# Chapter IV: Geodata Crawler: A centralized national geodatabase and automated multi-scale data crawler

Douglas R. Leasure

Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701 USA

## Abstract

There is now an unprecedented availability of GIS and remote sensing data that provides a powerful new tool for scientific research, but it is often difficult to acquire and process these data to generate tabular datasets that quantify landscapes at specific research sites using appropriate spatial scales for the questions being investigated. This can limit the number of sample locations and the variety of GIS data and spatial scales included in studies, and it may even prevent some researchers from utilizing GIS resources at all. Geodata Crawler contains a centralized national geodatabase with dozens of ecologically-relevant datasets including land cover types, soils, climate characteristics, hydrology, and human populations. The automated multi-scale data crawler delineates customized sample areas for user-locations anywhere in the continental United States and tabulates summary statistics from national geodatabases within those sample areas. Six spatial scales are available for delineating sample areas: point, local, watershed, riparian, local-watershed, and local-riparian. User options allow customization of these spatial scales by adjusting, for example, the site radius used by the local scale, or the stream buffer size used by the riparian scale. Geodata Crawler output includes 1) a project-specific geodatabase with all GIS layers required to collect user-requested data, 2) polygons representing sample areas delineated at each site, and 3) tabular data appropriate for most statistical analyses. Geodata Crawler can run on a single local machine or on a server allowing remote access by multiple users. Several time-saving features are available that include simultaneous processing of multiple projects on multiple processing cores, data archiving for rapid retrieval by other projects, and simultaneous processing of subsets of user-locations from a single project. Future directions for the Geodata Crawler project are discussed including web-based project submission and cluster computing.

## Introduction

The unprecedented availability of GIS (*i.e.* geographic information systems) and remotely sensed data in recent years provides a rich resource that supports scientific research in many disciplines including landscape ecology, eco-hydrology, climate change, and landscape genetics. Increased availability of spatial data has been coupled with increased access to powerful analytical techniques like machine learning and Bayesian statistics that can accommodate these high dimensional datasets. Although significant progress has been made leveraging these tools, several impediments remain to efficient acquisition and processing of large GIS datasets to generate site-specific tabular data appropriate for particular research questions and analytical methods.

Many national and global GIS datasets are freely available for download, but access is scattered among various websites and FTP servers, and data are often stored in disparate file formats and spatial resolutions. Processing these data requires expert knowledge, specialized software, and significant computing power. Samples must then be extracted from national datasets for specific study locations at spatial scales that are appropriate for the research question being addressed. For example, two research projects, one studying endangered beetle habitat and the other studying natural stream hydrology, both required information from national land cover datasets about forest cover, but one was interested in forest cover within multiple site radii, while the other was interested in forest cover within watersheds of study sites (Chapters 2 & 3). The time-consuming and difficult process of delineating multi-scale sample areas for particular research locations and then tabulating appropriate data can severely limit the number of sample locations and the variety of GIS data included in a study.

Geodata Crawler was developed to address these issues by providing a centralized national geodatabase and an automated multi-scale data crawler that can rapidly build project specific GIS databases, delineate customized sample areas for user-defined locations, and tabulate requested data that can be exported in spreadsheet format suitable for most statistical analyses. Only minimal GIS skills are required from the user, and no GIS data are required other than sample locations and project boundaries. Geodata Crawler was coded using Python programming language using the *ArcPy* package which provides command-line access to Esri’s ArcGIS toolbox (Esri 2013a, Python 2012). It requires ArcGIS for Desktop 10.2 with the advanced license and the Spatial Analyst extension. Exact system requirements have not been explored thoroughly, but known issues can occur on systems with less than 8 GB RAM or single core processors. Two primary computer platforms were used in the development of Geodata Crawler:

1. Windows 7 Professional (64-bit) with an Intel Core i7-2600 CPU (3.40 GHz, 16 GB RAM) and an external RAID 1 hard drive array (3 TB actual storage, USB 3.0), and
2. Windows Server 2008 R2 Enterprise (64-bit) with Intel Xeon CPUs (2 processors @ 2.13GHz, 24 GB RAM) and an internal RAID ?? SCSI hard drive array (4 TB actual storage).

Geodata Crawler can be setup to run on a local machine, or it can be setup to run on a server so that it can accept simultaneous data requests from remote users. Remote users can create a mapped drive to the Geodata Crawler server to submit jobs and retrieve output. This enables them to execute projects remotely without ArcGIS licenses on their local computers. Geodata Crawler can utilize multiple processors (or cores) to process several jobs simultaneously. Multi-processing functionality was implemented using Python package *multiprocessing*. Geodata Crawler can also be configured to archive site-specific data in a system geodatabase for rapid retrieval by future projects. This can drastically improve performance when a project requests a large amount of data that were previously archived by another project, but it can also decrease processing efficiency when previously archived data do not exist. The archiving function can be toggled on or off for individual projects (see *User Options* section). The archiving function is supported by Geodata Crawler’s site identification system that provides a unique 11-digit numeric identifier for points along a 30 m grid covering the continental United States. All user-provided locations are snapped to this grid prior to analysis.

## Input/Output

User-input generally consists of three items: 1) a Geodata Crawler user-options worksheet, 2) an ArcGIS polygon shapefile to identify the project boundary, and 3) an ArcGIS point shapefile to identify locations of-interest (optional). The user-options worksheet enables customization of a Geodata Crawler project to best meet specific research goals and to optimize processing efficiency. It contains a list of variables and spatial scales that can be toggled on or off. Geodata Crawler automatically builds project-specific geodatabases that contain all user-requested GIS layers clipped to their project boundaries. All data requested for a single project must use the same project boundary. User-provided locations will be used as data collection sites. If no sample sites are provided, random locations can be generated by Geodata Crawler. Users may also provide their own rasters or Landsat images for automated processing. All user-inputs must be projected in the Albers NAD 1983 datum.

Geodata Crawler output includes: 1) a project-specific geodatabase, 2) a point shapefile with an attribute table containing requested data for all sample sites, and 3) a polygon shapefile for each requested spatial scale containing boundaries for sample areas delineated at each site. The project specific geodatabase will contain all of the original GIS layers as well as intermediate products required to collect requested data. It is named *PROJECT\_NAME*/GEODATABASE\_#dm.gdb, where # is the spatial resolution in decimeters of rasters in the geodatabase. This geodatabase will be retained by Geodata Crawler to use in subsequent data requests for the same project. A separate geodatabase will be created for site-specific output data and it will be named *PROJECT\_NAME*/OUTPUT.gdb. Output data will include a point shapefile with an attribute table containing requested data. The attribute table can be exported to Microsoft Excel using the “TableToExcel” tool in ArcGIS. The exported table will contain a row for each sample site and a column for each requested variable. Column names begin with a designation for the spatial scale of data collection (Table 1). Appendix 3 provides detailed descriptions of output data. OUTPUT.gdb will also contain polygon shapefiles representing sample areas for each site, and their attribute tables will contain all data ever collected at that spatial scale for a project. These polygon features can be used to produce maps and to support custom GIS analyses. All output data is projected in the Albers NAD 1983 datum.

## Variables

Geodata Crawler’s user options worksheet (Appendix 1) contains a row for each variable with a column for user-input that allows custom variable selection. A Geodata Crawler variable is a combination of:

1. A spatial scale for delineating sample areas (see *Spatial Scales* section),
2. A GIS layer from the national geodatabase (see *National Geodatabase* section), and
3. A statistic to summarize data within each location’s sample area (*e.g.* mean, mode, standard deviation, coverage of a particular map unit; see *Appendix 2*).

Some variables can be toggled with a simple true/false response in the user-options worksheet, but other variables require selection of individual map units for data collection. For example, *average elevation within watersheds* can be toggled with a true/false response, but *percent coverage of pine forest within watersheds* requires the user to specify pine forest as the land cover of-interest. Variables that require specific map units to be identified accept a list of map unit codes so that data can be collected for multiple map units during a single Geodata Crawler run. User options may be available to customize the way in which data are collected for some variables (see *User Options* section and Appendix 2). Appendix 2 provides descriptions of all Geodata Crawler variables and map unit codes, and it can be used as a reference when filling out the user-options worksheet. Appendix 3 provides descriptions of all column names that will be output for each variable, and it can be used as a reference when interpreting Geodata Crawler output.

## Spatial Scales

Geodata Crawler automates the process of delineating sample areas associated with each user-defined location. It includes six flexible spatial scales used to delineate sample areas: point, local, watershed, riparian, local-watershed, and local-riparian. Several options are available to customize spatial scales to meet specific research goals (see *User Options* section).

The *point* spatial scale is the simplest because no sample area needs to be delineated for each user-defined location. The point coordinates themselves serve as the “sample area”. Variables collected at this spatial scale generally either extract a value from a GIS layer at each point, or they measure the distance from the point to a landscape feature (*e.g.* nearest forest patch greater than 10 hectares). Variables that measure the distance to a landscape feature can be customized using the user options *DIST\_TO\_RADIUS* and *DIST\_TO\_MIN\_PATCH\_SIZE*.

*Local* spatial scale (Fig. 1) is a circular sample area centered at a user-defined location. Its size is determined by a user-defined site radius that can be customized using the user option *LOCAL\_RADIUS*.

*Watershed* spatial scale (Fig. 2) contains all land areas that drain into a user-defined location. This is delineated using the ArcGIS watershed tool (Esri 2013) based on the flow accumulation and flow direction rasters of the National Hydrography Dataset Plus (USEPA & USGS 2012). Watersheds can be delineated for terrestrial and stream sites. It is important to use the *MOVE\_TO\_STREAM* user option to ensure that locations meant to be on a stream are exactly located on the GIS representation of the stream channel.

*Local-watershed* spatial scale (Figs. 3 & 4) is the area where the local and watershed scales intersect. This contains land areas that drain into a user-defined location, but that are also within a user-defined radius of the site. This spatial scale can be customized using the user option *LOCAL\_RADIUS*.

*Riparian* spatial scale (Figs. 5 & 6) contains areas within a site’s watershed that are also within some user-defined distance of a stream. Riparian zones cannot be delineated for sites that are not on streams. Non-stream sites will be skipped in riparian-scale data requests. Sites can be snapped to streams using the *MOVE\_TO\_STREAM* user option, and this will guarantee that riparian-scale data are collected for all user-locations. Streams are delineated based on the flow accumulation raster of the National Hydrography Dataset Plus (USEPA & USGS 2012). The riparian spatial scale can be customized using the user options *RIPARIAN\_BUFFER\_WIDTH* and *MIN\_SHED*.

*Local-riparian* spatial scale (Figs. 7 & 8) is the area where the local and riparian scales intersect. This contains land areas that drain into a user-defined location, but that are also within a user-defined radius of the site and within a user-defined distance from a stream. Local-riparian zones cannot be delineated for sites that are not on streams. This spatial scale can be customized using the user options *RIPARIAN\_BUFFER\_WIDTH*, *LOCAL\_RADIUS*, and *MIN\_SHED*.

## National Geodatabase

Geodata Crawler includes a set of national GIS datasets used to derive over 1,000 variables available in Geodata Crawler. These data describe landscape characteristics like land cover, topography, hydrology, and climate, and they are all publicly available free-of-charge. New datasets are regularly added to Geodata Crawler, and it was designed to streamline this process. Data are projected to the Albers NAD 1983 datum and clipped to include only the continental United States before being stored in Geodata Crawler’s system databases (ArcGIS file geodatabases). Both vector and raster data can be utilized, and rasters are stored using their original spatial resolutions. Some of these datasets may be used to derive additional datasets when project geodatabases are built. For example, flow accumulation rasters from the National Hydrography Dataset (NHD) are used to delineate streams for each project individually, rather than relying on NHD flow lines because this increases flexibility in how streams are delineated for each project and it also reduces the overall disk space required to store Geodata Crawler’s system databases (*i.e.* currently 68.2 GB). Geodata Crawler resamples rasters to a uniform spatial resolution using bilinear interpolation as it builds project geodatabases, and it can be configured to either detect the highest spatial resolution among requested datasets or to simply resample all rasters to a system-defined resolution (*i.e.* 30 m). Rasters are never resampled to a lower resolution.

Geodata Crawler currently includes the following national geodatabases:

*National Land Cover* (USGS 2010, Fry *et al.* 2011, Homer *et al.* 2007, Vogelmann *et al.* 2001, Price *et al.* 2007):

These digital maps depict the geographic distribution of land cover classes for five time periods: 1970-80s, 1990s, 2001, 2006, and 2010. All of these data sets use hierarchical land cover classifications with two or three “levels”. Each specific-level classification (*e.g.* pine forest) is nested within at least one broad-level classification (*e.g.* forest). The 1970-80s data have 37 classes and two levels. The 1992 data have 21 classes and two levels. The 2001 and 2006 data have 16 classes and two levels. The 2010 data have 583 classes and three levels. The 2001 and 2006 data include rasters that quantify percent impervious surfaces, and the 2001 data include a raster that quantifies percent canopy cover. See Appendix 2 for descriptions of all land cover classes and levels. These data are stored in raster format with 30 m spatial resolutions.

*U.S. General Soil Map* (USDA 2006):

This is a digital representation of the U.S. general soil map and it depicts the geographic distributions of 9,193 soil types. These data were originally stored in vector format with accompanying tabular data describing soil attributes. Soil data were originally mapped in 1-by 2-degree quadrangles. See Appendix 2 for names and some attributes of all soil types. Original data were converted to raster format for the Geodata Crawler system database and stored with 30 m spatial resolution.

*Geologic Rock Type* (Schruben *et al.* 1994, King & Beikman 1974):

This is a digital representation of a geologic map depicting the geographic distributions of 175 rock types. See Appendix 2 names of all geologic rock types. It was originally stored in vector format and map units were mapped at the 1:2,500,000 scale. It was converted to raster format for the Geodata Crawler system database and stored with a 30 m spatial resolution.

*Aquifers* (USGS 2003):

This dataset identifies the boundaries of the principal aquifers in the conterminous United States mapped at 1:2,500,000 scale. Aquifer names and their rock types are included in the attribute table, and they are listed in Appendix 2. These vector data are stored as a polygon feature in an ArcGIS file geodatabase.

*Baseflow Index & Groundwater Recharge* (Wolock 2003a, 2003b):

The base-flow index raster was created by interpolating base-flow index values estimated at USGS stream gage locations. Base-flow is the component of stream flow that can be attributed to ground-water discharge into streams. The groundwater recharge raster provides an index of mean annual natural ground-water recharge calculated by multiplying the base-flow index raster by a raster of mean annual runoff values. The mean annual runoff data used for this calculation were long-term averages (1951-1980) of stream flow divided by drainage area. These data are all stored in raster format with 1 km resolution.

*Soil attributes for hydrological modeling* (Carlisle *et al.* 2010):

This is a set of 35 layers describing characteristics of soil and geology that were originally processed and used by Carlisle *et al.* (2010) for flow alteration modeling. These data were obtained from James Falcone, a USGS GIS specialist and co-author of the hydrological modeling project for which the data were developed (Carlisle *et al.* 2010). The data were originally derived from the U.S. general soil map (USDA 2006) or the geologic map of Reed and Bush (2005). All layers are stored in raster format with 1 km spatial resolution.

*National Hydrography Dataset Plus* (USEPA & USGS 2012):

This dataset includes rasters that quantify elevation, flow accumulation, and flow direction that are essential for delineating streams, watersheds, and riparian zones. All rasters have 30 m spatial resolution and are stored as raster mosaics.

*National Hydrography Dataset* (Smiley *et al.* 2009):

Two products from the National Hydrography Dataset are currently used by Geodata Crawler: The flow lines that are classified as artificial paths or canals, and the “spring seeps” points that identify geographic locations of known natural springs. Both are stored as vector data (line and point features) in an ArcGIS file geodatabase.

*WorldClim* (Hijmans *et al.* 2005):

This geodatabase contains a set of 67 global climate layers describing current climate conditions (*i.e.* 1950-2000). It includes measures of precipitation and temperature (annual and monthly mean, minimum, and maximum), and it includes 19 BioClim variables that were designed to measure biologically relevant aspects of climate like mean diurnal range in temperature and precipitation during the driest quarter each year. All layers are stored in raster format with 30 arc-second spatial resolution (~800 m).

*PRISM U.S. Climate Normals* (Daly *et al.* 2008):

This geodatabase contains a set of 39 climate layers for the conterminous United States describing current climate conditions (*i.e.* 1971-2000). It includes measures of annual and monthly precipitation and temperature (minimum and maximum). All layers are stored in raster format with 30 arc-second spatial resolutions (~800 m).

*U.S. Census Grids* (Seirup and Yetman 2006):

These raster datasets describe human demographics based on U.S. Census data from 2000, including population density, household density, education levels, income, and ethnicities. All rasters have 30 arc-second spatial resolution (~800 m).

*Nighttime Lights* (NOAA 2012):

These data provide an index of intensity of nighttime lights viewed from space. The files are cloud-free composites made using data from the Operational Linescan System (OLS) of the Defense Meteorological Satellite Program (DMSP). The data were screened to minimize bias due to sunlit areas, glare, clouds, or the aurora. Two time periods have been incorporated into Geodata Crawler (2000 and 2010), but data are available for all years 1992 to 2012 (NOAA 2012). Several related datasets are available such as average visible light, average stable lights (discarding ephemeral lights such as wildfire), and average light intensity scaled by the frequency of light detections in each pixel. Scaling data by frequency of detections normalizes data to account for variations in the persistence of lighting. These raster datasets have 30 arc-second spatial resolutions (~800 m).

*Agricultural Pesticides* (Nakagaki 2007a, 2007b):

These data quantify pesticide applications (kg/km2) in 1992 and 1997. The 1992 data include a raster for each of 199 pesticides and the 1997 data include a raster for each of 219 pesticides. Data for both years include a raster with application rates summed across all pesticides. Non-agricultural uses of pesticides are not included in these datasets. All rasters have 1 km spatial resolution.

*Roads 2011* (USDC 2011):

This national roads layer was created by merging TIGER/Line road layers for primary and secondary roads in each state. Roads are categorized as primary, secondary, local/rural, or four-wheel drive roads (see Appendix 2 for all road types). These vector data are stored as a line feature with an attribute table in a file geodatabase.

*Forest Service Active Fire Mapping* (USDA 2011):

These data were obtained as a point shapefile that depicted active wildfires detected throughout each year 2000-2011 by the MODIS satellite. MODIS images have 1 km spatial resolution, 36 spectral bands, and they are acquired for every location on earth every 1 to 2 days. Points in the active fire shapefile represent the center of a MODIS image’s pixel where a fire was detected. These data were converted to raster format with 1 km spatial resolution for use with Geodata Crawler. Each pixel was classified as either burned (1) or unburned (0) depending on presence of any active fire points for a particular year were within a pixel. The dataset was also used to create a national fire frequency raster that measured the number of fires per decade (2000-2011).

*U.S. Historical Oil & Gas Development* (Biewick 2008):

These data depict oil and gas development throughout the continental United States for each decade of the 20th century, and also for the pre-1900 and 2005-2006 time periods. Due to the proprietary nature of exact well locations, the U.S. was divided into grid cells ¼ square mile and oil/gas development was categorized as absent, oil only, gas only, oil and gas, or unknown well type. These raster data have 804.7 m (0.5 mile) spatial resolutions.

*Arkansas Oil & Gas Development* (AOGC 2013):

These data contain point locations of 55,126 oil and gas wells in the state of Arkansas. Attribute information identifies each well as gas, oil, or oil and gas, and identifies its status as active, inactive, permitted, plugged, or spud. These data were obtained as a Google Earth map document and converted to a point feature in an ArcGIS file geodatabase.

## User Options

The user-options worksheet (Appendix 1) begins with 24 user options that allow users to customize each project. User-options can be manipulated to balance project needs with processing efficiency. User options require either true/false, text, or numerical responses, and default values are provided for all user-options. In addition to the user-options listed below, the user options worksheet contains a list of all variables that can be collected (see *Variables* section).

*PROJECT\_NAME* (text):

A name for the Geodata Crawler system folder that will contain all data for a project. Data for all Geodata Crawler runs associated with a single project will be saved into this folder. All subsequent data collection runs for a single project must have identical project boundaries. The default value is “PROJECT”.

*OUTPUT\_FILE\_NAME* (text):

A name for point shapefile that will contain requested data for each sample location. The output file will be *PROJECT\_NAME*/OUTPUT.gdb/*OUTPUT\_FILE\_NAME*. If a file with this name already exists in the project folder, it will be overwritten. The default value is “OUTPUT”.

*USE\_RANDOM\_POINTS\_YN* (T/F):

If “T”, a set of random points will be created within the project boundary and user-provided locations will be ignored. Two additional user-options control the number of random points and their spacing. The default value is “F”.

*RANDOM\_POINT\_COUNT* (numeric):

The number of random points created. The default value is 10.

*RANDOM\_POINT\_SPACING* (numeric):

The minimum distance (in meters) allowed between random points. The default value is 50 m.

*MIN\_SHED* (numeric):

The minimum drainage area (km2) used to define the smallest stream-of-interest for the current data collection run. This value will be noted in output spreadsheets for any variables that may be affected by its value, and separate columns will be maintained in output spreadsheets if the same variable is collected using multiple values of this user-option. Separate stream layers will be created in the projects geodatabase for each value that is used for this option. The default value is 3 km2.

*MAX\_SHED* (numeric):

The maximum drainage area (km2) used to define the largest stream-of-interest for the current data collection run. This value will be noted in output spreadsheets for any variables that may be affected by its value, and separate columns will be maintained in output spreadsheets if the same variable is collected using multiple values of this user-option. Separate stream layers will be created in the projects geodatabase for each value that is used for this option. The default value is *null* which will not impose an upper-limit of stream size.

*MOVE\_TO\_STREAM\_YN* (T/F):

If “T”, all sample locations will be moved to the nearest stream, centered on a raster cell of the digital elevation model used to delineate streams. This guarantees that points intended to be stream locations are not several meters away from the digital representation of the stream which can severely alter results, particularly for data collected from watersheds or riparian zones. This can also be used with random points to identify random stream sites. The default value is “F”.

*MOVE\_TO\_STREAM\_SEGMENT\_YN* (T/F):

If “T”, all sample locations will be moved to the nearest stream, and then moved to the downstream-most point of that stream segment. A stream segment is defined as a section of stream between two confluences. The default value is “F”.

*FIND\_STREAM\_SEGMENT\_POINTS\_YN* (T/F):

If “T”, all stream segments in the study area will be identified and saved as a GIS layer. Then, the downstream-most point of each segment will be identified and saved as a separate GIS layer. This will be automatically toggled on with *MOVE\_TO\_STREAM\_SEGMENT\_YN*. The default value is “F”.

*REMOVE\_DUPLICATES\_YN* (T/F):

This option allows duplicate locations to be removed prior to sample area delineation and data collection. Duplicates will be removed after random points are created, points have been snapped to the nearest Geodata Crawler grid point (always within 30 m), and points have been moved to streams or stream segments. This will improve processing efficiency, and all duplicate rows will be retained in the output spreadsheet even though only one of them was actually processed. The default value is “T”.

*RIPARIAN\_BUFFER\_WIDTH* (numeric):

Defines the distance (in meters) away from streams used to delineate sample areas for upstream riparian zones. Very large values may result in sample areas that are identical to the watershed of a point because riparian zones cannot go beyond the watershed boundary. This value will affect delineation of sample areas at the riparian and local-riparian spatial scales. The default value is 1000 m.

*LOCAL\_RADIUS* (numeric):

Defines the distance (in meters) used as a site radius to delineate local-scale sample areas. Sample locations that are within this distance from the project boundary may have incomplete data for local-scale variables. This value will affect delineation of sample areas at the local, local-watershed, and local-riparian spatial scales. The default value is 1000 m.

*DIST\_TO\_RADIUS* (numeric):

Some variables will measure the distance from user-locations to some landscape feature like a road or a patch of forest. This user-option sets the upper limit to search for the landscape feature and setting this value to a reasonably small distance (in meters) can drastically improve processing efficiency, particularly for projects with a large project area. The default value is 1000 m.

*DIST\_TO\_MIN\_PATCH\_SIZE* (numeric):

This option sets the minimum patch size (in hectares) to be recognized when measuring the distance from a user-location to a landscape feature like the nearest patch of forest. Any patches smaller than this value will not be considered when making measurements. The default value is 0 which will not impose a lower limit on patch sizes being assessed.

*LANDSAT\_FILE\_NAME* (text):

Some variables allow a user to provide a Landsat image for automated processing by Geodata Crawler. The image must be provided as a multi-band composite .tif file. This option allows the user to specify the filename of the Landsat image. The default value is “LANDSAT.tif”.

*USER\_RASTER\_FILE\_NAME* (text):

Some variables allow a user to provide a raster dataset for automated processing by Geodata Crawler. The raster dataset must be provided as a single band .tif file. This option allows the user to specify the filename of this raster dataset. The default value is “USER\_RASTER.tif”.

*QUERY\_TYPE* (0, 0.5, 1, 2, or 3):

0: Normal Geodata Crawler operation with automated sample area delineation for user-defined point features and data tabulation from within those sample areas based on user-defined variables.

0.5: Similar to query type 0, except user-defined points will be ignored if their sample areas do not already exist in the project’s output geodatabase. Processing efficiency is improved because no sample areas will be delineated.

1: All data from Geodata Crawler’s archives within the project boundary will be returned. User-defined points will be ignored, no sample areas will be delineated and user’s variable selection will be ignored.

2: Data from Geodata Crawler’s archives within the project boundary will be returned for sites that contain data for *any* of the variables selected by the user. User-defined points will be ignored and no sample areas will be delineated.

3: Data from Geodata Crawler’s archives within the project boundary will be returned for sites that contain data for all of the variables selected by the user. User-defined points will be ignored and no sample areas will be delineated.

*CREATE\_SAMPLE\_AREAS\_ONLY\_YN* (T/F):

If “T”, sample areas will be delineated for all user sites, but no data will be collected. The default value is “F”.

*CLONES* (numeric):

This can be an integer from zero to four that will allow user-locations to be sub-divided (up to four times) to accommodate simultaneous processing of subsets of locations. This will drastically increase processing efficiency with a multi-core or multi-processing computer, but provides no advantage on a single core computer. Each cloned project will contain a subset of user locations and replicates of project geodatabases. Although replicating project geodatabases requires additional disk space (*e.g.* dozens of GB per clone for large projects), it facilitates multi-processing without requiring multi-user editing capability which is not supported by file geodatabases. Multi-user editing is supported by ArcSDE geodatabases (unavailable with ArcGIS for Desktop), and this would circumvent the need to clone projects. The default value is zero, resulting in no cloned projects being created.

*TOGGLE\_IMPORT\_ARCHIVES* (T/F):

If “T”, Geodata Crawlers system archives will be searched for pre-existing data matching the data requested for the current project’s locations. Performing this search can drastically increase processing time for projects with many locations and large data requests that are not in the archives. Using the archive function has resulted in read/write conflicts when processing multiple projects simultaneously. The default value is “F”.

*TestMode* (T/F):

If “T”, archived data will be written to a geodatabase separate from Geodata Crawler system archives to prevent potentially erroneous test data from being archived. Console windows will also be left open at the end of each project’s run to allow manual manipulation of Python objects at the end of a run, or in the event of an error. The default value is “F”.

*ALL\_VARIABLES* (T/F):

If “T”, all possible variables will be collected at user-defined locations. This is included for testing purposes and should never be used with more than a few locations. The default value is “F”.

## Future Directions

Geodata Crawler is a powerful new tool that can help to overcome GIS bottlenecks in data analysis work flows across many different research disciplines. Five important next steps are envisioned to broaden research applicability, improve processing efficiency, and increase ease-of-use:

1. Develop a web interface for setting up projects and submitting them to a server,
2. Adapt to the Linux operating system for cluster computing,
3. Shift to ArcSDE geodatabases with multi-user editing,
4. Add path-based spatial scales including stream paths, linear paths, and least-cost paths between sample locations,
5. Continue to incorporate new national GIS datasets, particularly future climate data for climate change research.

The Geodata Crawler website (<http://www.geodatacrawler.com>) gives a general overview of Geodata Crawler, its national geodatabases, and its spatial scales, but it does not currently support online setup or submission of projects. Adding this important feature would broaden the reach of Geodata Crawler. Building this capability will require web-based forms that walk users through the user-options worksheet and translate their responses into an appropriately formatted text file (USER\_OPTIONS.csv). Users would also need the ability to upload their project files that may include a boundary shapefile (required), a user-locations shapefile, a Landsat image, and a custom user raster. An FTP server (<ftp://ftp.geodatacrawler.com>) is already available with password-protected user accounts where geodatacrawler output can be accessed. Output often requires large amounts of disk space and the FTP site provides remote data storage where users can browse data and download only what is required for their specific needs.

Transferring Geodata Crawler to a cluster computing platform would drastically increase its processing efficiency by allowing it to process many projects simultaneously. This would broaden its capacity to support the potentially large number of jobs being submitted through a web-based portal. Geodata Crawler already includes multi-processing functionality that allows simultaneous processing of multiple jobs on separate processing cores. This capability should transfer to allow utilization of dozens of processing cores available with cluster computing. The University of Arkansas High Performance Computing Center (<http://hpc.uark.edu/>) provides ideal computing platforms for developing and testing this capability. ArcGIS for Server (Esri 2013b) provides an appropriate software package.

A shift from file geodatabases currently used by Geodata Crawler to ArcSDE databases would allow multi-user editing capability to streamline the processes of data archiving and project subdivision for faster processing. Archiving is required to share data among projects and this can drastically improve processing efficiency. This would be even more important with potentially increased traffic through a Geodata Crawler web portal because benefits of sharing data among projects, and costs of re-collecting previously collected data, would increase with more data requests. Multi-user editing capability also enables large projects to be subdivided and then processed as multiple simultaneous sub-units. Sub-dividing projects and archiving are currently supported in Geodata Crawler only by using an inefficient work-around that requires duplication of project geodatabases, because file geodatabases (ArcGIS for Desktop) do not allow multi-user editing. ArcSDE databases with multi-user editing would allow projects to be subdivided across many processing cores (*e.g.* on a computing cluster) without requiring additional disk space for duplicating project geodatabases.

Geodata Crawler has the capability to delineate linear paths and stream paths that connect all pairwise combinations of user-locations, but these features are still being developed. These features were designed for the purpose of studying animal movements, but there are many other applications of GIS path analysis. Analysis of animal movements has a long history using radio or GPS tracking devices on individual animals (Rodgers 2001), and landscape genetics is an exciting new field that measures movement of genes (and individuals that carry them) among populations (Manel *et al.* 2003, Storfer *et al.* 2010). For both approaches, it is necessary to delineate paths among animal locations (*e.g.* linear or stream paths) and to quantify landscape features along those paths. Different buffers are often applied to create two-dimensional sample areas, rather than linear transects. Developing new spatial scales for data collection, such as stream paths or linear paths, is made easier by Geodata Crawler’s pre-existing infrastructure of national geodatabases and multi-scale functionality.

New national geodatabases are constantly being added to Geodata Crawler as they become available or become of-interest to new projects. Future climate data represent a significant body of available GIS data that have not yet been incorporated into Geodata Crawler. This would expand possible research topics to include climate change research such as modeling potential effects on hydrology, species distributions, and gene flow among populations. Future climate data (Hijmans *et al.* 2005) have already been acquired, but have not yet been processed for Geodata Crawler. Processing will include identifying preferred climate models or multi-model averages, clipping data to include only the continental United States, and re-projecting data to the Albers NAD 1983 datum. New climate-related variables will then be developed and included in the user-options worksheet and supporting documentation. Geodata Crawler was designed to simplify the process of incorporating new data.

Geodata Crawler was envisioned as a centralized national GIS database and automated data crawler that accepts data requests through a website, and serves data through an FTP server to support any research that could benefit from rapid access to custom multi-scale GIS data for locations in the continental United States. This powerful new tool provides a template for a system to better distribute custom multi-scale GIS data to non-GIS researchers in support of a broad range of sciences. Special emphasis has been given to my own research interests and the interests of my collaborators for applications in modeling species distributions, hydrology, and gene flow, but the basic concept and the tool itself could be applied in many disciplines. Many national GIS servers have been implemented, but Geodata Crawler differs in several important ways: the output data contain site-specific samples from larger GIS datasets, the spatial scales of sample areas can be customized to meet specific research goals, the national geodatabase contains data from many different sources, project-specific GIS databases are created in the process, and the unique data archiving system can drastically reduce processing time by sharing data among user’s projects. Widespread-use of such a system could improve comparability of data among projects and disciplines, encourage new research topics by increasing data availability, support multi-scale analyses, and foster interdisciplinary collaboration.

## Acknowledgments

Geodata Crawler has been developed with constant encouragement and feedback from Scott Longing and Pablo Bacon. The Center for Advanced Spatial Technology (CAST) at the University of Arkansas provided software licensing, technical expertise, computers, and facilities, particularly Dr. Jason Tullis who gave valuable advice as a member of my dissertation committee, introduced me to Python programming, and provided access to his Windows server for Geodata Crawler development. Gary Huxel, Dan Magoulick, and Steve Stephenson also served on my dissertation committee throughout Geodata Crawler’s development and regularly offered constructive criticism and encouragement. James Falcone generously shared his national GIS datasets developed for hydrological modeling. Many collaborators have participated by utilizing Geodata Crawler’s data, providing feedback, and recommending new datasets. Collaborators have included Scott Longing, Dan Magoulick, Pablo Bacon, Marlis Douglas, Brad Austin, J. D. Wilson, Jackie Guzy, Bill Wolfe, Wyatt Hoback, and Thea Christianson.

## Work Cited

[AOGC] Arkansas Oil and Gas Commission. 2013. *Arkansas Oil and Gas Well Map*. Arkansas Oil and Gas Commission: Little Rock, AR. Available at <http://www.aogc.state.ar.us/Maps.htm>

Biewick LRH. 2008. *Areas of Historical Oil and Gas Exploration and Production in the United States.* U.S. Geological Survey Digital Data Series DDS-69-Q. U.S. Geological Survey: Reston, VA. Available at <http://pubs.usgs.gov/dds/dds-069/dds-069-q/>

Birth GS, McVey G. 1968. Measuring the color of growing turf with a reflectance spectroradiometer. *Agronomy Journal* **60**:640-643.

Carlisle DM, Falcone J, Wolock DM, Meador MR, Norris RH. 2010. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications* **26**:118-136.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* **28**:2031-2064. Data available at <http://www.prism.oregonstate.edu/normals/>

Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**:475-497.

Esri. 2013a. ArcGIS 10.2 for Desktop. Version 10.2.0.3348. Esri Inc.: Redlands, CA.

Esri. 2013b. ArcGIS for Server. Esri Inc.: Redlands, CA. Available at <http://www.esri.com/software/arcgis/arcgisserver>

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Heold N, Wickham J. 2011. Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **77**:858-864. Data available at <http://www.mrlc.gov/>

Gao, BC. 1996. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of Environment* **58**:257-266.

Hamon WR. 1961. Estimating Potential Evapotranspiration. *Journal of the Hydraulics Division, Proceedings of American Society Civil Engineers* **87**:107–120.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978. Data available at <http://www.worldclim.org/current>

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J. 2007. Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **73**:337-341. Available at <http://www.mrlc.gov/>

Huete AR, Justice C. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of the Environment* **25**:295-309.

Hunt CB. 1979. *Surficial geology of the conterminous United States*. U.S. Geological Survey, National Atlas of the United States of America. Available at <http://www.nationalatlas.gov/>

Jensen, JR. 2005. *Introductory digital image processing: A remote sensing perspective*. Pearson Education Inc.: Upper Saddle River, NJ.

Kauth RJ, Thomas GS. 1976. The tasseled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings, Symposium on Machine Processing of Remotely Sensed Data*, West Lafayette, IN: Laboratory for Applications of Remote Sensing. p. 41-51.

Key CH, Benson NC. 2005. Landscape assessment: remote sensing of severity, the normalized burn ratio and ground measure of severity, the composite burn index. In ‘FIREMON: Fire Effects Monitoring and Inventory System’. Eds. Lutes DC, Keane RE, Caratti JF, Key CH, Benson NC, Gani LJ. General Technical Report RMRS-GTR-164-CD. U.S. Department of Agriculture, Rocky Mountains Research Station: Ogden, UT.

King PB, Beikman HM. 1974. *Explanatory Text to Accompany the Geologic Map of the United States.* United States Geological Survey professional paper 901. Available at <http://pubs.usgs.gov/dds/dds11/pp901_html/1_PP901.HTML>

Krug WR, Gebert WA, Graczyk DJ. 1989. *Preparation of average annual runoff map of the United States, 1951-80.* U.S. Geological Survey Open-File Report 87-535.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**:189-197.

Musick HB, Pelletier RE. 1988. Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. *Remote Sensing of Environment* **25**:167-184.

Nakagaki N. 2007a. *Grids of Agricultural Pesticide Use in the Conterminous United States, 1992*. Version 2.0. United States Geological Survey: Reston, VA. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest92grd.xml>

Nakagaki N. 2007b. *Grids of Agricultural Pesticide Use in the Conterminous United States, 1997*. Version 2.0. United States Geological Survey: Reston, VA. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest97grd.xml>

[NASA] National Aeronautics and Space Administration. 2011. *Landsat 7 Science Data Users Handbook*. NASA Goddard Space Flight Center, Landsat Project Science Office: Greenbelt, MD. Available at <http://landsathandbook.gsfc.nasa.gov/>

[NOAA] National Oceanic and Atmospheric Administration. 2012. *Global DMSP-OLS Nighttime Lights Time Series 1992-2012 (Version 4).* NOAA National Geophysical Data Center: Boulder, CO. Available at <http://ngdc.noaa.gov/eog/download.html>

Price CV, Nakagaki N, Hitt KJ, Clawges RM. 2007. *Enhanced Historical Land-use and Land-cover Data Sets of the U.S. Geological Survey: Raster Format Files.* United States Geological Survey: Reston, VA. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/ds240_landuse_raster.xml>

Python. 2012. Python v2.7.3. Python Software Foundation. Available at: <http://www.python.org>

Qi J, Cabot F, Moran MS, Dedieu G. 1995. Biophysical parameter estimations using multidirectional spectral measurements. *Remote Sensing of Environment* **54**:71-83.

Reed JC, Bush CA. 2005. *Generalized Geologic Map of the Conterminous United States.* U.S. Geological Survey: Denver, Colorado. Metadata available at <http://nationalatlas.gov/metadata/geology48.xml>

Rock BN, Vogelmann JE, Williams DL, Vogelmann AF, Hoshizaki T. 1986. Remote detection of forest damage. *BioScience* **36**:439.

Rodgers AR. 2001. Tracking animals with GPS: the first 10 years. *Proceedings of the Conference on Tracking Animals with GPS.* The Macaulay Land Use Research Institute: Aberdeen, Scotland.

Rouse JW, Haas RH, Schell JA, Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium* **1**:48-62.

Schreve RL. 1966. Statistical law of stream numbers. *The Journal of Geology* **74**: 17-37.

Schruben PG, Arndt RE, Bawiec WJ, King PB, Beikman HM. 1994. *Geology of the Conterminous United States at 1:2,500,000 Scale – A Digital Representation of the 1974 P. B. King and H. M. Beikman Map.* United States Geological Survey: Reston, VA. Metadata available at <http://pubs.usgs.gov/dds/dds11/>

Seirup L, Yetman G. 2006. *U.S. Census Grids, 2000.* NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY. Available at <http://dx.doi.org/10.7927/H4B85623>

Smiley JD, Carswell WJ Jr. 2009. *The National Map—Hydrography*. U.S. Geological Survey Fact Sheet 2009-3054, 4 pp. Available at <http://nhd.usgs.gov/>

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. 2010. Landscape genetics: where are we now? *Molecular Ecology* **19**:3496-3514.

Strahler AN. 1952. Hypsometric (area-altitude) analysis of erosional topology. *Geological Society of America Bulletin* **63**:1117-1142.

[USDA] United States Department of Agriculture. 2006. *Digital General Soil Map of U.S.* United States Department of Agriculture, Natural Resources Conservation Service: Fort Worth, TX. Metadata available at <http://catalog.data.gov/harvest/object/69d5a4d1-a4ef-4208-bf4b-898c0b67a425/html>

[USDA] United States Department of Agriculture. 2011. *Active Fire Mapping Program.* U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center: Salt Lake City, Utah. Available at <http://activefiremaps.fs.fed.us/>

[USDC] United States Department of Commerce. 2011. *TIGER/Line Shapefile, 2011, State, Arkansas, Primary and Secondary Roads State-based Shapefile.* U.S. Department of Commerce, U.S. Census Bureau, Geography Division. Available at <http://www2.census.gov/geo/tiger/TIGER2011/PRISECROADS/tl_2011_05_prisecroads.zip>

[USEPA] United States Environmental Protection Agency, [USGS] United States Geological Survey. 2012. *National Hydrography Dataset Plus – NHDPlus.* Edition 2.10. Available at <http://www.horizon-systems.com/nhdplus/>

[USGS] United States Geological Survey. 2003. *Principal Aquifers of the 48 Conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands.* Version 1.0. U.S. Geological Survey: Reston, VA. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/aquifers_us.xml>

[USGS] United States Geological Survey. 2010. *National Land Cover Version 1.* US Geological Survey, National Biological Information Infrastructure, Gap Analysis Program (GAP). Available at <http://www.gap.uidaho.edu/Portal/DataDownload.html>

Vogelmann JE, Howard SM, Yang L, Larson CR, Wylie BK, Van Driel JN. 2001. Completion of the 1990’s national land cover data set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **67**:650-662. Available at <http://www.mrlc.gov/>

Wolock DM. 1997. *STATSGO Soil Characteristics for the Conterminous United States*. U.S. Geological Survey Open-File Report 97-656.

Wolock DM. 2003a. *Base-flow Index Grid for the Conterminous United States.* U.S. Geological Survey: Reston, VA. U.S. Geological Survey Open-File Report 03-263. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml>

Wolock DM. 2003b. *Estimated Mean Annual Natural Ground-water Recharge in the Conterminous United States.* U.S. Geological Survey: Reston, VA. Open-file report 03-311. Metadata available at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/rech48grd.xml>

Wolock DM. 2003c. *Infiltration-excess overland flow estimated by TOPMODEL for the conterminous United States*. U.S. Geological Survey Open-File Report 03-310.

Wolock DM. 2003d. *Saturation overland flow estimated by TOPMODEL for the conterminous United States.* U.S. Geological Survey Open-File Report 03-264.

Wolock DM, Ayers MA, Hay LE, McCabe GJ Jr. 1989. Climate change effects on watershed runoff. *American Society of Civil Engineers, Hydraulic Engineering ’89 Proceedings*, p. 673-678.

Wolock DM, McCabe GJ. 1995. Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* **31**: 1315-1324.

Wolock DM, McCabe GJ. 1999. Estimates of runoff using water-balance and atmospheric general circulation models. *Journal of the American Water Resources Association* **35**:1341-1350.

## Research with Geodata Crawler

**Papers:**

Leasure DR, Magoulick DD, Longing SD. 2014. Natural flow regimes of the Ozark and Ouachita Mountain region. *In review* for *River Research and Applications*.

Leasure DR. 2014. Landsat-based monitoring of an endangered beetle: Addressing issues of high mobility, annual life history, and imperfect detection. *In review* for *Landscape Ecology*.

**Funded Proposals:**

Magoulick DD, Leasure DR. 2014. Quantification of hydrologic alteration and relationships to biota in Arkansas streams: Development of tools and approaches for un-gaged streams. State Wildlife Grant, Arkansas Game & Fish Commission. $53,001.00

**Professional Reports:**

Willson JD, Guzy JC. 2014. Occupancy and habitat relationships of stream-associated salamanders in intensively managed forests of the Ouachita Mountains Ecoregion. Progress Report for Weyerhaeuser NR Company.

Willson JD, Guzy JC. 2013. Occupancy and habitat relationships of stream-associated salamanders in intensively managed forests of the Ouachita Mountains Ecoregion. Progress Report for Weyerhaeuser NR Company.

**Professional Presentations:**

Leasure DR. 2014. *Invited Symposium Speaker—*A foundation for Arkansas eFlows: Hydrologic classification and flow alteration modeling. Symposium: Environmental Flows: What, Why, and How? Arkansas Water Resources Center, Annual Conference: Fayetteville, AR.

Leasure DR. 2014. *Invited Symposium Speaker—*“Big data” squared: GIS, hydrological, and remote sensing data coupled with gene flow and species distributions. Symposium: Big Data Science and Its Impacts on Fish Conservation and Management. American Fisheries Society: Quebec City, Quebec, Canada. Abstract available online at <https://afs.confex.com/afs/2014/webprogram/Paper16999.html>

Leasure DR. 2014. *Invited Seminar Speaker—*Habitat of endangered American burying beetle in western Arkansas: Addressing issues of spatial scale and detection for satellite-based monitoring. Department of Biology, University of Nebraska, Kearney: Kearney, NE.

Leasure DR. 2014. *Contributed Oral Presentation—*Natural flow regimes of the Ozark-Ouachita Interior Highlands region. USGS Cooperators Meeting: Mayflower, AR.

Leasure DR. 2014. *Contributed Oral Presentation—*Assessment of hydrologic alteration at un-gaged streams. USGS Cooperators Meeting: Mayflower, AR.

Leasure DR. 2013. *Invited Symposium Speaker—*The modern era of “big data” in GIS: Multi-scale modeling of species distributions, hydrology, and gene flow. Symposium: Finding Simplicity In Complexity: Matching Models To Data. American Fisheries Society: Little Rock, AR. Abstract available online at <http://afs.confex.com/afs/2013/webprogram/Paper12193.html>

Leasure DR. 2013. *Invited Seminar Speaker—*From NASA satellites to DNA microsatellites: 21st century conservation tools. Conservation Biology, Department of Biological Sciences, University of Arkansas - Fort Smith: Fort Smith, AR.

Leasure DR. 2013. *Contributed Oral Presentation—*Geodata Crawler: A centralized national geodatabase and automated multi-scale data crawler to overcome GIS bottlenecks in data analysis workflows. Ecological Society of America: Minneapolis, MN. Abstract available online at <http://eco.confex.com/eco/2013/webprogram/Paper43858.html>

Leasure DR. 2013. *Contributed Oral Presentation—*Landsat to monitor an endangered American burying beetle population. Arkansas Entomological Society meeting, Southern Arkansas University: Magnolia, AR.

Leasure DR. 2013. *Invited Speaker—*Finding endangered beetles from space: Landsat to monitor American burying beetle habitat. U.S. Fish and Wildlife Service American burying beetle science meeting, Oklahoma State University: Stillwater, OK.

Longing DD. 2013. *Invited Seminar Speaker—*Biogeography and conservation of diving beetles in the southeastern U.S.  Department of Biological and Environmental Sciences, Georgia College and State University: Milledgeville, GA.

Longing SD, Bacon PA, Harp GL. 2013. *Contributed Oral Presentation—*Distribution, conservation and current status of three endemic *Heterosternuta* (Coleoptera: Dytiscidae: Hydroporinae) in Arkansas. Arkansas Entomological Society, Southern Arkansas University: Magnolia, AR.

Longing DD, Wolfe GW, Leasure DR. 2013. *Invited Symposium Speaker—*Distributions, ecology and conservation of diving beetles across endemic hotspots in Arkansas and Tennessee, USA. Symposium: When a Blind Beetle Crawls Over the Surface of the Globe, or Under the Water:  Biodiversity and Systematics of Aquatic Beetles. Entomological Society of America Annual Meeting: Austin, TX.

Leasure DR, S Longing, PA Bacon, GR Huxel. 2012. *Contributed Oral Presentation—*A new automated tool for multi-scale sampling of spatial environmental data to predict the distribution of the Sulphur Springs diving beetle in northwest Arkansas. Entomological Society of America: Knoxville, TN. Abstract available online at <http://esa.confex.com/esa/2012/webprogram/Paper68051.html>

Leasure DR. 2012. *Contributed Oral Presentation—*Streamscapes: An automated GIS tool for sampling landscapes associated with streams. Ecomunch Seminar, University of Arkansas: Fayetteville, AR.

Leasure DR. 2011. *Poster Presentation—*Remote sensing and GIS to model endangered American burying beetle abundance across a landscape and to determine the optimal spatial scale for habitat samples. Ecological Society of America: Austin, Texas. Abstract available online at <http://eco.confex.com/eco/2011/webprogram/Paper31938.html>

**Student Reports:**

Beacher J, Magoulick DD. 2013. Effect of land use and hydrologic disturbance on crayfish assemblages in the Ozark Highlands. Research Experience for Undergraduates (REU) Program, University of Arkansas: Fayetteville, AR.

Coffman M, Longing SD. 2013. Monahans Sandhills: Land cover change from oil and gas production, 1900 – 2005. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Davis L, Longing SD. 2013. Pesticide use on cultivated crops in five Texas regions. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Gust S, Magoulick DD. 2013. Influence of natural and anthropogenic factors on stream fish assemblages in Arkansas. Research Experience for Undergraduates (REU) Program, University of Arkansas: Fayetteville, AR.

McKelvey D, Longing SD. 2013. Analysis of factors affecting the distribution of the red imported fire ant, *Selnopsis invicta*. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Parks S, Longing SD. 2013. A landscape comparison of two Texas olive orchards. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Poole J, Longing SD. 2013. Pesticide use in four Southern High Plains watersheds in Swisher County, Texas. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.