## Storytelling Across Time:
## Building a Community Around Provenance (B-CAP)

Rutgers University's School of Communication and Information requests a 24-month, **$150,000 Planning grant** to convene experts in library and information science to define, debate, and develop plans to resolve current issues in documenting provenance, addressing NLG program priority **Goal 5, Objective 5.1**. In collaboration with Simmons University, and the University of Connecticut, PIs Drs. Yi-Yun Cheng, Rhiannon Bettivia, and Michael Gryk will address the need to create more integrated approaches to provenance documentation. This will enable stakeholders from many disciplines to avoid duplication of effort, create shared vocabularies and practices, and tackle the institutional, reputational, and financial challenges that arise when sharing the stories of objects in an era of information overload and misinformation.

## 1.   Project Justification

### 1.1  The Current Significant Need
We define provenance as *the story of things:* how something has come to be, what it could have been, and what it will be in the future. This overarching definition encompasses origins and ownership concerns from historical domains, as well as workflows and recipes in the e-science domains. The *something* in question can be anything: physical, digital, or even an abstract concept or idea. Provenance is not innate to objects: telling stories plays a crucial role in whether objects can establish authority in the world. The *Connecting Communities with Libraries, Archives, and Historians through Oral Histories* [LG-252370-OLS-22] project notes that we all tell stories. These stories echo across time, back and forth between the past, present, and future. Workflows are plans for crafting the future; historical provenance attempts to recreate a past beyond the conditions of an object's origins.

**The Gap:** The PIs have a long-term partnership exploring provenance documentation from different LIS perspectives: taxonomy, digital preservation, and bioinformatics. We have collaboratively conducted workshops and panels about provenance challenges and models at the International Digital Curation Conference (IDCC), iConference, iPres, and the Association for Information Science & Technology (ASIS&T). We wrote a book on modeling provenance (Bettivia, Cheng & Gryk, 2022a). Participants in our workshops ranged from video game researchers, archivists, and data curators to reproducibility scholars. The common gap they all expressed is that current models for documenting provenance fall short. They shared challenges distinct to their respective fields. Increasingly, documentation for provenance is digital, even when the materials whose stories are being told are not. Subject specialists bring domain expertise to materials and records, but can struggle with digital infrastructure underpinning provenance documentations standards. One participant expressed that provenance issues had gone unresolved at her archive because she did not know how to speak with the IT department about implementing PREMIS[1]; after the workshop, she felt empowered to work with them on solutions. Other scenarios test what has been considered "good enough" provenance. Well-documented historical metadata might convince one expert that a painting is real, while another will remain unconvinced. Vaccine hesitancy during the COVID-19 pandemic era demonstrated that more data is not always a remedy to doubt and it can even be a hindrance. The provenance of vaccines came into question: how were they developed? Suspicions about vaccine origins influenced some people's decisions not to get vaccinated. Recent work by McDowell [IMLS #RE-250094-OLS-21] with the World Health Organization (WHO) and work being done on health campaigns in public libraries, whose staffs are often at the forefront of crises (Wallace, 2022; Flaherty, 2016), have addressed the public health crisis engendered by mass disinformation, partly by telling stories about how medicines are developed, or the provenance of vaccines.

---

[1] Preservation Metadata and Implementation Strategies (PREMIS): https://www.loc.gov/standards/premis/

**The Need Is Now:** The challenges with provenance need to be addressed now. Addressing the systemic issues will take time, and the current social and technical conditions are exacerbating these provenance issues: computers and AI are increasing the situations in which decisions are made by computers, at speeds that move too fast for human documentation. Vaccination rates are dropping. People are losing trust in the validity of cultural heritage collections. Newspapers, especially smaller, local ones, are closing at record numbers, with back issues and institutional data often winding up in the local library, which has to construct provenance to tell the story of local papers of record[2]. Finding the common threads among these scenarios will result in a community of practice whereby people can leverage experience from many fields to share solutions and work together to solve those wicked problems — the foundational issues that remain unaddressed.

**Solutions:** This planning grant will address immediate solutions by supporting the convening of experts from a variety of fields to share their challenges and their solutions, as well as to start the conversations that will form the foundations of a community of pan-disciplinary scholars who will work together to address provenance gaps and needs. We build on work like that of the WHO and McDowell towards solutions that will address the gaps in provenance practices and documentation. For longer term solutions, we need to develop comprehensive models for provenance that account for the need to persuade and contextualize stories; to collect enough relevant information but not too much; and to communicate across time, with past objects and future users. These models will underpin tools and best practices that will give practitioners the ability to do more efficient and satisfying provenance work: collecting the data, storing it, and telling stories with it.

## 1.2  Target Group and Beneficiaries
This grant brings together experts from a variety of Library and Information Science (LIS) subfields, such as archives, documentation, data curation, metadata, and others, to complement each other's theories and approaches to provenance. Experts will bring theories from their subfield approaches to provenance, which will be complemented by those of experts in other disciplines altogether. Bringing these participants together will be a unique experience afforded by this funding opportunity: the scholars and professionals would not be able to collaborate without these events. By fostering an environment for these diverse teams to interact, we bring into conversation the broadest swath of existing provenance theory and practice to advance new approaches that address foundational challenges that impact all provenance researchers. This project will lay the groundwork for better provenance models and tools, which will have a broad impact on information professionals who collect and curate provenance documentation all over the world. The key here is not to make yet another schema or add more metadata fields, but to address provenance at a foundational level towards developing **smarter solutions**.

We target two immediate groups: (1) practitioners from Galleries, Libraries, Archives, Museums (GLAMs) who are involved in provenance work; and (2) researchers from various disciplines who are working on provenance concerns.  The ultimate beneficiaries are all stakeholders who are intentionally or unintentionally working with provenance documentation. Our project establishes a community of practice to facilitate collaboration and knowledge sharing related to provenance after the grant concludes. Beyond this, our project has a broader impact. Each participant in our events will take this work and disseminate it to their fields, reaching a wider set of audiences. In this way, we establish a lasting impact, which creates awareness on the adoption of best practices for provenance documentation and metadata. The broadest impact of this grant will promote greater transparency and accountability in the management and curation of cultural heritage, digital data, and e-science artifacts.

## 1.3  Relationship to Existing Theory, Scholarship, and Practice
Provenance is a crucial concern in information science. Our ASIS&T panelists discussed provenance concerns with applications in curation, authentication, and long-term preservation (Bettivia et al., 2022b). Art historians use the Getty Provenance, the FBI National Stolen Art File, and auction catalogs to track an artwork's chains of custody [see Provenance

---

[2] https://urbanafreelibrary.org/local-history/collections/news-gazette-archives

Databases in the Supporting Document 1]. Forestry uses provenance to identify seed sources (Breed et al., 2018). In computer science, computational workflow processes and decision-making occur too quickly for humans to manually document (Davidson et al., 2008). There is a lack of collective conversation and scholarship across disciplinary borders. Silos prevent the development of solutions which are both generalizable across disciplines and specific enough to meet individual needs, when different disciplines may offer solutions they have found in their practices. Art history has long experience with the need for persuasion that can be shared with e-sciences; e-sciences have computationally-aided models that could assist archives with provenance documentation backlogs. The current isolation leads to 'band-aid' solutions, which in turn beget more band-aid solutions. The isolation also prevents the various stakeholders from appreciating that edge cases in one domain are commonplace in another.

Historically, provenance concerns itself with what has happened, or retrospective provenance. Provenance work also encompasses what will happen, or prospective provenance, particularly in e-sciences (McPhillips, Bowers, Belhajjame, Ludäscher, 2015). We argue for a third temporal dimension, subjunctive provenance, to document "what could happen" (Bettivia, Cheng & Gryk, 2023). Temporal provenance has primarily been modeled in e-Science research (Bettivia, Cheng & Gryk, 2020, 2021), and our aim is to expand these approaches and share them with other areas of research and practice. Existing models in provenance require considerable labor, and the results are often unsatisfactory. Matthew Gengler from Cleveland Art responded to a survey we conducted about provenance, saying, "But when you have a collection of 40,000 objects, hundreds of years of history, the work is never-ending. Moreover, [the] results of the work are not often satisfactory." Model developers often discard key avenues of research to avoid indeterminacy, for expediency, or to maximize machine interpretability. In so doing, they have effectively stopped a necessary conversation: the ground truths about what provenance is and how it operates have been inadequately explored. This is where our planning events come in, which will lay the foundation for examining the disciplinarity, temporal, and persuasiveness challenges in documenting provenance.

## 2. Project Work Plan

### 2.1 Project Phases and Activities

#### 2.1.1 Phase 1. Pre-Grant and Project Initiation (Current & from August -November 2024)
Pre-Grant Preparation: The PIs have conducted pilot panels at LIS conferences such as ASIS&T, iConference, and iPres to gather a variety of perspectives on the concerns of provenance documentation specific to the LIS field. We received insightful questions regarding provenance in digital preservation, archives, geographic information, and everyday information behavior. These recommendations proved invaluable in helping us plan and refine themes for future events. Of particular note were concerns about supporting competing provenance during ownership disputes for important relics as well as the provenance of missing items - that is the provenance of an item whose origins are known but current location is not.

Project Initiation: During the initiation phase of the project, we will hire a graduate student for administrative tasks. The student will support the PIs in surveying potential venues, arranging event logistics, arranging travel for participants, and updating information on our project website (metaprov.org).

#### 2.1.2 Phase 2. Project Events – Convening Experts (November 2024- May 2026)
The project will facilitate three in-person events and at least four webinars to share expertise on each of the three provenance concerns: disciplinarity, temporal, and persuasiveness concerns. Each in-person event will host five to seven experts from LIS and other domains, and each webinar will feature one to three speakers with open audience participation. We anticipate more than 100 total participants across all grant supported events. The combination of in-person and virtual events is an important feature of this project. The PIs have hosted a series of LIS-specific events on the topic of provenance [see Supporting

Document 1] both online and in-person. From our experience, they accomplish separate objectives that work to complement the objective of establishing a community of practice. The in-person events will last for two days, allowing for sustained and intensive work with participants. In-person meetings engender deeper participation and trust, to create the professional bonds that will form the core of the community of practice. Online webinars allow for broader participation, bringing in additional members who are unable to travel, particularly practitioners. Virtual events also function as bridges between in-person events, reaffirming networks.

In-Person Planning Intensives: Two to three months prior to each event, the PIs will recruit participants via targeted emails. We seek participants in higher education, GLAM, and other sectors. Conversations have already begun with potential participants. They are listed below, and they expressed their excitement to participate in our events and to recommend additional participants.

Before each event, the PIs will ask the participants to share perspectives, provocations, readings, and tools via the project website (metaprov.org). Each two-day event has four parts: (1) presentation and conversation; (2) design fiction activities to speculate, explore, and create solutions; (3) affinity breakout sessions; and (4) recommendations or goal-setting to realign subsequent events. During the presentations, experts will situate their work within the larger concerns of their discipline to help other participants understand the provenance challenges of different fields. The PIs will use the presentations to create design fiction groups, where participants will propose changes to provenance models and tools or suggest new ones using crosswalking, storyboarding, rapid prototyping, or scenario modeling methods (details are listed below). The PIs will draw on their workshop experience to provide scaffolding. Each in-person planning intensive will include shared reflection time leading to interest-related breakout sessions to discuss thematic questions and propose new research avenues. The three PIs will facilitate one breakout group each via a roundtable discussion method. Finally, a representative from each breakout group will present a three-minute lightning talk on the breakout groups' outcomes. Then, the PIs and the participants will make recommendations and engage in goal-setting for future events. The evaluation and dissemination of the project and project outcomes will be highlighted in later sections of this proposal.

Webinars: The first three webinars will be hosted as companions to each in-person event. The fourth webinar will be a project summarization event. Each webinar will be approximately 90 minutes, including 30 minutes of presentation time, and 60 minutes for audience interaction/Q&A/breakout room. We will seek high participation rates, with at least 25 participants in each webinar, about 100 participants in total. The webinars will engage a diverse audience consisting of participants from other events, past workshop attendees, and members of the public who are interested. We will distribute and publicly advertise the webinar event details via LIS-related listservs (e.g. JESSE, Smithsonian, digital curation), professional organizations' listservs (e.g. ASIS&T, ALISE, iConference, ACM), and the departmental listservs at universities.

The webinars will feature professionals from practice, industry, and academia who will share their provenance stories. The PIs will recruit potential webinar speakers from various disciplines such as higher education, GLAMs, and other relevant areas. We will prioritize seeking (1) practitioners with hands-on experiences using provenance documentation, tools, or workflow systems in public libraries, academic libraries, or archives; (2) authors of recently published books, papers, or reports relating to the three themes of provenance; and (3) developers of provenance related standards or models. Paid speakers will share their experiences, feature their books, and discuss workflow tools, relating to the three provenance concerns. Potential webinar speaker lineups include: Jonathan Farbowitz (the Metropolitan Museum of Art), Emily Maemura (Web Archiving), Nicole Martin (Human Rights Watch), Jeff Martin (Kramlich Collection). We will ask participants to recommend additional speakers and pursue connections with those whose work comes up in the course of the events.

Here are the timelines, locations, themes and details of each event:

<u>Event 1:</u>
- ❖ In-Person Planning Intensive #1 (Suggested location: Boston, Massachusetts; November 2024)
- ❖ Webinar #1: March 2025

Provenance Concern: **Disciplinarity**
- How are provenance needs in history, memory institutions, sciences, computer science, commerce, metadata, curation, sociology, arts, agriculture, and others similar and distinct?
- What can we learn from each other?

The first events aim to investigate provocations relating to the inter- and pan- disciplinary concerns of provenance documentation. This planning intensive will feature presentations from speakers from different fields. In particular, each speaker will first discuss their definition(s) of provenance in their respective fields, and their provenance documentation practices. In the design fiction activity, these speakers will be asked to use existing tools we provide to model provenance in their domains, with a crosswalking activity at the end of the session to determine what fell short of these models and tools. Potential in-person speakers represent a diverse range of fields, such as Swati Chattopadhyay (Professor of Art and Architecture, Urban Planning), Donna Cox (National Center for Supercomputing Applications), Francis F. Kilkenny (Research biologist at USDA Forest Service), Radhika Gajjala (Communication and Subaltern Cyberculture) and more.

<u>Event 2:</u>
- ❖ In-Person Planning Intensive #2 (Suggested Location: New Brunswick, New Jersey; June 2025)
- ❖ Webinar #2: October 2025

Provenance Concern: **Time**
- What scenarios do not fit easily into the standard categories of retrospective and prospective provenance?
- How do we classify these scenarios?
- What do you think about subjunctive provenance as one of the temporal dimensions of provenance?:

These events examine temporality concerns with provenance documentation. The in-person planning intensive will showcase talks delivered by individuals from LIS-focused or closely related domains. Specifically, each presenter will be asked to provide examples where their procedures for documenting provenance cannot be easily classified under the conventional groups of retrospective and prospective provenance. In the design fiction activity, participants will be asked to select a scenario from another participant, and create their own storyboard to demonstrate the alternative provenance stories. Possible in-person speakers lineup includes: Alex Chassanoff (Digital preservation), Ann Graf (Graffiti art analysis), Lise Jaillant (Archives), Mary Coffey (Art history) and more.

<u>Event 3</u>
- ❖ In-Person Planning Intensive #3 (Suggested Location: Farmington, Connecticut; November 2025)
- ❖ Webinar #3: March 2026

Provenance Concern: **Persuasion**
- What makes some forms of provenance more trustworthy than others?
- How do financial gain, ownership, insurance and other commerce-related considerations interact with trust?
- How do we center justice and equity regarding competing provenance claims?

The final in-person event explores (1) the trustworthiness of provenance documentation; and (2) how and why some forms of provenance stories are more persuasive than others. Participants will present provenance scenarios from their fields where the trustworthiness of provenance has come into question. In the design fiction activity, we will showcase provenance

documentation scenarios that invoke different facets of trust using the structure created as part of the IMLS funded project on data storytelling [IMLS #RE-250094-OLS-21]. The participants will discuss and rank the criteria they use to judge these scenarios. Possible in-person participants cover a range of researchers and practitioners from LIS and related fields, such as: Alison Abritis (Retraction Watch), Jessie Chin (Health informatics), Martin Kalfatovic (Smithsonian libraries), Sandy Littletree (Indigenous knowledge organization), Kate McDowell (Data storytelling), Craig Robertson (media archaeology) and more.

Project Summarization Event
❖ Webinar #4: May 2026
This final webinar will include a presentation from the PIs to summarize the discussions and provocations derived from the whole of the grant project. It will also serve as an opportunity to reconnect with participants from across the 2-year grant period to discuss the work that needs to be done to continue to nurture the community of practice, including seeking partners and advisory board members for a subsequent research grant. With consent from the presenters, the webinars will be recorded and edited and annotated recordings will be made available on the project website.

**2.1.3  Phase 3. Finalizing Project Deliverables** (ongoing through Spring 2026)
Metaprov.org will house recordings and any artifacts from the grant. The website currently houses artifacts from pre-grant events [see Supporting Document 1] and will be updated throughout the project. The PIs will create an *Info-Graphic Novel*, a highly visual and accessible narrative non-fiction text, to share the project's findings (see link to page 9 for further information). The PIs have prior experience writing and arranging such materials in a short time frame with limited editorial support (Bettivia et al, 2022a).

**2.1.4  Phase 4. Post-Grant** (Summer 2026 onwards)
Following the grant, we will focus on the publication of the works aggregated from us and the participants. We will continue with the curation of the project website and deliverables. To maintain the community of practice, the PIs will utilize the project website to disseminate information regarding public events and consolidate the work and progress updates from community members who participated in the events. Finally, the collaborative endeavors of the project will serve as the foundation for our proposal submission to the IMLS National Leadership Grants (NLG) program for an upcoming cycle. The subsequent research proposal will extend the materials from this planning grant and focus on developing and enhancing a provenance documentation toolkit that specifically tackles the key concerns highlighted in this proposal and throughout the project events. This toolkit will enable researchers and practitioners to more efficiently and effectively document provenance in real-world scenarios.

## 2.2  Project Management

### 2.2.1  The Project Team
The three PIs will collaboratively work on the project design, execution, event design, and dissemination activities.
- PI and Project Lead Dr. Yi-Yun (Jessica) Cheng, an Assistant Professor of Library and Information Science at Rutgers University. Cheng has worked on a NSF-funded project (NSF DBI-1643002) on taxonomy alignment research that resulted in publications in *JASIS&T, Journal of Documentation*, *Knowledge Organization Journal,* etc.; and she was a fellow on the IMLS-funded LEADS-4-NDP program.
- PI Dr. Rhiannon Bettivia is an Assistant Professor of Library and Information Science at Simmons University. Bettivia's grant work includes projects on video game preservation and digital forensics education. Her work has been published in leading information science venues such as the *International Journal of Digital Curation*, the *Journal of Documentation,* etc.

- PI Dr. Michael Gryk is an Associate Professor of Molecular Biology and Biophysics at UCONN Health. Gryk leads a technical research and discovery component of the NMRbox BTRR Center to foster the computational reproducibility and scientific data reuse of bioNMR data. Gryk is also the Associate Director of BioMagResBank, the international repository for bioNMR research data.

### 2.2.2 Other Personnel
- Administrative assistant: A Ph.D. or master's student from the diverse pool at Rutgers University will be recruited to arrange participants' travels as a graduate student assistant.
- Book designer: A designer will be recruited in the second half of the project to assist in the creation of the dissemination materials. The PIs have begun conversations with potential designers.

## 2.3 Evaluation, progress tracking, and perspectives of the target group

We will employ surveys at each event to assess immediate concerns, questions, and feedback. In-person events will entail (1) a brief pre-survey with open questions to collect participants' perspectives on provenance and their expectations for the event; (2) an intermediate survey at the end of the first day asking what participants find interesting, lacking, etc. We will adjust the second day's activities accordingly; (3) and post-event survey, wherein we contact each participant to solicit their feedback, interest in future collaboration, and recommendations for future activity design. This feedback will help us to understand the efficacy of each event, incorporating the perspectives and contributions of the target group.

Specific dates for events will be established by participants and venue availability. The Info-Graphic Novel will be completed with final grant reporting. With our participants, we will co-construct a community of practice, utilizing Educopia's Community Cultivation Guide. The grant will support remuneration for featured speakers to present at webinars, to help foster the formation and validation of the community of practicum, and to enable co-creation of works of shared authorship which will accommodate ongoing external input, validation, and consensus building.

## 2.4 Dissemination of Findings

The planning events serve as the basis for forming a community of practice that brings together researchers and practitioners who work with provenance in a variety of fields and who call it by a variety of names. This community of practice draws from three specific pools of project participants: in-person participants, virtual participants, and audience members.

Our work with this community of practice, including events and the project website, will help to distribute findings and curated deliverables. We will also disseminate findings within professional LIS networks via list serves, professional associations, and conferences. The pan-disciplinary nature of our project will broaden the reach of these dissemination efforts: we will solicit information from participants about the primary avenues of communication and collaboration in their fields and share along these lines. We envisage presenting materials or supporting our project participants to present their work at their own disciplinary events, including conferences such as the Association of Critical Heritage Studies, the International Communication Association, HASTAC, National Women's Studies Association, and others.

Our deliverables enable wide dissemination of the project findings. In addition to working with the community of practice to distribute materials, writing The Info-Graphic Novel will better help us to reach a larger audience: we will make it openly available through our project website and, more sustainably, through institutional repositories that offer open access (Illinois' IDEALS and Rutgers' RUCore). Deliverables will be shared in library catalogs via Rutgers repository, discoverable by search engines, and crawlable by web crawlers. Most deliverables will also be accessible at metaprov.org. The PIs will share materials and findings via listservs and social media, including but not limited to: AERI, AoIR, SAA, DLF, Code4Lib, JESSE, Instagram, LIS alumni Facebook groups, and additional social and professional networks as suggested by the project

participants. PIs will distribute the project efforts via research publications in journals such as JDoc, JASIST, and IJDC and conferences such as iConference, ASIS&T, ICA, SAA, ACHS, and iPres.

## 3. Diversity of Perspectives and Practices

(1) When *recruiting participants*, the PIs will solicit broad representation of cultural identities, including candidates who identify as Black, Indigenous, Latinx, and Asian. We also seek participants from diverse geographic areas. (2) We will financially and infrastructurally *support participation* by paying people to travel as well as to work from home, recognizing the limitations often placed on caregivers, immunocompromised people, and others with mobility challenges. Speaker support for webinars will fairly compensate featured presenters for their labor. (3) By engaging the plurality of participants, we explicitly invite non-dominant *ways of knowing* and persuasive, historical, and future-oriented storytelling. The PIs have already begun conversations with specialists in Latinx art and indigenous weaving.

## 4. Project Results: With this Plan, We Can!

### 4.1 Project Deliverables
The deliverables of this planning grant are structured to support future grant work, which will result in the creation of better provenance tools and models. The collaboration of experts in this grant is a necessary first step to address urgent provenance challenges. The PIs will aggregate and curate findings of the (1) events, resulting in: (2) an infrastructure for a pan-disciplinary community of practice; (3) an Info-Graphic Novel that will be used to disseminate findings and recruit future participants; (4) conference presentations and community reinforcement mini-meetups (CRMMs) to share work; (5) a project website to share findings and promote events; and (6) the foundations of a subsequent grant proposal to deliver provenance solutions. The PIs have a longstanding collaboration working in practice and on research involving provenance, and information gathered from the planning grant will form part of the PIs' collaborative research and practice as they continue their careers in LIS.

### 4.1.1 Events: Convening of Experts
An important deliverable is the collaboration of experts that is necessary to address urgent provenance challenges. The artifacts from the events take on many forms and will be disseminated in multiple ways:
- The PIs will aggregate and curate the findings, resulting in the Info-Graphic Novel and on-going publications and workshops that occur as part of the PIs' practice.
- Events will be recorded; edited recordings will be made available at metaprov.org. To reflect our commitment to culturally responsive open access, we will edit the videos and other digital products to respect the privacy and cultural practices of participants.
- Metaprov.org will showcase images of the crosswalking boards, design fiction, low-fi and high-fi prototypes, and presentation slides.

### 4.1.2 The Creation of Infrastructure to Support a Community of Practice
A community of Practice around provenance is a key result. Many practitioners are doing excellent provenance work and bringing these experts together will maximize the impact of their efforts while minimizing duplication. At virtual events, talks will feature a paid presenter and an opportunity for community members to gather to strengthen ties. We envisage the community providing a means for people from different disciplines to collaborate on solutions to provenance documentation challenges. We will nurture the community through on-going events, website posts, and an email list. We will use Educopia's Field Guide to Community Cultivation to set up foundations for community activities that will outlast the planning grant funding. A community of practice is defined as a group of people who share a common interest around which they build professional connections. Educopia's work serves as a useful resource in ensuring that we can build and grow an

interdisciplinary community of researchers and practitioners that will advance work around provenance. The Field Guide is based on work with communities in adjacent professional practice areas, such as the Coalition to Advance Learning in Libraries, Archives, and Museums; Dodging the Memory Hole; National Digital Stewardship Alliance; and others. The Field Guide also contains specific recommendations on transitioning communities to sustainability after the grant has ended, such as strategically surveying on-going interest, building shared and transparent governance structures, and selectively sunsetting some projects while beginning new ones.

### 4.1.3 The Info-Graphic Novel
The Info-Graphic Novel will consolidate research on provenance often siloed by discipline and disseminate it broadly. Inspired by LG-74-17-0134-17, *Reimagining the User Experience in Archives & Special Collections*, we will create a succinct and highly approachable book of 50-75 pages to share results in a way that is widely accessible across professional and disciplinary boundaries so our findings reach the widest audience possible. We call this deliverable the **Info-Graphic Novel**: a non-fiction text that will use pictures, infographics, and narrative text to explain lessons learned and share data from all the grant's activities. We will take advantage of LIS's multidisciplinary role to create a publication that functions as a central repository of works representing many fields, types of knowledge production, and epistemological perspectives. The PIs will curate lines of conversation between the different events to maximize perspectives and interdisciplinarity because each in-person event will have a discrete set of participants and each webinar will have different featured speakers.

### 4.1.4  Conference Presentations
We have chosen **conferences** that represent the broadest disciplinary edges of the LIS field (see page 7). At three of the conferences (ASIS&T, Society of American Archivists, and the International Communications Association), we will host community reinforcement mini meet-ups (CRMMs) to bring together any event participants present at the conferences to strengthen ties between community of practice members. These informal business meetings will be open to any interested participants and will serve as working sessions to further explore topics from the planning events.

### 4.1.5 Project Website
Metaprov.org currently houses findings and deliverables from previous workshops and the pilot [see Supporting Document 1]. During the grant, it will serve as a clearinghouse to share events and resources from the grant project with domain experts and workshop participants. The website domain has been built for long-term access and maintenance. To maximize access, the PIs will engage in low-cost search engine optimization to increase the findability of the content. This includes enabling crawl accessibility, keyword optimization, and snippet/schema mark-up for key project deliverables. In addition to selective markup, we will provide written descriptive tags for content and automated video transcripts for increased accessibility. The website will provide access to event announcements for the web events; presentation slides and artifacts from events; videos and images; links to any public grant documentation; and the Info-Graphic Novel. We will also maintain a vocabulary and terminology page on the website of terms and a running, annotated bibliography of sources recommended by participants. All content provided via the project website will be under a CC-0 license.

### 4.1.6  Planning for the Future: a Follow-up Research Grant Proposal
The accessibility and reach of the Info-Graphic Novel will also aid in the recruitment of participants for a follow-up research grant proposal to create and refine a provenance documentation toolkit that explicitly addresses key concerns uncovered by this planning project. We propose a toolkit, rather than additional standards, because of our own work in practice: We don't need more standards, but better and easier use of what is there. The toolkit will address foundational concerns in provenance and allow researchers and practitioners to document provenance in real settings in ways that are both more efficient and satisfactory: an audience member at the Digital Library Federation 2023 conference referred to this future model as functioning like a unifying "Dublin Core of Provenance." In the last six months of the planning grant, we will convene online

working group meetings with the community of practice to plan the next steps after the conclusion of the grant. In addition to soliciting feedback on how to sustain the community of practice, we will also discuss how to build on the findings of the planning grant, seeking collaborators for future projects such as another IMLS NLG grant. Upon completing this project, the PIs will submit an NLG grant application for a 36-month project that will answer key questions about provenance, building on prior empirical data gathered in the planning grant to transform practice in libraries, archives, and other institutions.

Finally, the PIs bring a long-standing research and practice collaboration around provenance-related themes, which has resulted in publications, books, workshops, and other engagements. The results of this planning grant will continue to enrich and inform the career trajectories of the PIs for years to come as they continue this work as part of their active practice.

## 4.2  Adaptable, Generalizable, and Usable Deliverables

By convening experts from a variety of fields and institutions, the digital products generated by the project should be legible and applicable to researchers and practitioners in a number of environments and disciplines. Because so many communities will be represented in this project, we anticipate that the findings will be more approachable than disciplinarily siloed works to institutions and communities nationally.

## 4.3  Sustainability Plan

Sustainability is woven into the choice of deliverables for the project. (1) The collaborators have maintained materials from pilot workshops and related works for a period of four years before the grant [see Supporting Document 1]. The PIs will build on this existing infrastructure to maintain digital deliverables, such as text, image, and video content. File formats are selected to privilege small sizes, limiting the need for storage contracts. The materials will also be deposited in institutional repositories associated with the PIs, which further guarantees the persistence of the digital products in the short and medium term.

(2) The creation of the community of practice is the most impactful deliverable. To sustain this community of practice, the PIs will continue to use the project website and will start an email listserv to share information about public events and to aggregate work and updates from community members.

## 5. Conclusion

Based on work with researchers and practitioners, it is clear that provenance work is often both challenging and unsatisfactory in its outcomes. At the same time, we know that another standard or more metadata fields won't fix the many challenges facing provenance makers across the many disciplines and domains that deal with provenance everyday. Smarter provenance solutions are needed to facilitate better, more trustworthy story-telling. In this $150,000 NLG Planning Grant project, we will convene experts from LIS and extended fields to discuss and develop plans of actions surrounding current provenance documentation issues. By conducting three core in-person events and a series of related virtual talks, we will be able to leverage expertise from various disciplines to devise an integrated solution for provenance work.

# Schedule of Completion
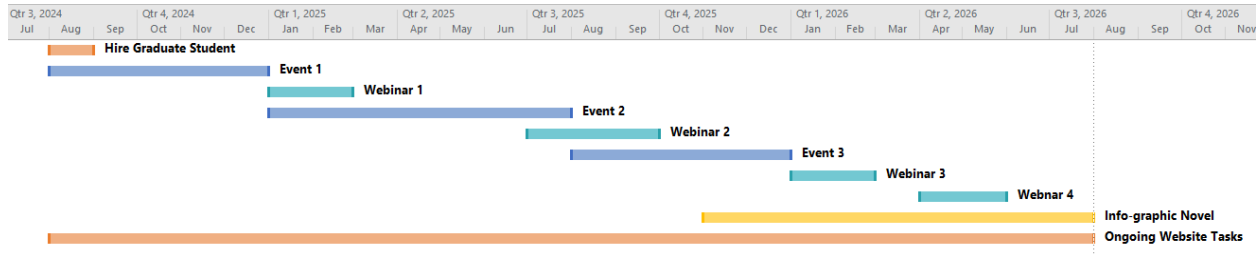


The major activities during this grant proposal can be broken down into four categories. There are the forum, the webinars, general administrative duties, and activities regarding the website, journal and white paper deliverables.

Each of the forums are structured such that there is already a putative list of potential participants. These and others will be recruited during the two months prior to each forum. One month prior to the forum the participants will share their perspectives, provocations and suggested readings. The forums themselves are two-days each. Each of the webinars will similarly have a two-month period of advertising prior to the webinar.

Logistics and administrative work will be under the charge of the graduate student who will be hired within the first few weeks of the grant period. Adding information to the website will be ongoing over the course of the grant as well as building the draft of the white paper which will be published at the end of the grant period. Further details are below.

## Graduate Student Hire

Rutgers is able to recruit and hire a graduate student hourly employee by October 15, 2024. This graduate student will perform administrative tasks associated with the grant including surveying forum venues, arranging for travel and lodging, forum and webinar logistics as well as managing the forum components of the project website.

## In-Person Event 1: Disciplinarity (November 2024)

August 2024 through December 2024. Pre-planning has begun and will continue over the summer. The target date for the forum is mid-November. Documentation will be compiled and uploaded to the website over the following month.

## Virtual Event 1: Disciplinarity (February 2025)

Advertisement for the webinar series will begin in January of 2024, including advertisement of the guest speaker for webinar 1. The webinar will be hosted in February of 2025 and contents uploaded to the website following the webinar.

## In-Person Event 2: Temporality (June 2025)

January 2025 through July 2025. Planning will begin at the start of the year, with adjustments made from our experiences from previous events. The target date for the event is late May / early June, considering the typical end of classes for university participants and summer family plans. Documentation will be compiled and uploaded to the website over the following month.

## Virtual Event 2: Temporality (September 2025)

Advertisement of the guest speaker for webinar 2 will begin in July, 2025. The webinar will be hosted in September of 2025 and contents uploaded to the website following the webinar.

## In-Person Event 3: Persuasiveness (November 2025)

August 2025 through December 2025. Planning will begin subsequent to the previous in-person events, with adjustments made from our experiences. The target date for this gathering is mid-November. Documentation will be compiled and uploaded to the website over the following month.

## Virtual Event 3: Persuasiveness (February 2026)

Advertisement of the guest speaker for webinar 3 will begin in January, 2026. The webinar will be hosted in February of 2026 and contents uploaded to the website following the webinar.

## PI's wrap-up Event (May 2026)

Advertisement of the wrap-up event will begin following the completion of the other planning grant events. This webinar will be hosted in May of 2026 and contents uploaded to the website following the webinar.

## Info-Graphic Novel (October 2025-July 2026)

Creation of Info-Graphic Novel will begin at Phase 3 of this project, starting from October of 2025 to July 2026.

## Website Maintenance and Deliverables

Website maintenance and publishing of web-deliverables will be ongoing throughout the project.

**Storytelling Across Time:**
**Building a Community Around Provenance (B-CAP)**

The Co-PIs share IMLS's commitment to the longevity and sustainability of the digital deliverables of this grant.

• **Type:** This planning grant will generate digital content. These will consist of born-digital products created by individuals, the Co-PIs, and via community gatherings, both online and in-person. The following chart contains an overview of most of the anticipated digital products associated with this project.

| Digital Content Type | Anticipated Quantity | File Format(s) | Additional Notes |
|---|---|---|---|
| Digital Videos | 5-10 videos; 25-50 minutes each | H.264, MP4 (Zoom) | Videos will feature edited content from in-person meetings |
| Zoom Video Transcripts | 4-6 | txt | |
| Digital Photos | 20-50 | jpeg | |
| Metadata records | 20-50 | xml, pyt | Provenance metadata records in PREMIS and PROV |
| Project News Updates | 10-15 | pdf | While such content will be posted to the website, it will be saved in pdf format |

File format selection: Selecting robust formats for digital content is a key step to longevity. In this case, formats were selected primarily for size: smaller files take less storage space, removing the need for paid storage options. Paid storage options require persistent funding, and thus should not be a key component of a digital preservation workflow in this project. Additional formats are determined by usage needs– python and xml for metadata records, for example– and endemic formats– the txt transcript output from Zoom, for example. In all cases, the formats are robustly supported and widely adopted, per the Library of Congress format recommendations.

• **Availability:** We will strive to make as many project deliverables openly accessible as possible. In this instance, we will use two primary venues that are not behind paywalls: the Rutgers institutional repository and the project website, metaprov.org. The vast majority of digital materials created by the project will be housed in these locations, with all reports and white papers, blog posts, video and other A/V content available on the project website and any reports and white papers also available through the Rutgers institutional repository.

Digital products such as metadata records and data models will be made available through the project website in forms that are editable via any text editing software. This means that the public may download the content and analyze and reuse it without charge and without the need for specialized software purchases. A note on the project website will make clear that materials can be manipulated using open source software, while also providing recommendations for additional tools such as the Oxygen XML editor which can be accessed for free via a 30-day trial license.

Limitations in our digital product plan include the fact that certain publications that may be generated through the project may appear in venues that require an institutional subscription for access. Another limitation to access is awareness of the materials: placing materials on a website does not accomplish access if key stakeholders are not aware of the site. We will use our professional networks and social media to make people as widely aware of the materials as possible, however a potential limitation is the fast-changing landscape of social media that may require new approaches in future to reach wide audiences.

• **Access:** We believe in responsibly open access that is culturally sensitive. Most products generated by the project will be licensed under a creative commons CC-0 license, effectively placing materials in the public domain in the US. Exceptions include sensitive information and some publications. For publications that are not open access, we will opt for the least restrictive access measures offered by the publisher, sharing pre-print versions on the project website where possible. To exercise responsible stewardship over potentially sensitive information, we will obtain informed consent from participants to share their materials and views and likenesses.

• **Sustainability:** Written and image materials will be stored in the Rutgers University institutional repository, guaranteeing access for approximately 10 years, considered long term in this case with respect to the estimated use value of the project materials.

The video materials and other project deliverables will be stored in a dedicated GoogleDrive associated with the project website at metaprov.org. The project team created and maintained the website for a year prior to the commencement of the project, and has maintained workshop and other educational digital products related to previous provenance research for more than four years, demonstrating our commitment and ability to maintain materials for the short- and mid-term. We do not anticipate the need to migrate materials within the timeframe of their potential use: formats like jpeg and pdf should ensure format-level access for the next 5-10 years, which is within the projected life span of the project materials. The PIs hope that on-going research will build beyond our deliverables in that time.