Niwot Ridge Metadata submittion instructions

Updated 2018-09-05

Reminders

- All LTER funded data are required to be publicly accessible within 2 years of collection (some exceptions possible on a case by case basis)
- To meet this timeline, your Niwot data must be returned - with metadata - to the IM on a timely basis
- Depending on the number of simultaneous submissions from other researchers, it can take a few days to a few weeks to get the data posted; please plan your timeline accordingly

DECIDE WHAT TO ARCHIVE

- Where file sizes are small, best practice is to publicly archive *quality-controlled raw data* (rather than calculated fields)
  - Exceptions may be made for large datasets, fine-scale temporal data; model outputs, and or other products derived from complex calculations)
  - In either case - keep a copy of all your own scanned datasheets & raw files (prior to cleaning or reformatting) - on your own computer. These can be critical to go back to in case of questions but the IM does not need a copy of these.
- Very large datasets (e.g. genomic data, drone data) are not supported by EDI. If you have one of these datasets - contact the IM for an alternate archiving solution.

ORGANIZE & CLEAN YOUR DATA

- [ ] Helpful tips here:
  - Organizing: https://environmentaldatainitiative.org/resources/five-phases-of-data-publishing/phase-1/
  - Cleaning: https://environmentaldatainitiative.org/resources/five-phases-of-data-publishing/phase-2/
  - Formatting: Use non proprietary tabular data (comma or tab delimited ASCII text) and geospatial data types.
    - UTF-8 encoding
    - Windows line endings
    - Quote all text fields
    - Avoid using special characters in column names
    - Remove empty rows at the end of the file (you can view this in a text editor such as notepad++ or textwrangler)
    - **If you need help achieving the above, contact the IM**
  - Consider grouping/packaging of datasets into user-friendly sets
    - a single 'dataset' can contain multiple 'entities', e.g. 2 shapefiles + 3 csvs + some code to analyze the two can comprise a single 'dataset'
  - Consider archiving code on EDI (especially if you are not already using github or another service to do this). EDI can accomodate a wide range of objects - including the code you use to write that amazing paper.

PREPARE METADATA TEMPLATE

- [ ] Download template from here:
  - http://niwot.colorado.edu/meta_data/NWT_metadata_form.xlsx
  - This form is machine-read to automagically generate EML. Do not change headers.
  - This form is the input to a basic pipeline to make EML and deliver NWT LTER datasets to the EDI Data Portal. It populates required elements but if there are additional pieces of metadata you'd like to add not captured by the fields here, talk to the information manager. It's fairly straightforward to augment. This template assumes that your work is funded by the NWTLTER grant. If you'd like to add an ADDITIONAL funding source, contact the IM.
- [ ] Fill in all tabs that begin with 'DataSet' (i.e. DataSet, DataSetEntities, DataSetMaintenance (OPTIONAL), DataSetAttributes, DataSetAttributeCodeDefinition (ONLY IF CODES ARE USED), DataSetKeywords, DataSetPersonnel, DataSetSites, DataSetTemporal)
- [ ] Some excel validation checks are embedded in the form, so it is necessary to fill out the DataSet and DataSetEntitites tab FIRST.
- [ ] Examples in ORANGE text can be deleted or kept as you choose

Hints

- View comments in the header cells for information on how to fill out each field
- Choices constrained by picklists, where appropriate
- DataSetEntities - add one line for each 'item' you would like to inlucde in your data package (i.e. each table, code file, rasterfile, etc)
- DataSetMaintenance - Used for documenting updates/changes to the data. Likely not needed for the first time something is archived.
- DataSetAttributes - Used for documenting contents of your datafiles. Generally you'll want to paste->transpose the column names of all tabular files in here.
  - storageType - see full list here: https://www.w3.org/2001/XMLSchema-datatypes or follow general rules
    - use 'integer' for integers
    - use 'float' for non-integer real numbers
    - use 'boolean' for binary variables
    - use 'date' for date
    - use 'dateTime' for dateTimes
    - use 'time' for times
    - *it is fine to have missing value codes of a different storageType than the data field, but they must be documented, e.g. you can have a missing value code of 'IMPLAUSIBLE' in a column of storageType 'integer', so long as you add 'IMPLAUSIBLE' to the list of missing value codes for that column
  - formatString - only required for dateTime fields, used to document the formatting of your dateTimes
  - unit - required for numeric fields; LEAVE BLANK for string and dateTime fields
  - if you have multiple missing value codes - send them to the IM separately; EML supports this but this excel template does not.
- DataSetAttributeCodes - Used only if you have codes as data values
- DataSetKeywords - To enhance data discovery, please select at least some LTER keywords from the LTER controlled vocabulary + at least one LTER core area
- DataSetPersonnel - Add as many people as you like
- DataSetTemporal - please use the YYYY-MM-DD formatting
- DataSetSpatial - please use decimal latitude/longitude. If you only a point location, not bounding box, use the same coordinates for N&S, E&W.

WRITE ABSTRACT AND METHODS

- [ ] Prepare two separate msword (.docx) files, one for the abstract, one for the methods.
- [ ] Title them 'Abstract.docx' and 'Methods.docx'
- [ ] For inline equations, use LaTex formatting and enclose the equations in single `$`, e.g. type `$\sum_{x=1}^{x=\infty}{x}{y}^{\frac{2}{3}}$` for $\sum_{x=1}^{x=\infty} xy^{\frac{2}{3}}$
- [ ] For equations on a separate line, use LaTex formatting and enclose the equations in double `\$$`, e.g. `$$\sum_{x=1}^{x=\infty}{x}{y}^{\frac{2}{3}}$$`

EMAIL lternwt@colorado.edu to let us know your data are ready to go!