its various third-party domains (Libert 2015). In this way, the websites that implement Google Analytics and other Google tracking services are participating in the extensive network of third-party trackers that are capable of sharing user data across sites. In many cases, the user has not given fully informed consent for this data sharing, nor does the website owner fully understand the capabilities and consequences of web analytics and other third-party trackers. In these cases, user privacy may be violated by web analytics trackers.

1.3 Web Analytics and Web Privacy

The library literature is well represented with articles and books that detail the implementation, application, and justification of Google Analytics for the purposes of web traffic analysis and service improvement (Hess 2012; Barba et al. 2013; Cohen and Thorpe 2014; Fagan 2014; Yang and Perrin 2014; Conrad 2015; Farney 2016). In these articles and manuals, however, privacy is not often mentioned. Yet the user data collected by Google Analytics, such as search terms, user-agent software, geographical location, and time of day, can potentially be leaked to other third-parties via extensive networks of web trackers. User privacy can be further undermined when third-parties match user behavior data with user profiles, thereby allowing users to be tracked and targeted across the web (Olejnik et al. 2012). Certain Google Analytics implementation practices can help reduce the capability of data collection and reduce the capability of cross-site data sharing. These mitigating techniques include IP anonymization, opt-out mechanisms, and secure HTTP connections. A secure HTTP connection, also referred to as HTTPS, can be activated via a secure digital certificate (Naylor et al. 2014; Askey and Arlitsch 2015). These practices can also signal to users that a site is aware of web tracking and concerned with user privacy.

Privacy has long been a concern of libraries (Million and Fisher 1986; Garoogian 1991; Johnston 2000; Nicols Hess 2015). Given the extent of third-party tracking, however, it is exceedingly difficult to implement analytics trackers like Google Analytics without compromising the privacy for users that libraries have championed. Library professional organizations have acknowledged the complexity of contemporary information privacy, and have released privacy statements to help guide action (see Table 1).

Table 1. Privacy Statements - National Library Organizations

Organization	Statement Title	Statement Excerpt
American Library Association (ALA)	Library Bill of Rights - Interpretation of Privacy	"In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one's interest examined or scrutinized by others."
Coalition of Networked Information (CNI)	Privacy in the Age of Analytics	"Libraries collecting data using Google Analytics are realizing they may be violating the ALA *Library Bill of Rights*...this is but one example of how easily convenient web-based service offerings can come with unexpected consequences."
National Information Standards Organization (NISO)	NISO Privacy Principles	The principles outlined in this document are a starting point. Additional community consensus work will be necessary to make some of these principles implementable by the spectrum of providers that supports library services.

The statements from ALA and CNI highlight the challenge of maintaining our profession's positions toward user privacy vis-à-vis third-party tracking, and the statement from the NISO highlights the need for a community-based and consensus-driven approach for this work.

A recent survey of librarians' attitudes towards privacy found that 97 percent of respondents agree or strongly agree that libraries should never share personal information and circulation or Internet records without authorization or a court order (Zimmer 2014). In the same survey, 76 percent of respondents feel that libraries are doing all they can to prevent unauthorized access to individual's personal information and circulation records; however, a different survey investigating the configuration of public internet terminals showed that many libraries have not installed ad-blocking and privacy-protecting features on web browsers, nor do they offer instruction to users regarding web privacy (Gardner and Groover 2015). Results from our IMLS-funded research also show that many pro-privacy practices are not widely-implemented on library websites. As other authors have neatly summarized, "many websites use [Google Analytics and other click-tracking mechanisms], and their utility within library systems is an ongoing debate as we balance the needs of reliable metrics with patron privacy" (Caro and Markman, 2016). The use of Google Analytics on library websites is ubiquitous, yet the tension between web analytics and web privacy demands further investigation to ensure that libraries are in fact doing all we can to prevent unauthorized and unwanted data sharing.

1.4 A National Forum on Web Analytics and Web Privacy
We propose hosting a *National Forum on Web Privacy and Web Analytics* (henceforth, the Forum) to address the challenges of user privacy on the web. While many librarians agree that libraries should not share user data, the use of web analytics trackers such as Google Analytics on library websites and the lack of privacy-protection features demonstrate that the library profession needs more resources, more dialogue, and more direction for addressing the issues.

Our discussion at the Forum will be framed broadly around key foundational questions. Given that usage measurements are necessary to communicate value and improve services, how can libraries continue to report web analytics while also protecting users' privacy? Can traditional library values be compatible with contemporary web analytics? Why haven't more libraries implemented alternatives to Google Analytics such as Piwik? Why haven't more library websites implemented cookie consent notices? Is it possible for the library

community to build a new privacy-focused web analytics platform? What are the tradeoffs and implications of building our own analytics solution versus advancing the conversation around third-party providers? What are the strategies for holding technology companies and vendors accountable for data collection and retention? How can libraries use web analytics while respecting and protecting user privacy?

The Forum will prove vital for advancing professional discourse, knowledge, practice, and policy regarding third-party web tracking on library websites. By bringing together stakeholders and participants from across disciplines, we will focus our profession's privacy conversation and provide an actionable roadmap for enhancing user privacy in the age of web analytics.

1.5. Relevance to Funding Category, Project Category, and Agency Goals and Objectives

This project exemplifies the National Forum Grant by proposing to convene expert practitioners, qualified researchers, and key stakeholders—including those from adjacent fields—to develop a cohesive national response to a critical issue in the library profession. By enhancing our professional understanding of web analytics and web privacy, and coordinating action towards a national privacy agenda, this Forum also realizes goals of the National Digital Platform project category, which seeks to research the need for and impact of investments in national digital library infrastructures and services. The Forum also realizes goals and objectives of the IMLS (as articulated in the Strategic Plan 2018-2022): by building professional capacity for practices related to privacy and intellectual freedom, the work of the Forum leverages and reinforces the distinct role of libraries as trusted sources of information; and the Forum outputs—a white paper, an action handbook, conference presentations, and open access publications—align with the IMLS agency goal of investing in professional development that supports libraries' ability to implement best practices and make informed decisions.

2. Project Design

2.1 Goals and Outputs

The primary goal of the Forum is to articulate a national agenda for meeting our professional standards of privacy while also meeting the need to implement and report web analytics. The activities of the Forum will be framed around three supporting goals: 1. Activating and focusing our professional discourse around privacy and analytics; 2. Improving our field's understanding of third-party web tracking and web analytics; and 3. Developing practical approaches and policy guidelines for implementing privacy-focused analytics within the contemporary landscape of the web. The realization of these goals will ultimately produce three key outputs that will offer immediate practical and scholarly advances for privacy and analytics in libraries:

> A post-Forum **scholarly white paper** that synthesizes Forum activities and articulates the direction for a national privacy-focused values-based agenda for web analytics. The white paper will include pre-Forum privacy worksheets completed by participants, a literature review, and an environmental scan. It will also include identification of partners for further development and sustainability of the agenda. The white paper will be informed by community input, with final synthesis and production by the project staff. The white paper will be licensed CC BY 4.0 to allow for wide reuse and dissemination.

> A practice-based **action handbook** that provides background, resources, and best practices to guide libraries in responsibly implementing web analytics. The action handbook will be licensed CC BY 4.0 to allow for wide reuse and dissemination. The action handbook will be informed by community input, with final synthesis and production by the project staff.

> **Conference presentations** and **peer-reviewed publications** that will further disseminate Forum activities and generate engagement and buy-in around Forum topics. These outputs will be submitted to

4

journals that are fully Open Access (sometimes called Diamond or Platinum OA) so as to allow for the widest possible dissemination at the lowest cost.

2.2 Personnel

Scott Young (Project Director), Assistant Professor and Digital Initiatives Librarian at Montana State University, will serve as Project Director. Through research conducted under a previously-awarded IMLS grant, Scott has developed an expertise in the area of privacy and analytics.[1] Scott's work focuses on user experience of the web and community building on social media. As a UX researcher and practitioner, Scott is attuned to the needs and expectations of users, including privacy, which is a critical factor that conditions the user experience. Scott publishes, presents, and teaches on user experience, social media, privacy, and ethics.

Jason Clark (Co-Project Director), Associate Professor and Head of Library Informatics and Computing at Montana State University, will serve as co-Project Director. Jason builds and directs the digital branch of the Montana State University (MSU) Library. His work focuses on Semantic Web development, digital library development, metadata and data modeling, web services and APIs, search engine optimization, and interface design.

Sara Mannheimer (Co-Project Director), Assistant Professor and Data Management Librarian at Montana State University, will serve as co-Project Director. Sara educates researchers on data management best practices and ethical data sharing. Sara has published and presented extensively on the ethics and privacy issues surrounding data-driven research and assessment.

Lisa Janicke Hinchliffe (Project Analyst), Professor and Coordinator for Information Literacy Services and Instruction in the University Library at the University of Illinois at Urbana-Champaign, will serve as Program Analyst. Lisa has presented and published widely on information literacy, teaching and learning, the value of academic libraries and library assessment, evaluation, innovation, analytics, and privacy.

With its breadth of experience, the project staff is well positioned to convene and coordinate a multidisciplinary group of stakeholders to catalyze effort towards formulating a national agenda around web privacy and web analytics.

2.3 Structure and Phasing

The National Forum Grant will support the creation of a community-driven strategic direction to be developed over a one-year period, from May 2018 to April 2019. To achieve the goals of this project, work will unfold in three three main phases: Phase 1, Forum event planning and preparation; Phase 2, Forum event; Phase 3: Post-Forum activities, including community feedback and final output development and delivery.

2.3.1 Phase 1: Forum Event Planning and Preparation

IMLS funding will primarily support the attendance of approximately 40 Forum participants. Participation is critical to the success of this project, and will be the central focus of our planning. We will reach across and beyond the profession to ensure that a diverse range of voices and perspectives are heard at this Forum, especially from traditionally underrepresented populations. The scope of participation is identified through the following following four related community groups:

---

[1] IMLS grant LG-06-14-0090-14, "Measuring Up: Assessing Accuracy of Reported Use and Impact of Digital Repositories." https://www.imls.gov/grants/awarded/lg-06-14-0090-14; and No Such Thing as a Free Lunch: Google Analytics and User Privacy.

I. **Librarians, Library Administrators, Library Faculty, Library Vendors, Library Funders**
Librarians and others working in and around libraries form the core of this project. We intend to invite representatives from many types of libraries, including academic libraries, public libraries, community college libraries, tribal college libraries, and school libraries. We furthermore intend to invite professionals from across the library spectrum, including administrators, developers, faculty, assessment professionals, funding agencies, and vendors. We have secured preliminary commitments of participation from: **Gretchen Caserotti** (Library Director at Meridian Public Library), **Bethany Nowviskie** (Director of the Digital Library Federation), **Andrew Asher** (Anthropologist/Assessment Librarian at Indiana University), **Susanna Galbraith** (Virtual Services Librarian at McMaster University's Health Sciences Library), **Kelvin Watson** (Director of the Broward County Libraries), **Maura Smale** (Chief Librarian and Professor at the New York City College of Technology), **Andrew Nagy** (Director of SaaS Innovation, EBSCO), **Patricia Hswe** (Program Officer for Scholarly Communications at the Andrew W. Mellon Foundation), **Michael Zimmer** (Associate Professor in the School of Information Studies at the University of Wisconsin-Milwaukee), and **Rhiannon Sorrell** (Instruction and Digital Services Librarian at Diné College). Other potential invitees include: **Erica Reynolds** (Director of Library Partnership Development at BiblioCommons), and **Karen McKeown** (Director of Product Discovery, Usage and Analytics at Gale). These and other participants who work in libraries will bring a depth of experience that will contribute critically to this forum.

II. **Privacy Researchers, Activists, and Lawyers**
The Forum will be an opportunity for our profession to learn more from privacy experts working in libraries and in related disciplines, including computer scientists, activists, and legal scholars. We have secured preliminary commitments of participation from: **Patrick Gage Kelley** (Senior Privacy UX Researcher at Google), **Alison Macrina** (Founder and Director of the Library Freedom Project), **Stephanie Wright** (Program Lead at the Mozilla Science Lab), **Gabriela Zanfir-Fortuna** (Data Protection Specialist and Fellow at the Future for Privacy Forum), and **Projects by IF**. Further potential invitees include: **Bonnie Tijerina** (Research Fellow and Data & Society), **Crystal Nwaneri** (Fellow at the Berkman Klein Center), **Omer Tene** (Senior Fellow at the Future of Privacy Forum), and **Arvind Narayanan** (Assistant Professor of Computer Science at Princeton University), **Laura Kalbag** (Co-Founder and Designer of the privacy-focused design firm Ind.ie) We will also seek participation from privacy-focused groups such as **Tactical Tech**, **The Engine Room**, and the **Electronic Privacy Information Center**. These and other privacy experts will provide invaluable perspectives to this conversation.

III. **Designers and Developers**
The privacy landscape includes those who design and build the systems that we use. Perspectives from this community will be essential for understanding the realities and pressures of systems design and development with regards to tracking and privacy. We have secured preliminary commitments of participation from: **Becky Yoose** (Library Applications and Systems Manager at The Seattle Public Library), **Tristan Denyer** (UX Designer, Product Designer, and UI Designer), and **Magera Holton** (Product Designer at Etsy). Further potential invitees include: **William Budington** (Security Engineer and Technologist at the Electronic Frontier Foundation), **Mark Matienzo** (Collaboration & Interoperability at Stanford University Libraries), **Thomas Johnson**, (current PhD student at the University of Washington iSchool and former DPLA Metadata and Platform Architect at the DPLA), and **Jason Griffey** (founder and principal at Evenly Distributed).

IV.   End-users

A final key stakeholder group is library end-users themselves. In the context of the academic library, students make up an important community of users. Project Director Young and Co-Project Director Mannheimer are scheduled to teach a Spring 2018 Honors College seminar at MSU focusing on information ethics and privacy. Students from this seminar will be invited to participate in the Forum to provide a voice and perspective from a key user group.

In addition to the above-named participants, we will organize an open call for participation. To focus the outreach and communication of the Forum, we will create a project website to describe the work of the Forum and to solicit nominations and self-nominations for participation (in proposing a project website and nomination process, we are guided by the example set by the <u>IMLS-funded National Forum Grant *Always Already Computational*</u> and its <u>corresponding website</u>, and by the website and programming for <u>Design for Diversity: An IMLS National Forum Project</u>). We will also solicit participation through email and social media. Criteria for inclusion will include experience or interest—interpreted broadly—in the topic areas of the Forum: privacy, analytics, third-party tracking, user experience, and systems design and development. Of the 40 participants in the Forum, approximately 20 will be included via direct invitation and approximately 20 will be included via the nomination process. With this inclusive balance, we aim to bring in participants through targeted recruitment and through a broader, more open nomination process that will allow us to include participants who might not otherwise have been known to the project staff. In bringing together a wide variety of participants, we aim to realize the IMLS theme of strategic collaboration.

To prepare participants for the Forum, the project staff will produce a set of pre-Forum resources to establish common grounding for the event: a privacy worksheet to be completed by each participant that outlines current privacy and analytics practices at their home institution or organization; an environmental scan that highlights the current landscape of privacy and analytics in and beyond libraries; and a literature review to set the foundation for discussion. The pre-Forum resources will be critically informed by community engagement at ALA Annual 2018. Project staff will attend this major national conference to strategically engage with diverse stakeholder communities: public librarians, academic librarians, administrators, web designers, web developers, and vendors.

Pre-Forum resources will be published through the project website for public review and feedback. These resources and the subsequent work of the Forum will then be organized and made openly available online through the project website and through the Open Science Framework (OSF). OSF integrates with external systems such as Google Drive, Box, Github, and Zotero, thus facilitating collaboration between forum participants. OSF also promotes transparency of the Forum and its outputs by facilitating open working practices. Project resources will be archived in <u>MSU ScholarWorks</u>, the open access institutional repository of Montana State University.

2.3.2 Phase 2: The Forum Event

The Forum will take place over two and a half days at Montana State University in Bozeman, Montana in September 2018. This Forum will include keynote speakers to frame the problem of privacy and analytics, and interactive breakout sessions for all participants to share ideas and advance discussion. Three rounds of breakout sessions will be sequenced to develop an agenda for addressing the problem. These three rounds will be organized around human-centered design principles, and coordinated around the two and a half days of the Forum:

Day One - Introduction and Exploration. The first day of the Forum will be dedicated to introductions and topic exploration. With the pre-Forum resources as a foundational guide, the first round of sessions will focus on exploratory exercises that will bring clarity to the problem from the viewpoint of each participant and stakeholder group.

Day Two - Idea Generation. The second day of the Forum will focus on creating ideas in response to the central problem. The second round of sessions will focus on generative exercises that will bring forth potential solutions to the problem.

Day Three - Evaluation and Selection. Day three will comprise a full morning of activities aimed at evaluating the ideas produced during the preceding day, with the goal of selecting ideas based on viability (sustainability), feasibility (technically possible), and desirability (user needs). The final round of sessions will focus on evaluative exercises that will gauge the overall merit of the proposed solutions.

By scaffolding the forum in this sequence, we will be able to bring a diversity of ideas together and build effectively towards a cohesive response to the problem. For those unable to attend the Forum, we will explore live streaming options such as Periscope, Facebook Live, or YouTube Live. We will also ensure that conversations are captured via collaborative note-taking to be uploaded later to the project website. We will also create a Forum hashtag so that conversation can be cultivated and sustained via Twitter and other social media platforms.

To accommodate these activities, we have made preliminary reservations to host the Forum in the meeting rooms and breakout rooms of the <u>Student Union Building on the Montana State University campus</u>. These spaces will allow participants to convene both as a large group and in smaller groups as discussion requires.

2.3.3 Phase 3: Post-Forum Activities and Project Communication

Following the Forum event, project personnel will review and synthesize the evidence produced by the Forum participants, ultimately producing the project outputs: a white paper, an action handbook, conference presentations, and peer-reviewed publications. These outputs will be informed by, and supplemented with, community input following the Forum event. Members of the project staff will travel to relevant library and technology conferences for the purposes of generating dialogue around web privacy and web analytics in libraries and communicating the activities and progress of the Forum. At each of these conferences, the project staff will engage with the library community through presentations, workshops, and/or working groups that meet and discuss Forum topics. We have strategically planned travel so that the project staff will be able to engage with a range of stakeholder communities at the following conferences: DLF Fall 2018, Code4Lib 2019, and CNI Spring 2019. The DLF community represents a diverse group of academic librarians at many levels of experience, while the Code4Lib conference will allow the project staff to engage with on-the-ground designers and developers from public and academic libraries. Final project outputs will be delivered at CNI Spring 2019, which will allow the project team to interface with academic library administrators, a key policy-making stakeholder group that can help shape analytics implementation at a high level. In additional to in-person community meetings, we will also solicit and cultivate virtual contribution to the project through web and social channels. Following the final output production, we intend to work with ALA, LITA, and/or DLF to develop an open, online workshop or webinar series so that Forum insights and best practices can be shared widely with the library professional community.

These community-based post-Forum activities will help ensure that the priorities and perspectives of a larger, national community of stakeholders are represented in the final project outputs, and that the Forum insights and outputs reach a broad audience.

2.4 Forum Collaboration

The work of our *National Forum on Web Privacy and Web Analytics* will be enhanced by <u>a related forum</u>— recently funded by the IMLS with Michael Zimmer and Bonnie Tijerina as Principal Directors—that will explore what the library value of privacy means in the digital world. The timing of these two forums is fortuitous, as it allows for complementary work and amplified impact. Zimmer and Tijerina's forum will precede our proposed forum, and aims the privacy discussion broadly at digital platforms and digital spaces. Zimmer and Tijerina's forum will help establish a foundation for our forum to build upon. Our forum aims precisely at web analytics and third-party tracking, so we will be in a position to integrate the work of Zimmer and Tijerina's higher-level forum in a way that allows us to even more quickly and deeply engage in our specific privacy topic. For instance, the privacy field guides created by Zimmer and Tijerina will inform the production of our pre-Forum resources, and their final report—which is planned to be released the same month (May 2018) that our grant period will begin—will be integrated into our forum discussions and exercises. To help connect our two forums and establish our complementary work, we have opened a channel of communication with Zimmer and Tijerina and have invited them to participate in our proposed *National Forum on Web Privacy and Web Analytics* (see supplemental material for email correspondence). In coordination with Zimmer and Tijerina's forum, we can build on their recent and relevant work in a way that will accelerate the overall advancement of privacy in libraries.

2.5 Project Timeline

Phase	Duration	Activities
1. Forum Planning and Preparation	May 2018 - September 2018	- Create Forum website and call for nominations - Identify participants - Produce pre-Forum resources with community input
2. Forum Event	September 2018	Forum event
3a. Forum activities reported to the community both in person and with remote participation, with opportunity for feedback and refinement	October 2018 - February 2019	- DLF Fall 2018 - Code4Lib Spring 2019 - Communication and engagement via web and social channels
3b. Final output production	February 2019-April 2019	- Analysis and synthesis of Forum and community input
3c. Final output presentation	April 2019	- CNI Spring 2019 - Online workshops and webinars - Project website

2.6 Assumptions

In proposing this forum, we assume that threats to privacy will only increase as third-party tracking becomes more widespread, embedded, and sophisticated. We additionally assume that privacy and analytics tracking will continue to resonate within the library community throughout the duration of the grant. For this reason, we assume that the outputs from this grant work are—and will continue to be—of value to the library community.

3. Statement of National Impact

A national forum on web privacy and web analytics will allow the library community to translate our principles into practice. In convening experts from across disciplines and professions, we can break down silos, learn new perspectives, focus our privacy discussions around analytics and third-party tracking, and produce a roadmap for enhancing web privacy with values-driven web analytics practices.

The Forum will bring together professionals with a range of expertise related to privacy, analytics, and the web. We intend to include representatives from across disciplines, including librarians, library administrators, systems developers, product designers, user experience designers, grant program officers, lawyers, technologists, lawyers, and privacy researchers. A gathering of these stakeholders is needed in order to develop an actionable agenda for implementing values-driven web analytics across a nation-wide spectrum of libraries: public, academic, community college, tribal, school, and special. The dialogue and outputs from the Forum will be targeted at two main areas of impact: practice and policy. In terms of practice, the Forum will produce an action handbook that will inform and shape practice by helping librarians implement and teach privacy for themselves and their communities. The action handbook can be integrated into library curricula and other teaching contexts. We recognize that in order to transform insights from the Forum into actionable outcomes, we will need to effect policy at the local, state, and national levels. Forum participants will notably include policy-makers, including library deans and directors, who can help ensure that recommendations developed during the Forum can be implemented.

A community-fueled effort focused around the Forum and its follow-up activities will enable our profession to take important strides toward a better analytics practice that protects our users' privacy from unwanted third-party tracking and targeting. In convening experts from across disciplines and professions, we can break down silos, learn new perspectives, focus our privacy discussions around analytics and third-party tracking, and produce a roadmap for enhancing our analytics practice in support of privacy.

4. Diversity Plan

The Forum will be designed for diversity and inclusion, especially from traditionally underrepresented populations. A depth of diversity will help ensure that a range of voices are included in this discussion, and that a range of perspectives and ideas are able to come forward, ultimately producing a better result. To help ensure diverse representation and an atmosphere of inclusion at the forum, we will consult with members of Montana State University's Cultural Attunement Committee, which was created by ADVANCE Project TRACS, a multi-year grant project funded by the National Science Foundation that aims to increase faculty diversity through research, policy, and training. This committee possesses expertise in recruiting and retaining diverse talent, and will use this expertise to provide guidance on the following matters throughout the project: drafting the call for participation via email and web; placing the call in appropriate forums; evaluating the participant pool; drafting the forum's code of conduct; and advising on post-Forum outputs.

Activities			2018								2019			
			May	June	July	August	September	October	November	December	January	Feburary	March	April
Phase 1:	National Forum Event Planning and Preparation	1. Create Forum website and call for nominations												
		2. Identify participants	██	██	██	██								
		3. Produce pre Forum resources with community input												
Phase 2:	The National Forum						██							
Phase 3a:	Dissemination and Communication with the Community	1. Communication and engagement via web and social channels												
		2. DLF Fall 2018						██	██	██	██	██		
		3. Code4Lib Spring 2019												
Phase 3b:	Final Output Production											██	██	
Phase 3c:	Final Output Presentation	1. CNI Spring 2019.												
		2. Online workshops and webinars.												██
		3. Project Website												

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.

> In order to maximize dissemination and reuse of project resources, all project resources will be assigned a Creative Commons License CC BY 4.0. The project team is committed to making all materials and resources available free of payment and access restrictions according to the terms of the CC BY 4.0 license. Resources will be made available through the project website, hosted by Montana State University. Project resources will be managed through OSF and archived in MSU ScholarWorks: http://scholarworks.montana.edu. The design and content of the website itself will also be licensed CC BY 4.0.

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.

> Project resources will be publicly available without any access restrictions. Use will be governed by a CC BY 4.0 license that will be noted on the materials. Published materials will be copyright by the authors and will be made open access via ScholarWorks, the open access institutional repository of Montana State University: http://scholarworks.montana.edu.

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.

> N/A

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.

> Project resources will include a website, literature review, environmental scan, approximately 40 privacy worksheets to be completed by each forum participant, notes created during the forum, a white paper, a privacy analytics action handbook, and conferences presentations and peer review publications. These resources will be distributed and archived as HTML and PDF documents.

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.

> Standard web services and word processing software will be used to create project resources, including Google Docs, Microsoft Word, and Box Notes.

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).

> HTML, PDF

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).

> The project will be led and managed by PD Young. Production and dissemination of project resources will also be led and managed by PD Young. Co-PDs Clark and Mannheimer will provide support throughout the duration of the project, and program analyst Hinchcliffe will provide reviews, evaluation, and consultation in the production of project resources.

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).

> Resources will be made available through the project website, hosted by Montana State University, with project resources managed through OSF and archived in MSU ScholarWorks, the open access intuitional repository of Montana State University: http://scholarworks.montana.edu.

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).

> MSU ScholarWorks runs on the DSpace software, which automatically creates metadata in accordance with Dublin Core standards.

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

> Project resources will be archived in ScholarWorks, the open access intuitional repository of Montana State University: http://scholarworks.montana.edu. The MSU Library is committed to preserving and providing access to all of the contents in ScholarWorks, as prescribed in the Guidelines and Policies of the IR: http://scholarworks.montana.edu/docs/#preservation.

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).

> MSU has expertise in search engine optimization and social media optimization, and will apply that expertise to the project website so that project resources can be discovered widely through search engines, databases, and social media platforms. Moreover, MSU ScholarWorks has been optimized for search engine discoverability.

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).

> All project resources will be openly available via the project website and MSU ScholarWorks. No special software or tools will be necessary.

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.

> N/A

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.

    N/A

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.

    N/A

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

    N/A

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

    N/A

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

    N/A

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.

    N/A

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

    N/A

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

N/A

Name of publicly accessible source code repository:

URL:

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):

N/A

Name of repository:

URL:

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