Proposed Activities and Rationale

A. Importance to the Organization, Its Constituencies, and the Field

Vision for an Art Information Commons

Like other encyclopedic art museums, the Philadelphia Museum of Art has been committed for more than a decade to the digital documentation of its art collection. To date, this systematic process has resulted in the web publication of digital images and basic catalog records for 140,000 works of art—about sixty percent of the art collection. This process will continue for the decade to come. Although this growing digital resource represents a tremendous shift in how the public can access the Museum's collection, it has also opened up a new possibility: linking it to the Museum's vast, largely unmined information resources that provide historic and interpretive context for the art collection.

The Museum seeks the support of the Mellon Foundation to help it define how information about individual works of art in its collection can be linked to related contextual information—such as archival materials, bibliographic references, exhibition histories, provenance records, conservation records, interpretive content, curatorial research files, and non-object images such as exhibition photography. Encompassing the many different informational resources maintained by the Museum, the Art Information Commons initiative promises to unite digitized and born-digital materials, digital images of works of art and relevant contextual information. Linking these information assets will dramatically enhance digital access to the Museum's core artistic and intellectual resources for Museum staff, scholars, artists, students, and the broader public.

Institutional data management practices tend to grow up around departmental units within an organization. Years ago, each Museum division—curatorial, conservation, education, *etc.*—adopted systems based on its professional standards and distinct departmental needs. Over the years, this has resulted in a lack of unified cataloging standards and workflows. As long as the majority of scholarly work was based on physical research and the use of print publications, it was appropriate for the departmental systems to be isolated and inward-facing. But that is changing with the movement toward digital scholarship, large-scale digitization, and open access to extensive research content. The management of data, content, and technology within cultural institutions has not yet evolved with these external shifts.

As a result, both internal and external researchers currently face significant barriers to accessing the full extent of the Museum's collection-related information resources. Through the search mechanisms currently available to them, they are likely to discover only a fraction of what the Museum has to offer. For instance, the most popular search term on the Museum's website during the past year was "Duchamp." Using this term, a researcher would readily discover

multiple works of art in the collection. In a separate section of the website that is not searchable in the same manner, they might discover some of the books that the Museum has published about his work. In yet another section of the website, they might find recent exhibitions organized by the Museum. If they are already aware of the Museum's archival holdings, they could open yet another separate search function for archival finding aids. To view the archival records themselves the user would need to request an in-person appointment to access those materials. Because of the siloed nature of the Museum's archives most users remain completely unaware of the rich informational resources held in those files.

These siloes prevent the researcher from seeing the larger informational context related to their topic and make it more difficult to find relevant items, such as a preparatory drawing, letter from the artist, historic installation photography, and scholarly publication that all pertain to the same work of art. Siloed systems place the burden on the researcher to spend time evaluating disparate systems that use different cataloging standards. Part of the vision for the Art Information Commons is to provide a searchable interface that enables staff, scholars, and the public to tap into the deep reservoir of related content that exists for the permanent collection in a manner that best suits them, from basic data to deeper scholarly information. Once implemented, it will enable researchers to conduct flexible searching across collections, organizing and combining art information in ways that are not possible at present. This will be achieved through linked open data, the glue that will bring these disparate collections together, a model described further below.

Beyond improving the research process, support of the Mellon Foundation will help the Museum undergo an organization-wide transition to the holistic management of its collections-related data and implement the changes it needs to create a sustainable information culture—an Art Information Commons—that spans the entire organization and will benefit data creation, access, and preservation for years to come. This planning phase will develop the important skills needed for quality information management that supports rich integration across departments, and flexible use and reuse of data for many applications (which in the long-term is more efficient). It will formalize holistic governance, training, and oversight on data standards across the different disciplines within a large institution.

This project will establish the foundation for integrating the Philadelphia Museum of Art's collections data with external data sets, allowing for relationships to be created with collections held by other cultural heritage institutions that also use linked open data. This integration between collections holds the promise to yield research that tells an exponentially richer story, one informed by a network of collections and multiple voices in the field. With proposed Mellon Foundation support, the Philadelphia Museum of Art will be among the first of this country's encyclopedic art museums to move towards a full integration of all of its collections using a linked data approach, and the first to create such a managed data center in its library and archives. Planning for this institution-wide shift will yield many lessons to benefit other cultural heritage organizations. These broader impacts are discussed further below, under *Relationship to the Field.*

Benefits of Linked Open Data

Open, standards-based data formats like linked open data and, for images, the International Image Interoperability Framework (IIIF), are well suited as forward-looking solutions to gather data from the existing departmental systems and provide a consistent, richly linked, integrated resource. Linked data is a method of publishing structured data that allows information to be interconnected and thus rendered broadly useful. Linked data is based on principles of publishing content for the web. It uses tagged relationships and unique identifiers to provide context and meaning. An unstructured search for a particular Bellini painting of the 'Virgin and Child' produces countless outcomes based on keywords "virgin," "child," and the name "Bellini." A researcher would have to comb through these results to determine which references were relevant. With linked data in place, Bellini is tagged as an artist/creator, while Virgin and Child is tagged as the title of an art work by Bellini, helping the researcher hone in on directly relevant search results.

Linked data creates a new fabric to bring disparate collections together and allow for computer-assisted analysis. For example, a human can easily understand that an artwork painted by Max Ernst and a letter written by Max Ernst are created by the same person. However, to make this connection in the digital realm, the relationship between the two artifacts must be explicitly spelled out in a format that machines can "read." When computers are able to connect the objects, additional contextual information about that person, the artwork, and the letter can be gathered together, making relationships explicit and potentially even discovering connections that humans may not have made. Linked data is a standards-based model so that it can be shared. With linked *open* data, information can be networked with external data sets in order to facilitate even greater possibilities for discovery and analysis.

The Museum has begun to explore methods of linking its art and archival collections through projects like the Duchamp Research Portal, now underway and described further below. The recent digital publication of a selection of European paintings in the John G. Johnson Collection has allowed the Museum to experiment with adding complete provenance, exhibition, and bibliographic records to object pages. These features empower users to more fully understand the history of an object, the details of its creation and provenance, how it has been studied and interpreted by others, and the people associated with it, such as relevant artists, collectors, and scholars. These exercises can be considered as demonstration projects that point the way towards the realization of a more comprehensive vision of integrated access to our information resources. They illustrate the promise of systematically integrating art works within the full context of the Museum's cultural data.

What was achieved with these projects can be applied to other resources that are of equal importance. These include archival holdings related to significant American artists such as Thomas Eakins and Paul Strand; documentation that traces the activities of galleries such as Julien Levy and collectors such as Louise and Walter Arensberg; and resources that illuminate how works came to the Museum and how their presentation has changed over the years,

ranging from records of the Museum's founding in 1876, to Fiske Kimball's development of a "walk through time" and curator Stella Kramrisch's pioneering work in South Asian art. This project will enable Museum staff, scholars, and the public to tap into the deep reservoir of informational context that exists for the permanent collection.

Current Data Condition

The Philadelphia Museum of Art's current strategic plan, adopted in 2013, embraces technology as a primary tool to engage new audiences, enhance the visitor experience, activate the collection in new and different ways, and contribute to the advancement of humanities scholarship. Over the past decade, the Museum has made a number of significant, phased investments in technology. Most notably, it has focused on the monumental task of building and populating its first and only art collection database, The Museum System (TMS) by Gallery Systems, which is used by many other art museums. Implementing TMS has formed the foundation for the current online catalog, which holds 140,000 object records, with more added daily. With basic cataloging standards and images in place, the Museum is ready to turn its attention to enhancing this resource and integrating it with other collections-related data, namely:

- Conservation records will soon be transferred to the Conservation Studio module of TMS. This data has been managed through Conservation Tracker System, a custom-built, standalone program developed beginning in 2000 with support from the Mellon Foundation. Conservation Tracker software remains available to the cultural heritage community, but in 2016 Gallery Systems introduced its Conservation Studio module which offers the advantages of integration with TMS.
- Information about the Library's 285,000 volumes is housed globally in WorldCat and locally in the ExLibris library system, Aleph.
- The Museum holds 4,200+ linear feet of archival materials, including approximately 70,000 digitized images. Archives data is held in the open source repository ArchivesSpace.
- In addition to object photography in TMS, the Museum also owns some 200,000 digital images such as exhibition installation and program photography. Beginning in late 2017, the Library and Archives has been leading the adoption of an institution-wide digital asset management system for these resources.

Evolving the Library and Archives Division

The goal of this planning phase for the Art Information Commons is to create an infrastructure and roadmap for adaptable, interoperable, and sustainable digital collections. The proposed Art Information Commons will aim for the full integration of the information resources listed above as linked open data, and will establish coordinated governance, training, and data management practices across the institution. This vision calls for evolving the role of the Library and Archives division and positioning it as the steward of the Museum's cultural data, working in close partnership with the curatorial, conservation, publishing, and technology divisions.

Service, accessibility, training, collaboration, data organization, preservation, digitization, and image and information management have long been the hallmarks of libraries and the services they provide. The Library and Archives staff of this institution have extensive knowledge of the research needs of the Museum's curators, educators, visiting scholars, and the public. The Library and Archives serves as the research center for museum staff and visiting scholars, and is founded on a commitment to information access. Curators and librarians jointly purchase books and manuscripts that relate to Museum collections and exhibitions. The Library maintains a scholar's study for visiting researchers, and it regularly hosts graduate art history classes and other groups conducting specialized research assignments. Librarians work closely with patrons to conduct database and archival research.

The librarian profession has its foundations in structured vocabulary work, cataloging, and sharing deep information resources at scale. Information management is different from technical management of systems: It is a discipline rooted in the needs of internal and external users. And it must have as its focus establishing consistent practices and vocabularies that will integrate across silos as well as extend to other institutions' data in order to support broad-based cultural heritage scholarship. In academic and cultural heritage institutions, librarians and archivists often are hired to work in digital centers, labs and media teams, on DAMS projects, and in a variety of digital scholarship projects that are not housed in library and archives units. The Art Information Commons evolves the library and archives into the hub and center of such work.

The Library and Archives is already responsible for the management and preservation of institutional records—analog and digital. It has implemented the Museum's first digital repository Preservica, as well as an institution-wide Digital Preservation Policy (only the second of its kind to holistically address preservation of all museum data, from archival email to art collections). In 2016, the Library and Archives division received an Institute of Museum and Library Services grant to create the National Digital Stewardship Residency program for art information professionals (NDSR Art). The NDSR Art program is strengthening the position of art information professionals as digital stewards. At the Philadelphia Museum of Art, the NDSR Art resident worked with contemporary art curators and conservators to establish digital preservation methods for the time-based media collection, using existing tools and competencies of the Library and Archives. Meanwhile, the NDSR Art resident at the Yale Center for British Art focused on preserving TMS data, a project that exemplifies the field-wide need for cross-institutional data management.

The Museum has gradually prepared the Library and Archives division as the institutional authority responsible for reorganizing its work processes and intellectual resources toward the goal of creating this Art Information Commons. As an initial step down this path, in 2014, the division integrated the Museum's Rights and Reproductions department, the historic office of record for photography at the Museum. Today, the delivery of images to internal and external users, along with rights management, is a reference service like any other research service, provided by the Library and Archives to internal users and to the public without fees. This

internal shift signals the Museum's commitment to open access and a service orientation with respect to its cultural data. In doing so, the Philadelphia Museum of Art may be the only major art museum library to function in such a capacity. Many art museums have maintained legacy Rights and Reproduction departments even though fair use and open access have entered into museum practices and images are now delivered electronically.

Continuing its role in helping the institution manage all of its resources, as noted above, in 2017 the Library and Archives division took first steps to lead the institution-wide adoption of an audio-visual digital asset management system (DAMS). The Art Information Commons will build on this service to create an even more extensive data service. The Library and Archives division will be responsible for building consistency of data across the entirety of Museum holdings (so systems work effectively), ensuring accountability (to effectively support all other departments), and facilitating consensus among different subject experts to support their user communities. This information management function acts as a complement to the technical functions, but is content/data-facing, because the need for rich, integrated, digital content grows daily. We are no longer in a world where our data is managed internally, and only tombstone data and label text is available to others. Everything has value within and outside the institution.

Demonstration Projects Using Linked Data

With leadership from the Library and Archives division, in early 2018, the Museum released its first online scholarly collections catalogue, *The John G. Johnson Collection: A History and Selected Works*. Among its several functions, this digital catalogue links information on paintings in this collection with archival materials that have hitherto not been accessible to scholars and individuals accessing the Museum's collection online. The Johnson project resulted in a new automated content management workflow for publishing scholarly essays, new processes for displaying IIIF-compliant images, and a new Elasticsearch repository. Over 6,000 archival objects were digitized, described, and mapped to JSON-LD based on the Linked.Art approach. The Johnson publication makes explicit the relationships between art and archival records by enhancing website object entry pages and integrating them with primary source documents, such as object provenance records that are linked to relevant correspondence. The process of creating this publication also revealed the challenges with exporting scholarly collection data from TMS. This is due to a lack of consistent style guides between departments—such as differences between curatorial, provenance, and publications standards—as well as formatting limitations within the system. TMS must be capable of exporting consistent scholarly data in order to make enhanced object entry pages a scalable endeavor for both digital publications and the website. The Johnson digital publication demonstrated that better information repository standards, management, and tools are needed to advance the Museum's digital publishing goals.

The Library and Archives division is taking a similar but more sophisticated approach to integrating archival materials in the Duchamp Research Portal, to be launched in 2020. The Portal will provide unified access to 65,000 archival records from the Duchamp collections of the

Philadelphia Museum of Art, Centre Pompidou, and Association Marcel Duchamp. The proposed Mellon planning project will be informed by the Museum's experience of linking the art and archival collections through the Duchamp Research Portal. As a demonstration project, it will help the Library and Archives develop requirements for managing integration and standards across all of the Museum's holdings in ways that also increase flexibility and access. The Museum is also learning through Design for Context's involvement as the John G. Johnson Collection and Duchamp Research Portal advisor and developer. Design for Context is also the consultant for the American Art Collaborative (AAC) and their involvement has conveyed some of the lessons, models, and tools gained from working with the AAC. Design for Context is proposed as the lead consultant to facilitate the proposed Art Information Commons planning process.

Relationship to the Field

Like many other art museums, the Philadelphia Museum of Art's cultural data was created in isolated systems, managed by different departments, and followed disparate standards with minimal authority control. Most of these institutions are constrained by a commitment to proprietary databases that limit their capacity to connect and publish the data held in them. Fortunately, new technologies are providing opportunities to expand research possibilities through enhanced access, description, linking, and sharing across databases through the aid of linked open data. Linked open data addresses internal and external needs for cultural heritage information—supporting data integration and application-building internally, as well as opportunities for collaboration and sharing across communities. Major precedents for the use of linked data in the cultural heritage field include:

- The Canadian Heritage Information Network, which published data on 85,000 items from eight museums.
- The American Art Collaborative (AAC), which tested and demonstrated the value of linked open data among fourteen institutions (the Smithsonian Archives of American Art and thirteen art museums; the Philadelphia Museum of Art is not a founding member).
- The Yale Center for British Art, which established a SPARQL endpoint for its linked data and created a user interface for TMS to improve curatorial workflow.
- The Carnegie Museum of Art's Art Tracks project, which established a model for structured data on provenance.
- The Georgia O'Keeffe Museum, which is building on linked data models to create an internal resource to integrate their collection, archives, and library materials, which are currently managed and described in disparate databases.
- The Pharos project, a consortium of fourteen European and North American art historical photo archives that is mapping their collections information to linked data in order to publish information about their collections to the British Museum's ResearchSpace.
- Europeana, which published data on seventeen million objects from 1,500 cultural heritage institutions.

- Linked Data 4 Libraries, a six-library collaboration that is focused on the transformation of library workflows and extension of BIBFRAME.
- Linked.Art, a community developing target models for linked open data about art based on CIDOC-CRM as the core ontology, influenced by the work done by AAC.
- The Mellon "Space" projects, in particular ResearchSpace and ConservationSpace, which use a linked data architecture as a core of their technical approach.
- Getty Research Institute and the Library of Congress have published their controlled vocabularies as linked data, making the alignment between vocabulary standards more feasible for museum, library, and archives data.

In sum, these efforts indicate progress toward the long-held aspiration for collections data-sharing across the cultural heritage field. The proposed Art Information Commons will be an important source of input for these projects, allowing the community to accelerate adoption and use.

Global integration of data about the various kinds of collections held within a single museum—objects and artifacts as well as archival records, library holdings, conservation, curatorial records, and interpretive materials—is where the field is moving, but no major, encyclopedic art museum has yet attempted this goal. More importantly, this project is the only one that is standards-based and open, incorporating a linked data approach. In other cases, integration is being achieved through the creation of APIs (Application Programming Interfaces) that may not present data in open standard formats or connect extensively with existing external vocabulary authorities such as the Getty Vocabularies or the Library of Congress Subject Headings.

The Georgia O'Keeffe Museum, under an IMLS grant, is working on integrating its art, archival, and library holdings; however, it is a mid-sized institution with a highly focused collection. The Smithsonian Institution is preparing to implement linked data within its individual, smaller museums, and is only beginning to consider a central approach. The Getty Research Institute is the only project working toward linked open data at the same scale as the Museum's proposed project. No art museum has assigned this responsibility to its Library and Archives division. And not all of those projects have the intention of identifying how their work becomes a model for other institutions interested in undertaking the same change processes.

Ultimately, the Art Information Commons will contribute new tools and strategies that will enable modestly funded and resourced cultural heritage institutions, and especially those outside the academic realm, to embrace linked data. This project will result in models in two senses of the word: First, a model focused on data cataloging standards and organizational change management will be adaptable for use by other institutions. Many of the working practices and considerations for how to describe holdings that cross department boundaries will be very useful to guide other institutions in considering guidelines and governance for holistic data management. Second, data models and technical code may be shareable, with one caveat: Even though TMS and ArchivesSpace are widely used by museums, they are often customized extensively by individual museums. This means that extracting data from them will

require case-by-case adjustments and local interpretation. Cataloging standards can also vary widely by institution. Source code, data models, and documentation will be made accessible and shareable through a Museum-affiliated platform such as GitHub. Conference presentations, journal publications, and Museum programs will be used to help disseminate the project to other libraries, archives, and museums.

B. Expected Outcomes

The ultimate outcomes of this work are to increase efficiency in data management, increase access to the full range of art information resources stewarded by the Museum for both internal and external researchers, and to lay the groundwork for sharing collections data with other cultural institutions adopting a linked data approach.

The planning for the Art Information Commons will be led by the Library and Archives division, but will require the involvement of multiple departments within the Museum. The structure for such an effort represents a logical development of the work of the Art Information Task Force, which was formed in 2013 and is co-chaired by the director of the Library and Archives and the director of Information and Interpretive Technologies (IIT). This task force oversees two subcommittees for digital collections initiatives: one for collection interpretation, called the Collection Presentation and Interpretation Group (CPIG), and the other for collection information management, the Collection Resources and Information Group (CRAIG).

Through the Art Information Task Force, in consultation with outside expertise, staff will create an art information content strategy establishing greater consistency, vocabulary standards, richer descriptive structured data, workflows, and the ability to adopt and establish linked data resources throughout the institution and across data collection types: curatorial, conservation, library, and archives. Responsibility for content creation and data ownership will continue to be distributed across the institution, but with new governance, direction, and training to enable staff to begin to develop and implement linked data models.

This proposal is focused on art information. However, with the scale and cross departmental nature of the project objectives, it is anticipated that this work will also inform and encourage holistic information and systems approaches for non-collection-based data as well. For example, during this time the Museum's IIT division will explore the integration of constituent information systems—such as admissions data, program calendar, and membership—in order to better serve and understand the Museum's various audiences.

Goals of Planning Phase

- Revise the mission and composition of the cross-departmental Art Information Task Force (AITF) and its two subcommittees.
- Working with lead consultant Design for Context as facilitator, conduct a comprehensive analysis and documentation of the Museum's current collections data ecosystem, encompassing: content creation workflows, standards, vocabularies, end-user needs, and systems infrastructure, including capabilities to export scholarly content.
- Make recommendations to enhance digital workflows and sharing and creation of scholarly content.
- Create an art information content strategy, which includes a holistic strategic plan for the creation, delivery, and governance of useful, usable content—from cataloging to web publication and encompassing all types of content, whether images, data, metadata, or other media—aligning internal workflows with the needs of different users and audiences. Some strategies for creating meaningful and useful data are consistent formatting, referencing external vocabularies, and delivering the data based on user experience needs.
- Series of cross-departmental training opportunities to build understanding and involvement across the Museum's staff in the potential benefits and emerging best practices related to linked open data, the integration of collections information systems, metadata creation and management, and digital publishing. This internal change process will be accomplished through an ongoing series of cross-departmental training opportunities and staff dialogue with advisors.
- Quarterly public programs (lectures, workshops, or other educational formats).
- Develop vocabularies and standardize terminologies
- Develop infrastructure for prototyping the exposure of collection management systems contents as linked data web resources.
- Evaluate the current array of tools available to art museums, archives, and libraries adopting a linked data approach to their collections and investigate development of lightweight extensions.
- Develop and refine a proof of concept for data integration

New Positions

The Museum will establish three permanent positions within the Library and Archives division to support project implementation:

- Project Manager to provide day-to-day project coordination, including project documentation, metadata creation, working with consultants, supporting internal and public awareness, and organizing internal and external meetings.
- Systems Developer to configure and support information-based systems; integrate existing enterprise-wide applications
- Taxonomist to develop a standards-based taxonomy that uses external and internal controlled vocabularies

Consultant
The Museum will contract with Design for Context, LLC, (DfC) the leading consultancy firm for linked data projects.

Advisory Board
The Advisory Board will meet annually in Philadelphia and consult as needed throughout the project.

Advisory Board members include:
- Emmanuelle Delmas-Glass, Collections Data Manager, Yale Center for British Art
- Elizabeth Ehrnst, Head of Research Collections and Services, Georgia O'Keeffe Museum
- Mark Matienzo, Collaboration & Interoperability Architect, Digital Library Systems and Services, Stanford University
- David Newbury, Enterprise Software & Data Architect, J. Paul Getty Trust
- Rob Sanderson, Semantic Architect, J. Paul Getty Trust
- Nik Honeysett, Chief Executive Officer, Balboa Park Online Collaborative

D. Schedule of Major Activities

Year One (October 2018 - August 2019)
Quarter One (October- December)
- Hire Project Manager.
- Establish communication channels, administrative workflows, and schedules.
- Develop and launch internal and external communication strategy.

Quarter Two (January - March)
- Hire remaining project staff.
- Convene first Advisory Board meeting.
- Map out public program and training topics/formats/speakers; potential topics include vocabulary standards, provenance, collection management, linked data, and digital publishing.
- Survey internal and external audiences to see how they use or would like to use Museum data.
- Work with internal stakeholders to begin analyzing and documenting current data practices, processes, and systems.
- Announce and promote program internally and externally (ongoing).
- Host quarterly public program and institution-wide training (ongoing).

Quarter Three (April - June)
- Begin creating strategic plan for data architecture, with a focus on how to reconcile internal and external standards and staff workflows.
- Evaluate audience research to inform target data model work.

- Work with findings from user research to begin to establish target data models that integrate data across subjects and Museum departments.

Quarter Four (July - September)
- Finish development of target data models.
- Begin mapping and planning for how existing Museum data will align with target data model.
- Begin to develop systems model for a linked data infrastructure.
- Begin evaluating available linked data tools, including hardware, software, and cloud services such as triplestore hosting.

Year Two (October 2019 - September 2020)

Quarter One (October - December)
- Create interim report on recommendations for data architecture, workflows, and infrastructure to guide proof of concept development based on work and research completed to date in order to receive feedback from the Advisory Board.

Quarter Two (January - March)
- Hold second annual Advisory Board meeting to assess progress against key objectives.
- Solidify roadmap and goals for Year Two.
- Develop infrastructure and test environments for the exposure of collection management systems contents as linked data web resources.

Quarter Three (April - June)
- Begin active development of the proof of concept, including lightweight extensions for existing systems.

Quarter Four (July - September)
- Continue development of the proof of concept, including lightweight extensions for existing systems.

Year Three (October 2020 - September 2021)

Quarter One (October - December)
- Compare feedback from proof of concept to previous research survey on Museum data user needs.

Quarter Two (January - March)
- Hold third annual Advisory Board meeting.
- Solidify roadmap and goals for Year Three.

Quarters Three/Four (April - September)

- Develop and refine a proof of concept for data integration, including detailed recommendations for implementation, costs, staffing requirements, target and crosswalk data models.
- Create detailed plan for implementation including development roadmap and vision for the future.