Victoria Kerr

Repository Profile

LIS4220 Data Curation

I searched for a repository based on subject, since the dataset chosen for this project is fairly specific to agriculture. The repository chosen is the Agri-environmental Research Data Repository from the University of Guelph (<https://dataverse.scholarsportal.info/dataverse/ugardr>) a public university in Ontario with a focus on agriculture and veterinary sciences. This repository was chosen because it is a smaller repository not connected to the United States government, and therefore has different data housed in it and different topics that the data concerns.

The collection scope of this repository is focused on specifically agricultural and environmental data produced by the University of Guelph. There are a number of academic programs at the university that tie in with these topics, so having a specific repository for the university is logical. Additional information available on the repository site also states that the repository “was established by the University of Guelph Library in 2012 with the assistance of the Ontario Ministry of Agriculture, Food and Rural Affair's (OMAFRA) Knowledge Translation and Transfer (KTT) Funding Program (University of Guelph, 2021),” and that research and projects funded by OMAFRA have special emphasis due to that relationship.

The submission process for this repository is thus: depositors must contact the library in charge for full instructions and templates; depositors must gather necessary dataset materials and ensure they meet Guelph deposit guidelines; depositors must complete a metadata template for their dataset, as well as a codebook or variable metadata template if necessary for the data; depositors must transfer their data files and completed templates to the repository staff; after this transfer, the staff review the materials, create a record, and ingest the dataset into the repository. After one final check to ensure that no other changes need to be made, the dataset is published.

SIPs must include the data files, of course, as well as all supplemental files and documentation generated for the data. Descriptive metadata must also be included; there are metadata templates provided by the repository that depositors must use in order to keep the metadata standardized. A more in-depth codebook template is provided as well for datasets that require it. In addition, it is encouraged for depositors to include raw and processed data according to their best practices.

No login or registration is required to access the data on this repository. Accessing the data is actually very simple, as each dataset has its own page with both files presented and a link to access the dataset by downloading it. There are two types of direct download available, as well as a searchable database of the files and access to historical versions of the data, if available. Metadata in the repository follows basic DublinCore standards, with additional descriptive metadata added to increase metadata available.

The Dissemination Information Package that this repository offers includes the dataset files, in whatever format they might come in—I found both .sas and .xlsx copies in the datasets I looked through—as well as a preservation package that offers preservation and archival friendly versions of the dataset files. This includes .csv files created from the spreadsheets, text and image files containing any charts and figures, and copies of the original formatting of spreadsheets to preserve that original format. By offering these additional formats, accessibility to the datasets is increased. Additionally, metadata downloads are offered as separate downloads.

Works Cited

University of Guelph. (2021). Agri-environmental research data Repository Dataverse. Retrieved February 25, 2021, from https://dataverse.scholarsportal.info/dataverse/ugardr