FINAL REPORT

FEDERAL AID GRANT NO. T-12-P-1

DIGITAL ATLAS OF OKLAHOMA FISHES: PREPARATION OF DATA FROM OKLAHOMA STATE UNIVERSITY

OKLAHOMA DEPARTMENT OF WILDLIFE CONSERVATION

OCTOBER 1, 2003 through SEPTEMBER 30, 2006

FINAL REPORT

State: Oklahoma Grant Number: T-12-P-1

Grant Program: State Wildlife Grants

Grant Name: Digital Atlas of Oklahoma Fishes: Preparation of Data from Oklahoma State University

Grant Period: 1 October 2003 through 30 September 2006

A, Abstract

We developed a computer-based system, the Digital Atlas of Oklahoma Fishes (DAOF), to manage and disseminate information about fish collections in Oklahoma. The DAOF website is comprised of four interconnected user interfaces: an information based website, a database management system, an interactive map, and a database query module that interacts with the interactive map and database management system. The information based website provides users with a variety of information about Oklahoma's fishes and aquatic resources. The database management system contains collection information from databases at the state's two largest museums at Oklahoma State University and the University of Oklahoma. Each database is separately maintained by curators at these institutions. The interactive map uses ArcIMS® to display the location of fish collections in Oklahoma, and several mapping tools are available for users to customize their maps. The query modules, which are accessed through the interactive map website, enable users to query data by waterbody type and species. Theses queries can be exported to Microsoft Excel spreadsheet files. We developed additional code to automate database entry, editing, and maintenance, which can be incorporated in the future. We recommend ongoing support of the website to conduct site maintenance and updating, and the addition of fish collection information from other credible sources including other museums and documents such as scientific articles, theses, reports.

B. Objective:

To develop a database of Oklahoma fish collections based on specimens housed in the Oklahoma State University Collection of Vertebrates, and to combine this database with a database of fish collections from the Sam Noble Oklahoma Museum of Natural History to produce a GIS-based digital atlas of fishes in Oklahoma.

C. Need:

General information is available on the distribution and abundance of fishes in Oklahoma (Miller and Robison 2004), and although specific information on fish collections made by several state and federal agencies was recently compiled into the Oklahoma Streams Information System (OSIS) by Tejan and Fisher (2001), records of fish collections housed in university museums are lacking. Compilation of these records into a georeferenced, digital database and geographic information system (GIS) would provide verifiable information on specific locations of Oklahoma fishes. This information would be of value to state and federal agencies for management and conservation efforts, the scientific community for biogeographical and ecological research, and the general public for educational purposes. Furthermore, integration of museum fish collection information into OSIS would enhance the utility of this information system for fish habitat improvement projects.

General information on range and abundance patterns of individual fish species in Oklahoma is available in the scientific and popular literature, but there is no comprehensive, detailed document for all fish species in the state. Such a document would be invaluable for management and conservation plans and would assist in identifying areas of the state that are in need of sampling. Such a document would also assist in public education on natural resources and conservation, and would be available to the scientific community for use in basic ecological and biogeographic research.

Other state fish and wildlife agencies or museums have a digital atlas of fish species available on the World Wide Web (see below) and the Oklahoma Biological Survey is in the process of developing a database and website for the vascular plants of Oklahoma (funded by the National Science Foundation). Information available on the websites ranges from a photo and brief description of habitat (e.g., A Guide to Arkansas Wildlife) to comprehensive information on distribution, life history, and conservation status (e.g., sites listed on the Fish and Wildlife Information Exchange).

Several examples of fish databases are listed and annotated below: The Fish and Wildlife Information Exchange (fwie.fw.vt.edu/WWW/spp.htm), developed by the Conservation Management Institute in the College of Natural Resources at Virginia Tech, has helped develop and/or manage websites for Colorado, Kentucky, Missouri, New Hampshire, New Jersey, New Mexico, Oregon, Pennsylvania, Tennessee, Virginia, West Virginia, and Belize. As an example, the data on fishes on the Kentucky website includes a photo, taxonomy, conservation status (global, national, state), distribution by state, distribution by watershed (HUC), both recent and historical, life history information and references for each species. Texas Freshwater Fishes Index: Images, Maps and Information (www.tmm.utexas.edu/tnhc/fish/na/txindex.html). This site includes spot maps, photographs, and links to the University of Texas Memorial Museum archived collection of fishes. Fishes of Illinois (www.inhs.uiuc.edu/cbd/ilspecies/fishsplist.html) – This page includes spot distribution maps with historical records (pre-1979) distinguished from current distribution. A Guide to Arkansas Wildlife (Fish) (www.agfc.state.ar.us/critters/wildlifee_fish.html) - This page focuses only on game fish species and is designed to inform the public about general biology and record catches for the species covered. A photo of each species is provided, but only limited notes on habitat and distribution are given.

Data are available for development of a digital atlas of fish species in Oklahoma, but those data are currently maintained in separate databases by the Sam Noble Oklahoma Museum of Natural History at the University of Oklahoma, and by the Oklahoma State Vertebrate Research Collection. Specimens archived in these museums represent collections made throughout Oklahoma since the 1920s. The fish collection at the Sam Noble Oklahoma Museum of Natural History comprises approximately 40,000 lots (a lot consisting of all individuals of a given species collected at a given time and place) representing an estimated 3000 collections. All counties of the state are represented, although some have a limited number of samples. The compilation of the SNOMNH and OSU databases will help to identify those drainages in need of further sampling.

D. Approach:

The digital atlas of Oklahoma fishes (DAOF) was a cooperative effort between investigators at Oklahoma State University (OSU) and the University of Oklahoma (OU). Data for DAOF are from Oklahoma State University Collection of Vertebrates (OSUCOV) and University of Oklahoma, Sam Noble Oklahoma Museum of Natural History (SNOMNH).

The DAOF website was developed with four interconnected user interfaces: an *information based website*, a *database management system*, an *interactive map*, and a *database query module* that interacts with the interactive map and database management system. A website, available in both ASP.NET® and HTML formatting, acts as the interface by which the other modules are accessed. This information based website also contains information about Oklahoma's fishes and provides users with resources to better utilize DAOF. Secondly, an interactive map displays database information, which are stored in a Geodatabase, from the two respective museums. This interactive map utilizes ArcIMS® and Visual Basic® programming language. The database is managed through Microsoft SQL in conjunction with Visual Basic® and ASP.Net®, which allows it to be queried, new data to be entered into it, and old data to be edited. Additionally, a museum loan information interface allows museum curators to better manage collections. Lastly, database queries can be built that allow users to search for specific information and results may be displayed in the ArcIMS® interface or saved as a Microsoft Excel® file.

The online database structure is designed for the long-term maintenance of the databases. The databases (OSU and OU) are maintained on a server that is maintained by OSU. This server is backed up daily at a remote site and has two forms of backup power. The Microsoft SQL® database format allows data entry and editing through online interfaces and provides a stable platform that is easily converted to Microsoft Access®.

Information Based Website

The information based website provides an interface to access the ArcIMS® and database management systems. Additionally, this site interacts with the other modules to provide users additional information about Oklahoma's fishes, their habitats, and other information about aquatic resources in the state. Incorporated in this website are supporting documents and information to allow users to better navigate the interactive map and database; educate visitors about important issues related to fishes, water resources, habitats, and angling opportunities in Oklahoma; and provide links to other informational materials such as NatureServe.org species summaries, ODWC web pages, and educational sites and programs.

The purpose of the information based website is to: (1) provide users with background information to complement the information contained in the database; (2) increase the awareness of users to issues and unique Oklahoma water resources; and (3) support OSU, OU, and ODWC education efforts. The documents on the website provide researched information with links to supporting evidence that allow users to examine a topic in greater detail. We developed an organizational structure for the website that emphasizes broad issues affecting fishes, habitats, and water quality in Oklahoma and provide educational resources such as links to educational websites and ODWC information on specific resources (i.e. angling opportunities, educational programs), information about Oklahoma fishes, and educational and career opportunities. Additionally, the website provides access to the museum databases and ArcIMS® mapping interface as well as an avenue for feedback and museum loan requests.

We envision this website to be dynamic as new information is developed and added and user feedback provides direction for changes and additions. Although we took great efforts to research and prioritize information included in the informational portion of the website, we invite user feedback to further update and improve this information.

Query Interface

The query interface allows users to interact with the databases to search, download, and/or view records in a GIS setting. The query interfaces with the SQL database enabling users to develop specific queries of fish species or by locations or develop cross queries that are species and location specific. Queries are returned in a tabular format and can be printed or exported to Microsoft Excel®.

Database Management

Several interdependent databases are used to store and disseminate information in the DAOF. First, OSU and OU museum collection databases provide information stored in a many-to-one relationship (Figure 1). We created a database structure that the two databases share. This structure uses "Collection_EventID" as a primary key (referred to as "FieldNumber") that links all information within a collection event together (Table 1, Figure 1). A collection event has a unique collection location and data and may contain several museum records that represent individual species catalogued from a collection event. There is a one-to-many relationship between FieldNumber and MuseumNumber, which represents each species collected within a single collection.

Interactive GIS Map and Database Interface

The interactive map utilizes ArcIMS® to display collection records in relation to a base map, allow users to navigate the map, access and interact with the database query function, and print and/or save their results. Museum records are stored in a Geodatabase that is developed from the OSU and OU databases. Collection locations were georeferenced and checked for accuracy and precision. Fish species information, place names, and other information were likewise corrected for taxonomic changes, identification mistakes, and spelling or typographic errors.

Collection locations were georeferenced (given geographic coordinates based on place name descriptions) with GEOlocate, a georefemcing program available through Tulane University (http://www.museum.tulane.edu/geolocate/), and each location was checked for accuracy and given an associated confidence value (CV), which describes the accuracy of the coordinates relative to their "true location". These discrepancies were the result of place descriptions that were estimates of the true location. For example, geographic coordinates of or a description that references a specific bridge near a collection site provides high certainty, but describing a location as a number of miles (or kilometers) a distance from a city produces less certainty as to where the true collection was located. This confidence value system will allow future data users with knowledge of how accurate the locations are. Further information on georeferencing methods and confidence values are in Appendix 1: DAOF_Georeferencing_Guide. Details regarding the database quality control and assurance are featured in the below section on *Database Management.*

E. Results and Discussion:

Information Based Website

The information based website is currently housed on a server in the GIS lab in the Oklahoma Cooperative Fish and Wildlife Research Unit (http://zoology434.lsw.okstate.edu/DAOF_Html/) and once approved and edited and/or modified, can be easily moved to either the Geography server or to an ODWC server (Figure 2). The files and computer code used in this website and the ASP.NET site will be provided to the ODWC.

Database Management

Database management was one of the more difficult tasks of the project. We initially developed an online data entry and editing system using Microsoft SQL® database. Earlier tests with an online system showed that it performed well but required significant computer programming knowledge to maintain the system when it failed and an information technology person to add and accept new users and maintain security. Therefore, to minimize the computer knowledge and expertise required to manage the museum databases, we decided to continue using Microsoft Access® that each museum is currently using to maintain their individual databases. Although access by museum managers to the stand alone data management system is more limited than an online data management system, it will be simpler to maintain. Code was created to develop a password protected interface for the stand alone system and will be provided to the ODWC for subsequent implementation as needed.

Using Microsoft Access will require two conversions for the database to be incorporated into the ArcIMS® site. The database will first need to be converted to Microsoft SQL® and the SQL database will then need to be converted to a geodatabase so it can be used by ArcIMS®. The online data entry and editing interface we developed required a manual conversion from SQL to a geodatabase. Future automation of the conversion from SQL to a geodatabase is recommended, but it will take a substantial programming effort. The database should be updated quarterly or semiannually, as needed, and will require the efforts of computer science or management information systems personnel. Databases from OU and OSU are maintained and displayed separately within the DAOF to allow curators maximum control over their own databases (Figure 3).

Interactive GIS Map and Database Interface

The ArcIMS® site (http://www2.ocgi.okstate.edu/daofnew2005/DAOF.aspx) provides a mapping interface for users to navigate in a GIS environment (Figures 4 and 5). Although the ArcIMS® site requires some GIS experience, we developed a brief "how-to" guide to assist users (Appendix 2). The ArcIMS® site is running on the Oklahoma State University, Department of Geography server, which is maintained, backed-up, and provided with emergency electrical generation by OSU.

Query Interface

The query interface is accessible through the ArcIMS® website (http://www2.ocgi.okstate.edu/daofnew2005/DAOF.aspx, Figure 6) and can be used to search the OSU database with queries. The OU database will be added to the site when it is completed later in 2006. Users can conduct simple queries (single species or common name, location description, or county) or complex queries (a species within a county or location/river). Query results are exportable to Microsoft Excel. Additionally, spatial queries can be conducted using the ArcIMS® interface functions. For example, users can zoom to a desired view and use the *Identify* tool to select records within the map that are then returned in a table in a new window (see "How to use the ArcIMS Guide" in Appendix 2).

Recommendations

- The DAOF will require future maintenance and updating not only by biologists but also by computer programming and information technology personnel.

Funding should be provided to maintain and update the site either at its current location (OSU) or at a location with a dedicated database and information management staff. We developed more code than is currently being used in DAOF. Management information systems/information technology personnel could use or modify this code to employ the online automated database management system we initially developed to make it accessible to multiple users.

- Collection records from sources other than the OSU and OU museums should be added to DAOF. These records reside at other museums (e.g., University of Kansas, University of Michigan), in agency databases (e.g., Oklahoma Conservation Commission, Oklahoma Water Resources Board), and in various documents (scientific articles, theses, reports). A committee of experts and interested individuals should be developed to review these sources and determine their suitability for inclusion in DAOF.

F. Significant Deviations: None.

H. Prepared By:

Jason G. Freund, Post-doctoral Fellow
Oklahoma Cooperative Fish and Wildlife Research Unit
Department of Zoology
Oklahoma State University, Stillwater, OK 74078

William L. Fisher, Co-Project Leader
USGS, Oklahoma Cooperative Fish and Wildlife Research Unit
Department of Zoology
Oklahoma State University, Stillwater, OK 74078

Anthony A. Echelle, Co-project Leader
Department of Zoology
Oklahoma State University, Stillwater, OK 74078

Allen Finchum, Project Advisor
Department of Geography
Oklahoma State University, Stillwater, OK 74078

Varun Chudiwale, Amit Jain and Himanshu Kale
Research Technicians/Computer Programmers
Department Management Information Systems
Oklahoma State University, Stillwater, OK 74078

I. Date: 15 August 2006

J. Cost: $233,344.30

K. Approved By:

Natural Resources Section
Oklahoma Department of Wildlife Conservation

John Stafford, Federal Aid Coordinator
Oklahoma Department of Wildlife Conservation

L. Literature Cited:

Miller, R. J., and H. W. Robison. 2004. Fishes of Oklahoma. University of Oklahoma Press. Norman, Oklahoma.

Tejan, E. C., and W. L. Fisher. 2001. Development of a stream fisheries management information system for Oklahoma. Oklahoma Department of Wildlife Conservation. Final Report, Federal Aid Grant Number F-41-R. Oklahoma City.

Table 1. Summary of the information provided in each of the three main database tables. The field number is the field that relates the three tables together. There is a single geographic and collection information table; however there may be several fish information tables for each field number (individual collection).

Geographic Information	Collection Information	Fish Information
Field Number	Field Number	Field Number
Country	Collector Names	Museum Number
State	Permit Number	Museum Name
County	Collecting Institution	Species Code
HUC2	Research Project	Family
HUC4	Collection Method	Genus
HUC6	Site Description	Species
HUC8	Species Released	Common Name
Water Body	Waterbody Type	Author
Sample Location	Field Notes Kept	NatureServe Link
Latitude	Field Notes Number/Location	Number of Specimens Vouchered
Longitude	Accession Number	Identified By
Confidence Value	Data Entered by	Biomass (g)
Geographic Notes	Date Entered	Size Range (Min-Max in mm)
Month	Data Edited by	Container Type
Day	Date Edited	Original Preservative
Year	Edit Description	Current Preservative
	Collection Notes:	

Figure 1. A design schematic of the structure of the OSU and OU museum databases. Fields within each table are presented in Table 1.

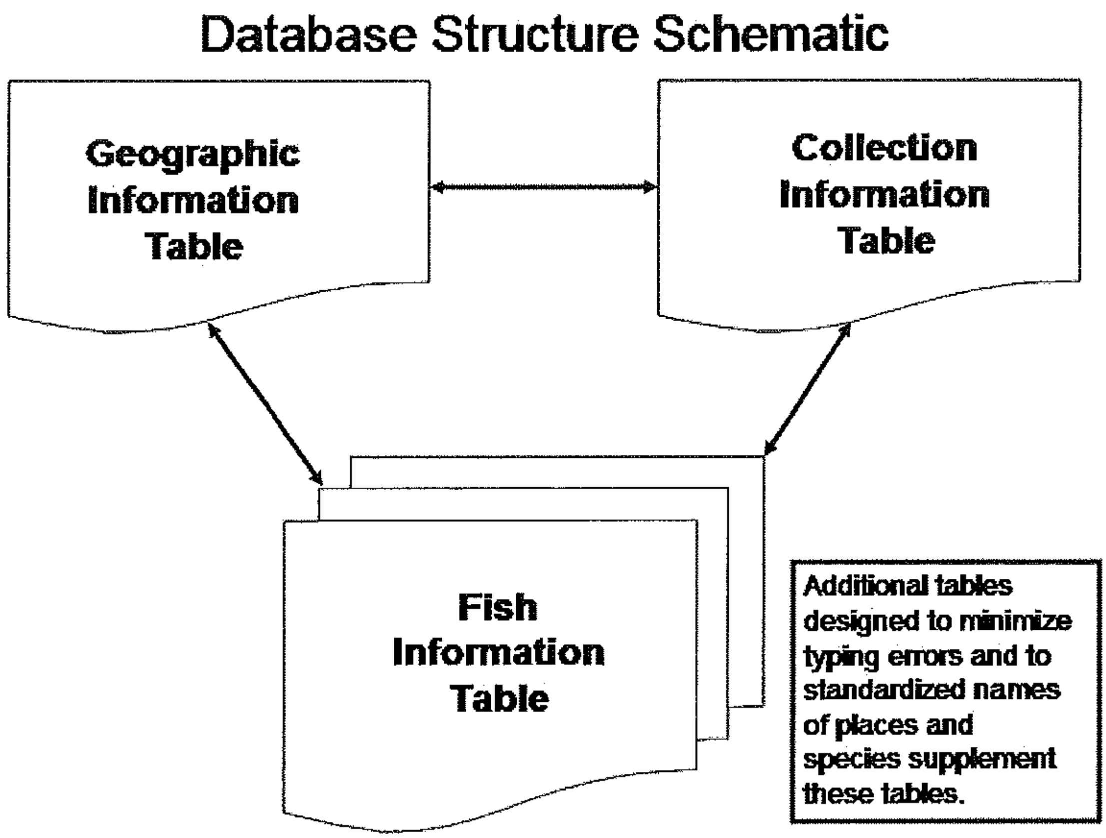


Figure 2. A screen capture of the information-based website that supports the DAOF ArcIMS mapping and query interfaces.

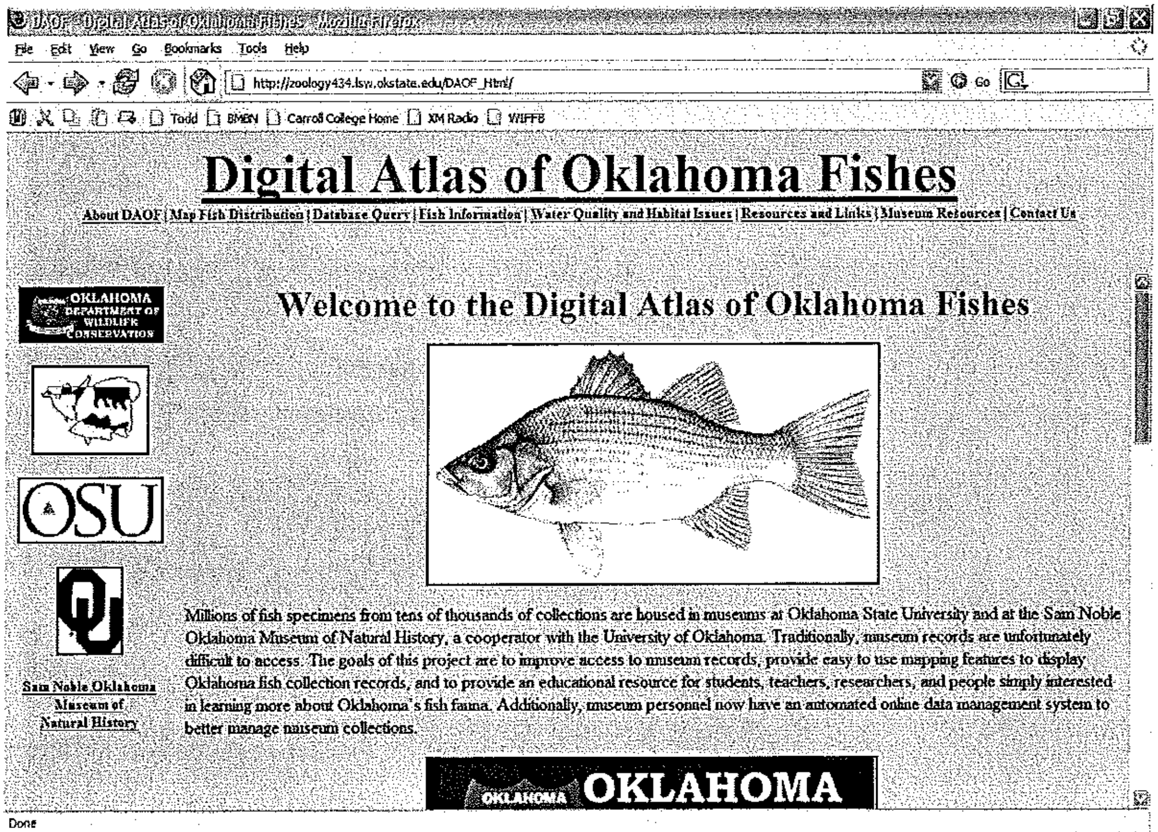
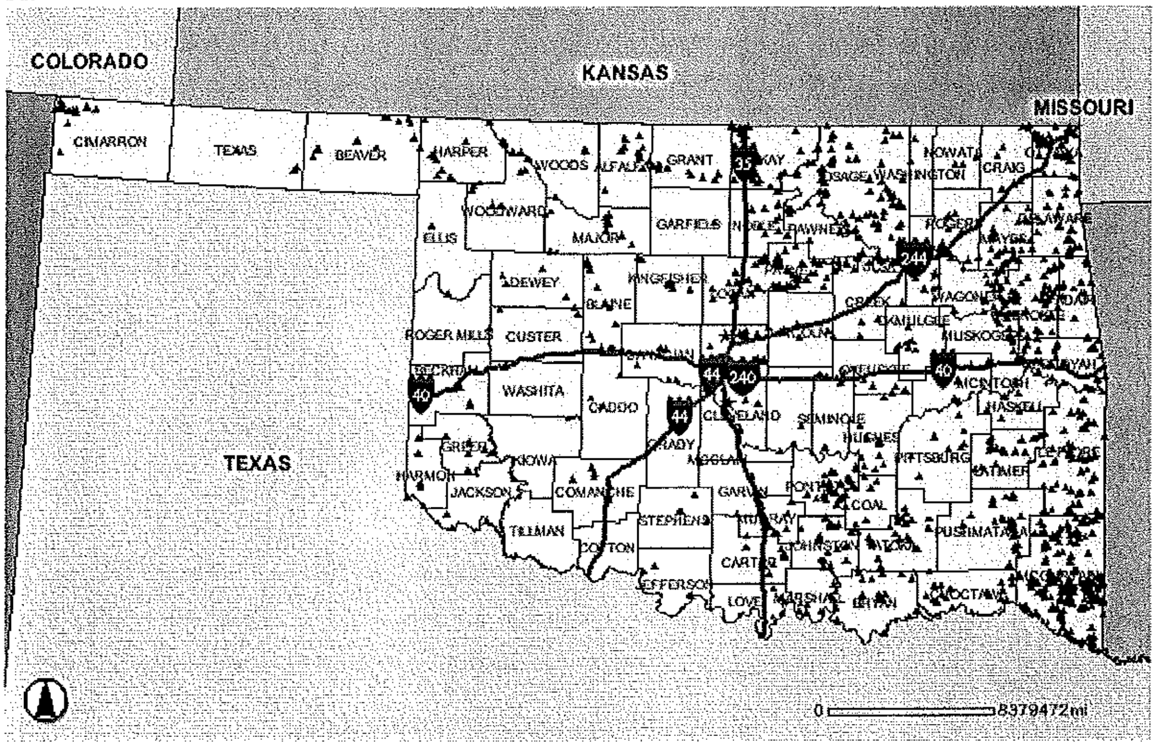


Figure 3. Collection records from OSU (A) and OU (B) museums that are included in DAOF.

A.



B.

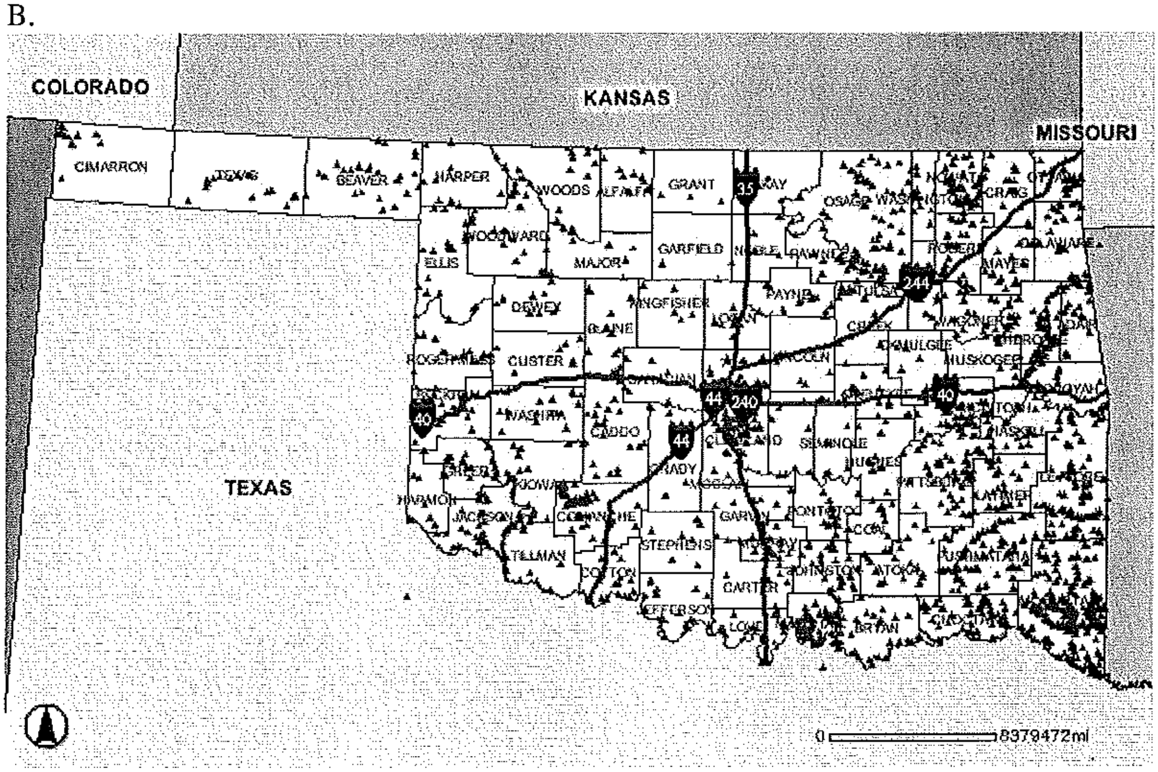


Figure 4. This screen capture of the ArcIMS® website shows both OSU and OU collection records. Information about the tools, legend, and GIS layers are in Appendix 2 and on DAOF website.

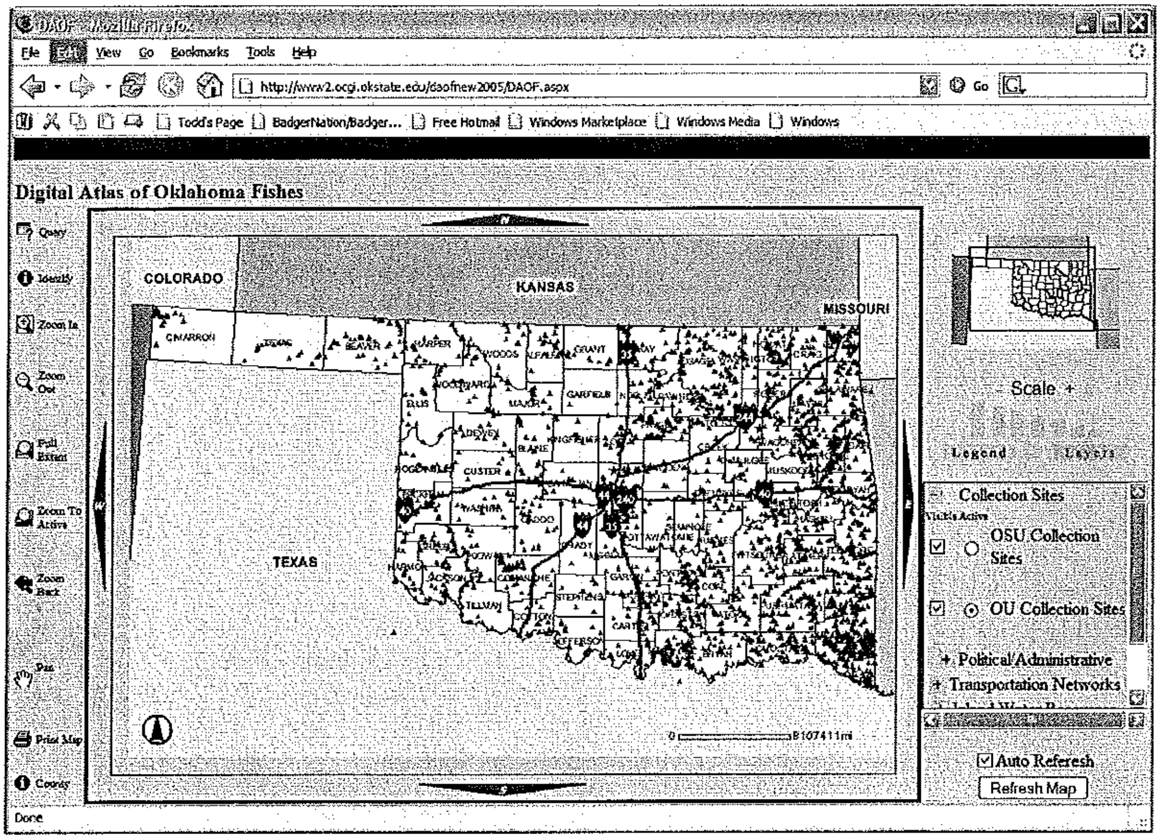


Figure 5. This screen capture shows the tools and layer options available for use within the DAOF. More information about their use is provided in Appendix 2.

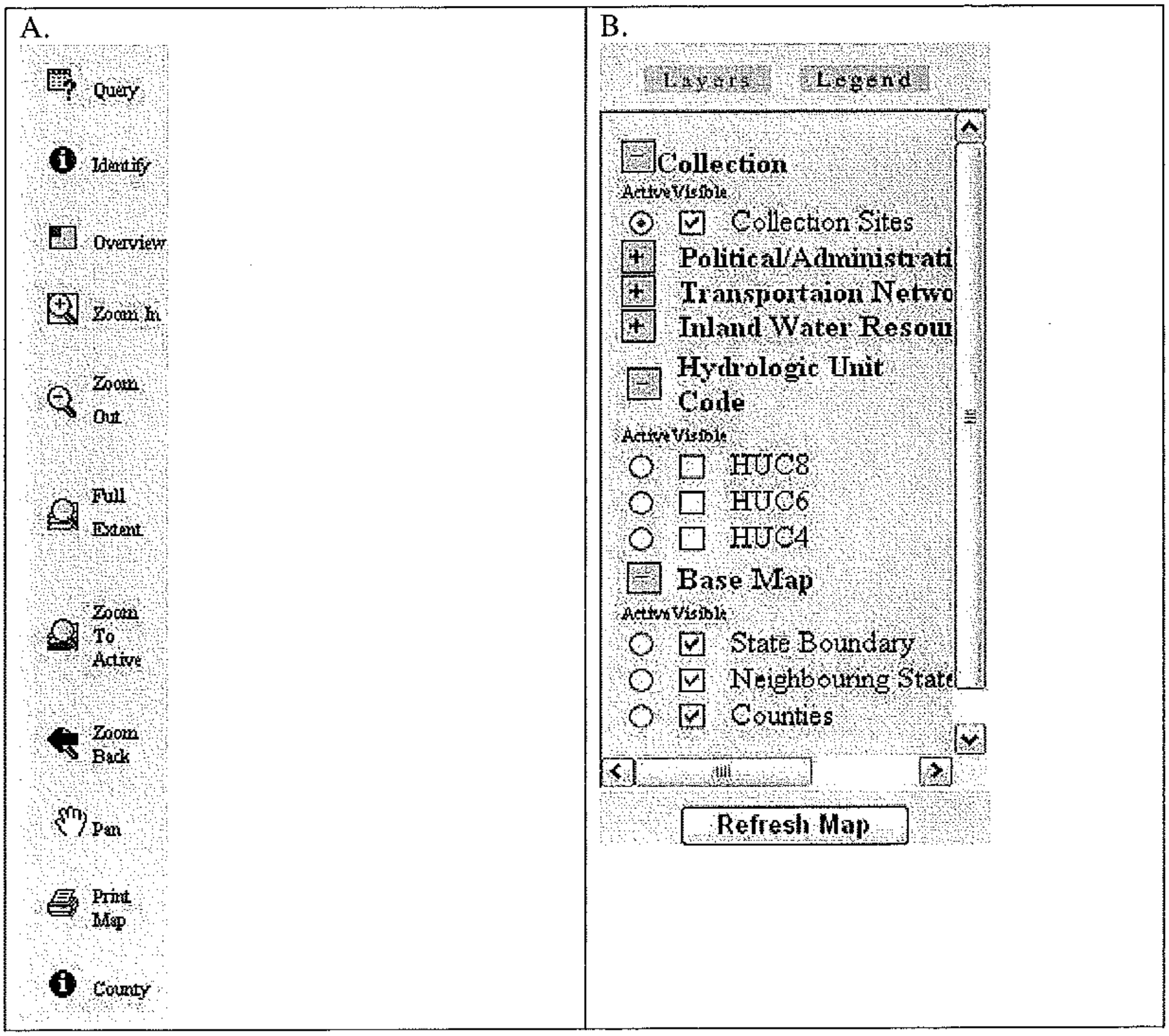
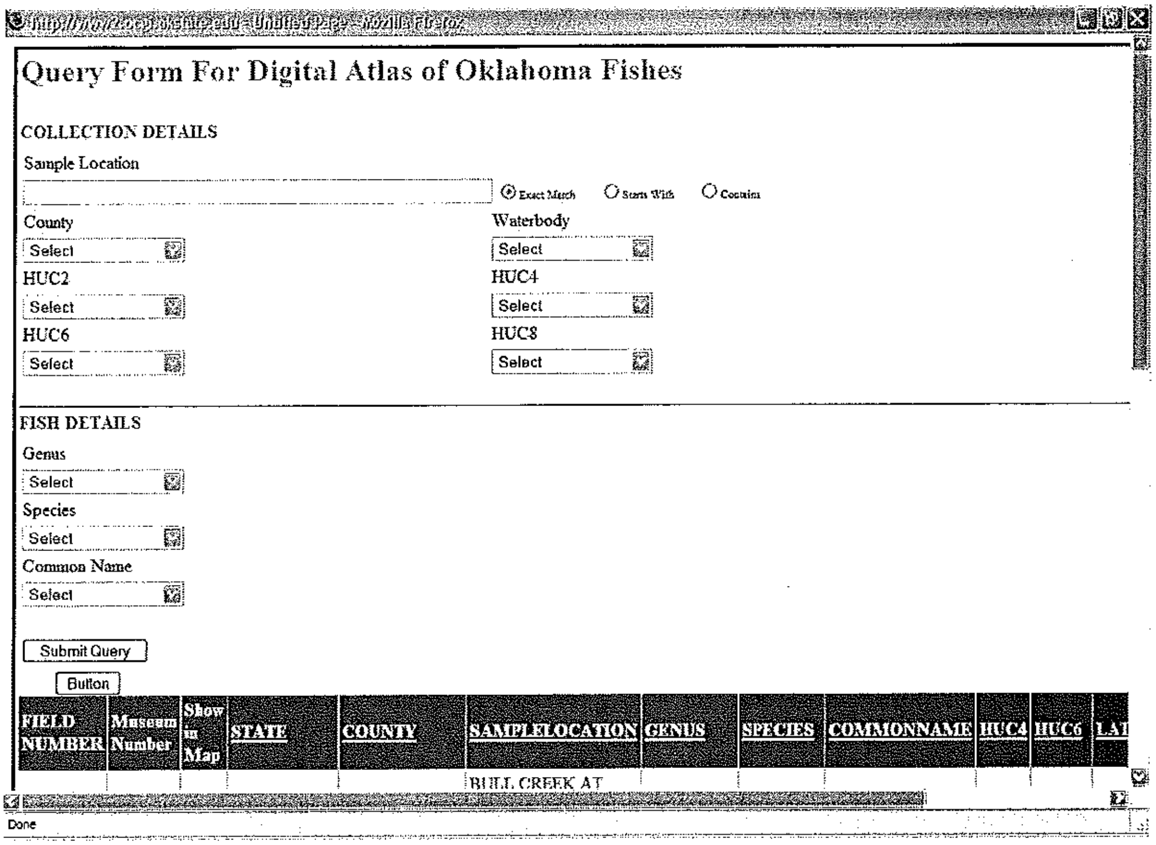


Figure 6. A screen capture from the query interface depicting the simple and complex queries that can be conducted through the interface.



Appendix 1

Digital Atlas of Oklahoma Fishes (DAOF) Project

Geocoding and Data Checking Protocols

Compiled by

William L. Fisher
Oklahoma Cooperative Fish and Wildlife Research Unit
Oklahoma State University

Introduction

The Digital Atlas of Oklahoma Fishes (DAOF) geocoding and data checking process is to assign geographic coordinates to museum fish collections at Oklahoma State University (OSU) and the University of Oklahoma (OU). Geocoding generates and validates spatial data for fish collections in Oklahoma, which will enable analysis of spatial and temporal trends in the distribution of Oklahoma fishes. It is critical that project participants (geocoders and data checkers) adhere to the protocols to ensure consistency and reproducibility. This document includes procedural instructions and standards for geocoding and checking museum records. Many of these procedures were adapted from the Mountains and Plains Spatio-Temporal Database Informatics Initiative (MaPSTeDI). Information about MaPSTeDI can be found at http://mapstedi.colorado.edu/index.html.

Why georeference?

One of the most important aspects of natural history collections is that they possess a detailed history of where specimens have been found. Almost every museum specimen collected includes some sort of locality description which describes the geographic location of the collection. Because of this detailed data, museum collections are the primary source for data for current issues such as biodiversity, intrusive species, and the changing environment. However, most museum collections do not include a specific geographic coordinate for the collection location of each specimen. This lack of geographic reference makes analysis of the data both time-consuming and restrictive. Georeferencing is the solution to this problem. Georeferencing is the assignation of geographic references that can be analyzed (such as geographic coordinates) to an item based on its location. With the advent of GIS, georeferenced specimens can be plotted and viewed easily, allowing scientists to perform complicated analysis upon their data.

Thus, the answer to the question, "Why georeference?", is simple: it will make a collection much more useful.

Before you start

There are several issues to deal with before beginning the georeferencing process. It is important to resolve these issues because ignoring them can lead to inconsistent georeferencing and useless data. The following questions are ones that should be answered before georeferencing begins:

- Is the collection databased?
- What is the current condition of the collection?
- What will the georeferenced data be used for?
- What sort of monetary resources are there for georeferencing?
- Who will do the georeferencing?
- Who will supervise the georeferencing?
- How much information should be collected?
- How long should georeferencing take?

The answers to these questions determine what path your georeferencing project will take. For instance, it is essential that your collection be databased. If it is not databased, there is little point to georeferencing it because it still cannot be analyzed. Likewise, if your database is not in particularly good shape, you may need to spend a little time cleaning up problems before beginning. Georeferencing incorrect or inaccurate data is wasted work. Hopefully, though, these questions will create a realistic framework of expectations for your georeferencing project as well as a solid base of standards.

It is very important to determine who will be doing the georeferencing. Georeferencing can be a tedious process. **By most estimates, a competent georeferencer can georeference one record every 5 minutes.** In reality, some records will take longer than 5 minutes and many will go faster depending on the skill and experience of the georeferencer. Thus, it is probably not feasible for collection managers and others with full-time positions to georeference databases. Ideally, georeferencers should be employed solely to work on georeferencing. However, for many institutions, this option is not viable. This fact is especially true for those institutions which rely heavily on volunteers. Strict and detailed guidelines combined with supervision by an experienced georeferencer will ensure that georeferencing is as consistent as possible.

There are five issues that must be addressed before beginning the georeferencing process:

1. Obtaining the proper materials
2. Setting up your database
3. Standards for georeferencing
4. Collaborative georeferencing
5. Quality checking

For the DAOF project, issues 1 and 2 have already been addressed; therefore, the focus of this document is on issues 3, 4 and 5.

Standards for georeferencing

Standards in georeferencing are important because they minimize the effect that a georeferencer's personal preference has on the data. They should be determined before beginning georeferencing to ensure this consistency in the data. A few standards to consider are:

- Units of measure
- Method and format of determining margin of error
- Degree of accuracy in determining points
- Fields that must be filled
- Format of recording coordinates (i.e. for lat/long, degrees/minutes/seconds, degrees/minutes.minutes, or degrees.degrees)
- Primary source of place names
- Dealing with typos and other errors in the existing database
- Number of decimal places to keep in decimal numbers

Determining these standards and then outlining them in some sort of document can help to maintain these standards.

Collaborative georeferencing

Determining a strategy for collaborative georeferencing is important for most institutions beginning a georeferencing process because most databases are georeferenced by more than one person. In general, it is not a good idea for two people to work in the same database on the same data at the same time for many database programs. For smaller projects, this strategy may be simple because there are only a couple georeferencers working. For other larger projects, however, the georeferencers working on the same database aren't even in the same building.

A few possible strategies are presented below:

- **One georeferencer at a time.**--This strategy involves setting the hours of each georeferencer so that no georeferencer is working at the same time. This strategy is very acceptable for small projects which have no more than a couple georeferencers. It may also be the only strategy possible for projects having access to only one computer. For larger projects, however, it is not practical because of the number of people involved.
- **One georeferencer per database.**--This strategy involves assigning only one georeferencer to each database. While it does work for institutions with many small collections, it does not work for larger collections because of the time it takes for a georeferencer to complete an entire database.

- **Import/Export.**--This strategy allows georeferencers to export sections of the database to a separate work database and then import them back when georeferencing is complete. The major advantage of this strategy is that georeferencers can work whenever they choose on small databases that can be moved easily from computer to computer. Another advantage is that georeferencers do not often work directly in the main work database, minimizing the risk of damaged data. The disadvantages are that it may be difficult to set up a good process and it may require georeferencers to be more familiar with the functions of the database program in order to prevent incorrect importing. However, the latter danger of import errors can be fixed by automating the import (and export) with scripts or macros. This solution is an excellent one for larger projects.
- **Work in the same database.**--Some database programs do allow safe access for more than one person. If your database program does allow this option, it may be worth it to investigate the strengths and weaknesses of it. However, it may still be best to assign georeferencers to different sections of the database to prevent any overlap of work.

There are several other possibilities for collaborative georeferencing that may fit your institution's needs more effectively. However, remember to choose one that maintains the integrity of the data to the utmost.

One additional aspect of collaborative georeferencing is communication between georeferencers. Often georeferencers notice patterns or problems that other georeferencers may not. In order to communicate these patterns or problems, it is important to have some method of communication between georeferencers. This communication can be in the form of daily logs in which georeferencers write observations and information about the records they complete. For more widespread projects, it can also be in the form of an email listserv for the georeferencers.

Quality checking

An important but often overlooked aspect to any georeferencing project is the ability to check the georeferenced data that goes into the database. Often this aspect is ignored because of lack of funds or personnel. However, because the point of any georeferencing project is to produce geographic coordinates linking a specimen to a place on a map, it is important that the coordinates chosen are truly the best ones for the location. Not only does it improve the quality of data, but it also identifies trends and habits in georeferencing that may need to be corrected. Often a graduate assistant, intern, or someone with more experience will do most of the quality checking.

The first part of quality checking is checking the work of each georeferencer. It is recommended that someone should check the work of beginning georeferencers closely to help train them. Until the georeferencer shows consistency in making correct choices for point location and margin or error, the quality checker should check every record that the georeferencer completes. This process may take two or three hundred records.

However, it is much easier to check the work of others than to georeference those three hundred records by yourself! After the georeferencer gains experience, their work can be checked less often. The work of experienced georeferencers can be checked at a rate of five or ten records per one hundred. No matter how experienced a georeferencer is, however, their work should still be checked occasionally. It is easy when georeferencing to acquire bad habits.

The second part of quality checking is making final decisions on problem records. Problem records are ones on which georeferencers are unable to make a definitive decision for one reason or another. Instead of guessing, the georeferencer can pass the record along to someone with more experience to make a final decision. Finding the correct coordinates for problem records can be like detective work and is often slower than regular georeferencing. However, it improves the quality of your data because it ensures that no record is ignored because of the difficulty in georeferencing it.

While georeferencing is not an exact science and no collection can be georeferenced 100% correctly, quality checking can drastically improve the percentage of the collection that is correctly georeferenced. Every project should take it into account when planning their georeferencing operation.

Georeferencing Procedures

I. Finding Coordinates

Assigning coordinates to museum specimens based on the locality information provided in the museum catalogue or collections database is the basis of georeferencing. The amount and accuracy of locality data provided in each record determines the difficulty of georeferencing it. This section is intended to provide a process by which most points can be easily found. All coordinates given are in Universal Transverse Mercator, North American Datum of 1927.

A. Coordinates given

Usually, there is no need to find coordinates for these records. However, it is necessary to verify that the coordinates do actually match the rest of the locality data provided. Occasionally, they will not match because of collecting errors or datum differences. Other sources of error include rounding coordinates when using UTM grids on maps and errors while using GPS units (which can be as much as 200 meters). However, the collectors did choose the coordinates for a reason, so be sure of the error before correcting it. It may be necessary to convert other coordinate systems such as latitude and longitude into UTMs.

Example: Boulder Falls, 465408E, 4428396N, Boulder County, Colorado (probably collected from a GPS unit or gazetteer)

Example: Boulder Falls, 465500E, 4428500N, Boulder County, Colorado (probably collected from a UTM grid on a paper map)

Example: Boulder Falls, 40° 0' 25" N, 105° 24' 19" W, Boulder County, Colorado

B. Township/Range/Section (TRS) given

If there is no other usable locality data, or if TRS is the most precise information provided in the locality description, place the point at the center of the TRS or ¼ section. Otherwise, TRS is only used as one factor in determining the final coordinates.

Example: Boulder Falls, PM 6 T1N R72W Sec.35 NW¼ NE¼ NW¼

C. Place Name Only

The majority of locality descriptions will reference a place name and may or may not mention some additional information clarifying where the specimen was collected within that place. Finding coordinates for these place names is usually simple. The Geographic Names Information System (GNIS) contains almost 2 million place names in the United States. Search GNIS in one of several ways to locate the correct place name and its corresponding coordinates. By using the GNIS coordinates, MaPSTeDI records with the same locality will have the same coordinates.
Example: Boulder Falls, Boulder County, Colorado

If the place name cannot be found in GNIS initially, it is worthwhile to check alternate spellings and partial names. If there is still no success, the georeferencer should search additional gazetteers as well as Internet resources and historical gazetteers for the place name. There are very few place names that cannot be found in one of these resources.

If additional information is provided about the locality, adjust the point within the boundaries of the place name to fit the locality most accurately.

Example: just below Boulder Falls on the west bank, Boulder County, Colorado

D. Place Name with Offset

Many locality descriptions will indicate that the specimen was collected a certain distance from a place name. This distance is called an offset. Collectors measure these offsets in two ways. An offset measured by air is measured in a straight line from the place name to the point of collection in a specific direction. Conversely, an offset measured by road or river is measured along a road or river in a general approximation of the indicated direction. Unless the collector indicates how the offset is measured, it can be difficult to determine which method to use.

1. Offset by Air

Offsets by air are measured by a straight line from the geographic reference point. This geographic reference point is almost always the GNIS coordinates for the place name in the locality description. An offset by air is occasionally indicated by a researcher but it is more often chosen by the georeferencer when there is no indication that the offset follows any specific road, river, or other linear feature. For instance, in the example below, the offset is almost definitely by air since there are no features that run northwest away from the falls. The GNIS coordinates for Boulder Falls are used as a beginning point for a line that measures 2.1 miles directly northwest away from the falls.

Example: 2.1 miles NW of Boulder Falls, Boulder County, Colorado

2. Offset by Linear Feature

Offsets by a linear feature are measured by tracing the feature for the distance and direction listed. The reference point should be the closest point to the place name that is located along the feature that is being traced. Often, the feature may travel only in the indicated direction for sections before turning a somewhat different direction. An offset by linear feature is used when a road, river, or other linear feature is mentioned in the locality description. It can also be used when a suitable linear feature clearly presents itself on the map. In the example below, CO State Highway 119 generally runs east-west right next to Boulder Falls. The offset would therefore be measured by tracing CO-119 for 2.1 miles east from the point on CO-119 that is closest to the falls.

Example: 2.1 miles E of Boulder Falls, Boulder County, Colorado

3. Offset with more than one direction

Offsets such as these are almost always measured by air. Occasionally, however, there will be references to roads that may indicate otherwise.

Example: 2 miles north, 1 mile east of Boulder Falls, Boulder County, Colorado

4. Undetermined Offsets

In certain cases, it is not possible to determine whether an offset has been measured by air or by linear feature. In such cases, the georeferencer should find the point equidistant between the two possibilities and address the difference by using confidence values (see II, Assigning Confidence Values).

5. Special Note for Offsets by River

Offsets on rivers are often measured with the words "above" and "below" instead of cardinal directions. "Above" is used when referring to upstream of the feature while "below" refers to downstream. The direction a river flows can be easily determined on a topographic map by looking at the contour lines and elevation. The contour lines will always point upstream as they cross the river. Also, remember that river drainages and lakes change size and thus location rapidly, and this is a possible source of error when attempting to georeference specimen records of aquatic taxa.

Example: N. Boulder Creek, 1.3 miles above Boulder Falls, Boulder County, Colorado

E. Other Modifiers

Other modifiers can adjust where the point is placed for a locality record.

II. Assigning Confidence Values

Confidence values (CV) are given to each set of coordinates to indicate their margin of error. Including a margin of error for coordinates is important because few historical localities contain coordinates obtained from paper maps or GPS units at the time of collection. As their name suggests, confidence values indicate the degree of confidence in the coordinates chosen for each record.

The scale of confidence values ranges from 1 to 9, with 1 being the most precise and 9 being the least precise in terms of geographic coordinates:

1. Exact (UTMs or Lat/Long to seconds or decimal minutes provided and no reason to suspect their inaccuracy)
2. Amended exact (UTMs or Lat/Long provided but were not accurate)
3. Public survey legal description (¼ Section, Section, Township, Range, Meridian)
4. Within 1 km radius
5. Within 5 km radius
6. Within 10 km radius
7. To county
8. To state
9. To project region (or country)

Confidence values are assigned after determining a point for a locality. The selection of the proper confidence value for each locality must take into account several factors. In addition, these factors change depending on the confidence value assigned.

A. CV of 1

Localities receiving a CV of 1 are often the easiest to georeference because the coordinates have already been provided by the collector. The only caveat to assigning a CV of 1 is that coordinates provided by the collector are not always accurate. It is important to check them against a topographic map before assigning a CV of 1.

Example: Confluence of North and Middle Boulder Creek, 465408E, 4428096N

B. CV of 2

A CV of 2 is rarely used, but it can be necessary in several cases. For instance, the collector may have collected the coordinates in a datum different than the NAD27 datum, which is what is used in MaPSTeDI. In this case, the coordinates would need to be converted to a new datum, which is easily accomplished with topographic mapping software. Without the CV of 2, the difference between the locality description and the coordinate fields would confuse later users of the data. Another example of a CV of 2 would be a locality in which the collector rounded off the UTM number. This imprecision should be corrected by the georeferencer and a CV of 2 should be assigned. Note that if a georeferencer is unable to resolve the inaccuracy of the given coordinates, the record should not receive a CV of 2 but, instead, a CV of 4 or higher.

C. CV of 3

A CV of 3 is assigned when the most precise indication in the locality is a legal description to the quarter section.

Example: Boulder Canyon, PM 6 T1N R72W Sec36, SE¼, Boulder County

D. CV of 4 to 6

The three CVs indicating a radius of error should be used when the georeferencer has enough information to choose a point but does not have specific coordinates provided by the collector. The CV used depends on the margin of error for the point chosen. This margin of error corresponds to the radius of the circle that can enclose all possible locations for the point. The center point of the circle is the point chosen by the georeferencer. For instance, the city of Boulder can be enclosed by a circle with a radius of 5.6 km. Therefore, the CV assigned to a locality indicating "Boulder" as the locality would be a 6 because the margin of error is less than 10 km but not less than 5 km.

Calculating this margin of error becomes more complex when the locality description is more complex. When determining the CV for an offset, it is necessary to take into account the CV for the reference point of the offset. For instance, a locality description indicating a specimen was collected 5 miles east of Boulder would require a CV of 6 or larger because of the uncertainty of the starting point. However, when the starting point can be more accurately determined, the margin of error can be reduced. For instance, if the previous example was 5 miles east of Boulder on Arapahoe Road, the margin of error could be reduced to the length of Arapahoe Road in Boulder (about 2.7 km), and the CV could be reduced to 5.

The margin of error must also take into account the uncertainty of the offset direction and length. In general, the georeferencer should assume that the distance and direction of an offset is accurate. However, there may be times when there is more than one possibility for the final location of an offset. In such cases, the margin of error must be increased to include all possible locations for the point. However, remember that if the CV of the reference point of the offset is large, many of the possible points are already accounted for.

E. CV of 7

There are two usages of this CV. The first is when a locality does not indicate anything beyond the county in which the specimen was collected. In this case, no point is assigned and a CV of 7 is assigned. The second usage is when the margin of error is greater than 10 km, but it does not extend out of the county. Often, localities indicating rivers or streams will receive CVs of 7 because their margins of error are so great.

Example: Boulder County, Boulder
Example: Boulder Creek, Boulder County, Boulder

F. CV of 8

This CV is used for records with little information beyond the state level. It is also used with localities lacking an indication of county with a large range of possibilities. It should not be used when the margin of error can also be described with a CV of 4 through 6. For example, a locality such as "Just east of Norfolk on I-25, Colorado" without a county reference could be located in either Larimer or Weld County. Although the range of possible coordinates spans two counties, it can be encompassed by a circle with a radius under 10 km. Therefore, the locality should receive a CV of 6 instead of a CV of 8.
Example: Eastern plains, Colorado
Example: South Platte River north of Denver, Colorado

G. CV of 9

T
This CV is used for records that have possible points in more than one state. It could also be used is the locality description lists only the country of origin. Like the CV of 8, it should not be used when the margin of error can also be described with a CV of 4 through 6.
Example: Rocky Mountains

H. Vagueness and Confidence Values

Vagueness in locality descriptions can often be addressed by increasing the confidence value by one level, or by adjusting the confidence value to account for the vagueness.
Example: Near Boulder Falls, Boulder County, Colorado
Boulder Falls would typically receive a CV of 4. Therefore, "near Boulder Falls" would receive a CV of 5. Note, however, that with larger margins of error, it may not be necessary to increase the CV. For instance, the CV of the City of Boulder would typically be 6. However, the margin of error is only 5.7 km. Since the maximum margin of error

for CV of 6 is 10 km, the CV of "near Boulder" can probably remain a 6 without any problems.

Example: 5 miles from Boulder Falls, Boulder County, Colorado.
The margin of error should be 5 miles (= 8.05 km) because there is no way to determine the direction of the offset. Therefore the CV would be 6.
Example: North of Boulder Falls, Boulder County, Colorado.

The margin of error for a record like this one is difficult to determine because there is no indication of the distance north from Boulder Falls. Often, an effective strategy for georeferencing these localities is to plot the point at the edge of the CV range in the direction of the offset and apply the same offset. In the example of Boulder Falls, the point would be plotted 1 km north of the GNIS point for the waterfall. While this solution disregards the possibility that the specimen was collected more than 2 km north of the waterfall, it is reasonable to assume that a new reference point would have been selected for the locality description if the point was a large distance away. Similar records are problematic, however, and should be taken on a case-by-case basis.

III. Quality Checking

To minimize the effect that georeferencing errors have upon the final data set, it is necessary to use a system of quality checking. Quality checking can take several forms, but it is recommended that it have two main purposes.

A. The first purpose is to check the accuracy of the georeferencing. This process involves checking a certain number of each georeferencer's records. Based on our trials, it is recommended that the first 200 records that a new georeferencer completes be checked for accuracy. Not only is this initial checking beneficial to the accuracy of the data, but it is essential to allow the georeferencer to improve and learn from making mistakes. If significant problems still exist after the initial 200 records, an additional batch of 100 records should be checked. After the quality checker, who is a highly experienced georeferencer, is satisfied with the new georeferencer's abilities, the quality checking is reduced to 10 randomly selected records out of every 100 completed. If more than 2 records are found to be incorrect within that 10, an additional 20 records should be checked. The quality checker may ask the georeferencer to redo the entire 100 if enough problems exist. After a period of few mistakes, the checking is reduced to 5 records for every 100 at the quality checker's discretion.

To summarize:

- Initial 200 records should be checked. If problems remain, check groups of 100 until satisfied with georeferencer's abilities.
- Regular checks of 10 randomly selected records for every 100
- If more than 2 incorrect records, quality checker should check 20 more records and can ask georeferencer to redo entire 100.

- After awhile, the regular checks can be reduced to 5 records for every 100.

B. The second purpose of quality checking is to allow georeferencers to refer difficult or confusing records to the quality checker for help or advice. The quality checker will then resolve these "problem records" as accurately and precisely as possible. Checking problem records can be like detective work. Historical records often have locality descriptions which do not appear on modern maps or in GNIS. To find these localities, it is often necessary to consult several different sources of information. These sources include, but are not limited to: catalog books, field notes, other records with similar localities, other collections, scientific and other publications, websites, online databases, specialty gazetteers, and historical maps. Bits of information from several places can often be used to establish the correct coordinates for a historical locality. In addition, some problem records do not make sense because of contradictions or missing or garbled information. These problem records may be the result of mistakes in data entry made in either the paper catalog book or the database. It may also be necessary to consult the curatorial staff or even the original collector is possible.

IV. Georeferencing Procedure

Experienced georeferencers develop their own preferences for the order in which they georeference. In general, we recommend this five-step procedure.

Step 1 - Locate and plot the locality point
The actions involved in this step are described in Finding Coordinates.

Step 2 - Assign a confidence value to the locality
The actions involved in this step are described in Assigning Confidence Values.

Step 3 - Record the georeferenced locality data
This is an important but often under-appreciated step. Most of the mistakes in georeferenced data come from incorrectly recorded data. It is important that all required database fields be filled in as completely as possible in the correct format. The database administrator should place constraints upon some fields to force correct format.

Step 4 - Document the georeferencing rationale for each record
This step is critical because it documents the decision making process for each georeferenced record. For problem records, as well as confusing or detailed records, this information is very important to permit quality checking personnel and museum database users to understand the rationale behind the locality point and confidence value selection. This information also serves as a daily log which permits georeferencing personnel to communicate ideas and report problems. This documentation should be databased with the georeferenced data. If databasing this information is not possible due to database software limitations, it should be kept in electronic documents.

Step 5 - Mark record for further review, if necessary

If the locality cannot be found or is confusing, it should be marked for review by quality checking personnel. This can occur in the database itself or however it is most convenient, but the georeferencer should attempt to complete the record if possible to expedite the quality checking process. The georeferencer should also collect as much relevant locality data as possible to aid the quality checker.

Appendix 2

Using the ArcIMS Site on the Digital Atlas of Oklahoma Fishes

Introduction

The interactive mapping site utilizes ArcIMS (Internet Mapping Service) to display collection records from Oklahoma State University Collection of Vertebrates (OSUCOV) and the University of Oklahoma and the Sam Noble Oklahoma Museum of Natural History (SNOMNH). The map displays over 30,000 collections covering nearly 100 years that are archived in one of the two museums (see Layers and Legend section below for details on how to change the display). These databases have recently been georeferenced, taxonomy has been updated, and many lots have been examined. Although we have examined many specimens, the museum data may still contain some errors.

The display has three major components, from left to right as you view the ArcIMS site: the toolbar, the map display, and the layer and legend section.

Using the Toolbar Functions

The toolbar provides several ArcGIS and ArcIMS tools that most users will find familiar. Below are brief descriptions of the functions and following the brief descriptions are more fully developed descriptions and guidelines for usage. Tools are shown below in Figure 1A.

- **Query** – Opens a new query window
- **Identify** – Provides information about the selected collections
- **Overview** – Hides or displays the overview inset map
- **Zoom In** – Draws a box that becomes the new zoomed view
- **Zoom Out** – Increases the amount of area viewed
- **Full Extent** – Changes the display to the entire state
- **Zoom to Active** – Zooms to include the full view of the active theme (see Layers and Legend for additional information)
- **Zoom Back** – Displays the previous view
- **Pan** – Allows the user to move the display without changing the zoom
- **Print Map** – Opens a new display window with a printable map
- **County** – Displays information about all collections for the selected county

The **Query tool** opens in a new window and allows users to create custom queries of the database. The query allow users to search based on date, collection location (stream name, hydrologic unit code (HUC)), and/or fish information (genus, species). The query builder allows the user to create dynamic queries encompassing date, location, and/or individual species (or genera).

The **Identify** tool can be used to either select an individual collection or all collections within a rectangular box that the user draws. If the user needs more specific collections to be returned (i.e.) collections within Lake Texoma, or samples from the Little River), the Query tool has its advantages. However, the Identify tool allows user to visually select records, some of which might not be selected using the Query tool if for example, the stream name is misspelled or misidentified. Collection information associated with the selected points are contained in a data table that is opened in a new window.

Within the newly opened data table, users have the ability to select all the records or select only a subset of the selected records. Use the top-most box to select all records or click the boxes next to the individual records to select a subset of the selected records. Once records are selected, the user can use the "***Zoom to Selected Records***" toolbar – the button is on the bottom of the page – to view the selected records at a scale suitable to viewing them. To view an individual record within the map, click on the "***Show in Map***" hyperlink. This will display the individual record within the map at an appropriate scale.

The **Overview** tool allows the user to either show or hide the inset map (upper right hand corner of the view) that places a crosshair within the map of Oklahoma to represent the location of the current view relative to the state.

The **Zoom In** tool allows users to zoom to a closer view of the map. This can be done one of two different ways. Simply selecting the tool and selecting a spot on the map, the view is zoomed to center on the selected location. Users can alternatively select an area to zoom to by drawing a rectangle while holding down the right mouse button.

The **Zoom Out** tool allows users to increase the amount of the map that they can view. In using the Zoom Out tool, the view will be centered around the point on the map where the user depressed the left mouse button.

Full Extent takes users to a view that includes the entire state of Oklahoma. This button is used to reset the view to the largest possible view which displays the entire state and portions of the surrounding states.

Zoom to Active allows users to view the map at a scale that includes the actively selected layer (See Layers and Legend section for details).

The **Zoom Back** tool takes users back to the previous view. This allows users to return to the exact previous view without trying to do so with the Zoom In / Zoom Out tools.

The **Pan** feature allows users to "grab" the map and move it. This is feature is best used once a suitable scale (level of detail) is obtained but the user would like to view areas that are just outside of the current view or they would like to re-center the current view without changing the scale. To use the tool, depress and hold the left mouse button and move the mouse to change the view.

The **Print Map** feature opens a new window that contains the map view at the current scale and includes a legend and scale. This map is ready to be printed or it may be saved using the right mouse button and "Save as" function.

The **County** feature selects all fish collection records within the selected county. The feature allows user to select a single county (you are not able to select multiple counties at once) and the collection information is displayed in a new window as it is with the Identify tool. Users may then save the resulting records as a Microsoft Excel file or they may further explore the information through the links the county query returns.

Layers and Legend

The layers and legend panel allows users to change the data layers that are visible and allows the users to view a legend containing the displayed features. The legend view simply illustrates the layers that are being displayed. Additionally, the legend view accompanies the Print function (see above). The layers view provides users with the opportunity to alter the information provided in the map. Many of the background layers are automatically scaled so that they are represented at an appropriate scale. Layers are depicted in Figure 1B. By checking the **Auto Refresh** box, the map will be automatically refresh the view when a layer is activated or deactivated. To expand any layer menu choice, click the plus box [+] next to the menu selection.

Collection Sites are shown separately for OSU (black) and OU (red; Figure 2). The current default view is to have the OSU records visible and active. To show the OU collection sites, check the "Visible" box next to the layer. To make that layer active, for use with the *Query* or *Identify* tools, check the bullet marked "Active". Either OSU or OU records may be represented at the same time and although the map is cluttered at a statewide scale, when users zoom to the county or sub-county level, the points are generally easily identified.

Political/Administrative layers depict municipalities at three different sizes.

Transportation Network layers depict roads ranging from interstate highways to county roads.

Inland Water Resource layers are streams, rivers, and lakes within the state of Oklahoma. Each of these layers are scaled so that they can be best viewed at the scale of view. Many details are not visible at larger scales as they would be difficult to read but at smaller scales they are depicted on the map.

Hydrologic Unit Code (HUC) were created by the USGS to identify watersheds across multiple spatial scales.

The **Base Map** layers depict Oklahoma and its counties as well as the surrounding states.

Figure 1.

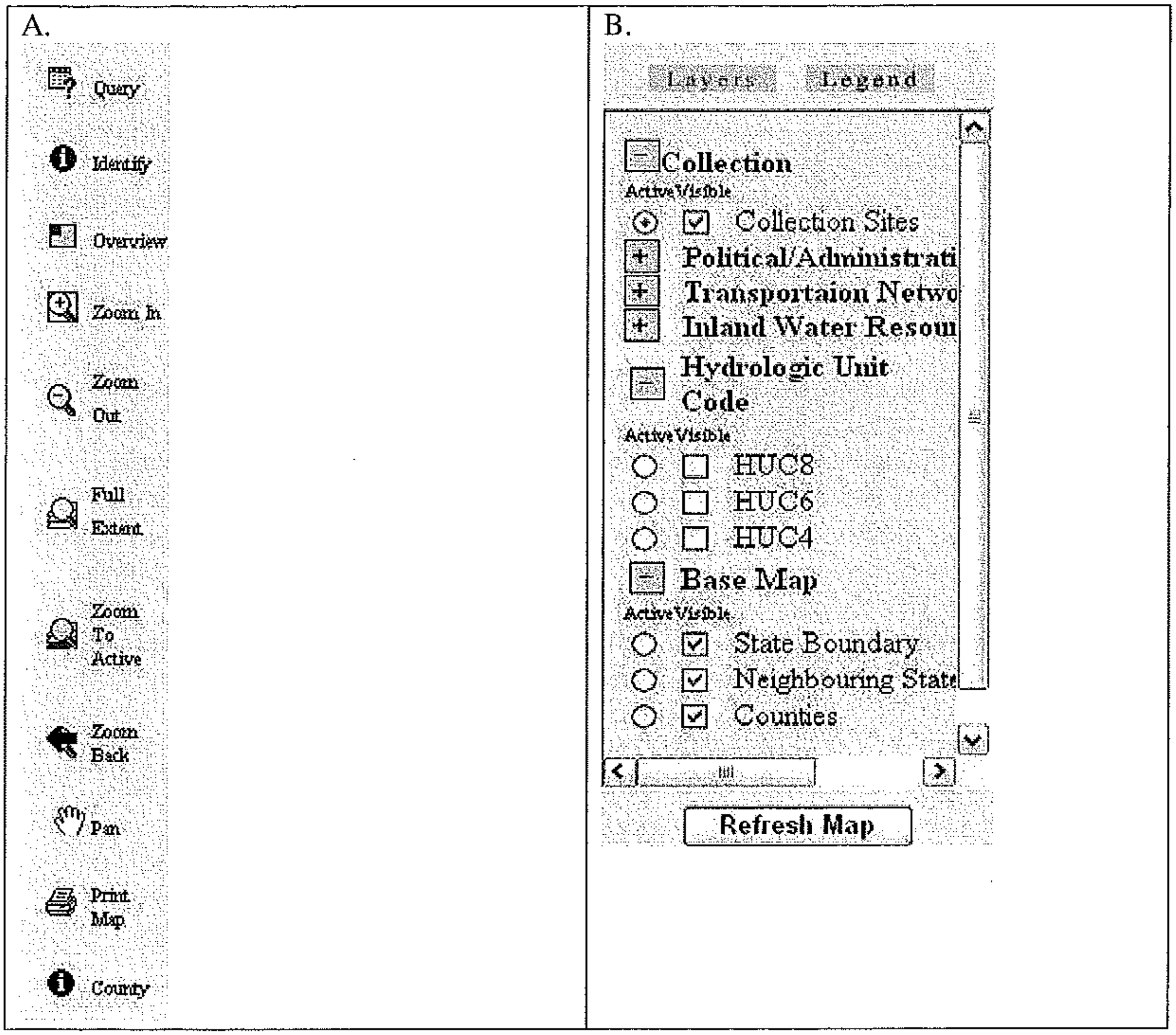
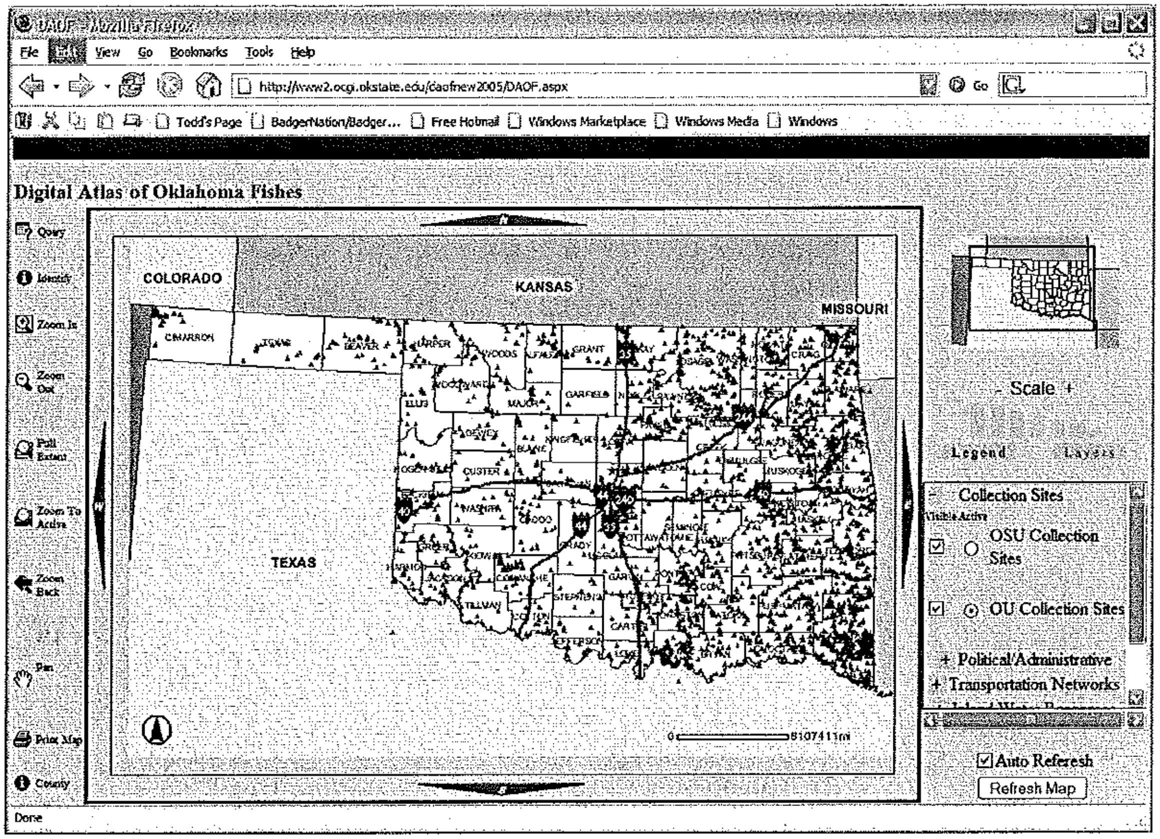

A.
Query
Identify
Overview
Zoom In
Zoom Out
Full Extent
Zoom To Active
Zoom Back
Pan
Print Map
County
B.
Layers
Legend
Collection
Active Visible
Collection Sites
Political/Administrati
Transportaion Netwo
Inland Water Resour
Hydrologic Unit Code
Active Visible
HUC8
HUC6
HUC4
Base Map
Active Visible
State Boundary
Neighbouring State
Counties
Refresh Map

Figure 2.


Digital Atlas of Oklahoma Fishes
COLORADO
KANSAS
MISSOURI
TEXAS
Scale
Collection Sites
OSU Collection Sites
OU Collection Sites
Political/Administrative
Transportation Networks
Auto Referesh
Refresh Map