Data Curation Plan

Bureau of Business and Economic Research (BBER)
Colonia Population and Housing Estimation

1. Project Review
    a. Initial project discussion with BBER staff
        i. A discussion between Jon Wheeler and Larry Compton on June 19th, 2012 to identify possible areas of collaboration between the University Libraries' Data Archives and Curation Services (DACS) and BBER resulted in the decision to use the New Mexico colonia estimation data as a pilot project. A sample of the data was already included in the *Colonia Population and Housing Estimation* publication record in LoboVault.
2. Research Documentation
    a. Transfer a copy of data to the University Libraries
        i. The complete dataset and accompanying GIS shapefile were downloaded from the BBER website (http://bber.unm.edu/colonias.htm) on June 29th, 2012.
    b. Assemble additional project documentation, published articles, etc.
        i. Besides the *Colonia Population and Housing Estimation* report itself, all available colonia maps were downloaded on June 29th with the exception of the map for the Las Palmeras colonia in Dona Ana County, NM. The link to this map was broken and remained so as of July 5th.
3. Dataset Conversion and Management
    a. Transform to application/platform agnostic format
        i. This action is understood to be approved by email from Kevin Kargacin, June 22nd, 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 CSV files were named by appending the name of the tab from which they were generated to the name of the source file. Thus from the source file *Colonias_2010Census.xls,* the following files were derived: *Colonias_2010Census_with_block_detail.csv* and *Colonias_2010Census_totals.csv*.
    b. Document provenance
        i. A narrative of actions taken on the dataset is provided by this document.
        ii. Derivation relationships are additionally documented by the METS metadata files which accompany both the source (XLS) and transformation (CSV) files.
    c. Apply standardized file system structure.
        i. Top level directory named after data provider, in this case BBER.
        ii. Next level directory named after project, in this case *Colonia Population and Housing Estimation*.

1. Subdirectory named 'documentation.'
   a. This directory includes a copy of the *Colonia Population and Housing Estimation* report.
   b. Additionally, this directory includes a copy of this data curation plan.
2. Subdirectory named 'data.'
   a. Subdirectory 'colonia maps.'
      i. This directory includes additional subdirectories, listed by name, of those NM counties with available colonia maps. Each of these directories includes PDF copies of the colonia maps for that county.
      ii. This directory additionally includes a 'meta' subdirectory containing an XML metadata file in METS format.
   b. Subdirectory 'originals.'
      i. This directory includes two subdirectories, 'census products' and 'gis.' These directories respectively contain the source dataset in MS Excel format and the NM colonia shapefile data archived in zip format.
      ii. This directory additionally includes a 'meta' subdirectory containing an XML metadata file in METS format.
   c. Subdirectory 'transformations.'
      i. This directory includes the subdirectory, 'census products,' which contains the CSV files described in section 3.a.ii.
      ii. This directory additionally includes a 'meta' subdirectory containing an XML metadata file in METS format.
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, including the METS metadata files and README files 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 the same URI as the publication to which it will be linked (http://hdl.handle.net/1928/17334).
  2. Individual items receive unique identifiers as generated by the Curator's Workbench.
- vi. Rights
  1. Data copyright 2011, the University of New Mexico, Bureau of Business and Economic Research.
- vii. Provenance
  1. File relationships and derivation information is recorded in the METS metadata files.
  2. This document details additional actions taken on the New Mexico colonia data.

5. Host and Publish Project Data
   a. Post data on University Libraries FreeNAS
   b. Publish via LoboVault