

# Documentation for GO on a Trail Data Challenge Data Set

The data processes described below were used to create the data set described in the paper:

Kohl H.A., Nelson P., Pring J., Weaver K.L., Wiley D.M., Danielson A.B., Cooper R.M., Mortimer H., Overoye D., Burdick A., Taylor S., Haley M., Haley S., Lang J. and Lindblad M.E. (2021). GLOBE Observer and the GO on a Trail Data Challenge: A Citizen Science Approach to Generating a Global Land Cover Land Use Reference Dataset. *Frontiers in Climate*, 3:620497. DOI: [10.3389/fclim.2021.620497](https://doi.org/10.3389/fclim.2021.620497)

The data set can be accessed at <https://observer.globe.gov/get-data/land-cover-data> (including supplemental materials). More information about GLOBE data as a whole is available in the GLOBE Data User Guide, <https://www.globe.gov/globe-data/globe-data-user-guide>.

## Steps used to create data table:

1. Receive data query from GLOBE database for all 'landcovers' records from 2019-June-1 to 2019-Oct-15. *Returned 3839 records.* (sheet=raw data)
2. Filtered to 'land covers:data source='GLOBE Observer'. *Returned 3748 records.* (sheet=onlyGLOBE\_Observer)
3. Cleaned up data table for easier end-user understanding. (sheet=onlyGLOBE\_Observer\_cleaned)
  - a. Rename fields to remove prefix 'landcovers'
  - b. Rearrange fields for better grouping
  - c. Added 'null' value to blanks in image URL fields
  - d. Create QA fields and data (see further documentation below.)
4. Perform secondary photo review
  - a. Identify photos for rejection. (sheet=Listof2ndReviewRejectedPhotos)
  - b. Replace URL for identified photos from 4a with 'rejected'(sheet=onlyGLOBE\_ObserverCleaned2ndRej)
5. Gather MODIS-based land cover data and add to sheet (see further documentation below)
6. Remove rows without any useable GO Land Cover data (sheet=Frontiers\_dataTable), (sheet=Frontiers\_removed)
  - a. `removeif [accuracy_m] > 100` (*removed =242, remain=3506*)
  - b. `removeif [QACompleteness] = 0` (*removed =154, remain=3352*)
7. Export data table to "*FrontiersPaper\_GLOBEMeasurementData\_Landcover\_2019-06-01\_thru\_2019-10-15\_withAccuracy\_and\_MODISLandCover.csv*"

## Explanation of columns in data file (columns A – AQ):

Variable Name	Explanation
landCoverId	A unique number for each land cover observation, assigned by the GLOBE database
dataSource	Descriptor indicating whether the data were reported by a GLOBE member at a defined land cover site or reported from the GLOBE Observer app. Value options are "Site Definition" and "GLOBE Observer." All of the data for the GO on a Trail data set are from the GLOBE Observer app.

Variable Name	Explanation
siteName	The name assigned to the site where the data were collected. The GLOBE Observer app automatically assigns a site_name to an observation based on the Military Grid Reference System. A brief overview of the MGRS system is available at <a href="https://www.maptools.com/tutorials/mgrs/quick_guide">https://www.maptools.com/tutorials/mgrs/quick_guide</a> .
site_id	The GLOBE ID assigned to the site.
measureLatitude, measureLongitude	The true latitude and longitude of the location of the observation, as measured by the Global Position System (GPS) reading from the observer's mobile device. (Given in decimal degrees.)
measureElevation	The elevation of the observer's measured location based on a Google web service that determines elevation based on WGS84 EGM96 geoid height. (The GLOBE Observer app does not measure elevation using GPS directly.)
SiteLatitude, SiteLongitude	The latitude and longitude associated with the site definition. For sites created automatically by the GLOBE Observer app, this corresponds to the lower left corner of the associated MGRS gridbox. If the site was created via the GLOBE online data entry tools or the GLOBE Data Entry app, this value is provided by the observer. (Given in decimal degrees.)
SiteElevation	For sites created automatically, the elevation of the site location based on a Google web service that determines elevation based on WGS84 EGM96 geoid height. (The GLOBE Observer app does not measure elevation using GPS directly.) If the site was created via the GLOBE online data entry tools or the GLOBE Data Entry app, this value is provided by the observer.
measureDate	The date only of when the data were observed in UTC.
measuredAt	The date and time when the data were observed in UTC, following the format yyyy-mmddTHH:MM.
locationMethod	The method for determining location can be automatic (determined by the device GPS), or manual, meaning that the observer shifted their location on the map within the app, or manually entered values.
accuracy_m	The reported accuracy of the measurement location, given in meters. The GLOBE Observer app pings a user's location repeatedly up to 10 times.
sc_SnowIce	User-reported TRUE/FALSE value for the presence of snow and ice on the ground
sc_StandingWater	User-reported TRUE/FALSE value for the presence of standing water
sc_Muddy	User-reported TRUE/FALSE value for the presence of muddy conditions
sc_DryGround	User-reported TRUE/FALSE value for the presence of dry ground
sc_LeavesOnTrees	User-reported TRUE/FALSE value for the presence of leaves on the trees
sc_RainSnow	User-reported TRUE/FALSE value for the presence of rain and snow falling during the observation

Variable Name	Explanation
upURL, downURL, northURL, eastURL, southURL, westURL	The URLs for each of the photos sent in by users. By default, each photo is given the name "original.jpg" in the GLOBE database. Other possible values are NULL if no photo was submitted or REJECTED if the photo was not approved for inclusion in the database (for issues of privacy, such as identifiable faces or license plates, or for inappropriate content).
upExtraData, downExtraData, northExtraData, eastExtraData, southExtraData, westExtraData	Indicates if the user uploaded the photo from their camera manually (source: camera) or took the photo inside the app (source: app (compassData.horizon: 10.049474657691292)). In the case of the app, this also includes the angle the camera was pointed from horizon. Blank indicates no photo.
field notes	User-submitted comments related to the observation
northClassifications, eastClassifications, southClassifications, westClassifications	User-assigned percent land cover classifications for each direction
muc_code	MUC code of the land cover site
muc_details	Added information to align MUC classifications with IGBP classifications. For example, MUC maps trees to Evergreen or Deciduous, while IGBP classified by Broad Leaved or Needle Leaved. In this case, the field includes "b" or "n" for broad or needle leaves.
muc_description	Description of the land cover site MUC code
organization_id	The GLOBE ID assigned to the school or organization
org_name	The name of the reporting school or other institution
observerNum	A unique user ID assigned to each individual observer

#### Quality assurance fields (columns AR-BD):

Variable Name	Process
QA_location_flag	CountIf 'accuracy_m' is greater than 100 (value=1) CountIf 'accuracy_m' is less than 100 or empty (value=0)
QAImageCount	CountIf [Images] are not blank (range = 0-6)
QAImgRejectCount	CountIf [Images] are 'rejected' (range = 0-6)
QAImgNull	CountIf [Images] are blank 'null' (range = 0-6)
QAClassificationDirectionCount	CountIf [Classifications] are not blank (range = 0-4)
QACompleteness	Score = (QAImgCount+QAClassificationDirectionCount)/10 (range of 0-1)
QAbit_upImg	CountIf [upImg] is not blank (1), else (0)
QAbit_downImg	CountIf [downImg] is not blank (1), else (0)
QAbit_northImg	CountIf [northImg] is not blank (1), else (0)
QAbit_eastImg	CountIf [eastImg] is not blank (1), else (0)
QAbit_southImg	CountIf [southImg] is not blank (1), else (0)
QAbit_westImg	CountIf [westImg] is not blank (1), else (0)
QAbits_Images	Sequence of values created from Concatenate of (QAbit_upImg, QAbit_downImg, QAbit_northImg, QAbit_eastImg, QAbit_southImg, QAbit_westImg)

Data from the Moderate Resolution Imaging Spectroradiometer or MODIS instrument (columns BE – CE):

1. Gather MODIS-based land cover data
  - a. Using AppEEARS (<https://lpdaacsvc.cr.usgs.gov/appeears/>), uploaded point samples using ‘measureLatitude’ and ‘measureLongitude’ coordinates. Requested most recent data for year 2018 of **MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid**. (<https://lpdaac.usgs.gov/products/mcd12q1v006/>)  
 Friedl, M., Sulla-Menashe, D. (2019). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-02-28 from <https://doi.org/10.5067/MODIS/MCD12Q1.006>
  - b. Request limit is 1,000 records so divided list into parts  
 (sheet=appeears\_format\_pt1\_1to1000), (sheet=appeears\_format\_pt2\_1001to2000),  
 (sheet=appeears\_format\_pt3\_2001to3000), (sheet=appeears\_format\_pt4\_3001to4000)
  - c. Downloaded data in csv files  
 Added results to separate sheets ((sheet=appeears-formatted),  
 (sheet=appeears\_format\_pt1\_1to1000), (sheet=appeears\_format\_pt2\_1001to2000),  
 (sheet=appeears\_format\_pt3\_2001to3000), (sheet=appeears\_format\_pt4\_3001to4000)
  - d. Add MODIS land cover to GO Land Cover sheet (sheet=onlyGLOBE\_ObserverCleaned2ndRej)

The full user guide for the MCD12Q1 Data Sets can be found at [https://lpdaac.usgs.gov/documents/101/MCD12\\_User\\_Guide\\_V6.pdf](https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf), which includes more information about units, data types, valid ranges and fill values, as well as legends for each of the properties and types listed below. Links directly to the relevant tables in the user guide are also provided below.

Variable Name	Explanation
MCD12Q1_Date	Moderate Resolution Image Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) data product with date
MODIS_Tile	10 degree by 10 degrees tiles on a sinusoidal tile grid identified by a coordinate system that starts at (0,0) in the upper left to (35,17) in the lower right as defined at <a href="https://modis-land.gsfc.nasa.gov/MODLAND_grid.html">https://modis-land.gsfc.nasa.gov/MODLAND_grid.html</a>
MCD12Q1_006_Line_Y_500m	Coordinates used in the MODIS Tile grid system. A tool to convert between this system and latitude/longitude can be found under MODIS Land Tile Calculator available at <a href="https://modis-land.gsfc.nasa.gov/tools.html">https://modis-land.gsfc.nasa.gov/tools.html</a> .
MCD12Q1_006_Sample_X_500m	
MCD12Q1_006_LC_Prop1	Land Cover Classification System (LCCS), LCCS1 land cover layer numeric value ( <a href="#">MCD12Q1 User Guide Table 8</a> )
MCD12Q1_006_LC_Prop1_label	LCCS1 land cover layer descriptive name
MCD12Q1_006_LC_Prop1_Assessment	LCCS1 land cover layer confidence
MCD12Q1_006_LC_Prop2	LCCS2 land use layer numeric value ( <a href="#">MCD12Q1 User Guide Table 9</a> )
MCD12Q1_006_LC_Prop2_label	LCCS2 land use layer descriptive name
MCD12Q1_006_LC_Prop2_Assessment	LCCS2 land use layer confidence

Variable Name	Explanation
MCD12Q1_006_LC_Prop3	LCCS3 surface hydrology layer numeric value ( <a href="#">MCD12Q1 User Guide Table 10</a> )
MCD12Q1_006_LC_Prop3_label	LCCS3 surface hydrology layer descriptive name
MCD12Q1_006_LC_Prop3_Assessment	LCCS3 surface hydrology layer confidence
MCD12Q1_006_LC_Type1	Annual International Geosphere-Biosphere Programme (IGBP) classification ( <a href="#">MCD12Q1 User Guide Table 3</a> )
MCD12Q1_006_LC_Type1_label	Annual IGBP classification descriptive name
MCD12Q1_006_LC_Type2	Annual University of Maryland (UMD) classification ( <a href="#">MCD12Q1 User Guide Table 4</a> )
MCD12Q1_006_LC_Type2_label	Annual UMD classification descriptive name
MCD12Q1_006_LC_Type3	Annual Leaf Area Index (LAI) classification ( <a href="#">MCD12Q1 User Guide Table 5</a> )
MCD12Q1_006_LC_Type3_label	Annual LAI classification descriptive name
MCD12Q1_006_LC_Type4	Annual BIOME-Biogeochemical Cycles (BGC) classification ( <a href="#">MCD12Q1 User Guide Table 6</a> )
MCD12Q1_006_LC_Type4_label	Annual BGC classification descriptive name
MCD12Q1_006_LC_Type5	Annual Plant Functional Types (PFT) classification ( <a href="#">MCD12Q1 User Guide Table 7</a> )
MCD12Q1_006_LC_Type5_label	Annual PFT classification descriptive name
MCD12Q1_006_LW	Binary land (2) / water (1) mask derived from MOD44W
MCD12Q1_006_LW_label	Descriptive name for above field, either "land" or "water"
MCD12Q1_006_QC	Product quality flag numeric code ( <a href="#">MCD12Q1 User Guide Table 11</a> )
MCD12Q1_006_QC_Name_Description	Product quality flag and description

## Images Files:

The images collected for the GO on a Trail challenge are available in the data file as links to the GLOBE database but have also been collected as image files that can be downloaded all at once. The files have been re-named using a naming convention that includes some important information such that it is not necessary to refer to the CSV file to get a sense for the location, date, direction, MUC value, IGBP value, surface conditions (=true), accuracy value, and the image size.

The naming convention is as follows:

```
GOLC_[measureLatitude]_[measureLongitude]_[siteName]_[SiteElevation]m_[yyyymmdd]
[landCoverId]_{directionNumber}[direction]_[MCD12Q1_006_LW_label]_[muc_code]_IGBP[MCD12Q1_
006_LC_Type1]_[MCD12Q1_006_LC_Type1_label]_{surfaceConditions=True}_[locationMethod][accurac
y_m]m_widthXheight.jpg
```

Example filename:

```
GOLC_n45p12345_w123p12345_12TVQ791547_1000m_20190814_00026634_3North_Land1_xxx_IGBP
13UrbanAndBuildUpLands_DryGroundLeavesOnTrees_automatic5m_1920x1080.jpg
```