

Abstract

Despite the short remaining time to rescue unique cultural and personal materials stored on many 20th- and early 21st-century audiovisual and digital storage media—experts estimate that 20 years remain to rescue unique video on VHS tapes—options for doing so are starkly limited. Building a rescue apparatus in-house requires significant expertise and expense and (since most of these collections are not large) is of limited continuing use. Outsourcing capture is extraordinarily expensive—a single floppy disk can cost upwards of \$100 to image—and therefore well beyond the resources of even well-funded academic libraries and archives, never mind public libraries, small archives, and local historical societies. Moreover, no established method exists to provide audiovisual digitization and digital-capture equipment (without which hands-on experience cannot be gained) to distance (online) students in LIS programs. The likeliest solution is well-documented, portable digitization and capture equipment kits, which might reside at professional touchpoints such as LIS schools and Digital Public Library of America Service Hubs and be shipped where needed.

Dorothea Salo, Faculty Associate at UW-Madison SLIS, will lead a project to build two portable, shippable information-rescue kits: PROUD (Portable Recovery of Unique Data) to rescue data on commonly-held obsolete digital-storage media such as floppy disks and Iomega ZIP disks, and PRAVDA (Portably Reformat A/V to Digital from Analog) to digitize common obsolescing 20th-century audio and video carriers such as VHS and audio cassette tapes. To maximize the nationwide impact of PROUD and PRAVDA, kit construction guidelines and advice, as well as workflow documentation, will be made publicly available for anyone wishing to build similar kits.

The kits will be sourced and built, and initial documentation completed, between December 1 2016 and August 2017. Usability tests and documentation improvements as well as outreach will be completed between August and November 2017.

PROUD and PRAVDA primarily address IMLS's strategic goal of “support[ing] exemplary stewardship of museum and library collections and promot[ing] the use of technology to facilitate discovery of knowledge and cultural heritage.” Secondly, increased online access to unique materials facilitated by the kits this project will create will also “promote museums and libraries as strong community anchors,” and increasing hands-on educational opportunities for LIS students and working professionals will “place the learner at the center.” The kits also hold potential for museums and libraries to help their communities rescue personal and family materials.

Narrative

1. Statement of Need

Dorothea Salo began teaching LIS 668 “Digital Curation” for the School of Library and Information Studies (SLIS) at the University of Wisconsin-Madison in 2010, resolving from the outset that the course should require hands-on service-learning with real-world clients. For several years, she regretfully had to turn down fascinating projects involving culturally-significant material because they involved either digitization of 20th-century audiovisual (A/V) materials or rescue of digital data from obsolete storage media, neither of which SLIS had equipment to do.

In 2015, with design help from Digital Curation students and generous equipment donations from the SLIS, UW-Madison, and city of Madison communities, Salo built RADD (“Recovering Analog and Digital Data”), an apparatus that can digitize several A/V carriers and rescue data from several kinds of obsolete digital storage media. (RADD’s temporary website is at <http://radd.dsalo.info/>; it will move to the SLIS website once SLIS completes its website-platform migration.) RADD is already proving popular with campus and Madison communities, and is making strides toward data-rescue partnerships with small cultural-heritage organizations in Wisconsin.

Problem solved? Not quite. RADD is huge, occupying an entire corner of the SLIS Laboratory Library, and therefore sessile. Neither distance (online) students at SLIS nor most working professionals across Wisconsin can use it to gain experience digitizing A/V and rescuing digital data. Sometimes RADD cannot even rescue collections a mere day’s drive away, given collection owners’ understandable reluctance to let them go off-premises.

Wisconsin unquestionably needs more A/V digitization and digital data-rescue expertise, equipment, and capacity if it is to rescue existing at-risk collections scattered across libraries, archives, museums, historical societies, and institutions of higher education. In this, Wisconsin is a microcosm for the entire United States, where the cultural-heritage sector possesses untold hidden riches on consumer-grade (that is, never designed for long-term preservation) A/V carriers and digital storage media. Oral histories; audio and video recordings of concerts, speeches, and other community events; community radio and television programming and advertising; research data; classroom demonstrations; business and government records trapped on obsolete digital media—just in RADD’s short one-year existence it has rescued examples of all these and more. The clock is ticking loudly for these unique mementos of American history. Many 20th-century consumer-grade A/V carriers will not be playable much longer, even employing heroic measures far too expensive for most cultural-heritage organizations, because of medium deterioration and/or scarcity of playback and digitization equipment. Most 20th- and early

21st-century digital storage media have risked data loss for similar reasons practically since their invention.

Moreover, these collections are often the *most* hidden among hidden American collections. Few are fully cataloged or described in finding aids, unsurprising when (for lack of equipment to access them) the only available descriptive metadata is whatever the creator scrawled on the physical carrier. Even traveling to the holding institution will not help patrons use these collections when, as is often true, the holding organization lacks playback equipment (for A/V) or working drives (for digital storage media). A/V digitization and digital data-capture are the only feasible ways to restore access to these collections before it is too late.

Expecting every cultural-heritage organization with at-risk, inaccessible A/V and digital materials to build a RADD is out of the question. RADD was expensive (despite generous equipment donations) and time-consuming to build; it is finicky to maintain and sometimes tricky to use. Given that many organizations' A/V and digital-media holdings are fairly small and not growing, the return on investment for RADD construction is simply not there. Outsourcing digitization/data-capture is similarly impractical for most such organizations, especially small ones, due to its steep expense and sometimes-dodgy vendors (see supplementary materials, filename "Outsourcing.pdf," for examples of both). American cultural-heritage institutions need a shareable solution that can be built and maintained in a single location but made available for use wherever it is needed.

To attack this dilemma, we propose to build, test, and document two portable, shippable kits as proof of concept. One kit, PROUD ("Portable Recovery of Unique Data"), will contain all necessary equipment to image and capture data from common 20th- and 21st-century digital storage media:

- 3.5 and 5.25-inch floppy disks
- 100 and 250MB Iomega ZIP disks
- several kinds of hard drives
- any other removable-media drive with a USB connection, if a kit user happens to have such a drive available

PROUD cannot make available heroic digital-forensics measures to rescue data from already-deteriorated media. (These typically require entire cleanroom laboratories, and are partly why professional data-rescue is hideously expensive.) Aside from that limitation, for the media it covers it is the equal of RADD and of similar constructions at (for example) the Maryland Institute for Technology in the Humanities. (RADD does handle a few more types of digital storage media for which drives are unpredictable or impossible to source, such as the IMation SuperDisk and the Iomega Jaz drive. We did not add these to PROUD because we could not be reasonably sure of sourcing working drives within the allotted grant time.)

The other kit, PRAVDA (“Portably Reformat A/V to Digital from Analog”), will contain all necessary equipment to digitize commonly-held at-risk 20th-century A/V carriers:

- audio cassette tapes
- audio microcassette tapes (sometimes used for oral histories)
- VHS videotapes
- Mini-DV videotapes (a common “camcorder” carrier, digital video on tape)

Obviously PRAVDA will not suffice for all A/V carriers in cultural-heritage collections. The above carriers were chosen for their commonness and the feasibility of shipping playback equipment. (U-Matic videotapes, for example, are not-uncommon holdings, but RADD’s U-Matic player weighs 40 pounds or so; shipping it is too expensive to be practical. Similarly, reel-to-reel audio players are typically too bulky to ship.) PRAVDA’s goal is to pluck low-hanging fruit so that cultural-heritage institutions can conserve preservation resources for other holdings while building confidence in their ability to digitize A/V carriers.

We also concede that PRAVDA will sometimes yield lower-quality results than best-practices documentation on A/V digitization prefer. (RADD, for example, added a time-base corrector for analog video fairly quickly to enable successful digitization of otherwise-unreadable VHS and Betamax tapes. Most time-base correctors are sized for rack-mounting in a professional broadcast center, so they are too big to ship.) Unfortunately, *many hidden A/V collections are out of waiting time*. An imperfect capture via PRAVDA and its imitators improves *immensely* on information loss from waiting too long for a perfect capture!

(We also anticipate that some of the same organizations willing to build and ship a kit will build an in-house RADD-like apparatus to handle additional A/V carriers and digital storage media. We certainly plan to use RADD to backstop PROUD and PRAVDA. Likewise, the configurations we have chosen for PROUD and PRAVDA are certainly not the only possible ones!)

It is vital that PROUD and PRAVDA be completely self-contained, not dependent on any equipment or software being present at its users’ sites. Not only is it likely that needed equipment or software will *not* be available, asking users to install and test software can be difficult and error-prone, and may even fall afoul of software license terms (e.g. for the proprietary software accompanying the video-capture dongle). These potentially deal-breaking barriers disappear with wholly self-contained kits.

Documentation for PROUD and PRAVDA will be of three types:

- Instructions for unpacking, setting up, and re-packing PROUD and PRAVDA

- Simple step-by-step “how to digitize/capture” instructions, written and short-video, for each carrier PROUD and PRAVDA handle, taking into account standards/best practices for resulting data quality
- Hardware/software lists, sourcing and cost information, and other advice on building PROUD- and PRAVDA-like kits

We have seen a number of promising “Scannebago”-style ideas around community-based text and image digitization founder on the expense of providing in-person training and support; we obviously do not want PROUD and PRAVDA to founder likewise. We believe that without detailed, careful, compassionately-composed, *well-tested* documentation, many librarians, archivists, curators, volunteers, and others will be unwilling or unable to use PROUD and PRAVDA, or will become a training and support burden for kit owners. Based on watching community members use RADD, as well as the success of Oscar Grady [Public] Library’s personal-archiving Digital Media Conversion Lab (http://www.oscargradylibrary.org/?page_id=6710) in Saukville WI, we do also believe that proper documentation of PROUD and PRAVDA should enable any reasonably-enterprising staff member, volunteer, or patron to make them work without in-person training.

To achieve this goal, we will test each kit and its written and video documentation for usability, both locally at SLIS and with partners across Wisconsin (not incidentally helping to digitize/capture materials from those partners’ collections). We will make what we learn from our usability tests, as well as our documentation itself, publicly available to help other kit constructors document their work; in addition to managing this information on RADD’s portion of SLIS’s website and submitting it to the local institutional repository for longer-term safekeeping, we plan to submit PROUD and PRAVDA user-facing documentation to the Library Workflow Exchange (<http://www.libraryworkflowexchange.org/>) for maximum attention.

2. Impact

PROUD and PRAVDA should hugely expand American cultural heritage organizations’ capacity and willingness to digitize A/V materials and capture data from obsolete digital carriers. In turn, this should greatly facilitate access to currently-hidden A/V and digital-media collections across the country. We anticipate that once this project proves the feasibility of creating, documenting, shipping, and using A/V digitization and data-capture kits, several types of state and regional organizations across the country will build them:

- State archives and libraries
- Digital Public Library of America Service Hubs
- Digitization programs in academic libraries and (especially) library consortia
- LIS schools/iSchools, especially those with distance (online) programs

- LIS schools and professional organizations with continuing-education programs for working professionals

We also anticipate that PROUD and PRAVDA will strengthen personal-archiving outreach and service programs (similar to Oscar Grady Library's) in archives, public libraries, and academic libraries. A public library consortium, for example, could build a PROUD kit for circulation among its branches to reach all its member libraries' patron populations with hands-on opportunities to rescue precious personal and family data from obsolete digital storage media.

The first necessary assessment of PROUD and PRAVDA kits will take place at SLIS, where volunteers from the UW-Madison community will perform usability tests aimed at improving the initial documentation. (As a carrot, we will allow volunteers to test the kits on their own carriers and keep the resulting data.) We will ask each volunteer if they are willing to give us a verbal or written endorsement for PROUD and PRAVDA for use in outreach materials. After a round of documentation editing, we will travel to the [REDACTED] and the [REDACTED] to perform more usability tests on mixed digital and A/V collections they have already approached us to discuss. Should these tests succeed and our partners prove happy with the results, we will have strong evidence that PROUD and PRAVDA represent a feasible approach to A/V and digital-media collection rescue in many types and sizes of cultural-heritage organizations.

The real test will be whether PROUD and PRAVDA inspire additional kit-builders. Because this grant is of very short duration, the best we will likely be able to do is assess interest. We will track and report back to IMLS expressions of interest in building kits at conferences as well as by email and social media. We will also track PROUD/PRAVDA website and documentation access logs and report back (anonymous) data. Obviously we will also track actionable plans to build kits, though (planning and budget cycles being what they are) we doubt we will see more than one or two in-progress or completed kits by the end of the grant term. Finally, we will report back volume and type of material digitized by PROUD and PRAVDA during the course of the grant.

3. Project Design

The project will begin, obviously enough, with sourcing equipment for PROUD and PRAVDA; much of the preparation for this has been done already via this very grant application, and Salo and Bhasin will handle the ordering with help from SLIS administrative staff. We will try to lower costs by soliciting equipment donations from local communities and sourcing equipment from lower-cost outlets such as Goodwill's online auction, which has already been a boon for RADD.

The other necessary start will be writing a position description and interviewing candidates for the Project Assistant (PA) for fall. Bhasin and Salo will jointly work on this;

Bhasin has significant hiring experience. As SLIS students highly covet PAships, we expect no difficulty finding an outstanding hire.

Once the equipment arrives, Salo will assemble the kits into their cases like jigsaw puzzles. (Salo has already done some measuring and testing in advance of this necessity; we anticipate some experimentation to find the best configuration, but little actual difficulty.) To aid in unpacking and repacking, Salo will use color-coded and labeled tape to mark which equipment goes in which place in each case.

Salo, given her experience with RADD, will install the necessary software atop the Windows operating system on both laptops. For PROUD, this will include:

- Device Side Data's FC5025 drivers and capture software (for 5.25-inch floppies)
- Drivers for the 3.5-inch and ZIP drives
- Drivers and software for the WiebeTech ComboDock
- VMWare Player and BitCurator (we anticipate few PROUD users will use BitCurator extensively owing to its excruciatingly poor usability, but we would be remiss not to include it and document its Guymager disk-imaging function), configured to allow drag-and-drop from the BitCurator environment to Windows

For PRAVDA, this will include:

- Drivers for ElGato's video dongle and the FireWire ExpressCard
- Open-source Audacity (with mp3 plugin to create access copies) for audio digitization
- Open-source WinDV for capture of digital video from mini-DV tapes (with freeware CaptureFlux as backup)
- ElGato's software for video digitization (with open-source VirtualDub AIO as backup)
- Open-source HandBrake for video transcoding

We anticipate that the kits will be complete and software installed by the end of February 2017; earlier would not surprise us.

With kits complete and labeled and software installation complete, we will begin kit testing, usability-protocol design, and video documentation (Salo), as well as soliciting local volunteers for usability testing (Bhasin). Building on Oscar Grady Library's success with short how-to screencasts, Salo will start with video documentation of the kit-unpacking and kit-packing processes, and continue by capturing screencasts of exactly how to digitize/capture each kind of carrier for each kit. The goal of this documentation is for PROUD and PRAVDA users to achieve as near archival-quality results as feasible with the fewest decision points possible. For example, we will not ask users to decide which software settings or file formats to use; we will simply inform them, in accordance

with known best practices. We will also document how to make web-ready access copies. These videos should be complete by the end of July 2017.

Once the PA has joined the team in August 2017, the PA will begin written documentation and usability testing. We plan one round of testing, with one to two tests per day, over a week's time within SLIS. After that, the PA will schedule and perform testing with our partners at [REDACTED] and in the [REDACTED] (we are leaving the timeline a little loose here in case documentation revisions after the initial round of tests are more cumbersome than we expect).

To interest likely organizations in building PROUD- and PRAVDA-like kits, Salo will demonstrate PROUD and PRAVDA at the 2017 Wisconsin Library Association annual conference (October), and will try to place a masterclass with the Special Libraries Association 2017 conference in June (not paid for by this grant; Salo is an SLA officer who receives registration and travel reimbursement from SLA). In addition to leveraging SLIS's substantial social-media presence and alumni contacts to spread the word, Bhasin and the PA will also reach out to Wisconsin organizations and publications that have already covered RADD, such as *Isthmus* (Madison weekly newspaper), *Madison Magazine*, *Recollection Wisconsin* (Wisconsin's Digital Public Library of America Service Hub) and the Department of Public Instruction, to cover PROUD and PRAVDA. Additional funding and conference location (not yet announced) permitting, we will also target DPLAFest 2017 for a poster or demo, since we believe DPLA Service Hubs are promising kit-builder candidates.

Salo, Bhasin, and the PA will use the project-management application Trello to forecast and track project progress. As noted on the budget, only the PA's salary, fringe, and tuition will be charged to IMLS; SLIS is contributing Salo and Bhasin's time. Evaluation plans have been described in section 2, but in brief:

- Usability-test results and analysis
- Tracking expressions of interest in kit-building from demos and email contacts
- Counting accesses to PROUD- and PRAVDA-specific documentation and website
- Counting amount and type of material digitized by usability testers and project partners at [REDACTED] and the [REDACTED]

Schedule of Completion

December 2016-January 2017:

- Solicitation of equipment donations
- Purchase and shipping of remaining kit materials
- Writing of Project Assistant job description
- Documentation of kit-sourcing process

January-February 2017:

- Kit construction
- Documentation of kit-construction process
- Software installation and testing
- Usability-test protocol design

February-July 2017:

- Short-video kit documentation
- Interviewing for Project Assistant position

August-September 2017:

- Initial usability tests (internal to SLIS)
- Test analysis
- Solicitation of press and partner coverage
- Video documentation revision
- Paper/PDF kit documentation

October-November 2017:

- Kit demo at Wisconsin Library Association conference (October 17-20)
- Usability tests with [REDACTED], [REDACTED]
- Written and video documentation revision
- Submission of revised documentation to Library Workflow Project
- Report to IMLS

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete ...	Part I
	If your project will create or collect ...	Then you should complete ...
<input type="checkbox"/>	Digital content	Part II
<input type="checkbox"/>	Software (systems, tools, apps, etc.)	Part III
<input type="checkbox"/>	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (<http://us.creativecommons.org>) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and/or collect, the quantities of each type, and format you will use.

A.2 List the equipment, software, and supplies that you will use to create the content or the name of the service provider who will perform the work.

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

B. Digital Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: You may charge the Federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the Federal award. (See 2 CFR 200.461).

C. Metadata

C.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, or preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, or PREMIS) and metadata content (e.g., thesauri).

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

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

D. Access and Use

D.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

D.2 Provide the name and URL(s) (Uniform Resource Locator) for any examples of previous digital collections or content your organization has created.

Part III. Projects Creating Software (systems, tools, apps, etc.)

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) this software will serve.

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software (systems, tools, apps, etc.) and explain why you chose them.

B.2 Describe how the intended software will extend or interoperate with other existing software.

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

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

B.5 Provide the name and URL(s) for examples of any previous software tools or systems your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

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

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository:

URL:

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

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

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

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

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

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

7. What is the plan for archiving, managing, and disseminating data after the completion of the award-funded project?

8. Identify where you will be publicly depositing dataset(s):

Name of repository:
URL:

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