NCHRP Project 20-97: Improving Findability and Relevance of Transportation Information

Part I: Project Overview
Gordon Kennedy, Washington State DOT

November, 2017
NCHRP is a State-Driven Program

Sponsored by individual state DOTs who:

Suggest research of national interest

Serve on oversight panels that guide the research

Administered by the Transportation Research Board (TRB) in cooperation with the Federal Highway Administration.
Practical, ready-to-use results

- Applied research aimed at state DOT practitioners
- Often become AASHTO standards, specifications, guides, syntheses
- Can be applied to planning, design, construction, operations, maintenance, safety, environment
Agenda

Introduction: Gordon Kennedy, WSDOT
Webinar Overview
Purpose: Why do an NCHRP Project on Findability?
Scope: What was NCHRP 20-97 About?
Product: What was Accomplished?
Next Steps: Implementing the Guidance
Introduction

Gordon Kennedy

- Information Resources Manager, WSDOT
- Member of NCHRP 20-97 Project Panel
NCHRP Project 20-97

Part I: Project Overview

Part II: Pilot Demonstration
  - Dr. Maureen Hammer, Battelle Memorial Institute (formerly with Virginia DOT)

Part III: Guide for State Transportation Agencies
  - Frances Harrison, Spy Pond Partners, LLC (Principal Investigator)
Why is Information Important?

DOTs Are Information Management Organizations

- Optimize the use of resources and assets
- Integration across disciplines, modes, and jurisdictions
- Data-driven decisions and priorities
- Accountability and openness
What is Findability?

…the degree to which relevant information is easy to find.

- Accessibility of data & information
- Search tools for discovering what is relevant
- Documentation & curation to identify authority and quality
Why is Findability Important?

Costs and Risks of Poor Findability

- Employees waste time searching
- Valuable information goes unused
- Employees re-create data that exists already
- Employees use outdated versions of documents
- Difficult and time-consuming to respond to information requests
- Difficult and time-consuming to produce evidence in response to claims
Example from WSDOT

Corridor Planning

- Consider a transportation corridor holistically to optimize resources for a practical solution
- Data should flow across jurisdictions and disciplines: local planning organizations > local WSDOT offices > central WSDOT offices.
- Assembling the information required is cumbersome:
  - Assembling data from disparate sources
  - Discovering information from disparate repositories
What can be Done?

Technology

- Content and document management systems
- Search tools
- Text analytics tools

Planning and Architecture

- What are your needs?
- What should be findable?
- Plan for searching within and across repositories
Project Purpose

What can be Done?

Policies

- What to store where and for how long
- File naming conventions
- Metadata standards

Processes

- Curation and designation of important content
- File clean-up
- Metadata assignment – manual or semi-automatic
Project Purpose

What can be Done?

Deliberate Intent

- It takes planned and coordinated work to improve findability
- It takes people, process and technology
Project Scope

Work Tasks of this project

**Literature Review** – preceding work and knowledge

**Interviews** – DOT’s and private sector

**Framework Development** – methods & techniques to find

**DOT Pilot** – trial of techniques at two DOTs

**Guide to Improving Findability** – publish findings
Project Scope

Pilot Study

Virginia DOT – primary agency

- Construction documents, mostly PDF, DOC and HTML formats
- Test text analytics software to automatically tag documents with project ID, construction issue, district, etc.
- Demonstrated capability to search by combinations of criteria

Washington State DOT – validation

- Applied classification rules to similar documents
- Rules were mostly transferable – some would require adjustment
Product

Guide to Improving Findability
Next Steps

Continuation Phase

Implementation

- Work with 2-4 agencies to apply the guidance and develop specific implementation plans.
- Document results and develop general guidance.
- Timing: Fall 2017 – Winter 2019

For Further Information…

- Contact Andrew Lemer, NCHRP alemer at nas.edu
NCHRP Project 20-97
Improving Findability and Relevance of Transportation Information

Part 2
Findability Demonstration at Virginia DOT

Dr. Maureen Hammer
Battelle Memorial Institute

November, 2017
Agenda

01 Introduction
   Dr. Maureen Hammer

02 Objectives
   Purpose of the Pilot

03 Context
   VDOT’s Findability Challenges

04 Approach
   Pilot Activities

05 Findings
   What was Learned?
Maureen Hammer

- Principal Research Scientist (Knowledge Management) at Battelle Memorial Institute

- Formerly served as Virginia DOT Knowledge Management lead

- Virginia DOT liaison for the NCHRP 20-97 Pilot
Objectives

Purpose of the Pilot

- Demonstrate and validate concepts and methods to improve findability
- Assess effort required and transferability to other DOTs
VDOT’s Findability Challenges

- Time consuming to find documents – with confidence that they are authoritative/final
  - New Construction Document Management System intended to address these areas – but won’t handle historical files

- Risk of document loss – e.g. when employee leaves and hasn’t filed everything

- Responding to FOIA, audits, NOIs, claims
  - Duplication makes application of retention schedules difficult

- Construction identified as major risk area
VDOT’s Parallel Efforts

- Corporate Documents Repository
  - One authoritative location for all critical documents (e.g., policies and procedures)

- FOIA Tracking System
  - Track requests and information located; minimize duplication

- Construction Document Management System
  - Track a project from pre-bid to completion
Pilot Activities

Approach

Assessment → Content Collection → Text Analytics Application → Solution Demonstration → Transferability Analysis
1. Find a Single Known Document for a Project (e.g. an estimate) – variety of search criteria

Find construction documents based on one or a combination of: tax map parcel, project number/UPC, project type (paving, bridge, etc.), fixed completion date (for active projects), construction document type, district, county, route, owner, contractor, subcontractor, cost range, types of material, item code/category, responsible charge engineer.

2. Find/Review All Documents for a Project (e.g. for FOIA Request)

Provide all records (or certain types of records) for a particular project (in response to a FOIA request, audit request, notice of intent (NOI), claim investigation, or Construction Quality Inspection Program check.

3. Search Across Projects – Find Projects with Item, Material, Construction Technique

Find recent projects that installed a Trinity Guardrail GR-9; Find projects that use a particular pay item. (to guide selection of pay items to include on a project currently being designed)

4. Research Reasons for Delays and Changes

Analyze systemic issues that contribute to construction projects not meeting goals for on-time, on-budget, environmental and quality scores. Currently very time consuming to do this research.
Content Collection

- Correspondence
- Meeting Minutes
- Contracts
- Work Orders (C-10)
- Daily Work Reports/Inspector Diaries (C-84)
- Material Documentation (C-85)
- Source of Materials Forms (C-25)
- Subletting Request (C-31)
- EEO Reports (C-64)
- Estimates (C-79)
- Starting and Completion (C-5)
- Vouchers
- Price Adjustments
- Blast Reports
- Environmental Compliance Reports
- Contractor Inspection Reports
- Certified Payroll
- Design Field Changes
- Job Mix Designs
- Materials Test Results
- Insurance Certificates
- Tracking Logs
- Notice of Intent
- Claims
Approach

Text Analytics Application

- Used commercial text analytics software
  - Developed rules for classification based on analysis of sample documents
  - Iterative testing and refinement

- User feedback
  - Validated with information experts
Approach

Text Analytics Rules

- Content Type
- Project No/Contract No
- Manufacturers and Suppliers
- Contractors
- Materials
- Equipment
- Pay Items
- Category of Work Order
- Work Issue
  - Drainage
  - Utility
  - Weather
  - Plan-Related
  - Work Zone-Related
Solution Demonstration

- Allow user to narrow down scope based on different categories
- All documents loaded to single repository – but similar approach could work across repositories – point search index at multiple locations
- Used standard commercial search engine
Solution Demonstration

Collected metrics comparing “Vanilla” search with faceted search on auto-classified documents

Examples:
Recall of Work Orders for specific project (in Top 30 Results)

Precision of work orders related to utility issues (for Top 20 Results)
Approach

Transferability Analysis

- WSDOT provided 3 content types – analogous to VDOT’s:
  - 134 change orders
  - 100 inspector’s daily work reports
  - 112 request for approval of material

- Analysis
  - Applied rules developed for VDOT directly to WSDOT content

- Findings
  - Some rules worked well “as is”
  - Others would take minimal effort to “tweak” to match WSDOT terminology
Findings

What was learned?

- DOTs have common “pain points” when it comes to findability
  - Public Disclosure response a particular concern

- Commercial tools are available that can partially automate assignment of metadata
  - Rules for tagging can be complex - look for combinations of terms, in specific parts of documents, etc.

- Rule development is time-consuming but once rules are developed they can be applied repeatedly
  - Rule development only makes sense for relatively large collections of similar document types – construction daily work reports were a good test case

- Because DOTs have similar business processes, rules developed at one DOT can be leveraged at another agency
  - This means that there is the potential for further collaboration on auto-classification of content across DOTs
NCHRP Project 20-97
Improving Findability and Relevance of Transportation Information

Part 3
A Guide for State DOTs

Frances D. Harrison
Spy Pond Partners, LLC

November, 2017
Agenda

01 Introduction
Frances Harrison

02 Overview
What’s in the Guide

03 Audience
Who are the Target Users?

04 Highlights
Tour of the Guide
Frances D. Harrison

- Principal Investigator for NCHRP Project 20-97: “Improving Findability and Relevance of Transportation Information”
- Chair of TRB Knowledge Management Task Force
- Founding partner of Spy Pond Partners, LLC
Overview

What’s in the Guide?

Understanding
What is findability?
What are the impediments to findability?

Improving
What strategies are available for improving findability?

Planning
What factors do we need to consider when planning findability improvements?

Implementing
How should we approach implementing findability improvements?
What management functions are needed?
Who are the Target Users?

People serving the following types of functions:

- Data and information management
- Internet/Intranet site management
- Knowledge management
- Records management
- Library management
- Document management
- Engineering content management
- Publications
- Staff development and training
Highlights

Tour of the Guide

DOT Information Findability

**Business Drivers (Chapter 1)**
- Reduced Time Spent Searching for Information
- More Re-use of Information, Less Re-work
- Ensure Use of Authoritative Information
- Efficient Response to FOIA Requests, Claims

**Planning (Section 4)**
- User Needs
- Information Landscape

**Implementing (Section 5)**
- Road Map
- Management Functions

**Information Management (Section 3.1)**
- Document/Content Management Systems
- Content Storage & Clean-Up Policies and Practices
- File Naming Conventions Scanning Practices
- Security and Access Controls

**Search & Navigation (Section 3.2)**
- Enterprise Search (within and across repositories)
- Faceted Navigation Auto-Suggest
- Search Monitoring & Tuning
- Search-Based Applications

**Metadata & Terminology (Section 3.3)**
- Standard Agency Metadata Elements & Content Types
- Standard Classifications (asset, work type, etc.)
- Terminology Resources (synonyms, related terms)
- Automated Metadata Creation
## Business Drivers for Findability

<table>
<thead>
<tr>
<th>Reduced Time Spent Searching for Information</th>
<th>More Re-use of Information, Less Re-work</th>
<th>Ensure Use of Authoritative Information</th>
<th>Efficient Response to FOIA Requests, Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Hourly Rate</td>
<td>Average Time Savings per Employee per Day</td>
<td>Annual Savings per 1,000 employees</td>
<td>Present Value of Savings over 10 years (3% discount rate)</td>
</tr>
<tr>
<td>$20</td>
<td>15</td>
<td>$1.2 M</td>
<td>$9.8 M</td>
</tr>
<tr>
<td>$30</td>
<td>15</td>
<td>$1.7 M</td>
<td>$14.7 M</td>
</tr>
<tr>
<td>$40</td>
<td>15</td>
<td>$2.3 M</td>
<td>$19.6 M</td>
</tr>
<tr>
<td>$40</td>
<td>30</td>
<td>$4.6 M</td>
<td>$39.2 M</td>
</tr>
</tbody>
</table>
Tour of the Guide

Understanding Findability

Information Need

Desired Results

Information Seeker

Search

Body of Available Content
Understanding Findability

- Users must know how and where to search
- Search tools must be in place and configured properly
- Content must be in a searchable location
- Metadata will always improve search results
Tour of the Guide

Understanding Findability

- Impediments to Findability
  - Lack of Disciplined Information Management Practices
  - Non-Searchable Information
  - Lack of Common Repositories and Search Tools
  - Lack of Metadata and Managed Terminology
<table>
<thead>
<tr>
<th>Information Management (Section 3.1)</th>
<th>Search &amp; Navigation (Section 3.2)</th>
<th>Metadata &amp; Terminology (Section 3.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document/Content Management Systems</td>
<td>Enterprise Search (within and across repositories)</td>
<td>Standard Agency Metadata Elements &amp; Content Types</td>
</tr>
<tr>
<td>Content Storage &amp; Clean-Up Policies and Practices</td>
<td>Faceted Navigation Auto-Suggest</td>
<td>Standard Classifications (asset, work type, etc.)</td>
</tr>
<tr>
<td>File Naming Conventions Scanning Practices</td>
<td>Search Monitoring &amp; Tuning</td>
<td>Terminology Resources (syonyms, related terms)</td>
</tr>
<tr>
<td>Security and Access Controls</td>
<td>Search-Based Applications</td>
<td>Automated Metadata Creation</td>
</tr>
</tbody>
</table>
Tour of the Guide

Planning for Findability Improvements

User Needs
Target Scope
Content Analysis
Info Management Practices
Holistic Plan for Improvement
Implementing Findability Improvements

Tour of the Guide

1. Create Architectural Vision for Findability:
   - Shared Repositories
   - Common Metadata Elements
   - Common Terminology
   - Master Data
   - Enterprise Search

2. Identify a Focus Area
3. Conduct an Assessment
4. Identify Candidate Improvements

5. Implement Quick Wins
6. Implement a Pilot
7. Expand and Formalize the Pilot

Refine and Build
Implementing Findability Improvements

Example Improvements

<table>
<thead>
<tr>
<th>Construction Project Information</th>
<th>Critical Corporate Documents Repository</th>
<th>Home Pages for Critical Job Functions</th>
</tr>
</thead>
</table>
| • Search for project information – across content types and repositories  
  • Contracts, change orders, correspondence, inspector notes  | • Search for authoritative copies  
  • Manuals, guidance, specs, plans  | • Find everything I need to know to do my job  
  • People, documents, data sets, training materials  |
Implementing Findability Improvements

Common Solution Elements

- Information Governance
- Terminology and Tagging
- Tools
- Training
- Content Conversion
Tour of the Guide

Resources

- Search Team Roles
- Commercial Products
- Glossary
- Performance Measures
- Info Org Schemes
- Tutorials
Search Team Roles

Manager
- Lead responsible for improvements

Technology Manager
- IT role – performance, backup, security, applications

Analytics Manager
- Search results monitoring and diagnostics

Information Specialist
- Metadata and taxonomy, best bets

Support Manager
- User training, usability, communication
Resources

Performance Measures

- Content Availability
- Search Time
- Search Precision and Relevance
- Search Success Rate
- User Satisfaction
- Content Processing Effort
- Metadata Quality
For Further Information…

http://www.trb.org/Main/Blurbs/176113.aspx

Contact:
Frances Harrison (PI)
fharrison at spypondpartners.com

Andrew Lemer (NCHRP Staff)
alemer at nas.edu