**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

## Abstract

PALCI and partnering library consortia are collaborating on a one-year planning grant to develop an open technology, prototype usage data platform for library consortia and their member libraries, for the collection, display and analysis of consortial library usage data. Partnering organizations include Jisc (UK), Couperin (France), Virtual Library of Virginia (VIVA, USA), University System of Maryland and Affiliated Institutions (USMAI, USA), Canadian Research Knowledge Network (CRKN, Canada), Statewide California Electronic Library Consortium (SCELC, USA), and California Digital Library (CDL, USA).

The proof-of-concept platform will be built by adapting the source code of an existing proven consortial library usage data solution, the JUSP service (Jisc Usage Statistics Portal), originally developed by Jisc for higher education institutions in the United Kingdom. Jisc previously shared this same source code with Couperin in their efforts to develop a similar usage data service for French libraries (MESURE). The international library consortium community has overwhelmingly indicated through surveys and community feedback a desire and mission-critical need for a usage data platform/service to facilitate collection development/purchase decisions; yet, there are no scalable, widely-affordable existing solutions that meet the needs of consortia in the United States. Pending commercial solutions due to market in 2017 have major disadvantages, including the loss of highly valuable competitive intelligence information vital to library negotiations for electronic resources and content.

This project will be conducted between June 2017 and May 2018. During this time, PALCI and project partners, with the help of an experienced application developer, will adapt Jisc source code to create a proof-of-concept platform, evaluate and test the platform, and develop a report and plan with actionable next steps for future platform implementation, sustainability and service. The report, which will be delivered to the library consortium community through the International Coalition of Library Consortia (ICOLC) meeting and communication channels, and via a project website, will provide recommendations and an opportunity for collective feedback.

This project aims to increase consortia and libraries' economies of scale by further developing a community of interest in this area through increased collaboration on issues related to library usage data challenges, joint vendor problem resolution, and community-based technical development. The availability of a standards-based usage data platform for library consortia will result in libraries and consortia positioned to serve as exemplary stewards, empowered to make evidence-based decisions and better informed investments in electronic resources.

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

# Introduction

With this grant, PALCI and partnering library consortia will develop and test an international, modular, open technology, proof of concept platform for the collection, display, and analysis of data about licensed electronic resources.

The CC-PLUS (Consortia Collaborating on a Platform for Library Usage Statistics) proof of concept platform will adapt software previously developed by a partnering organization (using generic components of the Jisc Usage Statistics Portal codebase) to collaboratively address community-identified usage data challenges. This project will result in a shareable platform deployed to consortia and member libraries to:
- increase libraries' analytic capacity;
- create staffing and cost efficiencies; and
- empower libraries and consortia to practice exemplary stewardship by making data-informed decisions regarding investments in electronic resources.

# 1. Statement of National Need

Consortial Need for Usage Data
Library consortia exist around the world to give member libraries a strategic advantage by working as a collective to obtain and create resources and services critical to their users. Consortia do this by pooling members' financial, staffing, and physical and electronic resources, providing libraries with a critical way to keep pace in the constantly changing information environment. With bigger buying power, consortia achieve improved discounting and purchase terms, and through centralization of systems and infrastructure, consortia create efficiencies, often providing services individual libraries may not be able to achieve on their own.

A core activity for many libraries and consortia is the purchasing and licensing of electronic resources such as databases, e-journals, and ebooks. Managing and providing access to these eResources is complex and requires extensive resource commitments. Libraries must justify both the initial and ongoing investments associated with eResources, especially in an era of widespread diminishing budgets.

One key component of ongoing eResources evaluation is data about the usage of those resources. Data about the usage of electronic collections is vital for library managers in making informed decisions, such as renewal and cancellation decisions, price negotiations, and decisions about library activities supporting the integration and use of those collections by patrons. While this data is vital, many libraries and consortia face significant challenges in collecting and analyzing eResource usage data, which is usually provided by the library vendors from which eResources are purchased or licensed. Standards and protocols, such as Counting Online Usage of NeTworked Electronic Resources (COUNTER) and Standardized Usage Statistics Harvesting Initiative (SUSHI) aid in the reporting and collection of usage data, but there is wide variance among content vendors in how, and to what extent, these standards are actually supported. Consequently, obtaining and analyzing reliable and consistent eResources usage data from multiple vendors is often a very difficult and time-consuming process.

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017


In addition to problems obtaining reliable and consistent usage data from vendors, libraries and consortia are challenged by a lack of adequate tools for doing this work on a large scale. More than 40 consortia representing thousands of libraries around the world articulated a mission-critical need for usage data systems in a 2015 survey conducted by the International Coalition of Library Consortia (ICOLC)[i]. ICOLC, a group comprised of approximately 200 library consortia from around the world, facilitates discussion and action on issues of common interest. Many of the individuals involved in this grant proposal were involved in developing and/or taking the ICOLC survey after on-going discussions of need in this area. The survey results demonstrated a tremendous shared and urgent need for usage data and usage data system solutions, highlighting the struggle facing consortia to provide this pivotal service to members, and revealing the need for further collective action.

Existing Solutions and Challenges
Some commercial usage data systems exist, though they do not sufficiently address consortial needs in several ways. First, systems available to consortia[ii] are integrated with ILS functionality or other products. Though these systems offer automated functionality similar to features of the solution proposed in this project, they require libraries to maintain and use the same ILS, whereas most consortia serve libraries using different ILS systems. Secondly, libraries and consortia are hesitant to risk managing this vital decision-making data with services offered by non-library entities due to the competitive intelligence the data provide. These data are key in negotiations and provide significant leverage when validating and evaluating the value of various products. Providing library vendors or third-party companies with this information puts libraries at a disadvantage. There is also a strong possibility libraries' own usage data may, in such a scenario, be repackaged and sold back to the very institutions from which it was derived.

One known existing commercial system not tied to an ILS, ConsortiaManager, still requires significant manual effort to acquire and display usage data from multiple libraries in a single interface. ConsortiaManager[iii] is a third-party ERM and consortium workflow management tool offering the management of multiple libraries' usage data through a single interface, but based on a review and demonstration of the service, it does not offer automated harvesting services, nor is it flexible enough for use by most consortia. Other commercial solutions exist for individual libraries (Intota, EBSCO Usage Data Service), but these offer no consortial functionality, and do not allow libraries/consortia to collaboratively address the shared problems mentioned above. Additionally, these products are costly, require significant management and staffing time at the local level, and still risk the loss of competitive intelligence information.

The lack of consortial system options in this area is not surprising; the volume of consortial data is substantial, and the efforts involved in automation of data harvesting, verification, ingest, and storage are more complex than the already complicated needs of an individual library. The lack of commercial solutions has led many consortia to create locally developed alternatives, while others simply cannot afford, or do not have the staffing, expertise, or resources to do this important work on their own. Some consortia hire staff who gather and normalize statistics for all of the member libraries; however, assembling, housing, and analyzing usage statistics for multiple libraries quickly becomes complex and presents significant challenges.

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

Widespread data-related challenges are faced by consortia and individual libraries, including: varying levels and/or lack of COUNTER compliance by electronic resource vendors; vendor responsiveness to data-related problems; the considerable time required to access and download data files across many unique platforms; and finally, the significant time and resources required to review and monitor the data for issues such as inconsistent and missing data. These challenges often leave even libraries and consortia able to collect the data with little time for true analysis.

Many publishers operate their own proprietary platform, and in the case where publishers have merged, it is common for journals, and data about usage of journals, to be distributed across multiple platforms. Accessing and downloading usage data requires logging in to each distinct platform, locating the usage files for the date range desired, then downloading the usage file in a usable format. For research-intensive institutions with large collections, this exercise is complex and time consuming and libraries and most consortia do not have the resources required to complete this more than a few times per year, if at all. The usage data downloaded from these platforms are often inconsistent, and so data normalization is necessary before the data are useful for analysis. In addition, there are often gaps in the data, where usage for periods of months or years is unavailable because of technical problems or other reasons.

Many times consortia complete this work through heavy use of manual collection processes and Microsoft Excel.[iv] While Excel provides excellent assistance with cleaning, processing, and viewing data, it cannot provide the large picture analysis or significant storage needs a database or integrated system could. In addition, workarounds are needed as the volume of data to be analyzed in a consortial data set often runs up against Excel's limitations on rows and columns.

For a variety of reasons, such as limitations of technical expertise/resources, only 20% of ICOLC Survey respondents reported making use of the SUSHI automated retrieval standard, further signifying much of the work of collecting data is done manually. The lack of an effective automated consortial usage statistics tool makes gathering consistent usage data complicated and unsustainable, and for those with disparate locally developed services, there is often duplicated effort in the resolution of shared data issues.

The timing of the CC-PLUS initiative is fortunately paired with the upcoming COUNTER Code of Practice Release 5. Early insights into Release 5 have shown it will bring significant changes to how libraries and consortia must process and analyze usage statistics. If the initiative continues as stated, it is likely to eliminate COUNTER Consortium Reports many consortia rely on so heavily for collated usage statistics. As Release 5 is planned to be distributed as a draft in January 2017 with required compliance by the end of 2019, this is the optimal time to create a shared system to accommodate the new version of the standard.

## 2. Project Design
The CC-Plus project is a collaborative expansion on an existing proven solution in the United Kingdom (the Jisc Usage Statistics Portal, also known as JUSP) and France (Couperin's MESURE service). This project provides for the development of a community-built proof-of-concept platform for the collection, display, and analysis of library usage data, the need for

which has been clearly delineated across consortia, both nationally and internationally. Initially developed by Jisc, the JUSP codebase has been similarly adapted for use in France by the Couperin consortium.[v] CC-PLUS intends to build on these successes, adapting the same platform for use by many other consortia interested in offering usage data services to member libraries.

The Jisc Usage Statistics Portal service collects, aggregates and presents usage data for online resources from publishers and aggregators as part of Jisc's offering to UK higher education institutions. Not only does JUSP offer a single point of access to usage data, it provides institutions with a variety of reports and analytical tools, supported by user guides and support materials.

JUSP uses the SUSHI protocol to gather COUNTER reports (currently JR1, JR1a and JR1 GOA, and shortly BR1, BR2, BR3). Through its SUSHI server and API, it enables institutions to re-harvest data for inclusion in their own library management tools for local analysis. Other important, though less tangible elements of the JUSP service include developing relationships with all users and vendors to ensure the service continually responds to user needs, delivering training and support in the area of usage data, and providing expertise to a range of suppliers, organizations and initiatives dealing with usage data. The JUSP portal is free for all UK higher education institutions and research councils. Elements of the service provided by Jisc, particularly the benefits of SUSHI harvesting and presentation of data via the portal, have clear benefits for institutions and consortia outside UK higher education.

The CC-PLUS project will adapt selected generic components of the JUSP service to extend its codebase, making it freely available as a software package to consortia and individual libraries in the United States and beyond.

The envisioned CC-PLUS consortial solution, adapted from the generic components of the JUSP codebase, will provide a working prototype software for the following:
- the robust management of many libraries' usage data through a single interface;
- an automation of data harvesting, verification, ingest, and storage;
- a shared tool among consortia for data analysis, preservation, and distribution of usage data to member libraries;
- and a collaborative approach to resolving usage data issues.

As a prototype, the first stage of software development will be limited to basic functionality scoped by the CC-PLUS Advisory Board in early technical planning stages. But to make effective use of library usage statistics, consortia need additional tools for collection and analysis of usage data, such as customized reporting functions, data visualization, and the ability to augment usage statistics by incorporating additional contextual data sources, such as cost data, holdings data, and institutional information, such as enrollment counts. Fortunately, these challenges can be addressed in future iterations and project phases. Jisc began developing its JUSP platform in 2008 and completed the first iteration in 2011.[vi] Since then, Jisc has continually updated the software, most recently including functionality for ebooks. The CC-PLUS project proposes to adapt the generic components of the JUSP codebase for use by

consortia, then contribute to the international community of consortia using this tool with additional customizations and functionality.

Anticipated Project Phases

| Phase 1: Proof-of-Concept | Phase 2: Future Pilot Implementation | Phase 3: Envisioned Hosted Service |
|---|---|---|
| <ul><li>Funded by IMLS NLG Planning Grant</li><li>Develop technical plan for adaptation of generic components of the JUSP codebase</li><li>Create basic prototype, scoped by the CC-PLUS Advisory Board for: 1) Management of multiple libraries' usage data through a single interface, 2) Automated harvesting/verification /ingest/storage, 3) Web-based interface for data analysis and distribution</li><li>Evaluate and report on long-term sustainability prospects for the product</li></ul> | <ul><li>May be self-funded or completed through future grant funding</li><li>Test software package by extending software for use and evaluation by several consortia</li><li>Extend capabilities by building or incorporating tools for data visualization, ability to import contextual data (enrollment data, cost data, holdings data, etc.), and enhanced reporting features and functionality identified in proof-of-concept phase</li></ul> | <ul><li>Creation of a service using developed software hosted and offered by a library consortium or possibly a separate non-profit organization</li><li>Aimed at consortia unable or unwilling to host/implement their own usage service</li></ul> |
| **Ongoing Project Development: Phases 1-3** <ul><li>Collaborative technical development on open software by consortial community</li><li>Development of a consortial community to resolve shared usage data issues</li></ul> | | |

Outcomes
The projected performance goals and outcomes for this project are as follows:
- To develop a modular, open technology, proof-of-concept platform for the collection, display, and analysis of library usage statistics, including technical documentation.

- To complete a full project report. This project report will include an evaluation of the proof of concept and recommended next steps for future development, such as a preliminary plan for how the platform may be locally deployed for use by any consortium and its member libraries. The report will also include a sustainability plan.
- To create a project webpage and communication forums facilitating stakeholder collaboration throughout the planning grant and beyond.

These goals and deliverables assume there is a base level of similarity among library consortia, making a shared usage statistics system meaningful.  The JUSP software is currently heavily customized for the Jisc environment, including needs specific to higher education in the United Kingdom. Similarly, Couperin's adaptation of the JUSP software is relevant and useful for the higher education environment in France; their experience showed the need for significant adjustment and customization, including the development of customized reporting. As the CC-PLUS project intends to create a product useful to many different consortia, care will be taken to ensure the final product is both general and flexible enough to be applicable to a diversity of consortial environments. It is likely early phases will limit the project's scope to standardized COUNTER data, but later phases may expand the scope significantly, and thereby expand the number and type of consortia to which this project may be of interest.

Project Communication and Dissemination
The results of this project will be made available to consortia through monthly updates to the ICOLC (International Coalition of Library Consortia) community email list, a project website, and a final report, with presentations planned at relevant conferences and venues (e.g., ICOLC conference, Charleston Conference, and ALA). Software will be made available through the project webpage, and will likely be available through public open development sites, such as GitHub.

Risks
The primary risk to this project was a commercial service meeting consortial needs related to usage data prior to the completion of this open, freely available, proof of concept platform. However, project partners have investigated where potentially parallel work is being developed, including a future Alma platform plugin[vii] and the newly launched ConsortiaManager tool[viii]. Based on research and conversations with these companies, project partners have determined neither of these products will provide the necessary solution in a complete or timely manner. Further, neither solution is freely available to consortia with limited funds, therefore mitigating any initial concerns about duplicative work in this field.

Ongoing risks to this project are 1) the software developed will not fully meet the needs of consortia wanting to manage usage statistics for their member institutions, or 2) the software will require too much technical expertise to use and maintain.  These risks are mitigated by involvement of Jisc, Couperin and other consortial experts who have experience directly developing and maintaining these systems. Communication and the active support of the ICOLC community throughout the project will ensure the platform's relevance and usability to this diverse audience.

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

Project Staff & Advisory Board
This project will be directed by Jill Morris, of the Pennsylvania Academic Library Consortium, Inc (PALCI). Jill has more than eight years of experience working for two large library consortia in senior positions ranging from Assistant Director and Interim Executive Director to Senior Program Officer. While working for a large multi-type consortium in a previous professional position, Jill managed a staff who developed and maintained a locally developed automated usage data system with a web-based interface for the display and analysis of library usage data. During this time, Jill also studied ways to make usage data more meaningful to libraries by contextualizing this data[ix], and encountered many of the same challenges articulated by her consortial colleagues in the 2015 ICOLC usage data survey. In addition, Jill has significant experience managing library consortium projects, and has a deep knowledge of library consortia needs and challenges.

In order to meet the goals outlined above, the project team will contract with a trusted application developer. Should this project be funded, Scott Ross, an independent application developer, has already agreed to lead the technology development effort. Scott Ross is an independent contractor who recently retired from working with NC LIVE, a multitype library consortium, for more than 17 years. As the primary application developer for NC LIVE, Scott oversaw all technical development and project management, and built a locally developed usage data system, including an automated/scripted harvest, ingest and web interface reporting system still in use today[x]. Scott is familiar with usage data standard protocols and has worked on several recent projects to modify COUNTER usage reporting for various consortia.

Central to the success of this initiative is creation of an Advisory Board composed of nine committed partner organizations experienced with and knowledgeable about journal usage data systems and formats.[xi] These organizations include consortia and non-profit organizations of different sizes and scopes representing many types of libraries, including academic, public, and research libraries. The Advisory Board members have deep understanding and hands-on working knowledge of consortial and library usage data needs, formats, systems, and workflows for managing this data. Jisc, developer of the initial software, is a key partner in the project, as is Couperin, the French consortium which previously adapted the Jisc code for its local needs. Many of the additional consortia, including VIVA, CRKN, CDL, and USMAI, track usage statistics centrally for their member institutions through a variety of methods and are familiar with the range of challenges in managing consortial data. They will be ideal partners for testing the solution as it develops. In addition, a member of the Advisory Board is on the COUNTER communication group and the NISO Committee for SUSHI. This will enable direct communication with these standards organizations as needed.

Project Schedule Overview
The twelve-month timeline for this project, outlined below, assumes an anticipated start date in June and approximately 17 non-consecutive weeks of product development time. Please see the Schedule of Completion for more information.
- Months 1-3: Project planning; set-up hosting environment and code repository based on the generic components of Jisc software; install and configure available Jisc software

> scripts, SUSHI code, SUSHI server, and database schemas; customize the schemas, page templates, and harvesting scripts, as needed.
> - Months 3-5: Construct a prototype consortial/institutional management interface; define resource and vendor connections to harvest data for testing purposes; begin harvest and ingest of vendor and/or uploaded datasets.
> - Months 5-12: Construct a prototype reporting interface; conduct a sustainability analysis for the software and envisioned service; release a draft report and documentation for ICOLC community review; create a final report incorporating feedback.
> - Throughout the project, partner consortia represented on the Advisory Board will communicate with additional stakeholders through the ICOLC community to:
>   - Refine community-based needs; perform testing and collect feedback on platform development; plan for future project phases and pilot implementations; and create and nurture a consortial community on library usage data, including developers and leaders for process and platform improvements.

Feedback on the platform evolution and project parameters will be regularly contributed by the Advisory Board and supporting partners, described earlier. This group will meet at least monthly via conference and/or video call to review project plans, timeline, feedback, and development of the proof of concept platform. The Advisory Board will also be the group responsible for ensuring the project is on-course to meet performance objectives and deliverables. This approach ensures the project remains a collaboratively developed project addressing a specific need, already clearly identified within the consortial community. As such, consensus building and buy-in are critical to its success. For those reasons, and because of the project's diverse partners, the timeline above accounts for a draft report to be distributed to the consortial community for feedback, and continuous communication and testing on platform phases as they are developed.

To that end, the consortial community, through outreach on the ICOLC email list, has already responded enthusiastically to this project, as reflected in the letters of support. A listserv for consortia members to provide regular feedback as the project progresses has already been developed and populated.

The work completed during this planning grant, including both the successful completion of the proof of concept platform and the development of a consortial community interested in library usage data, will situate this project well for further development grants to expand the reach and scope of the platform, including the investigation of additional content/data types. After a successful, viable prototype is created, future phases will include consortial pilots and testing. More information is available in the Sustainability section below.

Sustainability Beyond the Proof-of-Concept
The project team aims to create a successful prototype as a proof-of-concept, as well as project plan that will, in future project phases, result in the deployment of the platform across multiple consortia, in addition to the creation of a truly collaborative community aimed at improving consortial effectiveness in this area. As the platform is adopted by a greater number of consortia, efforts may be made to build out the ideal features of a usage data solution as described in the 2015 ICOLC survey results.

This project may also result in the eventual creation of a hosted service similar to that of JUSP in the UK and MESURE in France. Ideally, this platform will be lightweight and flexible enough for consortia to implement on their own, but for those with limited staffing and technical resources, an affordable hosted usage statistics service may be more beneficial. It is possible PALCI or a partner organization may develop and offer this service on a cost-recovery basis, or it may be preferable to create a non-profit organization solely for the purpose of managing and providing such a service to library consortia.

Further development of the platform and these services would likely become self-funded efforts, where developers can contribute to an open platform, creating efficiencies for those currently managing their usage statistics through other means, and allowing those with limited resources to take part in a collaborative approach to managing this type of data.

Budget
The estimated project budget is $50,000. See Budget form and Budget Justification document for more information.
- Contract Wages of $40,800 (approximately 680 hours at $60 an hour to contract with identified developer, no benefits or overhead, quote attached as a supporting document).
- Travel/Meeting attendance: $2,400
- Web hosting services: $1,800
- PALCI project staffing: $5,000
- In-kind contribution of PALCI staff time
- In-kind contributions of time, expertise and other resources from partners and those serving on the CC-PLUS Advisory Board. (Although no matching funds are required for this proposal, the contributions of partners to this project could be recognized as a competitive match amount.)

# 3. National Impact

The development of a modular, freely available open technology platform for consortial collection, display, and analysis of library usage data has the potential for large-scale impact across thousands of libraries, nationally and internationally. Specifically, by providing an open usage statistics platform to manage data for libraries simultaneously, this project will improve effectiveness of libraries by decreasing time spent managing systems, support data driven decision making for libraries without the infrastructure to maintain this type of information locally, and most importantly, allow for consortia and member libraries to place a greater focus on analysis and use of usage data for informed collection development activities, including vendor negotiation, collection management, and resource sharing.

By collaboratively addressing the community-identified usage data challenges surfaced in the 2015 ICOLC survey, the proposed project aligns with the IMLS priority of a national digital platform supporting exemplary stewardship of library services. In leveraging the collective knowledge of library consortia with decades of experience and vested interest in usage statistics management, this project brings a nuanced understanding of the digital infrastructure issues

faced in the creation of such a platform. This proof of concept platform will answer the challenges outlined in the statement of need by: building on existing technologies and work, engaging with consortial communities, and promoting computational analysis through adoption of an open, internationally accepted, standards-based tool that empowers libraries to make informed decisions about investments in electronic resources.

The sharing of an open, common codebase for this effort will further empower consortia and member libraries to create efficiencies in time and effort, and encourage innovation through the pooling of diverse needs, experience, and expertise.  The project's accessibility for use by any consortium and its member libraries is ensured by focusing on a proof of concept platform for the collection, display, and analysis of library usage statistics that may be locally deployed, and is both modular and open technology.  Finally, having an option for consortia that is conceived and developed by and for libraries is critical, as it removes the real risks, noted earlier in the narrative, associated with having this crucial decision-making data managed by non-library entities.

As outlined in the performance measures, the project webpage and communication forum will make the developed proof of concept platform for the collection, display, and analysis of library usage statistics freely available. Also included on this webpage will be the full project report including recommended steps for future development, a pilot service proposal, and a sustainability plan.  In addition to this public facing resource, this project boasts nine committed partners and Advisory Board organizations ensuring robust community feedback throughout this initial development stage. These diverse consortia and partner organizations represent hundreds of libraries across the United States, as well as organizations in Canada, the United Kingdom, and France. Project partners will communicate with stakeholders and refine community-based needs by: developing the functional needs for the prototype; collecting feedback on product development; and finally, by creating and nurturing a consortial community for process and platform improvements. Success will be measured by community engagement in the development of the platform, the completion of the proof of concept platform, and the broad dissemination of the report and tool to the community at-large.

This planning grant will create a project plan and prototype for the development of a needed standards-based usage data platform for consortia. The platform will then position consortia and member libraries to serve as exemplary stewards for evidence-based decisions and better-informed investments in electronic resources.


(Please see all endnotes available as a supporting document.)

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

## Schedule of Completion

| Date(s) | Activity/Milestones |
|---|---|
| June 2017 – May 2018 *(On-going throughout the project)* | ● Monthly CC-PLUS Advisory Board Meetings via conference call <br> ● Monthly CC-PLUS updates to the ICOLC community via project web pages and previously established email list (previously known as the ICOLC Usage Statistics Initiative Interest List) <br> ● Evaluation/Collection of feedback via surveys and user testing with project partners and stakeholders, including consortia of multiple types <br> ● Seek additional funding and contributions from stakeholder community, in addition to researching and planning options for long-term platform/service sustainability models |
| June – August 2017 | ● Hire identified application developer/technical project lead <br> ● Scope proof-of-concept features with Advisory Board and Application Developer feedback <br> ● Fully develop technical plan with application developer based on project requirements adapting Jisc Usage Statistics Portal (JUSP) codebase <br> ● Communicate project scope and development phases; <br> ● Identify appropriate licensing model for resulting digital product <br> ● Receive generic components of JUSP code base from Jisc <br> ● Set-up hosting environment <br> ● Install/configure Jisc software scripts, SUSHI code and server, database schemas <br> ● Customize schemas, page templates, and harvesting scripts <br> ● Develop test use cases |
| September – November 2017 | ● Construct a prototype consortial/institutional management interface <br> ● Define resource and vendor connections to harvest data <br> ● Begin harvest and ingest of vendor and/or uploaded datasets using defined test cases and anonymized data |
| December 2017 – February 2018 | ● Construct a prototype reporting interface <br> ● Conduct a sustainability analysis for the software and envisioned service |
| March - April 2018 | ● Release a draft report and documentation for ICOLC community review, including a sustainability plan for next steps <br> ● Begin drafting pilot project implementation plan |
| May 2018 | ● Create a final report incorporating feedback <br> ● Transition to future project phases with implementation of a CC-PLUS pilot and sustainability plan as needed |

**Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**
Submitted to IMLS by PALCI, January 2017

## 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.

> This project will develop an open software product by assigning a non-restrictive license, allowing for community development and modifications. A shared understanding of this goal has been developed through a series of conversations with Jisc, and part of the work outlined in this planning grant will be establishing which license is most appropriate. As the software product will be based on the generic components of the Jisc Usage Statistics Portal (JUSP) codebase, the proof-of-concept platform license will be mutually decided with Jisc.

**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.

> The proof-of-concept platform will be made publicly available to libraries and consortia once a viable product is available, and may be subject to terms of access and use complying with the selected open license model. Potential users of this product will be provided with written/digital notification of any limitations of use through a license agreement notice prior to use. PALCI does not intend to assert ownership other than to steward future life of the product.

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

> The proof-of-concept platform will be designed to harvest and ingest usage data. For this project phase, only anonymized test data will be used with explicit institutional permission, limiting privacy concerns.

> Future project phases will require on-going discussions with regard to privacy as institutions will want to be able to control who is able to see data about their institution.

**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.
> N/A

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

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

> Sample usage data and reports will be available in formats including XML, SQL, .xlsx, .csv, .tsv formats. In future, file formats may also include JSON.

**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).
> N/A

**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).

> Anonymized data will be used for testing the prototype and will not retained.

## 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).

> N/A

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

> N/A

**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).

> N/A

## 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).

> N/A

**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.

> This project will develop a prototype, proof-of-concept usage data platform adapting the generic parts of the JUSP source code and scripts currently available from the Jisc organization in the UK for local hosting and use by library consortia and individual libraries.

> This platform will be capable of harvesting, validating, storing, and presenting usage data from publishers and vendors for multiple libraries through a single point of access.  Using the SUSHI protocol, the platform will enable automated harvesting of COUNTER reports, and it will provide libraries and consortia with a variety of reports to use this data to inform their collection management decisions.

Through its SUSHI server, the platform will enable individual institutions to re-harvest data for inclusion in their own library management tools for local analysis.

This tool will be a stand-alone software package that can be locally deployed, and in future project phases, may be hosted/used by an individual library or library consortium, and could possibly become part of a hosted service.

Future phases may also include usage data performance metrics and analytics, additional/advanced custom reports, and an API.

**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.

- There are currently no other commercially-available or open software solutions that can harvest, store, and present COUNTER usage data for multiple libraries simultaneously.
- The following products are in development for consortial use, but do not meet the needs of this project for the reasons listed below:
  - Alma platform plug-in – ProQuest/Ex Libris intends to offer a single interface for consortium statistics that allows for the automated loading of COUNTER data as part of the Alma Integrated Library System (ILS), but it is not yet available. If it is successfully developed, it would require the use of the Alma shared ILS by a consortium, which is cost-prohibitive for most libraries not already using this system, and, it would risk providing competitive intelligence to a major library information vendor.
  - ConsortiaManager – ConsortiaManager, a newly available third-party tool that assists consortia with the management of their daily tasks, such as negotiation, licensing, and invoicing, does not offer automated loading of COUNTER data.
- Other services, such as EBSCO Usage Data Service, OCLC License Manager, and Intota, exist with data harvesting and reporting capabilities for individual libraries, but they do not function at a consortium level, risk providing competitive intelligence used in negotiations to major library information vendors, are costly, and result in duplicative work as multiple libraries maintain separate systems.
- Manual processing is an option available to libraries and consortia, using Microsoft Excel, but this process is time intensive and lacks automation capacity needed for sustainability on a large scale.

**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.

  - PHP, perl, XML, SQL, Javascript - Languages selected due to use in existing JUSP solution. Also selected because they are pervasive, open-source, and a flexible means for building web-accessible database-driven applications. These languages align with the expertise of our selected application developer.

**B.2** Describe how the software you intend to create will extend or interoperate with relevant existing software.
- o   This software requires generic components of the Jisc Usage Statistics Portal (JUSP) source code and will function similar to the JUSP service currently offered by Jisc to libraries in the United Kingdom, and Couperin's implementation of the JUSP service in France (MESURE). The CC-PLUS prototype will be separate from JUSP and MESURE, but will share the basic structure of the JUSP system, allowing for joint development work. Usage data will be harvested using standard protocols and in COUNTER format to allow libraries to integrate data into local systems.

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

Web/server hosting technology and a MySQL database will be needed for developing this proof-of-concept.

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

A complete technical plan will be developed as part of this planning grant. Project management methodology and the project's Advisory Board will be used to ensure adequate documentation and maintenance of this proof-of-concept platform, which will primarily be used by library consortia staff.

Project specifications and regular evaluations will be built into the technical project plan. During the development phase of the proof-of-concept, the product will continually be evaluated and tested to produce a viable product. At completion of this project, a report will evaluate the platform against stated goals, and describe plans for future improvements.

**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.

This project will develop an open software product by assigning a non-restrictive license, allowing for community development and modifications. PALCI does not intend to assert ownership other than to steward future life of the product. A shared understanding of the goal for open, community development has been established through a series of discussions and partnership with Jisc. Part of the work outlined in this planning grant will be establishing which license is most appropriate. As the software product will

be based on the generic components of the JUSP codebase, the open proof-of-concept platform license will be mutually determined with Jisc.

**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 via download from the project website, and possibly other public development forums, such as GitHub. This planning grant will be used to develop an agreement with Jisc allowing for libraries to use this platform more widely.

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

Name of publicly accessible source code repository: Not yet available

URL: Not yet available

**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.

This project will not be creating new datasets suitable for long-term use or re-use. Instead, the project will be using automated protocols to collect anonymized usage test-data from a variety of content providers. The collected test-data will not be retained for any purpose beyond the funded project period.

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

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

**A.4** If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.
N/A

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

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

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

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

> Name of repository:   N/A

> URL:  N/A

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