PRESERVING A **DATASET**

GREO has developed a procedure for preserving a dataset based on best practices in data management. This procedure was adopted to encourage sustainable, long-term preservation of datasets.

This schematic summarizes the ten step process implemented by GREO. For a detailed description of GREO's preservation process, including tools and methods used, contact us at **dataset@greo.ca**.

GET STARTED

STEP 1 — Create a space to store the dataset material

Create a main project folder on your storage device.

Create sub-folders inside the main folder to hold various dataset files. Make sure to give them logical names (e.g. data files, reports, user guides, codebooks).

STEP 2 — Gather and verify all dataset materials

Gather all data files and supplemental files (e.g. user guides, codebooks, reports).

Place a copy of each file into the appropriate folder created in Step 1.

Be sure that all materials necessary to understand and use the dataset have been gathered together.

DESCRIBE DATASETS (ADD METADATA)

STEP 3 — Develop study level metadata

Describe the project scope (e.g. title, investigators, time period, funders, purpose, keywords, geographic coverage, data type).

Describe the research methodology (e.g. sampling, instruments, processing).

Outline any conditions or restrictions on use of the dataset, including licensing.

STEP 4 — Describe variable level metadata

If you do not have a data dictionary or codebook for the data, create one.

The codebook should include variable names, variable labels, description of variable meaning, units of measure, sample weighting, and coding frequencies.

PREPARE FILES

STEP 5 — Check for missing and/or discrepant information

Ensure data files match the supporting documentation.

Ensure data is anonymized. Remove odd characters (e.g. @ # $ % ^ &) and consider coding missing values (i.e. blank cells).

STEP 6 — Create preservation friendly files

Ensure file names are explanatory and follow naming convention.

Convert files into formats that do not require proprietary software.

Do not encrypt or compress files.

ADD TO ARCHIVAL PLATFORM

STEP 7 — Decide how to present the dataset materials

Decide how to organize the dataset materials for preservation and sharing.

Create categories to group common files (e.g. by year, by file type).

STEP 8 — Upload dataset materials to preservation platform

Create a container for the dataset on the preservation platform.

Upload metadata.

Upload data files and set access permissions.

Apply GREO Terms of Use to the data.

VERIFY AND PUBLISH

STEP 9 — Review the dataset for completeness and accuracy

Check the metadata for completeness and accuracy.

Ensure all files are available.

Ensure appropriate access permissions and/or licenses have been applied to files.

STEP 10 — Make the dataset publicly available

Publish the dataset.

Check the dataset regularly to ensure data integrity, read ability and recoverability.

GAMBLING RESEARCH EXCHANGE ONTARIO
DRIVING KNOWLEDGE INTO ACTION