FINAL REPORT

Federal Aid Grant No. F16AF00296 (T-90-M-1)

Oklahoma Biodiversity Information System: An Integrated, Online Data System for Conservation Planning

Oklahoma Department Of Wildlife Conservation

1 April 2016 – 31 March 2020

FINAL REPORT

STATE: Oklahoma **GRANT NUMBER:** F16AF00296 (T-90-M-1)

GRANT PROGRAM: State Wildlife Grants

GRANT NAME: Oklahoma Biodiversity Information System: An Integrated, Online Data System for Conservation Planning

GRANT PERIOD: 1 April 2016 to 31 March 2020

PRINCIPAL INVESTIGATOR:
Bruce Hoagland
Oklahoma Biological Survey and
Department of Geography and Environmental Sustainability
University of Oklahoma

Dan Hough
Oklahoma Biological Survey
University of Oklahoma

Todd Fagin
Oklahoma Biological Survey
University of Oklahoma

ABSTRACT:

The Oklahoma Biodiversity Information System (OBIS) is an integrated, online data management tool designed to facilitate biodiversity conservation planning and collaborative data sharing between the Oklahoma Natural Heritage Inventory (ONHI), state and federal agencies, and non-governmental organizations in Oklahoma. In 2016, staff at the Oklahoma Natural Heritage Inventory began to collate data from existing, disparate databases into a central repository. These initial datasets included the Distribution of Oklahoma Amphibian and Reptiles by Recorded Sightings (DOKARRS), Oklahoma Mammal Specimen Records (OMSR), Oklahoma Bird Specimen Records (OBSR), Oklahoma Fisheries Research Laboratory database (Fisheries), and the Oklahoma Natural Heritage Inventory Biotics database (Biotics). We then created a database structure that was compliant with the Darwin Core data standard to consume these data. Subsequently, we integrated additional datasets that contain records for species of greatest conservation need into the existing data structure. Concurrent with the development of the database, we developed two separate web interfaces to consume and disseminate the underlying data. The first, developed in-house under the direction of the PIs, provides a use-friendly interface to search for taxa known to occur in the state of Oklahoma. It provides both basic taxonomic and distribution information for these taxa. This platform also provides an editing interface for select users (collaborators) to edit existing and add new records. The second platform was developed by a third-party contractor to provide more detailed individual occurrence information for select users. Both platforms underwent beta-testing by ONHI staff

and are now available. Additional beta-testing by external partners is on-going to ensure proper functionality.

OBJECTIVE:

To develop the Oklahoma Biodiversity Information System, an integrated, web-based application that draws upon multiple existing biodiversity databases in order to provide data for species of greatest conservation need, federally-listed threatened and endangered species, and other elements of biodiversity to members of the Oklahoma conservation community in real time for biodiversity conservation and planning.

INTRODUCTION:

Informed and accurate conservation planning relies on access to current, spatially explicit data for the occurrences of species. In Oklahoma, the Oklahoma Natural Heritage Inventory (ONHI), a program of the Oklahoma Biological Survey (OBS), is the designated centralized repository for biodiversity data. The OBS enabling legislation states that it must "acquire, archive, process and disseminate information on biological resources and natural areas that is or could be of value to policy and decision makers in the state (§70 3314)." To achieve this mandate, the OBS and ONHI have historically maintained a collection of databases related to the biodiversity of Oklahoma. These data are provided to interested state and federal agencies, NGOs, consulting firms, and the public. For example, the ONHI staff responded to 664 information requests in 2019 based on information stored in the conservation database.

The ONHI is not the only agency charged with inventory and monitoring of rare species in Oklahoma. Members of the Oklahoma Conservation Exchange Group (core members include staff from the ONHI, Oklahoma Department of Wildlife Conservation, U.S. Fish and Wildlife Service-Tulsa, and The Nature Conservancy) recognize the need for effective methods and protocols to integrate relevant conservation data collected by sister agencies and other conservation groups into the ONHI database, thereby enhancing the accessibility of these data to conservation planners.

At the onset of this project, data relevant to conservation planning in Oklahoma were distributed among several state and federal agencies and private entities. Collating those data into a centralized data repository had been a challenge, and past attempts, which were based upon completion of paper forms that were mailed to the ONHI database manager, were ineffective. Conservation Exchange Group members agreed that a web-based interface would be an effective way for partners to share and view data. Partners did not have the ability to view data in real-time, thus it was our intention to develop an integrated data interface to provide access to ONHI/OBS datasets, that also allowed users to contribute data, and to search and view data through a single portal.

APPROACH:

We divided this project into five, sometimes concurrent components:

1. Discuss OBIS development with collaborators and stakeholders;
2. Migrate and modernize the existing conservation databases housed at the OBS into a PostgreSQL/PostGIS database
3. Develop a web interface
4. Beta testing of the web interface by OBS staff and external collaborators
5. Database management and maintenance

1) Discussion with Collaborators and Stakeholders

In order to best serve the needs of our conservation partners, we consulted with multiple collaborators and other relevant stakeholders regularly throughout the development. Collaborators and stakeholders included but were not limited to, biologists and data managers at the Oklahoma Natural Heritage Inventory, Oklahoma Department of Wildlife Conservation, the U.S. Fish and Wildlife Service, Oklahoma Conservation Commission, Sam Noble Museum of Natural History, and The Nature Conservancy. We also made presentations regarding the project at a several regional conferences, including the Oklahoma Natural Resources Conference, in order to let conservation professionals know about this project and to seek feedback from potential stakeholders. In addition to discussing database structure, data integrity, and data sharing agreements, we sought input on access to sensitive species data. Such discussions occurred throughout the project duration and we did our best to incorporate ideas and address concerns from these individuals and agencies.

2) Migrate and Modernize Existing Conservation Databases

The initial source databases for OBIS were the DOKARRS, OMSR, OBSR, Fisheries, and Biotic databases. Each of these was designed at a different time, developed for a different purpose, and adhered to a different data standard. For instance, both DOKARRS and Biotics were observation databases, whereas the OMSR and OBSR were specimen-based databases (based on specimens located at state and regional museums). Additionally, the taxonomy used in the various databases were not always consistent or up-to-date. Lastly, there was not consistency in how similar data attributes were stored across the differing databases.

Based on a review of each of the source database schemas and discussion with collaborators, we developed a database schema to accommodate these disparate data sources and cross-walked the data into the newly created database over a two-year period. Additionally, we employed student workers to examine the spatial data associated with many of the individual records within these databases and to assign these to more precise spatial coordinates where possible.

3) Develop a Web Interface

Upon project initiation, we had hoped to develop a single, integrated web-based application to access the OBIS database in a seamless manner. The web interface was designed to provide multiple functions for the user community, such as data access, creation, export, editing, and analysis. While we have largely adhered to the initial goals, we discovered that a single web

interface to accommodate all user needs simply was not feasible. As a result, we developed two stand-alone web interfaces. The first is multi-functional and provides users the ability to search for taxa known to occur in the state of Oklahoma and acquire basic taxonomic and distribution information for these taxa. This interface also provides an editing platform for select users (collaborators) to edit existing records for which they have access and to add new records. The second interface was developed by a third-party contractor to provide more detailed individual occurrence information for select users. This interface is designed specifically for those collaborators who need to see the full range of underlying data, not just the basic taxonomic and distributional information.

4) Beta-testing of Web Interface

Because the OBIS depends on the contribution of data from collaborators to maintain the most current and complete conservation database, we invited select individuals to help beta test the web application. Beta testing occurred throughout all stages of database/web interface development but relied on OBS staff during the early stages of development. However, with the completion of the version 1.0 of OBIS and in preparation for a full public roll-out, we invited collaborators from state agencies (ODWC, OCC, Oklahoma Water Resources Board) to perform a rigorous, structured testing of the applications. This process is ongoing at the conclusion of the grant.

5) Database Management and Maintenance

OBIS is a living database because it is under a constant state of addition, maintenance, and modification. As referenced earlier, we used the help of OBS staff and employed student workers to validate the data in the initial databases used to create OBIS and refined the geospatial coordinates of these data where possible. We also sought out additional datasets that contained records and spatial data for species of greatest conservation need. One of these was a database of Oklahoma lepidoptera records that contained observational and photographic records for several butterfly SGCN. As we have sought additional data to include in OBIS, we have discovered that additional data fields and even tables may be required to better serve the conservation community. It is our intention to continue to update OBIS with new records throughout the lifespan of OBIS and beyond the completion of this grant.

RESULTS AND DISCUSSION:

1) Discussion with Collaborators and Stakeholders

Throughout the project duration, we were in discussion with collaborators and stakeholders about the database and web interface development. The earliest conversations were held primarily with OWDC staff, but as the project developed further, we had both direct and indirect conversations with collaborators and stakeholders. For instance, we routinely presented project updates in various forums, including at annual CEG meetings, where we received direct feedback from collaborators at TNC and the USFWS. Similarly, we presented our progress at several state and regional conferences that elicited responses from interested parties and biologists across the state who could contribute data to and/or benefit from OBIS.

2) Migrate and Modernize Existing Conservation Databases

During the first two years of the project, substantial time went into the design and implementation of a database schema that could adequately accommodate the data from the disparate source databases. To begin this process, we constructed four taxonomic databases to capture the various taxa known to occur in the state. The first of these tables, known as acctax, was designed to capture the currently accepted scientific name of each taxon, as well as other relevant information including its federal and state protection status (if any), its species of greatest conservation need tier (if any), its state and global Heritage ranks, its nativity, and more. The second taxonomic table, known as syntax, has a many-to-one relationship to the accepted name table and contains synonyms and other basic taxonomy information. The syntax table is related to the acctax table via a unique code assigned to each taxon. Similarly, a table known as comtax was developed to house the common names associated with each taxon. The comtax table, has a many-to-one relationship to the acctax table and contains the various common names associated with each accepted taxon. Lastly, the table hightax was created, which forms a one-to-many relationship to acctax based on taxonomic family and contains the higher taxonomic information (family to kingdom) and other vernacular taxonomic classifications. All of these taxonomic tables contain Darwin Core-compliant fields.

Though the tables are dynamic, at the conclusion of the grant and at the time that the Final Report was prepared, the acctax table contains 9,974 records, of which 517 are birds, 144 are mammals, 2,808 are invertebrates, 299 are herptiles, 234 are fish, and 5,879 are vascular plants. The remaining records are for various community types we have also captured. Similarly, the syntax table contains 11,194 records, of which 183 are birds, 496 mammals, 1,336 are invertebrates, 343, are herptiles, 759 are fish, and 8,077 are plants. Additionally, the comtax table contains 15,967 common names related to 8,997 records, while the hightax contains 463 records.

In addition to the four taxonomic tables, OBIS contains a table, called Occurrence, that is used to document individual observations or collections of a taxon. The Occurrence table contains 66 fields, most of which follow the Darwin Core standard. These fields are occurrence/record specific and include a column to indicate the taxon observed or collected, the event date, the observer(s) or collector(s), site specific information (such as county, locality, habitat, longitude and latitude, etc.), individual characteristics (such as count, behavior, life stage, sex, etc.), and information about the observation or collection (e.g. collector id, accession number, identification confidence, etc.)

The initial project plan was to develop a common database schema for each of the source databases to ensure consistency between them. Nonetheless, these were to reside in separate tables within the single database management system. However, after development of the schema and initial database testing, the project team decided that it would be most efficient to store all these occurrence records in a single occurrence table containing a source field to tie back to the original database. To date, 134,818 occurrence records have been included in the table from the original databases, of which 17,843 are from DOKARRS, 13,255 are from OMSR, 5,329 are from OBSR, 52,484 are from Fisheries, and 45,907 are from Biotics.

Since initial database development, we have incorporated other datasets into OBIS and added over 50,000 records provided by the Sam Noble Museum Oklahoma Museum of Natural History

(SNOMNH), the ODWC, as well as 29,556 fish records from the Oklahoma Conservation Commission (OCC). Additionally, we have received private collection data from state naturalist John Fisher, aquatic macroinvertebrate data from OCC, and fish collections from the Jimmy Pigg repository; however, these datasets have not been added to OBIS yet because they require additional work before a data crosswalk is feasible.

In addition to the core taxonomic and occurrence tables, the OBIS database is made up of multiple look-up tables to constrain allowable values added to field (to ensure data integrity), spatial tables, and recently added distribution tables to facilitate data exchanges with NatureServe. Ongoing discussions with collaborators that perform repeated sampling at monitoring sites identified a need to add a table for site-specific table for locations at which there have been repeat surveys. This change will be incorporated into a future iteration of the OBIS database.

3) Develop a Web Interface

The development of the OBIS web interface took many unexpected twists and turns. We had initially contracted with developers at the University of Oklahoma Department of Libraries to develop an application programming interface (API) based on the Django REST Framework. This API served as the "backend" of the web interface and the frontend was to be developed by ONHI. However, mid-way through development, the project lead at OU Libraries took an out-of-state job. Our project was shifted to his replacement, and more specifically, his graduate research assistant, putting us behind in our development schedule. We concurrently discovered some bugs and incompleteness in the API development.

At this time, we decided to approach a third-party developer to see if he could work with our existing API. After several discussions, we decided to use this developer to create a data mining application to access the tabular data in OBIS. This application is designed for select users with high level permissions. Concurrently, we hired a student programmer to fix the bugs in the API and to develop a public facing web interface. While the two interfaces access the same data and have some functionality in common, they have distinctly different purposes.

The basic viewer interface currently has two levels of access, public and authenticated users. The former provides a user-friendly search interface to search for taxa known to occur in Oklahoma. Users can search by currently accepted scientific name, synonyms, or common name. When a record is selected (regardless of the search string executed), the user is directed to an overview page, which provides basic taxonomic information, such as the currently accepted scientific name with authority, synonyms, common names, state and global conservation ranks, federal and state protection status, and higher taxonomic information. Users also can select an interactive distribution map to see both the counties in which the selected taxon occurs, as well as the distribution of geocoded occurrences with 5 km^2 hexagons. Users can view and export an occurrence table, which shows the counties in which the selected taxon occurs, the numbers of occurrences from each county, and the first and last observation dates. Additionally, users can access a more detailed map that, in addition to the county/5 km^2 hexagons, provides several additional data layers and locational and attribute search functionality. Users have the option to begin their search on the detailed map page. The search page is currently located at the following URL. We will soon acquire additional domain names to make access easier and memorable: http://obsvweb1.ou.edu/obis/search/

A select number of users will also be able to access the Collaborators page (http://obsvweb1.ou.edu/obis/collaborators/). Once logged in, users will have access to the same search functionality available in the public viewer. The major difference, though, is that the map views show actual occurrence locations (X,Y points) instead of 5 km^2 hexagons. Moreover, the attribute information associated with each X,Y point is more detailed.

The real power in the collaborator interface, though, is the ability to add and edit data. Authenticated users will be presented with a new menu item, Collaborators Menu, which provides functionality to add a new taxon, edit an existing taxon, add a new occurrence, and edit an existing occurrence. When adding a new taxon, the interface automatically connects to all of the underlying taxonomic tables, but in a seamless manner. As a result, individuals can add a new accepted scientific name, any associated synonyms and common names, and distributional data information. The editing is form based and enables users, depending on which fields are being added, to search the underlying database for the appropriate values or select constrained field values from a dropdown list.

Editing an existing taxon is similar. Individuals are presented with a search page that looks and is functionally identical to the basic search page. However, when an individual selects a taxon (whether searched by currently accepted scientific name, synonym, or common name), a page is displayed with all of the taxonomic information associated with the various taxonomic tables. Users can then edit any necessary values, add new synonyms or common names, etc.

Adding a new occurrence is similar to adding a new taxon, though it only connects directly to one table, the Occurrence table, plus any necessary look-up tables to constrain allowable values. The interface is form-based and provides mechanisms to search for values (such as the scientific name of the occurrence being added) or select allowable values from a dropdown list. To edit an existing occurrence, though, users are provided with an advanced search to find the correct occurrence record to edit. Users can search by county, family, genus, species, common name, collector, collection date, catalog number, or any combination thereof. A list of possible occurrences that meet the search criteria is displayed and users can then select the appropriate record to edit.

The data mining application (https://obis.coordinatesolutions.com/) has similar basic functionality, but with a different purpose in mind. Whereas the basic view is more taxonomic-based and shows basic distributional information, the data mining application is occurrence-based and shows detailed views of individual occurrences. This same information can be found on the other platform in collaborator mode when selecting individual occurrences to edit, but the data mining application allows for viewing more detailed information for multiple occurrence records much easier. In the data mining application, a user can search for a taxon in two ways, a free text search or an advanced search (based on family, genus, species, observer, common name, institution, catalog number, etc.). All occurrences meeting the search criterion/criteria are displayed on the map (XY point locations) and a grid view appears showing tabular information associated with each occurrence, such as family, scientific name, event date, dataset name, county, institution code, and catalog number. Users also can click on an individual occurrence record and get more detailed information about that occurrence. The data mining interface also has an interactive map with various layers and export tools to export tables. The data mining application has so-called reporting tools that have an advanced search interface that allows users to create detailed reports of occurrences or to create species lists for selected counties.

Although both interfaces are fully functional, we continue to work on enhancements and to identify potential bugs in each. Because of the unanticipated slowdown due to the COVID-19 pandemic, some of these enhancements and fixes have taken longer to implement than expected.

4) Beta-testing of Web Interface

Various iterations of both the basic viewer and data mining viewer have been available for nearly a year. During this time, we conducted extensive testing of each interface by the project's principle investigators and developers (alpha testing). However, we presented the project at several interagency meetings to demonstrate the interfaces for our collaborators and received valuable feedback in the process (beta testing)

As the grant ends, we continue the beta testing phase and have documented several extensive test cases. We continue the process of writing more test cases and working with our designated testers, while the web interfaces remain available for use.

5) Database Management and Maintenance

The OBIS is an ongoing development project. While the database and API backends are stable and the two web interfaces are functional, these are undergoing constant updates and modifications. For instance, as we opened up the database to consume data from more sources, we discovered some limitations to our current data model to capture the broad diversity of data that different agencies and individuals collect in association with their occurrence information. As a result, we continue to modify existing tables and add new tables to accommodate these data sources (we attempt to be sparing in this). Additionally, any changes in database structure, whether new columns or new tables, require updates to the API and occasional to one or both of the web interfaces. As such, all levels of development remain very much dynamic.

In addition to this maintenance, we have created extensive documentation record associated with all aspects of development. This includes, but is not limited to, a data dictionary, Federal Geographic Data Committee-complaint metadata, and web interface development notes. We also prepared a draft help document for users of the basic viewer and editing interface.

RECOMMENDATIONS:

As noted, OBIS is a dynamic data delivery system and is in a constant state of update. This includes the addition of new datasets to the database; updating of existing records in the database (such as geocoding, better parsing the data into appropriate fields to be more Darwin Core complaint, and removal of duplicates due to incorporating data from multiple data sources); and enhancing the user experience by altering the web interface based on user recommendations. Additionally, we frequently update the OBIS backend components because the underlying technologies, themselves, are also routinely updated. It is our goal to seek ongoing funding to support these maintenance operations and to support enhancements going forward.

SIGNIFICANT DEVIATIONS:

There were three deviations from the original project narrative, none of which we view as significant. First, we initially planned to use GeoServer and OpenLayers as the GIS data server

and the map interface, respectively. While the data mining application uses the latter, the basic viewer, at the request of ODWC, utilizes ArcGIS Server and the ArcGIS API for JavaScript for the GIS data server and map interface, respectively. Second, we had originally intended to host all source datasets in separate tables; however, we discovered this was not the most efficient way to store the occurrence data, nor was it a practical approach when incorporating future datasets into the OBIS (which we have already begun to do). As such, the backend database now uses a single occurrence table to capture data from multiple data sources. Lastly, and most significantly, we developed two separate web interfaces. Based on initial feedback and the realization that the basic web interface wouldn't appropriately accommodate the needs of a select number of users, we developed a second data mining application to access more detailed occurrence information. Despite these deviations, we have been able to successfully accomplish our project goals.

EQUIPMENT:
No equipment was purchased under this grant.

PREPARED BY: Bruce Hoagland and Todd Fagin
Oklahoma Natural Heritage Inventory
Oklahoma Biological Survey/University of Oklahoma

DATE: 10 August 2020

APPROVED BY: Russ Horton, Wildlife Division Assistant Chief
Oklahoma Department of Wildlife Conservation

Andrea Crews, Federal Aid Coordinator
Oklahoma Department of Wildlife Conservation