Abstract

Institutional repositories manage and disseminate a University’s scholarly output and provide a multitude of benefits to the organization and society. Rutgers University Libraries is actively expanding its repository to include materials with scholarly merit that are currently siloed in academic departments or otherwise unpreserved and unavailable to the public. This paper describes a collaboration between Rutgers Libraries faculty and Rutgers teaching faculty which is enabling discovery of a significant collection of video data relating to equine behavioral responses. The paper describes the research, the repository tools, and the process of developing and customizing metadata to bring this collection to light.

Introduction

The benefits of University-based digital repositories are numerous and well documented. (Scholarly Publishing & Academic Resources Coalition, Benefits, 2011). They provide an organized, secure, and permanent way to preserve data and scholarly works across media and disciplines (RLG & OCLC, 2002). Importantly for an academic environment, the repository makes faculty scholarship widely accessible, since it is available twenty-four hours a day, seven days a week, from any web-enabled device, and articles can be indexed by Google Scholar. Persistent URLs enable ready reference and access to electronic versions of articles and other data through citations in subsequent works. The end result is more frequent downloads and citations (Gargouri et al., 2010, and SPARC, Effect of open access, 2011), and greater impact. The repository can also help to establish prior discovery, facilitate rights management, and reduce faculty workload by managing a portfolio of scholarly works.

Rutgers University Libraries (RUL) is a national leader in digital repository development. It has developed its own institutional repository, RUcore (http://rucore.libraries.rutgers.edu), based on the Fedora Commons open repository application (http://fedora-commons.org), and has
Equine Research Videos in a Digital Repository

led collaborative development of the statewide repository, New Jersey Digital Highway (http://njdigitalhighway.org), and the statewide video repository, NJVid (http://www.njvid.net). In addition RUL has developed several open source tools to manage digital content and make it available.

Despite the multiple benefits offered by institutional repositories such as RUcore, the percentage of scholarly output that makes its way into a repository remains low relative to total output at a given institution (Zuber, 2008). At Rutgers, the use of RUcore is actively promoted to faculty through discussion forums and one-on-one conversations that describe the availability and benefits of an organized, secure, permanent repository for the institution’s scholarly output. Faculty deposits have been accepted since 2007, and RUL provides a variety of services such as metadata creation, customized search and display, statistical reporting to demonstrate impact, and an automated workflow for submission of theses and dissertations. RUL has an efficient, well-established workflow and self-deposit of materials is also an option.

There are a number of legitimate reasons why faculty contributions to institutional repositories are relatively low. Faculty authors need publication in peer-reviewed journals to gain tenure, and publishers of these journals impose restrictions to insure exclusivity of their publications. Concerns about scholarly credit, premature disclosure, plagiarism, and attribution are compounded by “confusion, uncertainty, and fear about intellectual property issues” (Zuber, 2008 and Van Westrienen & Lynch, 2005). In some fields, tenure-track faculty prefer a discipline-based platform that centralizes all publications to support claims of most downloaded or most cited article in the field. Even with automated faculty deposits, submission is often seen as just extra work, rather than a venue through which to advance scholarship (Xia & Sun, 2007).
Equine Research Videos in a Digital Repository

Part of the problem may be that institutional repositories have traditionally focused their support on journal publications and preprints. Rutgers is actively expanding beyond these boundaries to preserve the full scholarly record, focusing outreach efforts on forms of research ineligible for traditional publication, such as conference proceedings, data sets, and media. One example of this focus is the video data set produced by Dr. Sarah Ralston, through her Young Horse Teaching and Research Program of the Department of Animal Sciences at Rutgers, the State University of New Jersey.

The collaboration described here, between the Libraries and the Young Horse Program, provides a model for teaching faculty-library faculty partnerships that can expand discovery of underutilized research and is key to the library’s mission of preserving the scholarly record. Moreover, it demonstrates one way in which librarians may be integrated as partners into academic departments, and reflects the unique role technical services librarians can play in embedded librarianship initiatives. Finally, it demonstrates the benefits of an institutional repository to others in the organization, and particularly those with special collections of their own which should be preserved and made widely available to others.

The Horse, New Jersey, and Rutgers

The horse is the official State Animal of New Jersey. More than 83,000 equine animals reside in the state (American Horse Council, n.d.), representing a 1.1 billion dollar racing, breeding, and recreational riding industry (Rutgers Equine Science Center, 2007). Horses have a significant impact on the state’s economy, traditional agriculture, and the preservation and maintenance of open space, and New Jersey is one of the few states that recognize horses as agricultural animals. Consequently the New Jersey Agricultural Experiment Station and
Rutgers, the State University of New Jersey, have invested significant resources to support the horse industry through research, teaching and service activities since the early 1990s.

The Equine program at Rutgers has expanded dramatically since the 1980s when there was only one faculty line devoted to the species and fewer than 10,000 dollars annually in research grants and donations. In the 1990s two new full-time, tenure-track faculty were hired, dramatically expanding research, teaching and outreach efforts on behalf of the equine industry. In 2001 Rutgers established an Equine Science Center to educate policymakers and the public on equine issues. In 2002 a fourth equine faculty member was added as an extension specialist. Faculty affiliated with the Equine Science Center and Department of Animal Sciences conduct significant scientific research on many aspects of horse nutrition, behavior, management and well-being. Equine research at Rutgers, like much university research, has many different forms, many of them traditional, such as studies of metabolic or physiologic responses to drugs, nutrition and exercise physiology. However others, such as investigations of horse behavior, may be best understood and utilized by observing the responses in dynamic visual media rather than static photos or text. For this reason the Young Horse Teaching and Research Program has produced hundreds of videos documenting equine behaviors and behavioral responses in standardized behavior tests, looking at differences between breeds, ages and dietary manipulations. This paper describes the partnership between Animal Sciences teaching/research faculty and RUL faculty that is bringing this original scholarship to light for research, public, and classroom use.

The Research

The research videos produced for the Young Horse Teaching and Research Program are exemplary for several reasons. They offer the ability to objectively compare behavioral
responses for statistical analysis and evaluation. They can serve as teaching tools, clearly
demonstrating behaviors not easily conveyed through the written word, for example, resistance
versus compliance, anger versus fear, or calm versus agitation. Furthermore, the Program offers
students a unique opportunity to train horses that have had no previous handling, an unusual
situation in New Jersey, where the management on breeding farms tends to be very intensive,
and foals receive almost daily training from birth. None of the horses used in this Program had
been trained before arriving at Rutgers, and some of the videos demonstrate training techniques
the students learn as part of the Program.

The Research and Teaching program was conducted using two non-traditional types of
young horses that were in need of promotion and about which there was very little nutritional or
behavioral data. The young horses used for the first ten years were the progeny of Premarin
mares. Premarin is a hormone therapy drug derived from the urine of pregnant mares, and for
practical reasons, the pregnant mares were draft horses or draft breed crosses that had been bred
to non-draft horses such as American Quarter Horses or Thoroughbreds to produce a potential
“sport horse.” Draft cross “sport horses” were becoming increasingly popular in the eastern and
western coastal states in a wide variety of equine disciplines, such as show jumping, dressage,
driving and fox hunting and were being registered in the newly formed American Warmblood
and Canadian Sport Horse registries. However, virtually all of the Premarin ranches were in
North Dakota and central Canadian states where the above disciplines were not popular and in
1999 behavior and nutrition data were non-existent for that type of horse. About 30,000 foals
were being produced annually in the late 1990s and early 2000s and many were sent to auctions
where they were sold for slaughter. Therefore they were in need of promotion to insure sale to
good and productive homes, which was one of the focuses of the Young Horse Research and
Teaching Program and the North American Equine Ranching Information Council (http://www.naeric.org). By 2008 the marketability of the Premarin young horses had increased dramatically and the cost of a weanling purchased by the Young Horse Program had jumped from $500 in 1999 to $2000 in 2008. Beginning in 2003 there was a dramatic drop in demand for Premarin, marked by a 37% decrease in promotional spending and a 32% decrease in hormone therapy prescriptions (Majumdar, Almasi & Stafford, 2004). Indeed, by 2011 there were only 26 ranches left in the Premarin industry, all producing high quality registered foals of a variety of breeds (North American Equine Ranching Information Council, 2011). This dictated a dramatic reduction in availability and need for promotion of the foals.

Therefore in 2009 Dr. Ralston introduced one- and two-year-old feral mustangs into the Program. These are the horses that roam our western rangelands, managed by the U.S. Department of the Interior’s Bureau of Land Management (BLM). As a protected species they can’t be hunted; as an introduced species, they have few natural enemies. Without these controls, wild horse numbers can double in four years (U.S. Department of the Interior, Bureau of Land Management, *Wild horse*, 2011). If the herds are not adequately managed there is a real risk of serious overpopulation that would degrade the ecosystem, not only for the horses, but for all other wildlife. Consequently, since the 1970s the BLM has gathered the wild horses and burros and transported them to holding facilities to wait for adoption through the Adopt-A-Horse (or Burro) Program (U.S. Department of the Interior, Bureau of Land Management, *Gathers*). Despite being considered by most as “living legends” and symbols of the American West, mustangs have an undeserved reputation as horses that are difficult, if not impossible, to tame and train. Nonetheless there was a consistent and sustained demand for them by the relatively small population of people who recognized and treasured their true potential as athletes. There
were actually waiting lists of adopters at some of the horse management areas (HMAs)
(McFadden & Fontana, personal communication, August 20, 2010). However, in 2007, the last
horse slaughterhouse in the United States closed, and the resultant increase of domestic horses
available to the public meant direct competition with wild horse adoptions and sales.  (According
to the U.S. Government Accountability Office’s 2008 report, the number of domestic horses
killed in slaughterhouses from 2000 to 2005 ranged from about 40,000 to 75,000 annually.)  This
competition, coupled with an economic downturn, probably contributed to a decline in demand
for mustangs. “Thirty-six percent fewer animals were adopted in 2007 than compared to the
average adoption rates in the 1990s. As of June 2008, BLM was holding 30,088 animals in
holding facilities, up from 9,807 in 2001 (U.S. Government Accountability Office, 2008).” By
2011 there were an estimated 38,000-plus mustangs and burros on public HMAs, almost 12,000
more than the ecosystems could sustain, and an additional 41,000 of the wild equidae in holding
facilities awaiting adoption or direct purchase (U.S. Department of the Interior, Bureau of Land
Management, _Wild horse_, 2011).

The Young Horse Program therefore has dual benefits. First, the detailed scientific
information it produces about the temperament and nutritional needs of draft crosses and
mustangs helps to promote these non-traditional types of horses. Secondly, students gain
invaluable experience training and managing the young horses as part of the “hands on” learning
courses required by the Department of Animal Sciences. The horses are acquired as weanlings
or yearlings, then managed and trained during the academic year to produce well-trained young
horses with excellent ground manners and athletic potential in a variety of disciplines. The young
horses are shown in the Annual Rutgers Ag Field Day event in the spring and then sold at private
auction, with proceeds invested in the purchase and care of horses for the next academic year.
The program has been self-supporting since 2001 through private donor “sponsorships” of the horses as well as private foundation and horse industry support.

The Collaboration

The Libraries became aware of the behavior videos when Metadata Librarian Jane Otto attended the Equine Science Center’s annual Science Update, where Dr. Sarah Ralston reported her research, citing the use of videos but lamenting that they were not easily accessible, being stored on individual mini-DVD or AVI files on her office computer with limited ability to make them accessible through the student-run website (http://younghorse.rutgers.edu). Subsequent discussion made plain that these videos were prime candidates for the repository, where they could be organized, preserved and made freely available for research and classroom use. Here was an opportunity to combine teaching faculty’s subject expertise with library faculty’s information organization skills to enhance research, teaching and public use of the unique data collected.

In order to initiate and implement the partnership, library and teaching faculty had to reach a mutual understanding of both the data and how it would be cataloged and retrieved, through a full discussion of the following issues.

1. Data being collected, i.e., behavior videos of both standardized sequences of tasks asked of the horses and general behaviors such as play in the fields and responses to tranquilization, observer assessments of responses, growth and feed intake data, and descriptions of the horses themselves

2. Purpose and value of the various forms of data, to inform decisions on how to describe, index, and relate the different resources

3. Relationships between the video and any ancillary data
4. Number and format of individual records to determine storage and transcoding requirements and cataloging timeframe

5. Related information about the collection, such as previous publications or presentations

In the initial consultation, the above were established, and the various sources of metadata needed to enable management and discovery of these materials were assembled. Horse descriptions came from the Program’s website (for example, http://rci.rutgers.edu/~ruhorse/auction2009.html) as well as internal records; signed privacy releases for persons appearing onscreen were held in paper files; test, growth, and feed intake data were stored in spreadsheets. Additional information was often found on the DVD itself, in digital file names, on post-it labels, and in verbal identifications recorded on the video soundtrack during shooting.

After all the documentation had been compiled, it was determined, based on the target audience(s), what metadata would best support the basic user tasks of Find, Identify, Select, and Obtain (International Federation of Library Associations and Institutions, 1997). Two important points of this discussion concerned titles and controlled vocabularies.

Titles have to be descriptive for identification purposes, but they also serve a collocation function, since titles constitute a key sorting element in RUcore. The decision was made to supply hierarchically constructed titles to enable search results arranged by general category (Behavior tests, for example), then subarranged by test name, if applicable, then horse name, breed, age and gender. For example:

Equine behavior: behavior test, clippers test, Honky Tonk, draft-cross weanling colt

Equine behavior: behavior test, stepstool test, Canella, mustang two-year-old filly
Equine Research Videos in a Digital Repository

Equine behavior: behavior test, stepstool test, Marley, mustang yearling gelding

Equine behavior: Tranquilization before and after castration, Casanova, mustang yearling colt

Equine behavior: pasture play. Casanova, mustang two-year-old gelding, and Honky Tonk, draft-cross yearling gelding

Equine behavior: pasture play. Draft cross and mustang geldings playing with rubber tub

The second key point concerned access points and controlled vocabularies. The real question here was, how were users likely to search these materials. A demonstration of the partner portal search tool (described below) helped to clarify this question, and a preliminary search modality was agreed upon.

After this initial consultation, Otto began synthesizing all the data identified and reviewing the videos in conjunction with it. She then began creating sample descriptions and determining appropriate values for the technical, source, and rights metadata, utilizing the Open WMS system.

OpenWMS

In addition to its development of digital repositories, Rutgers University Libraries has developed a number of tools for managing and enabling discovery of repository resources. First and foremost of these is the open source, web-based Workflow Management System, or OpenWMS (http://rucore.libraries.rutgers.edu/open/projects/openwms). OpenWMS is a complete metadata creation system built on the Fedora repository. The WMS accommodates descriptions of resources in all formats, both digital files, and the physical (usually analog) source materials from which they are often derived. METS-based
Equine Research Videos in a Digital Repository

(http://www.loc.gov/standards/mets), it includes all the metadata necessary for end-to-end management of a resource, be it descriptive, source, technical, or rights.

An important additional feature of the OpenWMS metadata is its use of events. An event is defined as any action that occurs at a particular place and time, and OpenWMS accommodates five types: provenance event, preservation event, condition evaluation event, rights event, and descriptive event. Each event includes metadata elements for associated agents and associated objects. For example, a rights event can document a privacy permission. The associated agent would be the person granting permission (i.e. the person depicted in the video) and the associated object would be the signed permission form. The form can then be scanned and input to the repository and managed along with the resource itself, the metadata, and any other administrative documents. The events recorded in this collaboration include rights event (for both privacy permission and license agreement), as well as a provenance event to document the loan of the original source mini-DVDs used to create the digital files.

The WMS also features mapping and import/export utilities to bring in records in non-standard formats, including spreadsheets. Because of the diversity of metadata sources for the Young Horse Teaching and Research video collection, records were created from scratch, rather than mapped from existing records.

The Metadata

Descriptive, technical, and rights metadata were created for each resource in the collection. The descriptive metadata includes title, date, PURL, summary, notes on the character of the work, and access points for names and subjects.
Equine Research Videos in a Digital Repository

**Figure 1. Descriptive Metadata**

Technical metadata describes the digital characteristics of the master digital file and is “essential to the long-term preservation and management of audiovisual resources” (Otto, 2010). It documents the operating system and application used to create the file, frame rate, signal format, audio encoding, bit rate reduction data, etc., in other words, the information that allows the original image to be authentically reproduced.

Source metadata documents the physical characteristics of the original physical carrier (if there is one), and was input for the first batch of videos, which were transcoded from mini-DVDs. Subsequent batches originated as digital files in AVI format; since these digital files were not derived from a physical video source object, there was no source metadata.

Rights metadata enables management of a resource’s intellectual property rights. While there is no prevailing standard for rights metadata, OpenWMS draws from PREMIS (http://www.loc.gov/standards/premis) and the California Digital Library’s copyrightMD (http://www.cdlib.org/groups/rmg). The information recorded for each resource in this collection
Equine Research Videos in a Digital Repository

is derived, for the most part, from a set of controlled vocabularies, and includes a rights declaration (The author owns the copyright to this work), copyright status (copyright protected), availability status and rationale (open, via permission or license), publication status (unpublished), watermark (none in this case), and rights holder (Dr. Ralston). The signed license to make the material available in the repository, as well as the signed privacy permissions, are described in rights events, with the documents (associated objects) scanned and ingested along with the video and its metadata.

The metadata for this collection is typically voluminous. However, much of it remains constant from catalog record to catalog record, which allows utilization of OpenWMS templates to speed inputting. Source and technical metadata, and most rights metadata, for example, can be input one time in the template; once there, all subsequent records are created from the template, with very little additional keying. Use of templates, together with the synthesis of all documentation up front, has made for very streamlined inputting. Any given resource, even with scores of elements, can be described and ingested in less than five minutes.

Once Otto drafted the metadata, she submitted it to Ralston for review. At this point, the descriptive metadata was largely standard; some subject vocabularies had been established, but not all. Access points would need to work for both the general user accessing these materials through the main RUcore portal, and the expert user discovering them through a customized Equine Science Center search portal. Expert users would include Ralston herself, her students, and other equine science researchers. The customized “partner portal” could be made available in a number of places, including web pages outside the Libraries’ own website. Now another repository tool came into play: the Partner Portal Tool.
The RUcore Partner Portal Tool

The Partner Portal Tool has three key features: First, it provides targeted access to a single collection within RUcore. Within that collection, search options can be configured and displays customized. Third, it is completely portable. This little “search box” can be dropped into any web page, be it Ralston’s own page, a departmental page, a Young Horse Teaching & Research website page, or linked from the portals page in RUcore.

The easiest way to answer the abstract question of “how will users want to search this data?” was to demonstrate a working model of the Partner Portal Tool. (See Figure 2.)

![Figure 2. The RUcore Partner Portal Tool (Search tab)](image)

To the user, the Tool looks like a search box with three tabs, Search, Browse, and Help. Each of the tabbed pages can be customized. With the sample record in hand and a concrete example from another collection, Ralston readily identified her search categories. In addition to the more-or-less standard searches by title, name, and subject, she needed specialized searches by horse name, breed, age, and gender, plus searches by test name and test score. Retrieval on these metadata elements, among other things, would facilitate comparative analyses of test
performances by breed, age, and gender. Users can also pull up particular tests to compare individuals’ behaviors, or compare the highest and lowest scores for a particular test.

The tool’s ability to limit a search by a targeted collection is valuable. For example, Ralston and her students might wish to pull up a video by horse name, such as ‘Rose.’ A search of the term ‘rose’ in the RUcore general search portal yields 44 results; recall is high (100%) but relevance is low, since ‘rose’ occurs in personal and corporate names, and as a type of flower in subject headings. A search of the same term through the Equine Science Center Partner Portal tool yields three records describing videos with the horse named Rose: 100% recall, 100% relevance.

So Ralston’s chosen categories became pulldowns for the search-tabbed page, and for each category, actual controlled vocabularies were discussed and determined based on existing vocabularies and literature in the field. Again, choices tend to be collection specific. For the horse age category, for example, Ralston opted not for strictly numeric terms (1 year, 2 years, etc.), but rather for broader age spans, in this case, Weanlings, Yearlings, Two-year-olds, Three-year-olds, Adults, and Geriatrics.

Next, the Browse categories were discussed.

![Figure 3: The RUcore Partner Portal Tool (Browse Tab)](image)
Each of the hyperlinked natural language phrases in this display stands for a complex Boolean search running in the background, and provides great flexibility in searching, when neither the standard (generally broader) standard Library of Congress subject headings, nor the specialized subject terms alone, can retrieve the desired records. One application of these “canned” browse searches relates to class assignments. Take the stepstool test, for example. This test introduces the horse to a person standing on a stepstool. Most horses have a reaction of surprise, and look in curiosity at the person. The minority of horses immediately “figure out” that the unusual piece of the picture, the part that warrants attention, is the stool itself. It makes for an interesting cross-breed analysis to compare mustang responses with draft cross responses. For an assignment, Ralston might ask her students to pull up selected stepstool tests and compare responses by breed and gender. On the back end, a search is constructed to pull up the relevant records; they are assigned the natural language label “Assignment One,” and a link is created on the Browse page.

In addition to configurability of Search and Browse searches, the Partner Portal Tool allows configuration of record displays. As with most library systems, a search yielding multiple “hits” will always display brief records, so as to maximize the number of records displayed per screen. Collection owners such as Ralston can determine the fields to display, and their order, for both the brief and full records. One can also configure fonts, colors, and introductory text, as well as help screens.

In short, the RUcore Partner Portal Tool is customizable, configurable, portable and easy to use. It makes a specific research collection available through vocabularies and in locations familiar to the expert community, yet available to a lay audience through the main RUcore search interface. Designed to work on even the simplest websites, anyone can integrate search
and retrieval into their webpage by simply adding a single line of HTML code. Both faculty and the University reap all the benefits of the institutional repository (organized, secure, permanent, etc.), while integrating and exposing their data in ways never before possible.

**Next steps: the RUanalytic Tool**

The next tool that will come into play in this collaboration is Rutgers University Libraries’ RUanalytic tool, currently in testing. This tool will allow users to combine objects from the repository into seamless presentations and enable scholarly conversation about the resources through analysis and annotation. For the Young Horse program videos, this means clip reels that can be used in the classroom to illustrate and contrast a variety of behavioral responses, handling techniques, and veterinary care. The tool can be used not only with video clips, but with any unrestricted objects in the repository, including audio files, images, text, and data sets, so that Ralston can illustrate her own or others’ research with videos from her own collection. The resulting annotation can then be licensed by the creator and published in the repository for preservation and wide distribution.

**Conclusion**

The institutional repository preserves and makes available scholarly works and offers benefits that are numerous and far-ranging. It has been called the essential infrastructure for scholarship in the digital age (Lynch, 2003), yet faculty contributions, limited as they often are to journal publications and preprints, are relatively few. Rutgers University Libraries is focusing its development and outreach efforts on forms of scholarship not eligible for traditional journal or monograph publication, offering innovative repository tools to customize metadata and facilitate use by multiple diverse constituencies. This project involved determination and customization of metadata and displays and development of a search portal for expert users. The video data set
Equine Research Videos in a Digital Repository

produced by Rutgers’ Young Horse Teaching and Research Program is exemplary in that it can be used for classroom training as well as statistical analysis and evaluation that serves horse types in need of promotion and supports the Animal Sciences curriculum. This partnership between the Libraries and the Young Horse Program is enabling greater discovery of research data, demonstrating the repository’s value to others in the University, and serving as a model for collaborations between library faculty and teaching faculty in academic departments. As outreach is expanded to include other University departments, work will continue with this project as we incorporate an analytic tool to analyze and annotate a variety of repository resources to advance digital resources in the University curriculum.

REFERENCES


