



17 May 2012

***National Transportation Knowledge Network Digitization Working Group:
Current Activities and Next Steps***

NTKN Digitization Working Group

Remembering Mary Ellen Tucker



2

Good afternoon.

Before we get started today, the NTKN DWG wants to take the time to acknowledge the contribution and legacy of Mary Ellen Tucker to our group and activities. Mary Ellen was an enthusiastic, engaged WG member who contributed all of the resources she could to advance our efforts. She constantly sought opportunities to link our work with that of other activities in transportation, such as a panel presentation at the 2012 Transportation Research Board Annual Meeting on digitization activities in the transportation community sponsored by the Library and Information Science for Transportation committee and chaired by Mary Ellen.

As a broader supporter of the national transportation knowledge network, she was also host for the establishment of the Eastern Transportation Knowledge Network in 2008. I have only been in the field for 5 years and been impacted by her. I am sure there are other examples of her dedication to the success of transportation libraries and information centers.

For all that, we say thank you and hold her in our memories and thoughts.

NTKNG
DWG

National Transportation Knowledge Network Digitization Working Group

Current Activities and Next Steps
Transportation Librarians Roundtable

17 May 2012

In memory of Mary Ellen Tucker, NTKN DWG Member, June
2009 -April 2012



Agenda

- NTKN Digitization Working Group (DWG) overview
- Current Activities
 - Transportation digitization activities, facilities, & skills
 - Transportation community needs
- Next Steps
 - Products
 - Partnerships

4

This afternoon John and I are going to cover three broad topics:

First, an overview of the evolution of the Digitization Working Group from the first NTKN Annual Meeting in June 2009. Second, a review of the results of an environmental scan conducted by the Mid-Ohio Regional Planning Commission to scan the current state of digitization within the transportation community and delivered to the group. Third, a look at what next steps the group envisions from our results.

This is going to be an interactive presentation because we would like to get feedback from you all on what we found as it relates to your current environment, as well as thoughts on how we can most effectively serve the community in the area of digitization and access.

About the NTKN Digitization Working Group

- Who:
 - John Cherney, WisDOT Research Library
 - Alicia Kehrig, Mid-Ohio Regional Planning Commission
 - Louise Rosenzweig, University of Texas, Austin
 - Mary Ellen Tucker, University of North Carolina Highway Safety Research Center
 - Amanda J. Wilson, National Transportation Library
- What Purpose:

Facilitate a coordinated digitization agenda and approach for transportation information collections and resources to increase access for the transportation community and decision-makers

WHAT IS THE DWG:

The NTKN Digitization Working Group (DWG) is a subgroup of the National Transportation Knowledge Network (NTKN) that is focused on digitization efforts to improve access to transportation information. The group, formed in June 2009, has a broad charge to develop a digitization agenda for transportation information collections.

DWG formed out of 1st NTKN meeting. Digitization happening often at transportation libraries, but so much variation in personnel as well as the types of tasks and processes performed. Transportation librarians are asked more and more to perform digitization types of activities and ensure access to these items. The original purpose was to identify and perform “targeted digitization projects” to uncover hidden collections in transportation and increase access to high-value resources.

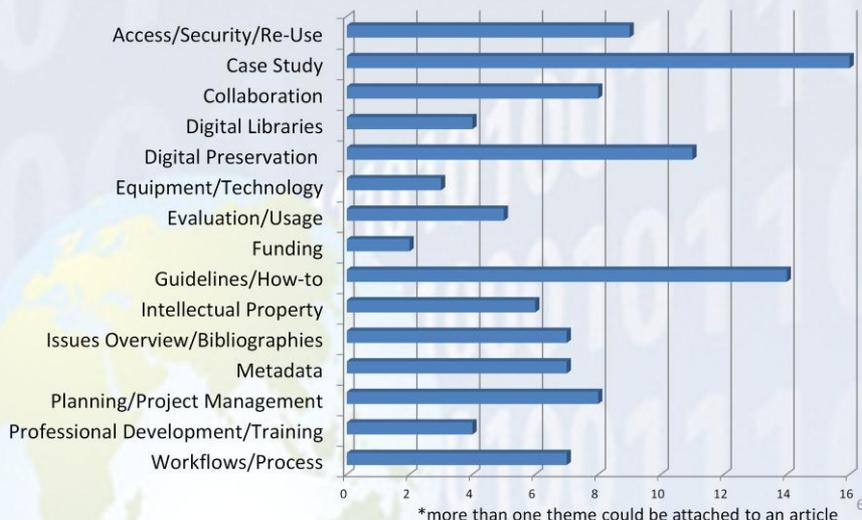
Why a DWG right now?

Many reasons... for one, the activities of such a group fit in with the ideas and plans called for in SR 284 and NCHRP 643. Ensuring standardization and greater access of digitized information is not only desirable but critical for today’s transportation information infrastructure. Ultimately, any standards or recommendations that the DWG proposes in the near future could be a sound foundation for a more formal (codified?) strategy as envisioned within any national (NTKN) official documentation or strategy. Also, with multiple digitized efforts happening today (Google, Hathi Trust, local partnerships, OCLC ContentDM, Dspace, etc. etc.) a group focused on these issues and making sense of them is needed.

Because of many issues that are perhaps not being implemented uniformly: Is metadata being applied uniformly? Are folks just scanning instead of digitizing. Are files searchable and optimized for the web? Different types of materials (photos, archival documents, current research reports, data sets, etc.) Wanted to get a handle on who was digitizing and what they were doing, hence the environmental scan. Examine different materials digitized, different formats produced. Of course, there are access issues as well. Is access available to everyone or

just internal to a particular organization on an intranet.

Current Activities: State of Digitization Research (themes*)



To better frame the information received from the environmental scan, Mary Ellen Tucker conducted a literature review of digitization activities in the broader LIS community from 2009-present. This product also serves as a method to identify best practices, lessons learned, and tools that the DWG could adopt, adapt, or otherwise incorporate into any product or activity we provide to and for the transportation community.

Without going into an exhaustive summary of the literature, the key themes or topic areas of the literature are listed on this slide.

Just taking a look at this brief analysis gives us a barometer reading for the overall state of digitization practice.

Please keep these themes in mind as we walk through the results of environmental scan of transportation activities.

Current Activities: State of Transportation Digitization

- Who is digitizing what?
- How often?
- For what purpose?
- Who has access?
- With what tools and skills?
- How are resources allocated to support digitization?
- Is it project-driven, ad hoc, or regular part of the work of transportation libraries and information centers?
- Who else besides library and information center staff are involved?
- What standards are being used?
- What challenges does the library face regarding digitization?
- What role(s) can the NTKN DWG fulfill to support identified needs?

Digitization (who, what, why, how much, etc.) and **Access** (description, availability, etc.) were the two areas for which the NTKN DWG needed information

7

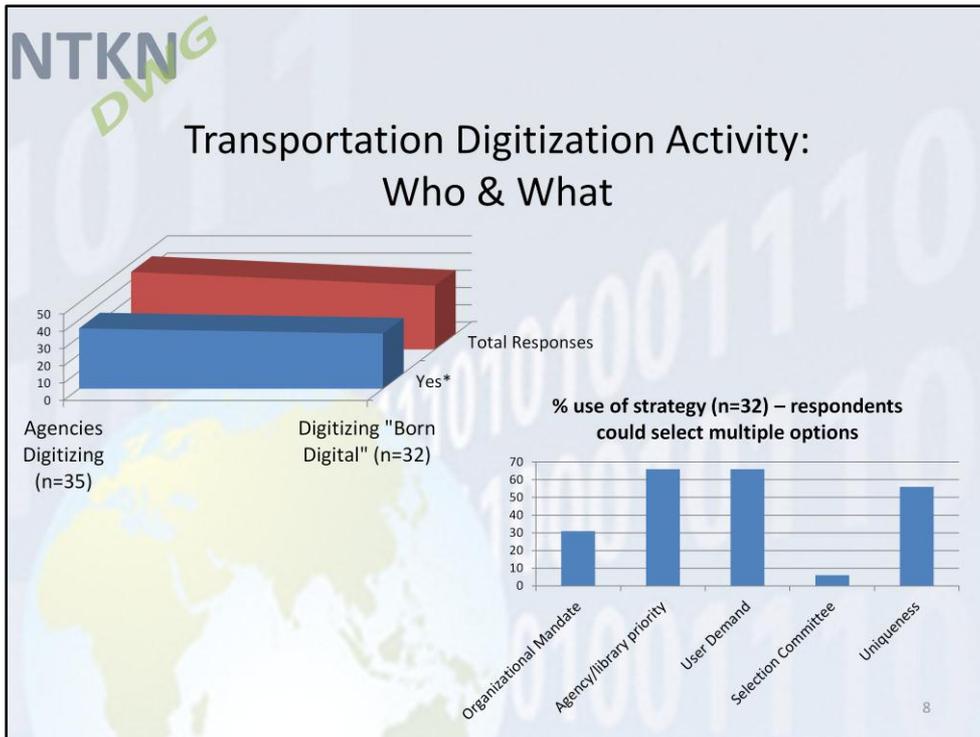
To start to address our charge, the group did not have a good information on the landscape what was happening regarding digitization in the transportation community. The Mid-Ohio Regional Planning Commission conducted the survey because little is known about what digitization activities are currently being undertaken in transportation libraries and information centers. They provided the results to the DWG for our review and analysis.

The model was a FLICC (Federal Library and Information Center Committee) Digitization Survey of federal libraries used to gauge digitization activities and, secondarily, to create a “family-tree” of federal libraries.

The survey is attempting to collect information about:

- What is being digitized
- Why is the library digitizing
- How is digitization happening
- What challenges does the library face regarding digitization
- How can NTKN better support digitization efforts

The DWG is looking for trends across the transportation community -- what motivates digitization and what types of organizations are able to support these projects. We hope to find best practices and advice that can be shared across the community so that transportation organizations can use the lessons learned from each other’s projects to further their own digitization efforts. The DWG is also interested in the barriers to digitization as well.



Anecdotally, we know that A variety of libraries and transportation organizations performing digitization activities – range in scale from sophisticated to simple. Northwestern, Iowa DOT, etc. – more sophisticated... have written digitization plans and standards – other one person libraries, more simplistic due to lack of resources – could be more of a scan-as-you-go approach. Also, organizations other than transportation libraries are doing this (Alicia at MORPC).

Well –established transportation libraries
(Northwestern, Texas DOT, etc.)

Established programs at smaller libraries (Iowa DOT).

Small one person shops – scan on demand and no time to do much else

Transportation information producers (not libraries) – like MORPC.

3rd party scanning for transportation libraries

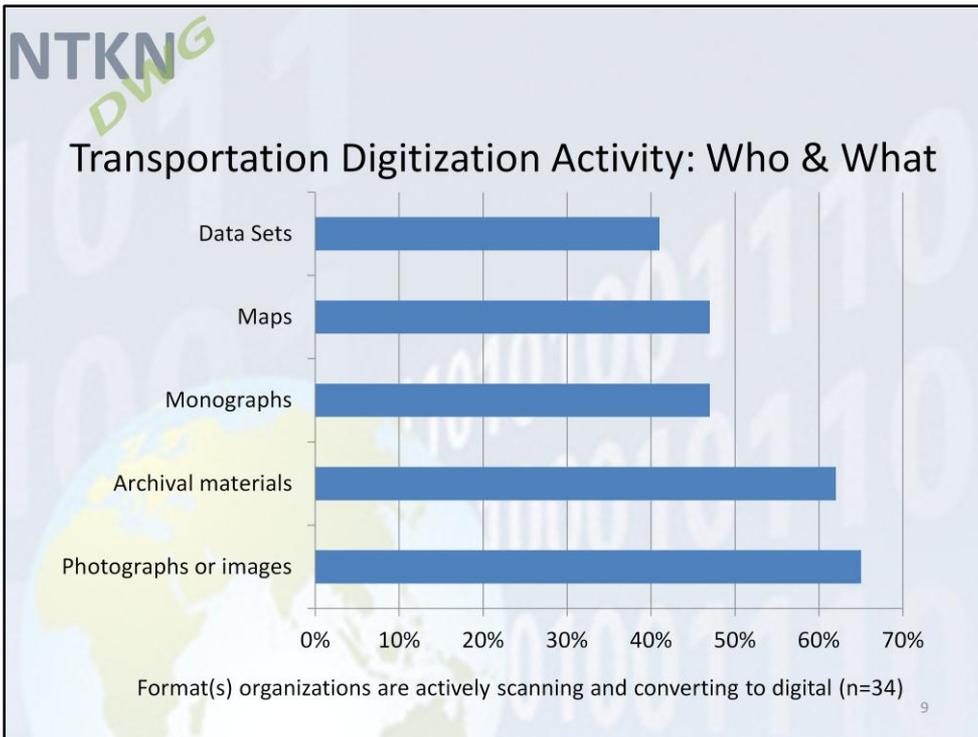
and organizations, like Google digitization projects, etc.

According to the results delivered from the 35-question Survey Monkey instrument, this is accurate. MORPC sent survey via email to the Transportation libraries and information centers listed in the Transportation Library and Information Center Directory. Sent to 311 names. 45 responses. RESPONSE RATE of 14%.

Of 33 respondents to the question, digital projects are considered to be part of regular operations for 33%, finite, special projects for 27%, and 39% consider digital projects as both.

In the survey, almost 80% responded that they are digitizing materials; 40% of respondents reported they have been digitizing between 5 and 10 years... 13% have been doing it more than a decade.

When asked about selection strategies, 32 respondents listed agency/library priority and user demand as the largest driving factors. A drawback of the survey is that there was not an "other" category to capture what was not asked. 92% (of 36 respondents to the question) reported that they are digitizing materials produced by their organization; 58% of those respondents are not digitizing info from other organizations.



Organizations are digitizing in the most commonly known formats. Depending on the type of resource (text, web, or audio/sound/moving image) various levels of respondents reported formants:

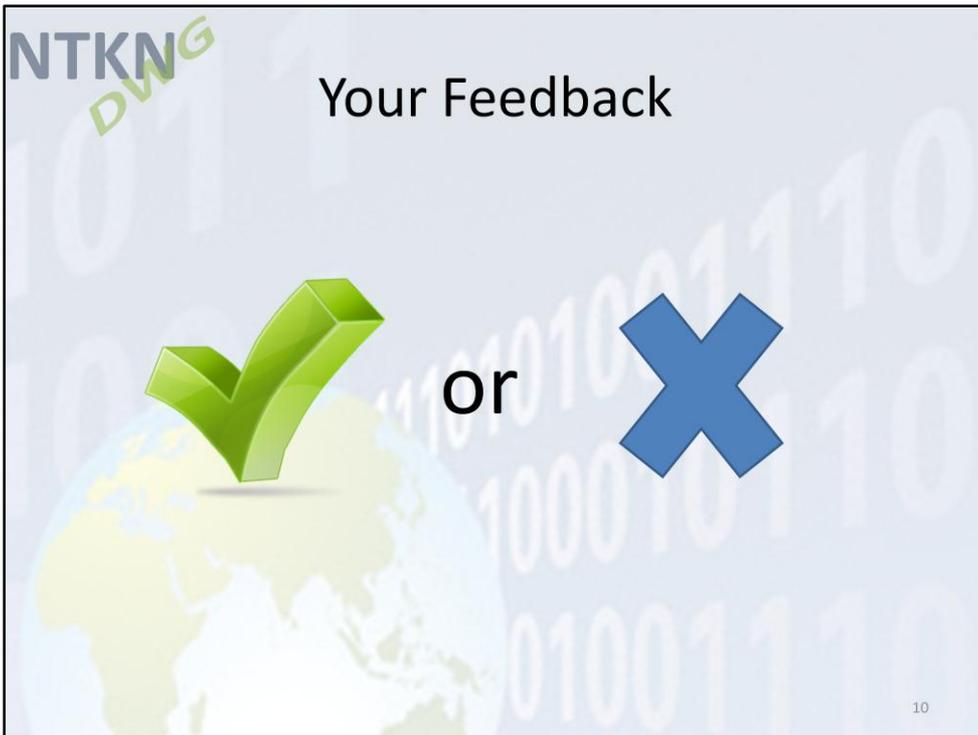
Text Resources: 97% PDF; 11% TXT (n=35 respondents)

Still/Photographic Images: 100% JPG; 55% TIFF (n=27 respondents)

Web Files: 95% HTM/HTML (n=22 respondents)

Audio/Sound: 66% MP3/ 66% Microsoft WAV form/ 55% WMA Windows Media Audio File (n=9 respondents)

Moving Images: 77% MPEG – Moving Pictures Experts Group (n=9 respondents)



What and Who Summary:

What is being digitized---

Variety of materials (documents, photos, film, maps, etc.)

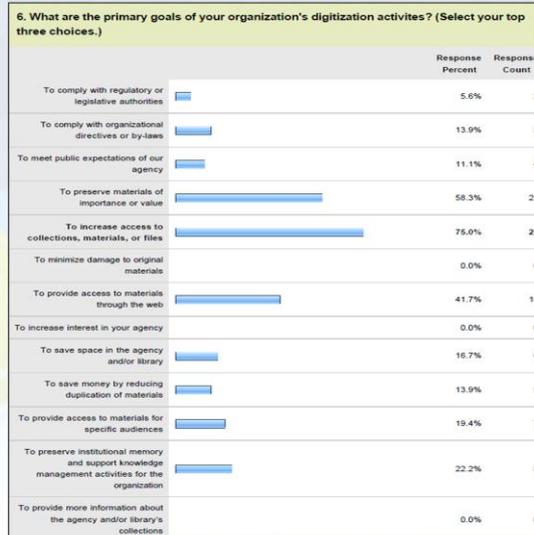
Variety of formats (pdf, mpg, jpg/tiff, mp3, etc.)

Do we partner with NTKN's cataloging workgroup for advice on some of these metadata issues?

Have we missed anything from our survey that is happening in your library (in terms of digitization)

What is your digitization experience? What activities are you doing?

Transportation Digitization Activity: Why & How



11

#1 reason for digitizing is more for the enhancement of access than for archival storage. 75% report increase access to collections, materials or files. Numbers 2 and 3 at 58% and 42% respectively are preserve materials of importance or value, and increase access to materials via the web.

Related -- Audience:

59% of respondents to this question (34) reported that they were proving access for internal agency customers. In another question, 83% of 36 respondents indicated that agency staff, were primary audiences for digitization activities. Transportation professionals (64%) and reserachers (57%), agency stakeholders (52%), and students and general public (both at 44%) round out the top 5 audience categories for digitization.

Digitizing for archival purposes-

0% - To minimize damage to original materials

22% - To preserve institutional memory & support knowledge management activities for the organization

In a related question, 84% - Report the audience for the digitized materials is the Agency staff via the intranet!

Transportation Digitization Activity: Why and How -- Access



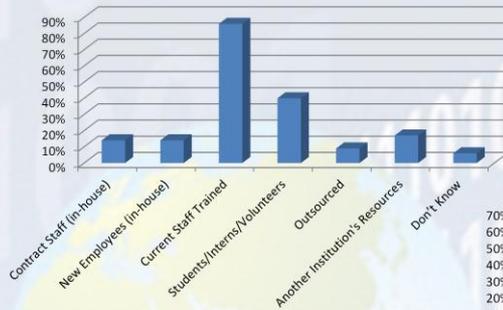
searchable 74% yes
 available to agency 84%
 available to public 61% yes

Digitized materials are provided to audiences primarily through local network drives (internal access only) – 77% (n=30). Next closest is in a digital repository (30%), followed closely by external at a commercial site or shared institutional site (27% for both).

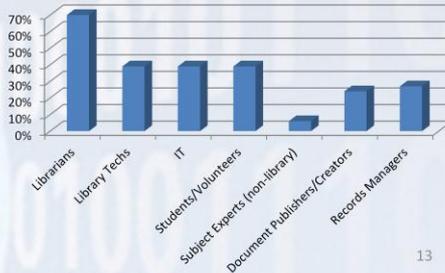
Coincides with stated earlier primary audience and goals for digitization.

Transportation Digitization Activity: Why and How – Process & People

Institutional Digitizing (n=35)



Positions Digitizing (n=33)



13

Institutional:

Contract Staff (in-house) 14%
 New Employees (in-house) 14%
 Current Staff Trained 86%
 Students/Interns/Volunteers 40%
 Outsourced 9%
 Another Institution's Resources 17%
 Don't Know 6%

Who:

Librarians 70%
 Library Techs 39% IT 39%
 Students/Volunteers 39%
 Subject Experts (non-library) 6%
 Document Publishers/Creators 24%
 Records Managers 27%

Transportation Digitization Activity: Why and How – Capability in Resources

| | Deficient | | | Fully Capable | Rating Average | Response Count |
|---------------------------------|------------|------------|------------|---------------|----------------|----------------|
| Staff skills and expertise | 11.8% (4) | 17.6% (6) | 41.2% (14) | 29.4% (10) | 2.88 | 34 |
| Equipment and software | 8.8% (3) | 35.3% (12) | 35.3% (12) | 20.6% (7) | 2.68 | 34 |
| Funding | 35.3% (12) | 29.4% (10) | 23.5% (8) | 11.8% (4) | 2.12 | 34 |
| Writing digitization plans | 30.3% (10) | 27.3% (9) | 24.2% (8) | 18.2% (6) | 2.30 | 33 |
| Establishing quality standards | 17.6% (6) | 35.3% (12) | 26.5% (9) | 20.6% (7) | 2.50 | 34 |
| Writing policies and procedures | 21.2% (7) | 27.3% (9) | 27.3% (9) | 24.2% (8) | 2.55 | 33 |

Out of 31 respondents, 94% dedicate less than 10% pdf budget to digitization; none spend over 50%

| | Recruited New Staff (employees or contractual staff) | Existing Staff (existing skill set) | Trained Staff (existing staff, trained for augmented skills) | Response Count |
|--|---|--|---|----------------|
| Archival or preservation practices | 10.7% (3) | 82.1% (23) | 32.1% (9) | 28 |
| Cataloging and metadata standards (e.g., AACR2, MARC, Dublin Core) | 16.0% (4) | 88.0% (22) | 16.0% (4) | 25 |
| Collection development practices | 8.3% (2) | 95.8% (23) | 8.3% (2) | 24 |
| Conversion of digital files from one format to another | 12.0% (3) | 80.0% (20) | 32.0% (8) | 25 |
| Digital repository software (e.g. CONTENTdm, Fedora, DSpace, ArchivalWare) | 9.5% (2) | 71.4% (15) | 33.3% (7) | 21 |
| Full-text markup languages (e.g. HTML, SGML, XML) | 9.1% (2) | 86.4% (19) | 22.7% (5) | 22 |
| Management of digital assets | 11.5% (3) | 88.5% (23) | 30.8% (8) | 26 |
| Project management | 7.4% (2) | 92.6% (25) | 14.8% (4) | 27 |
| Subject area expertise | 8.7% (2) | 91.3% (21) | 17.4% (4) | 23 |
| Taxonomy construction and maintenance | 9.1% (2) | 90.9% (20) | 9.1% (2) | 22 |
| Web authoring tools | 9.5% (2) | 81.0% (17) | 28.6% (6) | 21 |
| Web and server technologies | 12.5% (3) | 91.7% (22) | 20.8% (5) | 24 |
| Writing or editing (e.g. policies, marketing materials etc.) | 7.4% (2) | 100.0% (27) | 18.5% (5) | 15 27 |

Transportation Digitization Activity:
Why and How – Resources in Skill Sets

Q: How have you (respondents) filled the following knowledge or skill areas to conduct digital projects

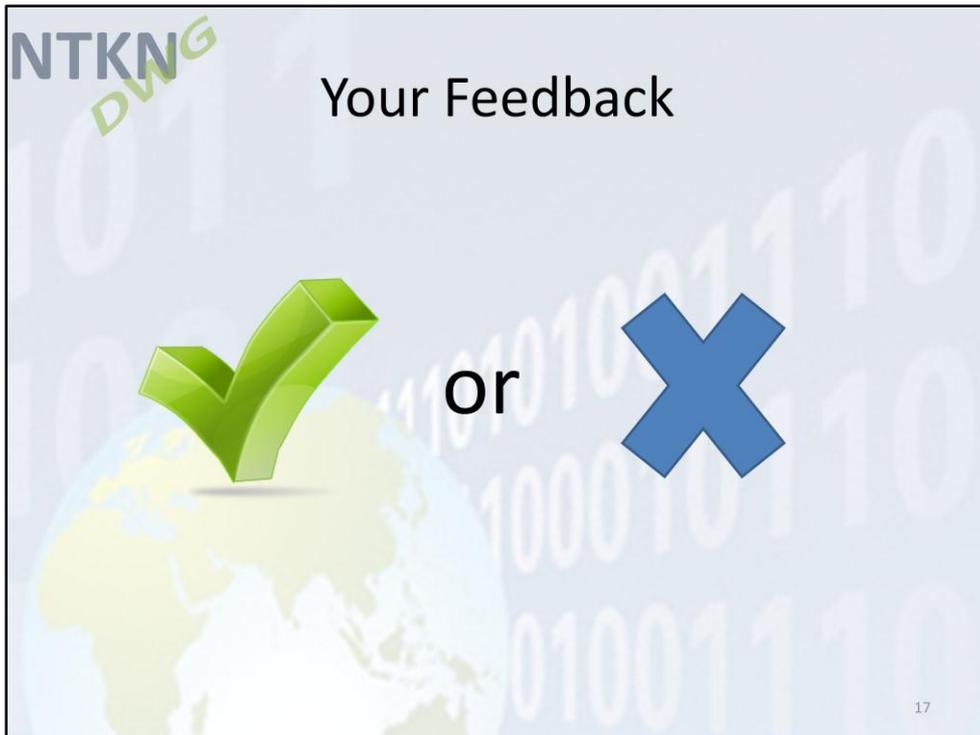
Regarding outsourcing (n=30):

- 60% do not outsource
- 17% long-term storage
- 6% System or software housing
- 0% creation of metadata
- 7% OCR
- 33% scanning

Regarding processing of digital files after scanning, any modification to files? (n=32) –
63% do not

Transportation Digitization Activity: Why and How -- Metadata

- 48% (n=31) currently embed metadata
- 26% (n=31) plan to embed metadata in the future
- 26% (n=31) do not embed. Reasons (n=8):
 - Lack human or fiscal resources (76%)
 - Lack system capability (25%)
 - Haven't thought about it (0%)
 - Don't know (38%)



Why and How Summary:

varying levels of sophistication (i.e. how much metadata is being applied) ...
example, dumb vs. smart .pdfs.

=>Louise can discuss dumb vs smart PDFs – how to convert (give case example)

Focus on access to internal agency staff

Metadata applied in-house (if at all)

Scanning projects done in-house and staff retrained to learn new skills and carry out digitization projects

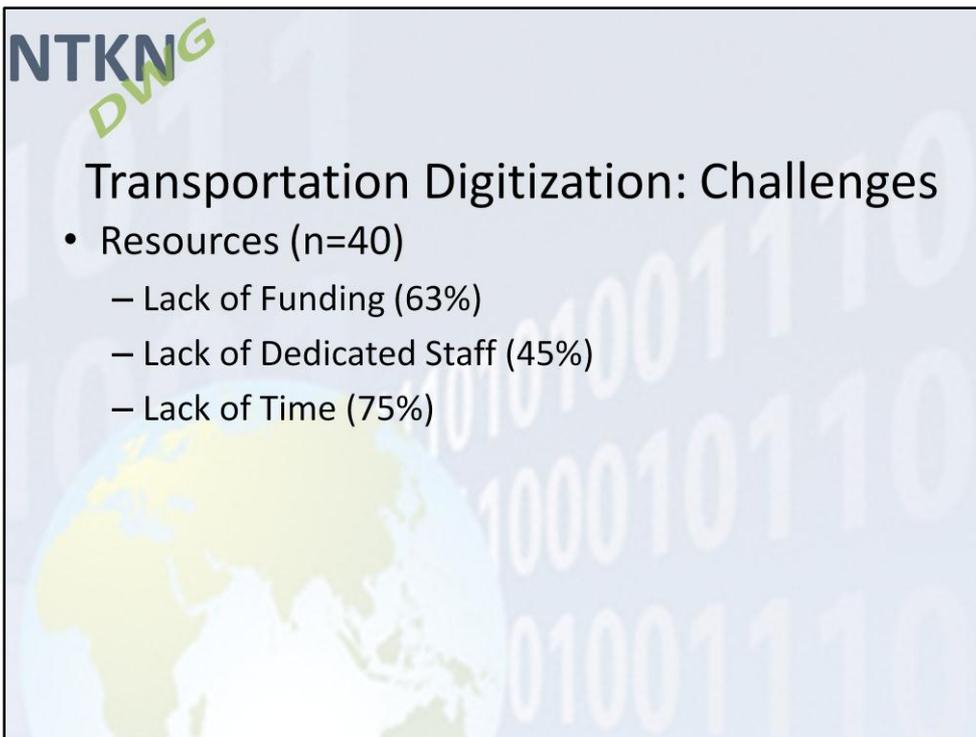
Done for as little resources as possible (volunteers, current staff, in-house equipment)

Major resource deficiencies are in writing plans, funding.

Major capabilities are in staff skills (only 30%)

Have we missed anything from our survey that is happening in your library (in terms of digitization)

What is your digitization experience? What activities are you doing?



NTKNG
DWG

Transportation Digitization: Challenges

- Resources (n=40)
 - Lack of Funding (63%)
 - Lack of Dedicated Staff (45%)
 - Lack of Time (75%)

How do we address the lack of resources for many of these libraries/organizations to do this work?

What are the copyright issues we need to be aware of? For example, too much access to proprietary documents, especially standards and specifications, or non-free TRB publications, etc.

Access issues mentioned previously present challenges.

Transportation Digitization: Challenges

- Copyright
 - Is item in the public domain?
 - Composing the correct rights statement

What are the copyright issues we need to be aware of? For example, too much access to proprietary documents, especially standards and specifications, or non-free TRB publications, etc.

Transportation Digitization: Challenges

- Copyright
 - <http://historicalphotos.iowadot.gov/> - great example of attribution statement, terms of use, and general usage guidelines
 - Paul Burley, Northwestern – excellent copyright checklist example



What are the copyright issues we need to be aware of? For example, too much access to proprietary documents, especially standards and specifications, or non-free TRB publications, etc.

Transportation Digitization: Challenges

- Access (internal)
 - Over 80% → primary audience is agency staff
 - Almost 60% → digitizing materials for internal access only

Variety of access to the materials being digitized – is some information sensitive and can only be viewed within the producer's organization (i.e. on an intranet or restricted internal network drive).

Are others using available 'free' digital repositories, such as DSpace, Hathi Trust, OCLC repositories using contentDM, such as MnDOT and WisDOT partnering with their state's historical digital collections. etc.

How do we incorporate all the work being performed by Google with our transportation collections?

Are we taking full advantage of the NTL's digital repository enough?

83% state that agency staff was a primary audience for their digitized materials and almost 60% stated that they are digitizing materials for internal access only. This is significant in that a large amount of digitized transportation information is inaccessible to the larger transportation information community. Of course, due to restrictions/privacy issues, some of this information isn't meant for public consumption, but perhaps much of it should be. Many of the respondents stated that their digitized materials are searchable...but a question we must ask is 'how searchable?' – is it searchable internally within an organization or to a larger audience.

Transportation Digitization: Challenges

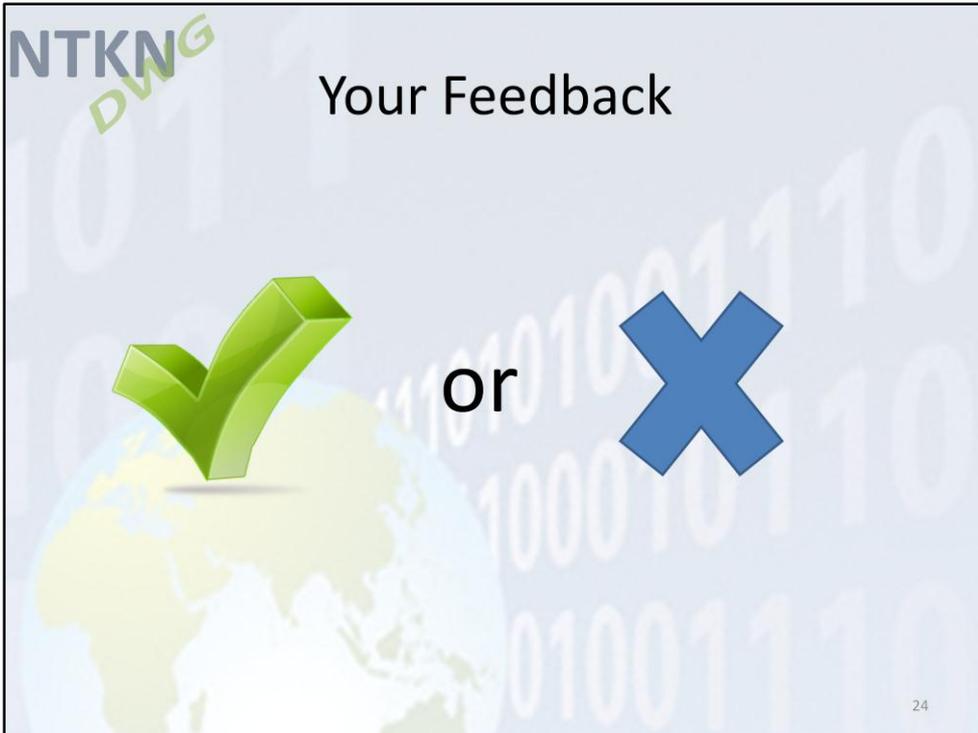
- Access (external)
 - Iowa Dept. of Transportation Historic Archives Digital Collections:
<http://historicalphotos.iowadot.gov/>
 - Northwestern Transportation Library Menu Collection:
<http://digital.library.northwestern.edu/transportation-menus/>
 - Minnesota Dept. of Transportation Collection:
<http://reflections.mndigital.org/cdm/landingpage/collection/mdt>
 - L.A. Metro Transportation Library and Archives photostream:
<http://www.flickr.com/photos/metrolibraryarchive/>
 - L.A. Metro Transportation Library and Archive film and video:
<http://www.youtube.com/user/metrolibrarian/>

In times of declining resources, the 'free' social tools approach used by Matt Baret is a great example of delivering access to a broad audience (YouTube, Flickr, etc.)

Transportation Digitization: Challenges

- Access (external)
 - Google digitization:
<http://books.google.com>
 - Hathi Trust Digital Library:
<http://www.hathitrust.org/>
 - Internet Archive:
<http://www.archive.org>

In times of declining resources, the 'free' social tools approach used by Matt Barrett is a great example of delivering access to a broad audience (YouTube, Flickr, etc.)



Challenges Summary:

Have we missed anything from our survey that is happening in your library (in terms of digitization)

What is your digitization experience? What activities are you doing?



NTKN DWG – Next Steps

- How can NTKN better support digitization efforts (n=40)
 - 67% - Identify & promote best practices
 - 49% - Take a leadership role in the development of guidelines and manuals



25

Slide = Q34

Q7. - 67% of respondents report have written policies and procedures for digitization projects.
See Q23 – Not sure how to represent the data for “writing policies and procedures”

However, when asked how capable your organization is in writing digitization plans, 30% indicated that were deficient in writing digitization plans and 35% felt they were somewhat deficient in establishing quality standard.

Q7 & 8 – not many people have written documentation and procedures

NOTES:

-bp: digitization and access

-product: toolkit to address and post on website (link to LIS best practices, and exemplars from transportation community)

NTKN DWG – Next Steps

- Identifying Best Practices

- Look at the successful transportation projects of Northwestern, L.A. Metro, Iowa DOT, Virginia DOT, Minnesota DOT and others.
- Contact national transportation organizations about their practices (TRB, AASHTO, FHWA, etc.)

You **Tube**



NORTHWESTERN UNIVERSITY LIBRARY

Transportation Library Menu Collection

NTKNG
DWG

NTKN DWG – Next Steps

- Best Practices of Professional Partners
 - Use the expertise and guidelines of established partners for transportation collections
 - WisDOT ☐ Wisconsin Heritage Online (<http://wisconsinheritage.org>)

27

The DWG group into the future – can we better connect with other interested groups within the national transportation information infrastructure (TRB LIST, AASHTO TKN RAC task force, TKNs, SLA) and other interested parties – other committees in TRB and elsewhere, such as Knowledge Management, Data/IT, History, etc..

NTKN DWG – Next Steps

- Best Practices of Professional Partners
 - Example: Scanning guidelines – Minimum Specifications and Resolution Chart

| Original material | Scanning resolution | Bit depth | Approximate scanned dimensions | Approx. size of preview image |
|----------------------|---------------------|---------------------------------|--------------------------------|-------------------------------|
| <i>Photographs</i> | | | | |
| 16" x 20" + | 200 ppi | 24-bit color | 6400 x 8000 pixels | 146 MB |
| 8 ½" x 11"–16" x 20" | 300 ppi | 24-bit color | 3200 x 4000 pixels | 36 MB |
| 8" x 10" | 400 ppi | 24-bit color | 3200 x 4000 pixels | 36 MB |
| 5" x 7" | 625 ppi | 24-bit color | 3200 x 4000 pixels | 36 MB |
| 4" x 5" | 800 ppi | 24-bit color | 3200 x 4000 pixels | 36 MB |
| 4" x 2 ½" | 1200 ppi | 24-bit color | 3200 x 4000 pixels | 36 MB |
| <i>Text</i> | | | | |
| Print—no images | 600 ppi | 1-bit bitonal | Varies | Varies |
| Print—with images | 300 ppi | 8-bit grayscale or 24-bit color | Varies | Varies |
| Manuscript | 400 ppi | 8-bit grayscale or 24-bit color | Varies | Varies |

NTKN DWG – Next Steps

- Connect with TKN and other Transportation Partners
 - NTKN Cataloging and Metadata Initiative
 - TRB LIST
 - AASHTO RAC TKN Task Force
 - Repository Working Group
 - Report Distribution Working Group
 - SLA Transportation Division
 - Other TRB Committees?

The DWG group into the future – can we better connect with other interested groups within the national transportation information infrastructure (TRB LIST, AASHTO TKN RAC task force, TKNs, SLA) and other interested parties – other committees in TRB and elsewhere, such as Knowledge Management, Data/IT, History, etc..

NTKN DWG – Next Steps

- Best Practices – Minimum Standards
 - Need to be easy enough for small outfits
 - Need to be standardized enough to create a quality product
 - Includes selection criteria, scanning, metadata, and others.

NTKN DWG – Next Steps

- Best Practices – Training
 - Software clients currently in use
 - DWG as a vetting agent?



Transportation Digitization – The Future

- Communicating Our Experience
 - Greater connectivity with our partners
 - Various report mechanisms – online, toolkit, others?

NTKN DWG – The Future: Some Ideas

- Link digital repositories for greater benefit and access to all – can we help identify and give access to previously internal-only information
- Publish minimal or standard set of best practices (toolkit idea)
- Suggest training resources, especially of the self-guided variety. Can the DWG be a vetting agent?
- Suggest resources and/or processes for monitoring copyright issues.
- For experienced digitizers, what can DWG offer?
- How do we maintain long-term access? (particularly with many being put up for internal use only)
- Who are our different audiences?
- Further investigation on the access issue for digitized content

Partners for Access

- AASHTO RAC TKN TF Report Distribution TG
- AASHTO RAC TKN TF Repositories WG
- NTKN Cataloging and Metadata Initiative
- Others?

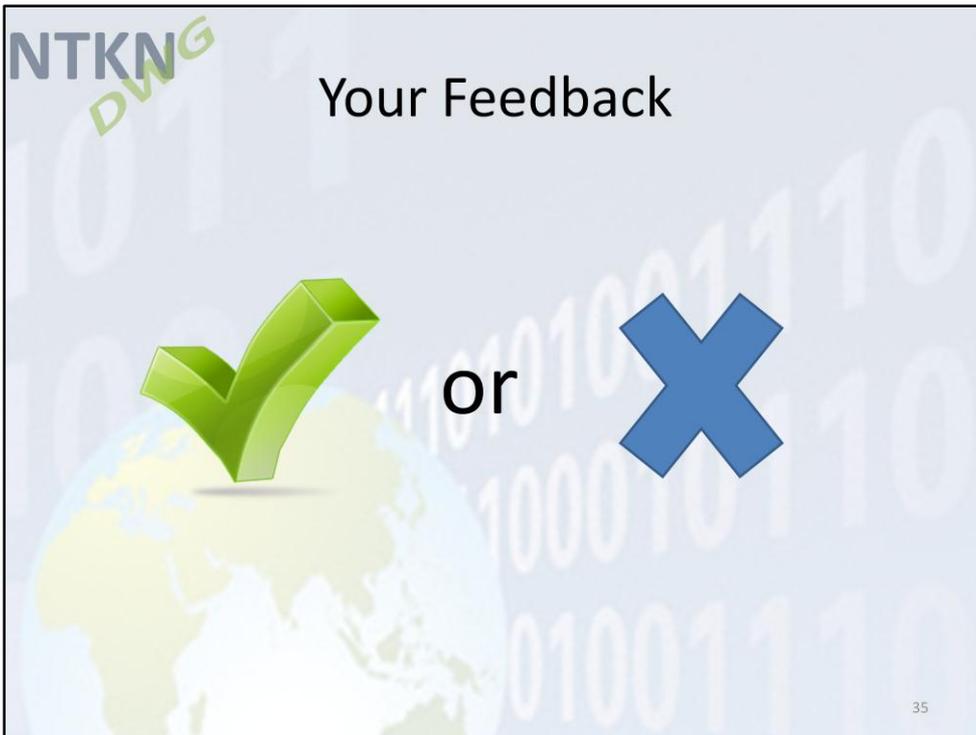
DWG – how can we help? – identify and link digital repositories for better overall access; creating a (minimal?) set of best practices for the resource-challenged digitizer to at least follow to get some standardization and better usefulness (perhaps a transportation digitization toolkit, similar to the transportation librarians toolkit- can contain standards for resolution, metadata standards to the item itself as well as policies such as written digitization plans, maybe marketing ideas.); suggest some training resources to help bridge the gap between those with no real digitizing experience and those who have been doing it ‘professionally’ for some time (Could we establish some sort of mentor-mentee program?). If we have training ideas or products to look at, could the DWG establish a baseline vetting process to determine whether any particular training resource is worthwhile?

Can we do a more in-depth scan of the types of information existing in silos and give better access to it, while remaining sensitive to copyright and other related issues.

Transportation Digitization – The Future

- New NTKN DWG Members
 - Do you see yourself in some of the data presented here?
 - Do you have processes, ideas, strategies, lessons learned, etc. to share?
 - Want to be a part of creating the DWG products for the transportation community?

We need YOU to become a part of the NTKN DWG!



What do you think?

What should be our next steps?

Have we missed anything from our survey that is happening in your library (in terms of digitization)

What is your digitization experience? What activities are you doing?

NTKN
DWG

Question? Comments? Contact any member of the NTKN DWG!

THANK YOU!

36