

Data Curation Plan

David Cavazos

Federal Aviation Administration Rulemaking, Media and Corporate Influence Data

1. Project Review
 - a. Initial project discussion with Dr. Cavazos
 - i. Several librarians met with Dr. Cavazos on March 15th, 2012 to discuss the archiving of multiple data sets, including FAA rulemaking data, vehicle crash data from the NHTSA and other government sites, as well as research on business school accreditation and career paths of PhD recipients relative to the prestige of the granting institution.
2. Research Documentation
 - a. Transfer a copy of data to the University Libraries
 - i. Spreadsheets containing data related to the FAA and NHTSA regulatory processes were received via email on April 25th, 2012.
 - b. Perform Data Interview with Dr. Cavazos
 - i. The data interview occurred on June 7th, 2012. While generally covering both sets of regulatory process data (FAA and NHTSA), the bulk of the discussion was focused on the FAA rulemaking data. During the interview a determination was made to treat each data set as a separate project.
 - c. Assemble additional project documentation, published articles, etc.
 - i. Additional documentation included verification of variable definitions and clarification of acronyms. With regard to the FAA data, links to particular dockets within www.regulations.gov were documented. Where available, links to federal records detailing contributions of named political action committees were also documented.
3. Dataset Conversion and Management
 - a. Rename files for clarity, description
 - i. This action approved by agreement with Dr. Cavazos on June 7th, 2012.
 - ii. *FAA.xls* renamed *faa_regulation_comment_final.xls*.
 - b. Transform to application/platform agnostic format
 - i. This action approved by agreement with Dr. Cavazos on June 7th, 2012.
 - ii. Source spreadsheet in MS Excel format included two worksheets, one for data and another for variable definitions. Comma delimited (CSV) format does not support multiple worksheets, so each worksheet was converted to a single CSV file. The data file is *faa_regulation_comment_final.csv* and the variable definition file is *faa_regulation_comment_variable_definitions.csv*.
 - c. Document provenance
 - i. A narrative of actions taken on the data set is provided by this document.
 - ii. Derivation relationships are additionally documented in the METS metadata files which accompany both the source (XLS) and transformation (CSV) files.

- d. Apply standardized file system structure
 - i. While delivery of additional, earlier versions of the data set and SAS files is anticipated, as of June 26, 2012 the data set consists of the one MS Excel source file and two CSV transformations described above. These files have been organized according to Data Archives and Curation Services policy as follows:
 - 1. Top level directory named after data provider (*'cavazos'*).
 - a. Next level directory named after project (*'federal aviation administration notice and comment'*).
 - i. Files in this directory include this data curation plan and a narrative report of the data interview.
 - ii. Subdirectory *'data'*.
 - 1. Subdirectory *'sources'* includes MS Excel format file and an additional metadata directory (*'meta'*).
 - 2. Subdirectory *'transformations'* includes CSV format files and an additional metadata directory (*'meta'*).
 - 3. Subdirectory *'links'*. Includes a metadata directory (*'meta'*) and files with http addresses of FAA dockets(*'rule_links.txt'*) and Political Action Committee records where available (*'pac_links.csv'*).
 - ii. File organization approved by agreement with Dr. Cavazos on June 7th, 2012.
4. Create Documentation
 - a. Additional metadata
 - i. Descriptive
 - 1. Qualified Dublin Core metadata included with METS metadata files, located in *'meta'* directories referenced above.
 - ii. Context
 - 1. Contextual information is included in multiple files, include METS metadata files, the data interview narrative, and README files included where appropriate.
 - iii. Representation Information
 - 1. PRONOM registry information for MS Excel and CSV files is included in the METS metadata files.
 - iv. Fixity
 - 1. MD5 checksums for MS Excel and CSV files are included in the METS metadata files.
 - v. Reference
 - 1. Upon upload to LoboVault the collection will be assigned a URI. This URI will be added to the METS metadata files.
 - vi. Rights

1. Data copyright 2005, David Cavazos. Dr. Cavazos has agreed to make the data publicly available on the University of New Mexico's institutional repository, LoboVault.
2. Application of an appropriate Creative Commons license has been recommended, TBD.

vii. Provenance

1. File relationships and derivation information is recorded in the METS metadata files.
2. This document details additional actions taken on the FAA rulemaking data.

5. Host and Publish Project Data

- a. Post data on University Libraries FreeNAS
- b. Publish via LoboVault