

## Plan Overview

---

*A Data Management Plan created using DMP Tool*

**Title:** SMART Counties in Kansas

**Creator:** Deb Ohlde - **ORCID:** [0009-0002-3602-8868](https://orcid.org/0009-0002-3602-8868)

**Affiliation:** United States Department of Transportation (DOT) ([transportation.gov](https://www.transportation.gov))

**Principal Investigator:** Dr. Qingbin Cui

**Project Administrator:** Debra Ohlde

**Contributor:** Dr. Yunfeng Zhang, Natnael Taye Tsegaye

**Funder:** United States Department of Transportation (DOT) ([transportation.gov](https://www.transportation.gov))

**Funding opportunity number:** NOFO #20.941

**Grant:** 69A3552341015 SMARTFY22N1P1G24

**Template:** SMART Grants Stage 1 Data Management Plan (DMP)

### Project abstract:

Prototype a GIS based platform that integrates local infrastructure assets to provide innovative data, monitor asset conditions, and inform technological solutions in Kansas.

**Start date:** 08-01-2023

**End date:** 02-01-2025

**Last modified:** 11-10-2025

---

## SMART Counties in Kansas

### Dataset and Contact Information

---

Please provide as much of the the following information as possible:

1. Name of the project;
2. Grant number;
3. Name of the person submitting this DMP;
4. ORCID of the person submitting this DMP (need an ORCID? Register here: <https://orcid.org/>);
5. Email and phone number of the person submitting this DMP;
6. Name of the organization for which the person submitting this DMP is working;
7. Email and phone number for the organization;
8. Link to organization or project website, if applicable; and,
9. Date the DMP was written.

1. SMART Counties in Kansas
2. 69A3552341015 SMARTFY22N1P1G24
3. Debra Carlson Ohlde
4. 000--0002-3602-8868
5. communitydevelopmentadvisor@ncrpc.org - 785-275-2499
6. North Central Regional Planning Commission
7. contact@ncrpc.org - 785-738-2218
8. <https://www.ncrpc.org/>
9. December 14, 2023

### Data Description

---

Please provide as much information as possible:

1. Provide a description of the data that you will be gathering in the course of your project or data from a third party that you will re-use, if any;
  1. If there will be no data collected or re-used from another source, state that this is case;
2. Address the expected nature, scope, and scale of the data that will be collected, as best as you can at this stage;
3. As best as you can, describe the characteristics of the data, their relationship to other data, and provide sufficient detail so that reviewers will understand any disclosure risks that may apply;
  1. If data might be sensitive, please describe how you will protect privacy and security, if you know that now;
  2. You may need to update your DMP later to add more detail;
4. Discuss the expected value of the data over the long-term.

This project will generate and use visual inspection data of bridge infrastructure collected through Unmanned Aerial Vehicle (UAV) systems. The primary data consist of high-resolution images and video recordings captured during field inspections. These visual datasets document concrete surfaces and structural components where material deterioration or damage may be present (e.g., cracking, spalling, rusting, exposed reinforcement, efflorescence, moisture staining). The imagery will be collected in a systematic manner to ensure consistent coverage of the inspected structures.

The project will also produce derived data products from computer vision and machine learning workflows. These include: (1) pixel-level segmentation masks identifying eight concrete defect types, (2) bounding box detections and classification labels for images and videos, and (3) defect count summaries per image. These datasets will be provided in non-proprietary formats including text files, JPG/PNG, and MP4.

In addition, the project will generate GIS-based visualization products, including ArcGIS StoryMaps and associated geospatial layers that integrate inspection imagery, defect annotations, and structure-level attributes (e.g., National Bridge Inventory (NBI) condition ratings and structure identifiers). Where possible, these geospatial products will be

exported into open formats (e.g., GeoJSON, shapefiles, or web-published map layers). Original ESRI layer packages may also be included to preserve full map functionality.

Where applicable, publicly available reference data (e.g., bridge location information and structure identifiers) may be re-used to support geospatial accuracy and situational context. No personally identifiable information (PII) or human subject data are collected as part of this project, and therefore no disclosure risk is anticipated. The data hold long-term value for transportation agencies, researchers, and practitioners developing automated inspection methods or conducting longitudinal monitoring of infrastructure condition.

## Data Format and Metadata Standards Employed

---

**Please provide as much information as you can:**

- 1. Describe the anticipated file formats of your data and related files;**
- 2. To the maximum extent practicable, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future;**
  - 1. If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.**
- 3. Identify the metadata standards you will use to describe the data.**
  - 1. At least one metadata file should be a DCAT-US v1.1 (<https://resources.data.gov/resources/dcat-us/>) .JSON file, the federal standard for data search and discovery.**

This project will generate several categories of research outputs. The primary data products consist of annotated images in JPG/PNG format and annotated video files in MP4 format, where model-detected concrete defects are visualized through pixel-level overlays. Defect classification results and per-image defect count summaries will be provided as plain text (TXT) files to ensure direct accessibility without requiring proprietary software.

Geospatial visualization outputs are provided through ArcGIS StoryMaps, which integrate inspection imagery, defect annotations, structure identifiers, and National Bridge Inventory (NBI) condition attributes. These StoryMaps are maintained and published in ESRI GIS formats to support interactive map layers and spatial visualization. For long-term openness and reusability, corresponding layers will also be exported in open geospatial formats such as GeoJSON and, where appropriate, Shapefiles, to ensure access outside of ESRI environments.

The trained concrete defect segmentation model will be shared in ONNX format, which is an open, framework-agnostic model representation standard. The deployed inference pipeline is containerized and may be executed as a FastAPI web service without requiring proprietary machine learning software. The complete training and deployment workflows will be archived and publicly released through Zenodo to support transparency and reproducibility.

All data releases will be accompanied by documentation describing dataset organization, variable and class definitions, coordinate reference information (for geospatial data), and model processing steps. A DCAT-US v1.1 compliant metadata JSON file will be included to satisfy USDOT public access requirements and support dataset discovery and indexing. Additional supporting documentation will be provided in TXT and PDF formats.

## Access Policies

---

**In general, data from DOT-funded projects must be made publicly accessible. Exceptions to this policy are: data that contain personally identifiable information (PII) that cannot be anonymized; confidential business information; or classified information. Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take to protect privacy and confidentiality prior to archiving your data, and any additional concerns. In general, in matters of human subject research, your DMP should describe how your informed consent forms will**

permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB), may be used to protect privacy and confidentiality. Additionally, when working with, or conducting research that includes Indigenous populations or Tribal communities, researcher will adhere to the CARE Principles for Indigenous Data Governance <https://www.gida-global.org/care> and make an explicit statement to that effect in this portion of the DMP.

Please provide as much information as possible:

1. Describe any sensitive data that may be collected or used;
2. Describe how you will protect PII or other sensitive data, including IRB review, application of CARE Principles guidelines, or other ethical norms and practices;
  1. If you will not be able to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, you should describe the necessary restrictions on access and use;
3. Describe any access restrictions that may apply to your data;
4. If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators or other project staff.

#### Access Policies

Data generated from this project will not be inherently proprietary in nature, however the specific aspects of bridge conditions, construction and location may be considered sensitive information. There will not be any data that would be related to individual persons collected as part of this project.

Access to data would be governed by individual county policies similar to other road and bridge data currently maintained. Data collected as part of this project will be shared among all those participating as partners in the project, subject to policies to be established.

#### Re-use, Redistribution, and Derivatives Products Policies

---

Recipients are reminded:

1. Data, as a collection of facts, cannot be copyrighted under US copyright law;
2. Projects carried out under a US DOT SMART Grants is federally funded; therefore, as stated in grant language:
  1. Recipients must comply with the US DOT Public Access Plan, meaning, among other requirements, project data must be shared with the public, either by the researchers or by US DOT;
  2. That by accepting US DOT funding through this grant, recipients have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all project outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,
  3. In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where grant recipients elect to retain title to any invention developed under this grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.

Please provide as much information as possible:

1. Describe who will hold the intellectual property rights for the data created or used during the project;
2. Describe whether you will transfer those rights to a data archive, if appropriate;
3. Identify whether any licenses apply to the data;

- 1. If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP;**
- 4. Describe any other legal requirements that might need to be addressed.**

Re-use, Redistribution, and Derivative Products Policies:

The Smart Counties Project encourages the re-use of its data for various purposes, including research, education, and community planning. The open accessibility of the data fosters collaboration and innovation, inviting users to leverage the dataset for their specific needs. The redistribution policy encourages users to share and distribute the data, extending the project's impact across diverse applications. Furthermore, the project supports the creation of derivative products, enabling users to develop publications, applications, or visualizations based on the original dataset. While users are free to build upon the data, proper attribution to the Smart Counties Project is required, ensuring due credit for the source material. This combination of re-use, redistribution, and derivative products policies is designed to facilitate widespread access and application of the project's valuable dataset, maximizing its potential impact and contributing to broader advancements in transportation safety and infrastructure management.

## Archiving and Preservation Plan

---

**Please provide as much information as possible:**

- 1. State where you intend to archive your data and why you have chosen that particular option;**
- 2. Provide a link to the repository;**
- 3. You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability;**
  - 1. Whatever archive option you choose, that archive should support the capture and provision of the US Federal Government DCAT-US Metadata Schema**  
<https://resources.data.gov/resources/dcat-us/>
- 4. In addition, the archive you choose should support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data;**
- 5. Your plan should address how your archiving and preservation choices meet these requirements.**

At the completion of the project, the full dataset, trained model, geospatial layers, documentation, and associated workflows will be archived in Zenodo (<https://zenodo.org/>), which is a USDOT-preferred, open-access repository that supports long-term preservation and assigns persistent Digital Object Identifiers (DOIs). Zenodo meets the requirements for public accessibility, version retention, dataset citation, and long-term discoverability.

The archived dataset will include:

Annotated images (JPG/PNG) and annotated videos (MP4);  
Defect label and count summary files (TXT);

Documentation and reports will be archived in PDF format to ensure accessibility. Metadata standards compliance will be maintained to ensure the integrity of descriptive information associated with the datasets. Access controls will be implemented to regulate data access and ensure compliance with re-use and redistribution policies. The archiving and preservation plan will be periodically reviewed and updated to align with evolving data management standards and best practices.

---

## **Planned Research Outputs**

### **Dataset - "20+ Annotated UAV Bridge Inspection Imagery and Videos"**

This dataset contains more than 20GB of UAV-captured bridge inspection imagery (JPG/PNG) and video recordings (MP4) annotated with pixel-level defect overlays generated using a trained YOLOv11 segmentation model. These files document surface-level concrete deterioration patterns such as cracking, spalling, rusting, efflorescence, and exposed reinforcement.

### **Dataset - "Defect Classification Labels and Count Summaries (TXT)"**

This dataset consists of plain text files that record detected defect types and defect instance counts for each annotated image or video frame. These outputs support quality assessment, model validation, and condition trend analysis.

### **Dataset - "ArcGIS StoryMap and Geospatial Layers for Bridge Condition Visualization"**

This dataset includes the published ArcGIS StoryMap and associated geospatial layers integrating inspection imagery, defect annotations, structure identifiers, and NBI condition ratings.

### **Software - "Trained YOLOv11 Segmentation Model and Containerized FastAPI Inference Service"**

This research output includes the trained YOLOv11 segmentation model in ONNX format and the accompanying containerized FastAPI inference service and workflows. This enables reproducible model deployment without requiring proprietary machine learning software.

---

## **Planned research output details**

<b>Title</b>	<b>Type</b>	<b>Anticipated release date</b>	<b>Initial access level</b>	<b>Intended repository(ies)</b>	<b>Anticipated file size</b>	<b>License</b>	<b>Metadata standard(s)</b>	<b>May contain sensitive data?</b>	<b>May contain PII?</b>
20+ Annotated UAV Bridge Inspection Imagery and Vi ...	Dataset	2025-11-09	Open	Zenodo		Creative Commons Attribution 4.0 International	DCAT-US	No	No
Defect Classification Labels and Count Summaries ( ...	Dataset	2025-11-09	Open	Zenodo		Creative Commons Attribution 4.0 International	DCAT-US	No	No
ArcGIS StoryMap and Geospatial Layers for Bridge C ...	Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Trained YOLOv11 Segmentation Model and Containeriz ...	Software	2025-11-09	Open	Zenodo		Creative Commons Attribution 4.0 International	DCAT-US	No	No