Rules Governing Data Quality, Preservation, and Transfer

Datasets received into this repository should follow the format guidelines for Data Deposit. Once submitted into the collection all data will be stored in .csv files. Any transformation required to convert the data into .csv from the datasets received in the submission information package (SIP) will be undertaken by the data curators for the repository. The repository will also undertake processes to normalize the data within the SIP. The depositor is only required to ensure data existence and accuracy for the metadata points required by the repository (see Metadata and Standards). Normalization will include:

1) Listing author as [last name],[first name]
2) Listing pub_date as [YEAR] and removing any day or month data
3) Standardizing publisher name based on publishing industry standards
   a) I.e. Simon & Schuster will be normalized to "Simon & Schuster" from any other representation such as "SS", "S&S" or "Simon and Schuster"

Any data provided by the depositor outside of the required metadata fields will be included as "Additional Information" for each observation that additional data is provided for.

Given that book data is inherently not sensitive there is no cleaning of sensitive data out of the datasets submitted to the repository.

Datasets submitted to the repository will be stored and preserved by the repository in the format that it is submitted in the SIP. When data is accessed through the DIP, however, the data across all datasets stored in the repository will be collated into a single .csv file with no depositor information tied to the observations.