Victoria Kerr

Data and Metadata

LIS4220 Data Curation

The data chosen for this project are statistics on wheat and rye from 1973 to current day, published by the United States Department of Agriculture (https://www.ers.usda.gov/data-products/wheat-data/). The source of the data is seven different agencies. Those agencies are the United States Department of Agriculture, World Agricultural Outlook Board, Agricultural Marketing Service, Farm Service Agency, Foreign Agricultural Service, the United States Department of Commerce, and the United States Census Bureau. In terms of where the data is hosted and found online, the United States Department of Agriculture hosts it on their Economic Research Service website.

Stakeholders could include the creators, users, managers, funders, or other relevant individuals/organizations. The stakeholders for this data are numerous. As the data itself is relevant to agriculture, related stakeholders would be those involved in wheat and rye production, including farmers and seed sellers. In addition, much of the data concerns the economic values of wheat and rye, and those stakeholders would be import-export businesses and the United States government, international grain markets, and the farmers growing and selling the wheat and rye. The data is so granular that even consumers who purchase wheat and flour in the United States would be considered stakeholders.

The data comes in three Microsoft Excel files (.xls and .xlsx), with two of the files including data up to 2021 and the third containing the majority of the historical data. The third file also has a more granular breakdown of the data by both quarterly and annual totals. This data set does not have any restrictions; as data collected and produced by the United States government, it is free to access and use. The only required software would be a spreadsheet reader of some sort, such as Microsoft Excel or Google Sheets.

The metadata is fairly simple and can be split into two types: the classification of the grains and the classification of economic factors. The grains are split into five specific types: hard red winter wheat, hard red spring wheat, soft red winter wheat, white wheat, and durum. The economic factors are more numerous, but cover the basic topics of import-export, production, and stock information. The metadata is fully integrated into the data files; as stated above, the metadata offered in these reports is fairly basic. As the information presented in these reports is rather simple, the metadata is similarly simple. Additionally, the metadata does not conform to any metadata standard or schema.

In terms of enrichment, there would be many opportunities for this dataset. The metadata, in particular, could be enriched in both discoverability and accessibility by conforming to a standard or schema, which would increase the coverage of the metadata. The data itself is very applicable and fairly discoverable since it is a product of the United States government, and therefore free to access and use.

This data has been published and used in publications, and variations and subsets of the data are also used in publications. As this dataset is collected and published by the United States government, it is physically published annually in Wheat Yearbook (Yearbook data is likewise compiled in two of the files provided for this dataset). Regular reports are published, typically updated monthly, annually, or quarterly. It is difficult to track all the publications that have been written on this dataset, as the set is so large and widely applicable to a range of areas of study. I performed only some cursory Google searches on the topic of “Wheat Yearbook”, which is one of the primary parts of the dataset, and found a number of publications from the USDA, the EPA, and other government agencies.