

# White Paper

Report ID: 115031

Application Number: PW-228261-15

Project Director: John Renaud

Institution: University of California, Irvine

Reporting Period: 5/1/2015-4/30/2016

Report Due: 7/31/2016

Date Submitted: 7/21/2016

White Paper

PW-228261-15

Piloting Linked Open Data for Artists' Books

White Paper Lead Author: Laura J. Smart

Project Director: John P. Renaud

University of California, Irvine

7/22/2016

## “Piloting Linked Open Data on Artists’ Books: a Case Study in Interoperability and Sustainability”

### **Abstract:**

Artists’ books are a common component of many art libraries, and are of great interest to artists and art historians because of their highly visual, interactive and sculptural qualities. However, many of these art-like qualities remain under-described when only represented in the typical library catalog. UCI Libraries completed a National Endowment for the Humanities - Humanities Collections Research Resources Foundations grant to extend interoperability and discoverability of artists’ books through the use of linked open data (LOD). We implemented processes of transforming legacy metadata from our Library catalog to linked open data while enhancing records with Visual Resources Association (VRA) Core elements. In addition to publishing linked open data with digital surrogates of artists’ books in our special collections, we built a prototype visualization tool to allow researchers to traverse relationships within and between the works, discovering connections between artists, genres, techniques, and materials. This article will describe the behind the scenes processes and challenges in making project interoperable with an emphasis on the metadata aspect of the project, and offer ways to sustain the project’s growth, through the recommendations and toolkits gathered from the Council on Library and Information Resources (CLIR), Digital Library Federation (DLF), and Ithaka S+R. The issues of interoperability and sustainability are huge issues to digital humanities’ continued growth; once applied to art information, digital humanities projects are even more likely to suffer complications related to these issues. This article will describe a digital humanities project using art information that directly discusses the lessons learned and recommended resources for tackling these issues head on.

### **Introduction:**

Artists’ books are well known in the art information community to defy easy categorization. According to scholar, critic, and book artist Johanna Drucker, an artists’ book "interrogates the conceptual or material form of the book as part of its intention, thematic interests, or production activities." This artistic intervention can include "fine printing, independent publishing, the craft tradition of books, conceptual art, painting and other traditional arts, politically motivated art activity and activist production, performance of both traditional and experimental varieties, concrete poetry, experimental music, computer and electronic arts, and last but not least, the tradition of the illustrated book, the *livre d’artiste*."<sup>1</sup>

Many artists’ book collections are held in libraries, in both art libraries and special collections libraries. As an item in the library’s collection, artists’ books typically are described in the library’s catalog using standard bibliographic description. A standard description includes title, author, imprint, place of publication, year, notes, subjects, and genres, but doesn’t

---

<sup>1</sup> Johanna Drucker, *The Century of Artists’ Books* (New York: Granary books, 1999): 3, 2.

necessarily convey the complex and nuanced meaning and associations triggered by interaction with the object, or even give much of a sense of what the artists' book looks like. Cataloging descriptive standards from the Rare Books and Manuscripts Section (RBMS) of the Association of College and Research Libraries (ACRL) help a great deal by providing more detailed and nuanced descriptions, but are more favorable towards rare book objects than art objects. Furthermore, they do not have subject terms that indicate more nuanced concepts that artists grapple with, through the interplay between form and content.

Many institutions that house artists' books have published images of them in their respective repositories, ensuring visual access to the collection, which is highly crucial not only for the types of objects being portrayed, but also the user community most likely to use the objects, i.e., visual artists and art historians. A number of artists' book collections have been at least partially photographed and enhanced with VRA Core and published in digital asset environments, such as the Joan Flasch Artists' Book Collection of School of the Art Institute of Chicago, or Reed College's digital library collection of artists' books. Others have gone even further in developing research tools: Johanna Drucker has created Artists' Books Online, a portal that describes and illustrates an impressive number of artists' books; Book Artists Unbound from University of Miami, which enhances artists' books records through Encoded Archival Collections – Corporate bodies, Persons, and Families (EAC-CPF); and Artists' Books DC, a tool that geolocates collections in the Washington, D.C. area. UCI's project pushed these advances further to develop workflows and tool recommendations so that other libraries may implement these collection enhancements. Our pilot has valuable takeaways for a variety of stakeholders, from visual resource professionals to art librarians, catalogers, and those who work in special collections, and digital scholarship centers.

## **Linked Data**

As Christian Bizer, Tom Heath, and Tim Berners-Lee state,

The adoption of the Linked Data best practices has led to the extension of the Web with a global data space connecting data from diverse domains such as people, companies, books, scientific publications, films, music, television and radio programmes, genes, proteins, drugs and clinical trials, online communities, statistical and scientific data, and reviews.<sup>2</sup>

Data is linked through the use of Uniform Resource Indicators (URI's) and serialized in a Resource Framework Description (RDF) format. A number of authority files have published their data as linked data, including Library of Congress Name Authority File (LCNAF), Library of Congress Subject Headings (LCSH), and Getty Authorities, including Art & Architecture Thesaurus (AAT), Thesaurus of Geographic Names (TGN), and Union List of Artist Names

---

<sup>2</sup> Christian Bizer, Tom Heath, and Tim Berners-Lee (2009) [Linked Data - The Story So Far](http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf). International Journal on Semantic Web and Information Systems, (Vol. 5(3), Pages 1-22. DOI: 10.4018/ijswis.2009081901): 2. Accessed as a preprint here: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>:

(ULAN). UCI Libraries staff believed linked data could be used in a browse/discovery layer to expose the complex and nuanced meanings hidden in the existing metadata.

Linked data has the promise to harness authoritative metadata in real time to dynamically display related data on a particular topic. Linked open data, has all the qualities of linked data while also adhering to the qualities of openness - available for others to use with the highest amount possible of technical ease with the clearest license available. Staff across a wide variety of departments at University of California, Irvine Libraries wanted to try using linked data for themselves, using a known issue in the art information community: the impossibility of conveying artists' books meaning through a typical library catalog.

### **Scope of the project**

UCI Libraries was awarded a Humanities Collections Research Resources Foundations from the National Endowment for the Humanities grant to pursue Piloting LInked Open Data on Artists' Books (PLODAB). We identified the following high level procedures as necessary for our grant work:

- Choose books for pilot based on differences and similarities; take pictures of books; upload to digital asset management system (DAMS)
- Identify systems needed for DAMS, application, and interface layers
- Extract, enhance, and transform legacy metadata from library's catalog to a form of LOD that visualization tool can use
- Build a visualization tool using linked data to display the relationships between books
- Test visualization tool on user community and gather audience for project launch and next steps

To focus on concepts of interoperability and sustainability, this article will focus the metadata and user testing aspects of our work.

### **Identifying processes for transforming metadata to VRA linked data**

We had three guiding principles for designing the workflow for creating linked open data (LOD). Our metadata should be interoperable with metadata from other artists' books collections and with linked data practices used within the visual arts community. Our processes should be scalable, meaning they could be used for very large sets of metadata beyond the small sample of 41 records we used for the pilot. They should be extensible, meaning others could add-on processes to further automate workflow or metadata enhancement.

We designed our processes based on several assumptions. We believed other libraries would have their artists' books metadata encoded in MARC and stored in a relational database or integrated library system (ILS). We expected that the people applying our process may have varying levels of technical training. And, we anticipated that interoperability of our metadata would be based on our selection of VRA as a descriptive standard. We presumed the community

would trust our evaluation and choice of standard and guessed that many others were already using VRA.<sup>3</sup> Finally, we assumed we would only be able to advocate for a high level generic process due to a wide amount of local variation in applying descriptive standards. We identified basic areas of transformation with associated tasks. The ordering of phases is malleable. Local context determines what steps to perform in which software tool. There are a variety of ways of making the resulting linked data available. How one does the work depends on which choices are made. One may be able to do the transformation using the database in which the legacy metadata resides. Or, one might have to use a variety of tools within a complex systems architecture. All steps should be automated wherever possible to meet the guiding principles of extensibility and scalability and minimize potential for introducing variability and error.

Table 1: Tasks for translating legacy metadata to linked data

Functional requirements	Create list of desiderata for features of transformed linked metadata Select descriptive standard and serializations based on desiderata Select tools for storing and exposing metadata
Analysis and cleaning	Review existing metadata for anomalies Fix inconsistencies (normalize dates, correct spelling errors, etc.)
Mapping	Review and select and adapt crosswalks between source descriptive standard and target descriptive standard or create new crosswalk if required
Translation	Convert legacy description to new descriptive standard Convert legacy encoding to new encoding
Enhancement	Identify appropriate linked data vocabularies Match data to associated URIs
Testing	Try out process for translation and/or enhancement Revise if required

We had a complex systems architecture and specific linked data vocabulary requirements which impacted our selection of tools and the design of our work process. Our legacy Resource Description and Access (RDA)/Anglo-American Cataloging Rules (AACR2) Machine Readable Cataloging (MARC) metadata was managed in our Millennium ILS. Our works were stored in a Nuxeo digital asset management system hosted by California Digital Library (CDL), the University of California-wide centralized service platform for digital services and tools. And, our user front end was a bespoke visualization tool requiring our data to be provided as a static RDF-

<sup>3</sup> We created a rubric to analyze RDA, CDWA, and CCO encoded in MODS as potential alternatives to using VRA. There are documented questions for considering context of use when choosing a descriptive metadata standard. See for example, Chapter 10 in Miller, Stephen J. (2011) *Metadata for digital collections: a how-to-do-it manual*. New York, Neal-Schuman Publishers.

XML file. We needed to ensure our linked data utilized the Library of Congress and Getty vocabularies predominately used by the visual arts community. Our choices were also affected by the technical knowledge of our staff. Our subject expert had little training on working with MARC metadata within the ILS.

The choice of Nuxeo impacted the mapping because it introduced the need for multiple maps. CDL uses the Digital Public Library of America (DPLA) application profile of Dublin Core (DC) for metadata ingest and storage in Nuxeo. The metadata team therefore had to map to that standard in addition to ensuring the metadata could be exported to VRA for the ultimate creation of linked data in RDF. Fortunately, the team didn't have to work from scratch. There are several metadata crosswalks available from other institutions, including Library of Congress. The team reviewed these crosswalks and made adaptations based on our local context.

We decided to work with our metadata outside of the ILS environment. Millennium doesn't have internal conversion tools. It would be easier for staff to manipulate the metadata in spreadsheets due to familiarity with that software. And, Nuxeo required data to be ingested in delimited form as did the tools we were considering for cleaning and enhancement.

Our final process went roughly as follows:

1. Create functional requirements which specify the content and encoding of the metadata you need and how it is to be handled
2. Export source (MARC) records from ILS
3. Perform initial MARC enhancement. We used MarcEdit Linked Data tool to automatically add URIs from the Library of Congress vocabularies to the source MARC file.
4. Convert MARC to delimited form. We used MarcEdit Export Tab Delimited tool
5. Cleaning and enhancement of MARC-ordered delimited data to create "canonical" source file
  - a. Clean data in spreadsheet form. We tested OpenRefine and considered Excel macros but found it more efficient to do the clean-up manually due to the small size of our pilot data set.
  - b. Reconcile the data against the Getty vocabularies<sup>4</sup>
    - i. We tested OpenRefine SPARQL reconciliation tool. It was unsuccessful due to lack of matches within the source metadata
    - ii. We manually added Getty URIs to our spreadsheet
6. Create metadata map for converting descriptive content (RDA/AACR2 data to VRA and DC)
7. Ingest "canonical" tab delimited file to Nuxeo and convert description information to Dublin core
  - a. Give metadata map and file to CDL to program batch loading script
  - b. Test load
  - c. Review quality post-ingest
  - d. Revise "canonical" source data and load production version in Nuxeo.

---

<sup>4</sup> The MarcEdit OpenRefine integration tool was not available at that time, which might have accelerated our work

8. Convert description to VRA and RDF for linked data<sup>5</sup>
  - a. Convert tab delimited “canonical” data to comma delimited data
  - b. Run conversion
  - c. Edit XML files. We found the output from the Heidelberg tool to be excellent, but insufficient. It didn’t include enough repeating subject-related elements to match our source data. We also wanted to use some VRA attributes which weren’t included in the tool. We did the hand editing using oXygen XML editor.
9. Publish linked data – in our case, it’s a static RDF-XML file posted on the UCI Libraries github

### *Hindrances to metadata interoperability*

Our experience creating linked data uncovered some of the well documented barriers to making interoperable metadata.<sup>6</sup> For example, our results were unsatisfactory when we tested enhancement with the OpenRefine Getty reconciliation tool. There was little keyword string matching due to synonymy and term granularity differences. Synonymy causes an obvious mismatch in search terms. Granularity problems occur when concepts are in the same classification category but at a broader or narrower level. The AACR2 and RDA descriptive cataloging rules dictate that catalogers describe the work at a level of specificity reflected by the “item in hand” and with literal transcription of some elements from the piece. So, two works may be in our politics collection but use the subject terms war and imprisonment. We might prefer to use the Getty vocabularies at the broader level to facilitate grouping concepts in the browsing interface.

We also had a GIGO problem when we attempted to automate conversion from AACR2/RDA MARC to VRA.<sup>7</sup> We wanted to use the VRA material, technique and cultural context elements. We had difficulty because of semantic issues in our subject terms and hidden information within the records. The Library of Congress Subject Headings (LCSH) used in our source records created prior to 2011 conflated subject and genre terms. In addition, our local subject cataloging practice was to put terms describing technique and material into general subject fields. Cultural context was buried as implicit data within the geographic location of publication.

The GIGO issue was not easily addressed. The subject librarian had to very carefully review MARC subject and imprint fields to manually match concepts with Getty vocabularies if we were to expose the richness of the works with linked data triples. Detailed subject analysis and remediation required her domain expertise in visual arts. This is time consuming original

---

<sup>5</sup> We tested a XSLT provided by Northwestern University but found the source XLST would need too much customization to work for our needs. We tested the VRA Core 4 XML Transform Tool designed by Heidelberg University and generated acceptable baseline RDF. <http://kjc-ws2.kjc.uni-heidelberg.de:8081/exist/apps/csv2xml/index.xq>

<sup>6</sup> The literature on metadata interoperability and the issues in metadata quality which affect interoperability go back to the early days of digital library development. See for example the published works of Thomas Bruce, Diane Hillmann, Naomi Dushay, Sarah Shreeves, and Jenn Riley.

<sup>7</sup> GIGO, or “garbage in, garbage out” is a phrase used in computer and information science to describe how computers can only work with the data as given. If there are problems with input, the output will reflect the same problem.



cataloging work, akin to creating records from scratch. It would not be generalizable to other institutions nor scalable to collections with a large number of legacy records.

*Lessons learned and suggestions for scaling metadata work?*

Our experience produced many lessons and shaped our thinking about what we might do differently to make it easier to work with a large number of records when transforming legacy metadata to linked data.

Lesson 1: Choose tools and partners based on current capacity

Our choice to work with CDL and use Nuxeo as DAMS was based on our long term aspiration to increase participation from other campuses within the UC and to incorporate more collections into a single artists' books resource for access and discovery. Nuxeo met our functional requirements. Yet, CDL had only recently migrated their in-house DAMS and hadn't yet implemented all aspects of the software including features for exporting metadata as linked data. There is risk in selecting metadata management systems based on un-developed or as-yet-to-be implemented features. The delay in developing export features caused us to create our own static linked data. This resulted in silos of data without an easy means to keep those silos synchronized. It would be advisable to create Memorandums of Understanding for expected deliverables and deadlines when working with external partners for development.

Lesson 2: Prioritize your functional requirements in detail.

We generally considered features to be required or desired without any detailed ranking because it was difficult for us to reach consensus and we had a limited amount of project time. It would have helped to spend time prioritizing features up-front. We would have saved time doing serialization conversions if we had ranked metadata transfer features higher and selected our systems accordingly. For example, if we could have imported raw MARC data into our DAMS and let the system make it Dublin Core, we would have had less work in configuring our delimited data.

Lesson 3: Follow best practices for data migration

We are aware that this is a well-known best practice to review, clean, and enhance metadata prior to transforming its content or structure during the migration process. We made the choice to work with the metadata in spreadsheets outside of the ILS, however, because some of our team members were not trained to use traditional cataloging rules or systems. It made it easier for the metadata neophytes to do the work but it added complexity to managing different versions of the metadata. We may have saved time by training team members to work with metadata at its source.

Lesson 4: Use version control tools

Our proliferation of working files caused confusion, delay, and extra work. Version control may have helped us avoid that.

Lesson 5: Legacy metadata is messy metadata. Accept semi-automation and good-enough records

There is no fully automated way of converting/migrating metadata without lossiness or error. There are scaling and granularity issues within the legacy metadata as a result of descriptive cataloging rules, semantic interoperability, and local practice and/or cataloger error. There are tools which can handle serialization changes or descriptive standard mapping for thousands of records. Variations in source data are not so consistent as to be covered by algorithms. Thus, software tools or programming scripts can only semi-automate clean-ups and quality control. Humans will need to manage metadata mapping and review quality control prior to ingest and after migration. There may not be additional time, budget, or labor for fixing the complex issues within the source data. If that is the case, then one will need to accept less-than-perfect description within the target application.

## Sustainability

In the midst of working on the grant, other environmental factors made an impact on how we measured its success. The Digital Scholarship Services unit (members of which were the main source of expertise for project management, metadata, and systems for the grant) was charged to develop business cases for implementing digital scholarship services for the UCI campus community, including data science, digital preservation, scholarly communication, and digital humanities. Creative lead on the grant project Emilee Mathews was appointed the digital humanities coordinator and charged to develop the business case for why a project like PLODAB should continue to be supported by the libraries. Using this charge as a springboard, our main questions were:

- If we continue to build the tool, will our campus community use it in research and teaching?
- Will other libraries, archives, and museums find our project and its associated documentation useful? Would they want to partner with us to develop it further?
- What are fruitful sources of additional internal and external funding?

Several publications developed by the library community served us in good stead. “Fit for Purpose: Developing Business Cases for New Services in Research Libraries” lays out the components of a business case tool set to justify the launch of a new service to library administration, and provides guidelines on how to go about doing so.<sup>8</sup> Ithaka S+R Report “Sustaining the Digital Humanities: Host Institution Support beyond the Start-Up Phase” discusses the results of the survey, site visits, and interviews conducted to determine how institutions can prepare themselves to sustain digital humanities project and derives recommendations.<sup>9</sup> Additionally, team members attended workshops by the Ithaka team which

---

<sup>8</sup> “Fit for Purpose: Developing Business Cases for New Services in Research Libraries.” <http://mcpress.media-commons.org/businesscases/>

<sup>9</sup> Nancy L. Maron and Sarah Pickle: “Sustaining the Digital Humanities: Host Institution Support Beyond the Start-up Phase,” June 18, 2014. [http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR\\_Supporting\\_Digital\\_Humanities\\_20140618f.pdf](http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR_Supporting_Digital_Humanities_20140618f.pdf)

were formative in establishing documentation, including “Building a Business Case Toolkit” and “Finding and Keeping an Audience in a Competitive Environment.”<sup>10</sup>

### **Findings from Ithaka and CLIR reports**

Throughout these materials, recurrent themes of institutional priorities, culture, and capacity, robust user community, and scalability of efforts were pointed to as main factors for the establishment or continuation of a proposed service or project.

### **Mission**

A sustainable project needs to add value to the institution’s mission, and to be strategically tied into the institution’s larger priorities. As “Sustaining DH” points out, “the system that will work best for an institution, its faculty and staff, is the one most closely tailored to the goals that institution holds dear.”<sup>11</sup> “Fit for Purpose” is influenced by social entrepreneurship, which emphasizes developing services in accordance to the institution’s mission, and further recommend that the team proposing potential services “examine the library’s mission to determine if it would need to be modified to include a proposed service.”<sup>12</sup>

### **Organizational capacity**

The organization’s ability to engage a project is crucial to the project’s success, both in its initial development and its ongoing support. “Fit for Purpose” recommends that the institution reflect on whether it has “sufficient physical, human, and financial resources available to consider embarking on new initiatives at the present time.”<sup>13</sup> “Sustaining DH” notes that in-kind institutional support is the highest percentage of ongoing support funding.<sup>14</sup> These findings indicate that institutions need to have prepared mechanisms in place, while project teams should have a clear sense of what happens to their project after the initial content has been created and the technical infrastructure developed are completed.

### **Goodness of fit**

Goodness of fit between institution, project, and timing is key to success. “Sustaining DH” outlines several different models of institutional support for DH projects: the service model, the lab model, and the network model. The service model tends to be a centralized unit on campus “where the unit seeks to support faculty and students in their work...” by offering consultations, talks, workshops, and meetups, emphasizing support and education.<sup>15</sup> Unlike the “lifting all boats” method embraced by the service model, lab models have a greater “focus on innovation and project development.”<sup>16</sup> While the lab model excels at developing projects, “even

---

<sup>10</sup> Building a Business Case Toolkit <http://www.sr.ithaka.org/services/workshops/#library-business-planning> And Finding and Keeping an Audience in a Competitive Environment <http://www.sr.ithaka.org/services/workshops/#keep-an-audience>

<sup>11</sup> “Sustaining DH,” 56.

<sup>12</sup> “Fit for Purpose,” 4.

<sup>13</sup> Ibid., 15.

<sup>14</sup> “Sustaining DH,” 19.

<sup>15</sup> Ibid., 23.

<sup>16</sup> Ibid., 31.

the most successful labs cannot absorb the costs of long-term hosting and support.”<sup>17</sup> By contrast to both service and lab models, in a network model of supporting DH work, “there are multiple units whose services have developed over time, in the library and IT departments, but also visualization labs, centers in museums, and instructional technology groups, each of which was formed to meet a specific need.”<sup>18</sup>

Understanding what model your institution fits into, or if there is no model currently in place, what model would fit most organically with institutional culture and current priorities, will help to determine if your institution has the capacity to take on the administrative and overhead costs associated with ensuring your project has a long life span.

### **User base**

The project planning team should conceive of what success looks like, how it is measured, and how to achieve penetration with its audience, particularly in projects based out of the library: “Those focused on library-based digital collections will want to gain a strong sense of who is using the materials, and how, in order to make a strong case to administration for future support.”<sup>19</sup>

“Fit for purpose” recommends talking to potential stakeholders throughout the development process to ensure that the proposed service has a defined audience and will contribute concrete impact. The report talks about value propositions as a way to demonstrate exactly “the intended benefits or value...to reaffirm the rationale for moving ahead.”<sup>20</sup> The Ithaka workshops provide crucial frameworks for developing value propositions and other user-focused brainstorming exercises.<sup>21</sup>

### **Scale Solutions**

In “Sustaining the Digital Humanities,” the authors recommend institutions “figure out how to use scale solutions, without overly limiting the creativity and research aims of project leaders.”<sup>22</sup> However, “Does Every Library need a Digital Humanities Center?” cautions that “Some large-scale projects to create comprehensive technical solutions for DH have demonstrated the danger of de-contextualizing scholarship and producing a homogenizing effect.”<sup>23</sup> Scale solutions also promote interoperability.

### **Applying these recommendations to PLODAB**

---

<sup>17</sup> Ibid., 33.

<sup>18</sup> Ibid., 34.

<sup>19</sup> Ibid., 54.

<sup>20</sup> “Fit for Purpose,” 24.

<sup>21</sup> Ithaka S+R workshops, “Building a Business Case Toolkit” and “Finding and Keeping an Audience in a Competitive Environment.”

<sup>22</sup> “Sustaining DH,” 51.

<sup>23</sup> Jennifer Schaffner and Ricky Erway, “Does Every Research Library Need a Digital Humanities Center?” (Dublin, Ohio: OCLC Research, 2014. <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-digital-humanities-center-2014.pdf>): 8.

Mission: PLODAB augmented the library's mission to advance research and innovation while preserving the cultural record.<sup>24</sup> UCI Libraries University Librarian encourages library staff to pursue external funding opportunities; PLODAB was one larger scale recent grant project.

Organizational capacity: UCI Libraries has staff with expertise in metadata, web services, systems, special collections, and visual arts; the grant allowed us to hire a short-term programmer to develop the visualization tool. The project had a large impact on several project team members' workloads and took up a lot of time from the members of the Digital Scholarship Services team, which hindered their ability to take on other projects. Frankly, the technical complexity paired with the aggressive timeline of the project was a stretch for the team members. While the project definitely had positive effects on the team's skills and abilities, the project was also challenging to incorporate into daily workflow.

Goodness of fit: In considering the various models of digital humanities work being done on campuses that "Sustaining DH" highlights, it became clear that the project was not a perfect fit for UCI's institutional culture. With its emphasis on project development and spurring innovation, the project was more suited towards a lab model for supporting DH; meanwhile our institutional culture would be a more natural fit for the service or network model, with its emphasis on faculty research and success. The project was not initiated out of a particular faculty or other campus group, it was conceived of within the libraries. The end user was considered throughout the project's lifecycle, however none were involved at the outset. So, unless we parlay the research and development we've done to apply directly to faculty and graduate students, or our culture majorly shifts through leadership or money, it is less likely that we will continue the project as currently scope.

User Base: The project team conducted iterative user testing during the prototype building phase of the project, serving three main purposes. During the grant project, we were able to gain early feedback from potential users and incorporate some of their suggestions into the tool, while keeping other suggestions for longer term development. We also learned how faculty anticipated using the tool, and discovered that it was primarily understood as a pedagogical tool. And lastly, we began to build an audience for the continuation of the tool, while identifying the types of content that would be most useful to add in in terms of our faculty's interests. Project members were strategically gathering data on what content would be most useful to add, using a model inspired by the Center for Primary Research and Training (CPRT), a unit of University of California, Los Angeles' Special Collections. CPRT staff have successfully managed a workflow based on faculty interest and use of material to selectively digitize pedagogically crucial materials.

We also considered the library, archive, and museum community to be a key audience for the project. We purposely built materials and tools to not only be interoperable but also adaptable to others' situations. Project team members started to present the project via blog posts and conference presentations in January, 2016. We gained enthusiastic feedback for potential partnerships among a wide variety of institutions.

Scale solutions: the project's use of authoritative metadata, institutional repositories, and open source tools with robust communities all weighed heavily in favor of its scalability.

---

<sup>24</sup> The UCI Libraries drafted a new strategic plan with new vision, mission, and values at the same time that the project was taking place. <http://www.lib.uci.edu/sites/all/files/docs/UCI-Libraries-Strategic-Plan-5-31-2016.pdf>

In sum, in judging the PLODAB project's ultimate sustainability by UCI Libraries and its staff, it has both strengths and weaknesses. Its strengths include its contribution to mission, the interested user base (particularly the library, archives, and museums community), and the use of scale solutions. However, its weaknesses include UCI's organizational capacity to keep the project afloat, the goodness of fit, and the lack of faculty champion for its continued existence. The project's team members believe that the best chance for the project's sustainability is to a.) find a faculty champion who would be willing to take on a lead role in the next phases of the project; and b.) to pursue external partnerships with institutions that can actively contribute to the project.

### **Lessons learned from sustainability**

- Understand your audience
- Understand your institutional readiness and capacity, and ability to provide institutional support at the outset and ongoing. Our project's identity as more of a lab model project than a service model project illuminates some ways that we should have understood our own organizational predilections better
- Understand the workload, staff expertise
- Understand the grant landscape

### **Conclusion**

While the PLODAB project experienced both successes and challenges in its mission to achieve interoperability and sustainability, it greatly enriched team members' experience which has been beneficial to the institution's greater capacity for digital scholarship.

### **References**

Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far." In *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 1-22. DOI: 10.4018/jswis.2009081901.

Drucker, Johanna. *The Century of Artists' Books*. New York: Granary books, 1999.

"Fit for Purpose: Developing Business Cases for New Services in Research Libraries."  
<http://mcpress.media-commons.org/businesscases/>

Maron, Nancy L., and Sarah Pickle. "Sustaining the Digital Humanities: Host Institution Support Beyond the Start-up Phase," June 18, 2014. [http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR\\_Supporting\\_Digital\\_Humanities\\_20140618f.pdf](http://www.sr.ithaka.org/wp-content/uploads/2015/08/SR_Supporting_Digital_Humanities_20140618f.pdf)

Miller, Stephen J. *Metadata for digital collections: a how-to-do-it manual*. New York, Neal-Schuman Publishers, 2014.

Schaffner, Jennifer and Ricky Erway. *Does Every Research Library Need a Digital Humanities Center?* Dublin, Ohio: OCLC Research, 2014.

PW-228261-15: Piloting Linked Open Data for Artists' Books  
White Paper  
UC Irvine Libraries

<http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-digital-humanities-center-2014.pdf>.

University of California, Irvine Libraries, “ National Endowment for the Humanities Final Report: Piloting Linked Open Data on Artists’ Books.” [URL forthcoming]

\_\_\_\_\_. “University of California, Irvine Libraries Strategic Plan,” May 31, 2016.

<http://www.lib.uci.edu/sites/all/files/docs/UCI-Libraries-Strategic-Plan-5-31-2016.pdf>