# Digital Repository Task Force Planning Meeting

Attending

Jordan, Jimmy, Mark, Alyssa, John Major, Nancy G, Nancy T, Jackie S., Pascal, Jo N. Matt Alleman, Nicole LeBouf

# Task force charter/governance

See Marmot committees and task force page marmot.org/node/16

Goal – Come up with a plan, make a recommendation to the board so we can spend money.

Money – Prospector was the last big grant

Focus on a winning grant application?

Statement from website:

*The DRTF was formed to act on a critical mass of interest in digitizing unique local resources, collecting born-digital resources, and making these discoverable and accessible. From September 2014 the work of this task force will include evaluating digital repository products; setting up hardware, software, and procedures for capturing, storing, and cataloging digital items; and advising developers on enhancements for the Marmot VuFind discovery layer.*

Need human resources, monetary resource, etc.

# Grant

Will help to organize work, good goal

Want to shoot for LSTA grant – Mid Summer 2015

Some things we may need

* Theme
* Product Evaluation
* Prototype
* How do we unify diverse libraries?

Can also re-use the LSTA grant for smaller applications

# Funding

* Need to be able to monetize some of the content.

# Organization

* Is this Marmot assisting with member libraries
* Or is it Marmot leading as a shared organization

# Subcommittees

## Larger Group

1. Get more information on how to do a project like this successfully
   1. Liz Bishoff – Lessons learned – Colorado Digitization Project
   2. Jordan – Kansas City History Project
   3. More tbd

## Digitization / Content

### Members

* Matt - Chair
* Jordan
* Mark
* John
* Nancy G
* Nancy T
* Jo N.
* Jaci
* Kevin W.

### Task List

1. Determine list of content types to load into the repository (is this based on file format or is it based on content? i.e. is Oral History separate from Music recording?)
2. Determine what existing content is already digitized for each content type at a higher level. (I.e. X Library has 250 Oral Histories on CD).
3. Determine what content can be easily acquired for each content type (i.e. we can get 500 photographs from x local museum).
4. Determine other sources of content that are more difficult to acquire.
5. Suggest priorities for content types
6. Visit locations that are already handling each content type
   1. DPL – Photos
   2. Jeffco – Oral History
   3. Jaci – Photos, Slides, Negatives, Documents
   4. Matt – Video, Large Format Images, etc
7. Review digitization vendors
   1. Lyrasis + CCU
   2. Archive.org?
8. Document process for digitizing/collecting each type of content
   1. Determine file format(s) for each type of content
   2. Determine file naming conventions for each type of content
   3. Review Copyright issues for each type of content
   4. Determine minimum standards for each type of content
   5. Determine alternative resolutions needed (i.e. thumbnail, full quality, etc).
   6. Determine what hardware is needed
   7. Determine if hardware needs to be purchased by Marmot and what needs to be purchased locally
   8. Determine outsourcing/insourcing options for each type of content
   9. Determine time to allocate per object split up by who is doing the work (i.e. what can be done by volunteers, what has to be done by a librarian, other?)
   10. Metadata
   11. Determine cost per object
9. Determine how to convert existing data to meet standards
10. Collect at least one example of each content type following the process

### Potential Content Types

* Bound Document Book/Magazine
  + Legal or smaller
  + Large
* Unbound Document
  + Legal or smaller
  + Large
* Handwritten Document
* Clippings
* Mixed Text/Photos
* Photo
* Flat art work
* Map
* 3D Sculpture
* Oral History
* Video
* Music
* Archeological Artifacts
* 3D Scans

## Repository Selection & Digital Preservation

### Members

* Jimmy - Chair
* Jordan
* Mark
* Nicole

### Task List

1. Likely tools to evaluate
   1. Fedora Commons
   2. BPress – CU is moving there
   3. DSpace
2. Tools to not consider (unless we really hate everything else)
   1. ContentDM – Proprietary OCLC
   2. DigiTool – End of life, no longer supported
   3. PastPerfect? – Likely does not scale well enough for us
3. Look at support organizations for repository
   1. III / Virtua (Fedora)
   2. Colorado Alliance (Fedora)
   3. Marmot (Fedora)
4. Create evaluation matrix for software
   1. Digital Preservation
   2. Repository Administration
   3. Ingest Process
   4. Metadata customization
   5. Metadata formats
   6. Backups
   7. Replication
   8. Hosting
   9. APIs
   10. Indexing
   11. DRM
   12. Monetization of content
   13. Ability to integrate with aggregators (i.e. DPLA)
5. Demos for all tools we want to evaluate
6. Request for Quotes
7. Select a tool
8. Define Supplemental Utilities that need to be created
   1. Off the shelf tools to buy
   2. Homegrown by Marmot or local libraries
9. Review other locations to copy data
   1. YouTube
   2. Archive.org
   3. Flickr
   4. LibHub
   5. Google
10. Determine mechanisms to automatically push data to other repositories

## Metadata / User Interface

### Members

* Alysa - Chair
* Jordan
* Mark
* Betsey
* Jo
* Jaci
* Western Cataloger TBD

### Task List (Metadata)

1. Read Dublin Core standard
2. Determine overall standard (DublinCore)
3. Determine minimum required fields for all objects
4. Determine additional required fields for each content type
5. Determine additional optional fields for each content type
6. Determine additional project fields
7. Determine fields to define relationships between content
8. Review standards from aggregators (i.e. DPLA, State Library)
9. Optimize metadata for Google Indexing
10. Determine how much oversight the metadata creation needs
11. Define data input standards
12. Create authority files for some fields

### Task List (User Interface)

1. Define interactivity for each file type. i.e. zooming images
2. Relevance of digital repository as it relates to books/movies
3. Should digital repository be integrated within “standard” catalog or should it be a separate repository similar to Genealogy?
4. How many portals do we need?
5. Need to be able to easily exclude/include digital repository?
6. What local customization options will need to be added?
7. What facets are needed? Should there be different facets for repositories?
8. Need to be able to have Browse Categories for Digital Repository.

# Possible Themes

* Water
* Local history
* Recreation
* Voices of Colorado
* K-12
* Rocky Mountain Life
* The other side of Colorado
* Colorado Flipside
* Rural Colorado
* Colorado Focus

## ­­Possible subthemes

Select a few based on content availability for Grant Proposal. Additional sub themes can be added later.

* Ranching
* Agriculture
* Native Coloradans
* Hispanic Population
* Art & Culture
* Heritage
* Outdoor Life & Sports
  + Fly-fishing
  + Mountain Biking
  + Skiing
  + Rafting
  + Rodeo
* Mining
* Oil & Gas
* Ecology
  + Forestry
  + Water
* Education
  + K-12
  + Higher Education
* Entertainment
* Business & Economy
* Religion
* Wildlife
* Local Lore

# Possible Partnerships

* Local Museums
* Local Historical Society

# What we have already

* Images (several libraries have them)
* Audio (oral histories) – several libraries have
* Video (Mesa & EVLD & Bud, & Western) – programs, local productions
* Digitized Documents
* Theses & Dissertations
* Administrative Records of the library
* Maps
* Misc. Historical documents – pdfs (typed transcriptions EVLD)

# Deliverables (short term)

* Grant application
* 1 page project description (elevator speech)

# Potential issues

* Transfer of rights (initial rights) – Copyrights? Are there restrictions on format.
* Local societies that are worried about
* DRM – can we say everything is non-DRM?
* Scalability
* Sustainability of content

# Other things we may need to pay for

* Copyright attorney – initial consultation w/ ground rules

# For the 18th meeting

1. Subcommittees report with draft timelines now to July 1st rough draft of Grant

# Robin Dean – Digital Repository Director

## Alliance Digital Repository

* Digital preservation – takes a lot of care and feeding. Need to attend to them. Can’t lock them in a dark closet ☺
* Why preserve?
  + Unique and irreplaceable digital content.
  + Do not have time or money to replace (from original material).
* Their members do their own scanning. They don’t work with it directly.
* Why a digital repository?
  + Sustainable – don’t want to be left in the lurch if a free service closes
  + Shareable – everything in one place
* Indexing through Google (and other search engines) is important
* Need to handle different qualities of files (super high resolution vs. access)
* The content will be used in different ways
  + Some people want high quality
  + Some people want low quality
  + Some people want just metadata
  + Some people want a combination of things
* Persistent links to files are important for research especially
* About half the Alliance Members use the ADR
* Pricing varies based on the amount of data in the repository
* Have changed software for ADR
* Use Islandora for display/access as well as Ingest & metadata creation
  + Upgrading from Drupal 6 to Drupal 7 version
* Fedora Commons – for repository. Some people use that directly for ingest
  + Running Fedora 3.7.1 (current release)
  + Will migrate to 4.0 eventually, but probably a few years away before they really migrate.
  + Fedora 4.0 will be significantly different (more linked data etc)
  + No real front end
* Had Fez before as a front-end.
* Hydra is another option for a front-end (they don’t use)
* Everything is open source
  + 98%? - One backup server may be Windows
* Hosted at Univ. of Wyoming, replicated at Univ. of Denver, additional backup in DuraCloud
* Some people use OAI/PMH to ingest into discovery layers
* Some people use OAI/PMH to load into WorldCat
* Not a “trusted digital repository”. Have not gotten TRAC certification due to expense and time, but have done most of the work.
* Have done ~3 migrations in 7 year.
  + Some data intentionally left behind (previews etc).
  + All original content has been preserved.
  + Long, careful slow process for migrations.
  + Have not needed to migrate data from one format to another (use pdf, jpg, tiff, open video/audio)
  + More like major upgrades than migrations (have continued to use Fedora as the back end).
* No built in e-commerce
* Do not search across institution, but could

# Assumptions / Decisions

* Everything is DRM free / Open Access
* Need to be able to monetize some of the content, especially images & b-roll video
* Will encompass both modern and historical content