Data Management Plan

INSpecting and Preserving Infrastructure through Robotic Exploration (INSPIRE)
University Transportation Center (INSPIRE UTC)

Data description

Within INSPIRE UTC, data is defined as any systematic collection of information that can be used by researchers for analysis and/or by practitioners to guide their practice. Data will be collected from sensors, nondestructive testing devices, cameras, robots, data mining and analysis algorithm, physical and numerical models, and field investigations. It is mainly grouped into:

- *Observational data* that is captured in real-time and usually irreplaceable/perishable. Examples include sensor data, telemetry, survey data and sample data during an extreme event.
- *Experimental data* that is obtained from laboratory equipment and often reproducible. Examples include the magnetic field of magnets by a magnetometer and the wavelength change of fiber optic sensors by an optical interrogator.
- *Simulation data* that is generated from test models where model and metadata (inputs) are more important than output data. Examples include the deterioration model of bridge components, and the fragility curve of bridge systems.
- *Derived or compiled data* that is reproducible. Examples include test and data mining, compiled database, 3D visualization models, and data gathered from public documents, such as the Bridge Portal managed by Federal Highway Administration.

Specifically, the types of data include ASCII numbers, images, videos, test sequences, metadata for illustration, and executable code of software. The period of time data will be collected and the frequency of update of data sets will be given together with a necessary document of data-use guidelines.

In the long term, the collected and recorded data can potentially be used by the general public to characterize and verify the performance of various sensors and devices, develop and validate numerical models of different physical phenomena, develop the preservation strategies of bridges, and understand the effectiveness of workforce training programs. In the case that intellectual properties are involved, associated data will be made available to the public after provisional patents have been filed by respective partner institutions.

Scholars' Mine at Missouri S&T is a permanent data repository that will be used to store all the data collected by INSPIRE UTC. A scholarly communications librarian (e.g., Roger Weaver) will manage the institutional repository, check for adherence to this plan, and approve the final submission of data sets prepared by all principal investigators and uploaded to the repository by the Administrative Coordinator (e.g., Abigayle Sherman) at INSPIRE UTC. Scholars' Mine is an open-access digital archive for the scholarly output of the Missouri S&T community. This institutional repository is part of the Digital Commons network of repositories and provides a perpetual archive, indexed by major search engines. The data and metadata are managed by professional librarians with technical assistance from information technology professionals. Users will be asked to provide appropriate attribution if any of the data is used.

Data format and metadata standards

Data will be presented in the format of text descriptions, numbers, tables, figures/illustrations, photos, images, and videos. Texts, numbers, tables, figures, photos, images, and derived equations will be prepared and/or summarized in Word, Excel, and Power Point Presentation according to American Society of Civil Engineers (ASCE) journal standards due to their wide acceptance among potential users in civil engineering. They will be saved in pdf format for curating. Various versions of data sets will be signified with a number in sequence added at the end of file names. Numbers are specified with no more than 3 significant digits. Images are prepared in JPEG format with a minimum of 300 dpi. Video clips are recorded up to 5 minutes at a minimum rate of 30 fps. Text descriptions in Word document and metadata schema in pdf document will be used to summarize and explain data sets for easy understanding by readers.

Policies for access and sharing

Data will be posted on the Missouri S&T institutional repository website (https://ntl.bts.gov/publicaccess/evaluatingrepositories.html). They will be made available within 6 months of their completion or on the date of publication, whichever comes first. Exception to timely posting is data that is closely related with intellectual property. In that case, the data will be posted after the provisional patent has been filed.

Interested users may access to the above website to download data. All requests for downloading data from the website will be free of charge. However, potential users of the original data that the center investigators use and include in website posting are expected to acknowledge the original data producer. Potential users of the processed data posted on the website are expected to acknowledge the principal investigator (PI) of each project and the U.S. Department of Transportation's Office of the Assistant Secretary of Research and Technology support under the auspices of INSPIRE University Transportation Center.

To ensure that there are no issues with protection of privacy, confidentiality, and intellectual property, each project PI will be required to sign the disclosure statement at the end of this plan before any project starts. Another disclosure statement will be signed by the PI before actual data sets are posted on the Missouri S&T repository website.

Policies for re-use, redistribution, derivatives

There will be no permission restriction placed on the processed data from the Center in order to encourage widespread uses in research community. Although PIs hold the intellectual property rights to the data they generate, the INSPIRE UTC has the right to use the data to promote the Center. Once PIs sign the data posting disclaimer, the Center Administrative Coordinator will post the data on the repository website. In case data sets are collected with copyrighted instruments, the collected data may be copyrighted, depending on the prior agreement between the PIs and vendors. In that case, data will be posted on the repository website only after a disclosure statement has been signed by both the PIs and the vendors.

Plans for archiving and preservation

Scholars' Mine at Missouri S&T is a hosted and cloud-based service provided by Berkeley Electronic Press (Bepress) (https://www.bepress.com/reference_guide_dc/safeguarding-content-

digital-commons/). Bepress maintains a robust infrastructure, multiple archival options, regular maintenance schedules, and an array of services protecting hosted content against a variety of threats, from the passage of time to direct, malicious attacks. All production servers are maintained at a high availability colocation facility with multiple backbone connections and backup generators. The facility is secure and requires physical tokens (badges) for access. Bepress maintains failover web, database, and storage servers to continue to serve content in case of failures. All databases have real-time redundancy that runs continuously. Daily backups of the entire database are made and stored. The nightly backups are stored away from their colocation facility in a separate physical location. All uploaded files are stored in triplicate in a redundant storage cluster, as well as backed up offsite to a third-party cloud service, Amazon Glacier, that specializes in data archiving and backup. Monthly backups of all other data is sent to third-party archival service, an industry leader in data protection and recovery services. The archival service maintains backup tapes for one year.

In addition to the services provided directly by Bepress, Curtis Laws Wilson Library at Missouri S&T maintains a separate Amazon S3 server service. This service mirrors the Bepress servers in real-time and adds an additional level of security and redundancy. Also, any data or documents submitted to Curtis Laws Wilson Library for inclusion in Scholars' Mine are maintained on a third set of servers managed locally by Missouri S&T's IT Office until they are uploaded into Scholars' Mine.

The long-term strategy for maintaining, curating, and archiving data involves data deposition in Scholars' Mine for long-term preservation and continuing data access by the research community. Scholars' Mine provides a perpetual archive, indexed by major search engines, follows accepted backup and archival practices, and is managed by professional librarians with technical assistance from information technology professionals. Data cleaning and/or anonymization will be done prior to preservation and/or sharing of the data. Along with the data and metadata, additional documentation to be deposited will consist of archival journal papers and conference papers after all copyright and publisher requirements are met.

Disclaimer

I have read the above INSPIRE UTC data management plan and agreed to prepare data derived from my projects funded by the INSPIRE UTC according to the plan requirements.

Signed by Project Principal Investigator: _____

Name of the Principal Investigator (print):   _____

Date: _____