Utilizing Existing Museum Collections and GIS for Paleontological Site Assessment and Management

by

Sara Evanoff

A Thesis Presented to the
FACULTY OF THE USC DORNSIFE COLLEGE OF LETTERS, ARTS AND SCIENCES
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2021

Copyright © 2021 Sara Evanoff

To Alex Van Ningen, my husband

Acknowledgements

I am grateful to my mentor, Jenifer Bernstein, for giving me the direction I needed and for the assistance of the other faculty. I would like to thank Alex Hastings, Gregory Liggett, and Mark Nebel for being willing to provide valuable information and assistance, without which this project would never have been created. I am grateful for the data provided to me by iDigBio, GBIF, BLM, the Science Museum of Minnesota, Museum of the Rockies, the Chicago Field Museum, and the Smithsonian National Museum of Natural History. To all those who helped edit this thesis, thank you.

Table of Contents

List of Tables

List of Figures

Abbreviations

ABCD	Access to Biological Collection Data
BLM	Bureau of Land Management
DOI	Department of the Interior
ERD	Entity-Relationship Diagram
GBIF	Global Biodiversity Information Facility
GIS	Geographic information system
NPS	National Park Service
PLSS	Public Land Survey System
PRPA	Paleontological Resources Preservation Act of 2009
PYFC	Potential Fossil Yield Classification
TDWG	Biodiversity Information Standards
USDA	United States Department of Agriculture
UGS	Utah Geological Society
USGS	The United States Geological Survey
USC	University of Southern California

Abstract

Regulation of paleontological resources began with the creation of the Antiquities Act of 1906, but it was not until the passing of the Paleontological Resources Preservation Act (PRPA) of 2009 that the regulation was clarified on the following: permitting, penalties, inventorying, amateur collecting, curation, and education and research in certain departments within the Department of Interior (DOI) and the Department of Agriculture. This served to further protect paleontological resources within public lands through the use of scientific expertise for the purpose of education and research.

Museum collections are held within public trust for the purpose of education and research and are curated to the highest scientific standards. These institutions work with federal institutions to preserve paleontological specimens found on public lands in accordance with the PRPA. Accessing collections protected by the PRPA requires written permission and location data is exempt from being revealed. Digitization of paleontological collections that are not protected by the PRPA has been a slow process.

Online collaborative databases such as Integrated Digitized Biocollections (iDigBio) and the Global Biodiversity Facility (GBIF) offer museums the chance to combine with other museums and institutions to develop a global database that shares highly curated biological and paleontological collections for the purpose of education and research. This evolution of museum collections can benefit spatial research as biodiversity data standards become more developed. However, this evolution has also been slow, and research incorporating paleontological collections with GIS is limited.

The objective of this thesis is to present a case study focusing on Petroleum County, Montana, where a dual geodatabase system was created incorporating both museum and field

collections to assist with site management, assessment, and inventorying. A customized model toolbox allows for paleontologists to perform analyses related to their hypotheses. These geodatabases are designed for the sole purpose of simplifying and aiding in management practices through the incorporation of museum collections. Museum collections are already incorporated with these assessments and inventorying. With the addition of GIS, museum collections' involvement will bring new possibilities for how their collections are applied.

Chapter 1 Introduction

This chapter will provide an overview of these methodologies and discuss their impact on the geodatabase created to perform the site assessment and management for Petroleum County, Montana. Section 1.1 discusses the significance of fossil resources. Section 1.2 discusses the threats to these resources. Section 1.3 discusses advancements in digitization of paleontological collections. Section 1.4 introduces the potential fossil yield classification system and section 1.5 introduces the case study, Petroleum County, Montana.

1.1. The Significance of Paleontological Resources

Interest in fossils within the United States is connected to the growth of the United States and its survey of the Western Territory (Liggett et al. 2019). Expansion and exploration fueled the development of modern museums, federal collections, and the management practices currently employed within archaeology and paleontology (Liggett et al. 2019). As museum collections have grown, public interest has grown with them (G. A. Liggett et al. 2019). Programs incorporating museum collections, developed through exhibits and outreach programs, use scientific research to both engage and educate the public.

The only documentation of life over the past 3.6 billion years is the fossil record (Murphy et al. 2015). Once a fossil is destroyed or removed, all information pertaining to its location, environment, and biology can be lost if not properly documented. The scientific importance and educational benefits of studying the fossil record can further benefit other areas of science.

The study of the fossil record in conjunction with biology can assist with identifying relationships to modern species, environment impacts and stressors, climate change, geographical distribution, behavior, bias, relationships to biostratigraphy and geology, reconstruction of paleoenvironments, and species distributions and movements (Murphey et al. 2019; Murphy et al. 2015). For example, research combining the fossil and biological record to look into the impact of environmental indicators and stressors that cause extinctions have been the focus of Stigall and Lieberman (2006) and Gavin et al. (2014). Their research into the Devonian extinction and the Quaternary Period uses the fossil record to map geographical distributions, viewing the environmental and geographical impact on species distribution and movement.

Scientific knowledge gleaned from the fossil record is frequently presented in scientific publications, educational materials, and outreach programs designed to educate the public about the earth's past biological life and environments. Public educational programs in state and national parks introduce visitors to the local geology and paleontology found within the park. Various agencies offer public digs that allow anyone to excavate with paleontologists. Museums and institutions display collections, design exhibits, and otherwise open their collections to the public to provide an interactive experience.

1.2. The Threats to Paleontological Resources

Environmental and human factors have an impact on fossil preservation (Santucci and Koch 2003). Environmental factors such as weathering (chemical and physical), erosion, or biological factors can impact the stability of fossil sites. Fossils can erode or weather out of a formation. Instability within a fossil can cause the fossil to become brittle or fragment, which can increase the difficulty to preserve it or even destroy the fossil entirely (Santucci and Koch 2003). Human factors such as illegal excavations, poaching, souvenir collecting, and destruction of sites also adversely impact fossils (Santucci and Koch 2003). With the introduction of site management and inventorying of federal parks, identification of at-risk sites has allowed management to increase monitoring of highly impacted sites by providing resources and directing them to where they are needed (Santucci and Koch 2003; Santucci and Tweet 2020; Murphy et al. 2015).

Another threat to paleontological resource management described by Santucci and Koch (2003) and Liggett et al. (2019) are inadequate data repositories. Collections and libraries are essential to maintaining accurate documentation of paleontological resources, however, lack of funding, poor maintenance, and other problems often plague collections. Not all fossils collected from public land are kept in federal collections but are stored within museum or institutional collections. Inadequate or improperly kept collections can adversely affect how management of the parks' or museums' resources are distributed. Lost records, poor record management, or the inability to maintain or transport collections to new facilities pose a threat to these collections. Ways to fix these problems through data standardization practices and data driven enterprising solutions are currently being developed.

1.3. Advances in Digitization with Paleontological Collections

Museum collections are slowly developing methods to open their collections to the public through open-sourced sites, such as iDigBio and GBIF. Collections are becoming more accessible to researchers and the public through digitalization. Digitalization of paleontological and biological collections has led to the movement of biodiversity informatics, which uses geography, ecology, and the classification of organisms into taxonomic groups to address datum preserved within paleontological and biological collections (Berendsohn et al. 2011). Standardization of data to address data dictionaries, protocols for data exchange over the internet, ontology, and schema are the responsibility of the Biodiversity Information Standards (TDWG) and the Access to Biological Collection Data (ABCD).

TDWG was the original standard for defining the language for taxonomy by developing data models and domains (Berendsohn et al. 2011). ABCD has become the new, accepted standard by the TDWG and allows for a global exchange of scientific collections that range from paleontological to biological collections and is used by the GBIF and iDigBio (Canhos et al. 2004). These standards are necessary to allow for the integration of data from distributed sources in real time (Canhos et al. 2004) and standardizes specimen and observational information provided by collections databases (Berendsohn et al. 2011).

Defining how information will be standardized also applies to the problems of hierarchy and domains. A museum database needs to be adaptable to the wide ranges of requests and queries. Granularity within the data can cause problems if information entered into the databases have multiple descriptions for a particular specimen. For example, there is a request to examine "phalanges" for a particular species. There are multiple ways this can be entered into the database and if the queries are too granular the request will not turn up variants of the word such as "phalanx" or "carpals" (personal communication with Amanda Millhouse). Pham (2015) and Burrows (2016) discuss this problem within their own theses, explaining how complex queries resulted in information missing within the resulting tables. To minimize this problem for the database, domains were used within specific fields to help standardize some of the language.

1.4. The Potential Fossil Yield Classification

Geology is important for the development and preservation of fossils, and the Bureau of Land Management (BLM) has developed a system for site managers to classify these formations through the Potential Fossil Yield Classification (PFYC) method. Liggett (2020) provides a classification for the geology of Montana, North, and South Dakota that is a guideline for identifying the state's paleontological resources using geological bedrock maps to depict occurrences (Liggett 2020). Each geologic formation is given a classification value between one and five and “U” to denote formations with the potential for producing fossils. The lowest classification, the value of one, is often given to geologic units that are igneous and metamorphic, excluding volcanic ash layers (Liggett 2020). Management concerns should be low since mitigation is unlikely unless specific cause is determined (Liggett 2020).

A classification value of two is given to geologic units where fossils are rare or not present. This value is often given to sedimentary units that exhibit characteristics where fossilization cannot occur (Liggett 2020). Management and mitigation are only necessary if found and further assessment is needed (Liggett 2020).

A value of three are given to geologic units that have a moderate potential for predicting the presence of fossils and are often marine in origin (Liggett 2020). Significant fossils can be found in these units, and mitigation and management practices should reflect any concerns that may be present (Liggett 2020).

Four is classified as high, and five is very high. Any significant finds should be documented and watched for illegal activity. Management and mitigation practices should be investigated and action plans proposed (Liggett 2020). Surveys should be performed to assess the region for other significant resources and it should be constantly monitored and controlled to prevent illegal activity (Liggett 2020). For formations with an unknown potential and little documentation to prove any fossils are present, a designation of “U” is given (G. Liggett 2020).

1.5. Overview of Methodology

Construction of this project focuses on the diverse needs that accompany paleontological resource management and inventorying. First, a thorough understanding of the rules and regulations within the field of paleontology and museum data standards and practices was necessary for the construction of the project. Paleontology can be broken down into multiple disciplines and practices, such as biology and geology, and this breakdown often refers to more than one field of knowledge. If it is to benefit management practices, the dual geodatabase system must allow for interactions related to research in the office, in the field, and across the various disciplines.

Secondly, these geodatabases must allow for museum information to be modeled and queried based on different hypotheses that are constantly being updated as new information is entered into museum collections. The results of these analyses and queries must be accessible to the paleontologist in the field for their reference.

Lastly, the geodatabases needed to be designed with non-GIS end users in mind. A database is of little use if the end user is incapable of using the database. Tasks allow for the creation of folders to house various tasks and are designed to walk users through each step without needing knowledge of GIS. Tasks allow for workflows to be placed in a centralized location so the end user does not have to understand how to use or operate the software.

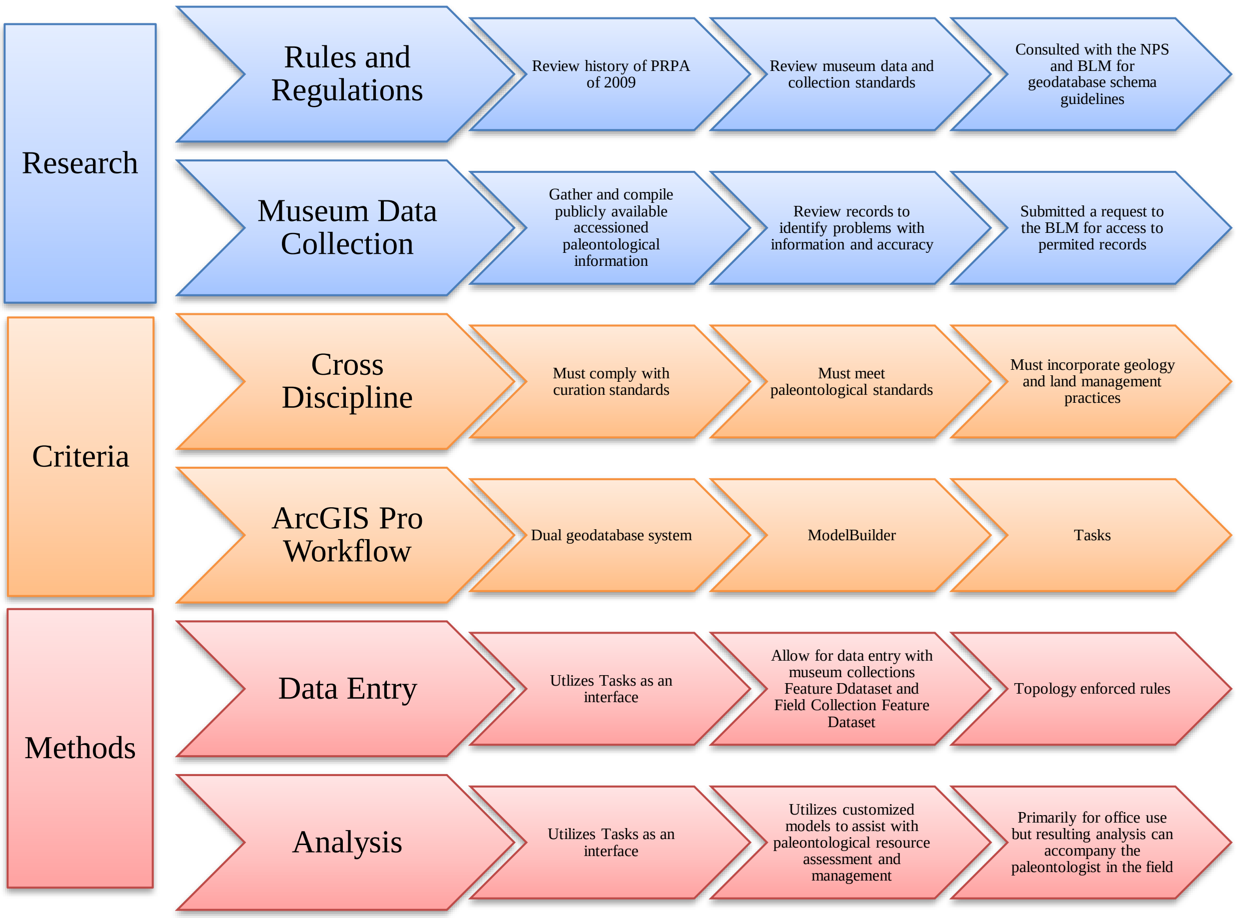


Figure 1 Method Overview

Figure 1 breaks down the process for the construction of these geodatabases into three stages, focusing on the research, criteria, and methods used to construct this project. Research focuses on the rules and regulations of working with museum collections' information and what is to be expected out of this project. The expectations define the criteria for the overall project and the GIS workflows. From there, the dual geodatabase system, its models, and its Task interface were constructed to match these criteria.

3.2 Dual Geodatabase System

Field accessioned fossils contain similar records to accessioned specimens. However, as each specimen is documented and prepared, information associated with the specimen is changed to reflect new findings. Since the purpose of this project is for paleontological resource management, assessment, and inventorying, it is necessary that the project follow office and field practices so the information will supplement the needs of paleontologists.

A dual geodatabase system was created for this purpose so the records for each fossil found covers surveying, excavation, and accessioning into museum collections. Allowing for a "chain of custody" procedure to be set up enables paleontologists to follow newly entered data as the information is adjusted to reflect each find. With one geodatabase designed to house museum collections and the other designed primarily for field work, each geodatabase comes with a specific set of tools and tasks to assist the paleontologist with data entry in the field and perform analysis within the office. This also allows for data to transport these findings into the field to assist with on-site management discussions and inventorying practices.

3.2 Modelbuilder

Modelbuilder is an extension in ArcGIS Pro that allows for models to be constructed to aid in workflows. For the purpose of this project, ModelBuilder was used to create queries and analysis tools that will assist paleontologists with management practices performed in the office. Constructed to be reusable and the results taken into the field, these models will aid in paleontological site assessment and management practices both in the field and in the office.

3.2 Task Designer

Tasks were designed using Task Designer in ArcGIS Pro to assist non-GIS users through data entry and performing analysis. Tasks were designed to centralize the necessary tools and models so the end user would not need to have an understanding of GIS or ArcGIS Pro. These tasks offer a simple interface for performing data entry, topology rules that can be run in the background to ensure data integrity, and analysis that can be performed with the accompanying models without needing to know the toolbox's location.

1.6. Case Study

Case studies provide insight and are used for preliminary studies (Rowley 2002). Due to the limited nature of research into incorporating museum collections and GIS for site management and assessment, a case study was developed to show the potential of this process. Petroleum County, Montana was chosen for investigation because of limited research and was examined in conjunction with the Science Museum of Minnesota's current study of the region. This offered the benefit of testing museum collections and assessing data quality of these records in relation to current research. Any adjustments needing to be performed on the records will present an opportunity for correction during field season.

Chapter 2 Related Work

This section delves into the history of the legislation, museums, and geology of Petroleum County, Montana. Section 2.1 provides an in-depth review of the legislation and its history. Section 2.2 discusses the role of GIS and paleontological management practices, and section 2.3 provides a geologic and paleontologic discussion of Petroleum County, Montana.

2.1. Paleontological Resource Management and the Law

The Paleontological Resources Preservation Act (PRPA) of 2009 explicitly describes federal ownership, management, permitting, collection, inventorying, and penalties. Building upon the Antiquities Act of 1906, this act increases protection, management, and penalty practices relating to paleontological resources located within public land, while also entrusting the preservation and storage of these resources through the use of scientific principles for educational and scientific endeavors. This section discusses the history of resource management, the importance of curated collections, and the threats posed to the nation's paleontological resources.

3.2 The History of Paleontological Resource Management

Protection of paleontological resources began with the passing of the Antiquities Act of 1906, however failure to specify the language pertaining to fossils were ultimately the cause of the Antiquities Act's inability to protect national paleontological resources (Cronin 2014). This left federal institutions such as the National Park Service (NPS), Bureau of Land Management (BLM), United State Fish and Wildlife Service (FWS), and the United States Forest Service (NFS) to develop their own regulations and management practices while leaving private collectors with little guidance for permitting, collecting, and the penalties for illegal collecting (Lazerwitz 1994; Santucci and Koch 2003; Murphey et al. 2019).

Despite the failures of the Antiquities Act to protect paleontological resources within public lands, the Act allowed for the creation and protection of monuments such as Florissant Fossil Beds National Monument, Agate Fossil Beds National Monument, and Hagerman Fossil Beds National Monument (Lazerwitz 1994). In 1916, these monuments were placed under the protection of the NPS and the newly created National Park Service Organic Act (Sax 2016; Santucci and Koch 2003). In a similar act, Federal Land Policy and Management Act of 1976 allowed the BLM to provide protection of all paleontological resources within its jurisdiction and manage them through the use of scientific, historical, and archaeological methodologies (Sax 2016). However, vague writing failed to provide specificity to how the management, collection, and preservation of these resources led to the development of the Omnibus Management Act granting increased powers to the NPS and BLM.

The creation of the Omnibus Management Act of 2009 began with the passing of the National Parks Omnibus Management Act of 1998 and the Federal Land Management Act. These granted the Secretaries of the Interior and Agriculture, mainly the National Forest Service, the power to inventory, monitor, and manage paleontological resources using scientific principles (Santucci and Koch 2003; PRPA 2009). In 1998, the passage of the National Parks Omnibus Management Act authorized the NPS to preserve, manage and inventory paleontological resources within their jurisdiction, using scientific principles and decision making to determine management practices within the park system (Santucci and Koch 2003). However, this act was specific towards the NPS, leaving the DOI and the NFS to develop their own guidelines for managing paleontological resources.

In 2009 the Omnibus Public Land Management Act, which contains the PRPA of 2009, was passed and solidified the ability of the DOI and the NFS to use scientific principles and expertise to enforce the collection, curation, permitting, inventorying, and management of paleontological resources found within public lands. In accordance with the Omnibus Public Land Management Act, the Departments of Interior and Agriculture were granted an increased authority to monitor, manage, and curate while developing and enforcing permitting and the penalties of illegal collecting (Bonnie 2015). The PRPA enforced that all fossils collected on public land belonged to the government and must be held within trust for the people, and it included provisions to promote education and scientific research to further the understanding of past life (Liggett et al. 2019).

The passing of the Paleontological Resource Preservation subtitled within the Omnibus Public Land Management Act paved the way for the passage of the Paleontological Resources Preservation Act (PRPA) in 2009. According to Bonnie (2015), the PRPA was modeled after the Archaeological Resources Preservation Act. The PRPA directs the Secretary of the DOI to manage and protect paleontological resources using scientific principles and expertise, develop educational and outreach opportunities to educate the public, and define permitting and collection of fossils for scientific, commercial, and amateur collecting (PRPA 2009). The PRPA also directs that the collection of fossils, and any data or media accompanying the fossils, must be held within approved curated repositories (PRPA 2009).

One of the differences between the PRPA and past acts are the increased penalties for illegal collection (Cronin 2014). According to Cronin (2014), this increase in power and description in penalties for illegal collection grants permitting and concessions for amateur collectors. This specification, according to Cronin (2014), now allows the courts to have a basis for establishing cases of legality in regards to confiscation of fossils and illegal collection.

However, there is still contention with the PRPA and its failure to address different types of paleontology outside of the academic profession. There are several different types of paleontologists: academic, consulting, and commercial. Academic paleontologists acquire specimens for the purpose of research and are often associated with universities and museums. Consulting paleontologists are hired to perform mitigation and compliance for companies prior to construction. These paleontologists assess the potential area of interest to make sure the site falls under compliance with state laws. If fossils are found, mitigation measures are taken to ensure that the potential area of interest will not result in damages to the fossils. Commercial paleontologists belong to private organizations that collect paleontological specimens and sell their finds to collectors and museums, examples being the cases of the *Tyrannosaurus Rexes* Stan and Sue. Murphey et al. (2019) states that the PRPA fails to protect consulting and commercial paleontologists by placing the burden of mitigation practices upon the client, opening up impacted sites to neglect and poor management practices. Thus, it is left up to the client to understand the conservation practices that are necessary, which could potentially cause the loss of valuable information if the company fails to hire a consulting paleontologist.

Another contention with the PRPA is only specific to public lands. The PRPA only enforces the conservation of paleontological specimens within the DOI and the Department of Agriculture, specifically the NFS. Cronin (2014) states the PRPA completely ignores fossils found within Native American Territory, suggesting that this may be an intentional oversight to give complete authority of paleontological resource management over to the Bureau of Indian Affairs. In a personal communication with Gregory Liggett, he stated that the purpose of the

PRPA was to focus on the main public land managers. Other departments outside of the Bureau of Indian Affairs were also intentionally excluded because the scope of the PRPA was strictly limited to public land managers.

3.2 Importance of Curated Collections

Museum and institutional collections are held in public trust for both the public and the scientific community. These museums and institutions are entrusted with the responsibility for the collections' protection as well as education and the maintaining of academic standards for curation and conservation (American Alliance of Museums 2012; Breithaupt 2016). Fossils found on public lands are housed within both federal and non-federal repositories, with thousands of fossils being added annually for the purpose of research (Liggett et al. 2019). The older collections, known as legacy collections, are constantly being updated or rediscovered (Rosenberg et al. 2019).

Collections are curated to the scientific standards and principals of their specific scientific fields and are constantly evolving as scientific standards evolve (American Alliance of Museums 2012). These curation practices are specified in the PRPA Sections 6302 and 6305, which state that the management and protection of paleontological resources on federal land must use scientific principles and expertise (PRPA 2009). They also state that any paleontological resource, along with its accompanying data and records collected under a permit, should be stored in an approved repository and the government allowed the opportunity to enter into an agreement with non-federal collections to house paleontological specimens (PRPA 2009).

There are, however, some problems that can come with museums holding federal collections. Before the standardization of curation and management practices, older collections were poorly curated and the loss of information currently plagues modern collections, affecting modern management practices (Santucci and Koch 2003). The loss or orphaning of federal collections is also a problem according to Liggett et al. (2019). With the universal attempts at data standardization through the digitization of collections, these problems are being addressed and have been associated with new finds as museums turn their attention to legacy collections.

2.2. GIS and Paleontological Resource Management

Inventorying of paleontological resources within public lands requires documentation of all specimens, both collected and *in situ*. This process of documentation requires inventorying specimens within collections, specimens' corresponding literature, and referencing each specimen with geological maps and written documentation to verify the accuracy of the information. A review of the region's geology is the next step for performing the inventory. The BLM has a classification for fossiliferous geologic units known as the Potential Fossil Yield Classification (PFYC). The PFYC rates the sensitivity of the geologic units, with formations with fewer fossils being given a rating of one while abundant fossiliferous formations are rated between two through five. Application of GIS in paleontological management and inventorying practices is growing. Geodatabases are created by assessment, evaluation, and management within the public lands under their domain. The BLM also incorporates the PFYC when mapping the geological units as a method to catalog geological units within public lands. However, the PFYC does come with limitations in sections of public lands that are extensively non-fossiliferous (Liggett 2020).

3.2 GIS and Inventory of Federal Lands, Parks, and Monuments

The NPS, BLM, and UGS have developed workflows built upon paleontological and geological surveys. They incorporate geographic information systems (GIS) through the construction of geodatabases to perform assessments and inventory the parks' and federal lands' nonrenewable paleontological resources. Murphy et al. (2015) used the geodatabase to correct and add fossil locality information to the Royal Gorge Fossil inventory and management. The NPS developed a geodatabase for field exploration and inventorying of fossil localities for the Grand Canyon National Park. Their geodatabase was also used to assess the correctness of the collections as well as to update information pertaining to legacy data (Santucci and Tweet 2020; Nebel 2020). The UGS has developed a geodatabase for the same reasons as the NPS and the BLM. Due to confidentiality, published data about their geodatabases is limited and direct communication with the NPS and BLM is necessary to develop a greater understanding of their geodatabases.

Similarities within the geodatabase schema can be found between the NPS and the BLM. Both define site and locality as point and polygon feature classes. While there are differences within the attributes, they follow traditional standards of defining taxonomic groupings, geology, identifying dates of collection, collectors' names, documentation of photographs, field notes, elements defining morphology, and so on. What is different is the classification of geologic units. The UGS and BLM use the Potential Fossil Yield Classification (PFYC) rank geologic units between one and five. This ranking defines the sensitivity of fossiliferous geologic units by their fossil bearing potential.

3.2 GIS and Museum and Institution Collections

Protection of paleontological resources within the United States are protected under the Paleontological Resource Preservation Act of 2009 which gives the Secretary of the Department of Interior the ability to develop plans for site inventories and the management of paleontological resources. The statute, 16 USC § 470aaa-4, states that any permitted collection must be stored within approved repositories (PRPA 2009). Approved repositories that enter into agreement must follow approved curation methods for fossil resources, data, and media (PRPA 2009).

Museum collections are held in trust to the public for scientific research and public education (American Alliance of Museums 2012). The curation of museum collections is constantly being held to the scientific field standards and requires the ability to expand and adapt to changes within the scientific community (Graham et al. 2004; Lieberman, Bruce S. and Kimmig 2018; G. A. Liggett et al. 2019). Collections are consistently curated to remain current to each field's classification and need to have curators equally qualified to maintain data integrity (Graham et al. 2004; Pham 2015; Marshall et al. 2018; Voss 2018). These repositories can hold over 2.5 billion specimens and each collection recorded to that department's criteria (Graham et al. 2004; American Alliance of Museums 2012). Collections provide historical, or legacy data, which can be used to map specimen distribution and movement (Graham et al. 2004; Stigall and Lieberman 2006; Gavin et al. 2014).

To get from the field to museum and institutional collections, indexing and identifying fossil specimens goes through a process from field identification to laboratory processing to obtaining an accession number. Processing specimens for the collection is done in a series of steps including accessioning, documentation, conservation, maintenance, and research of the specimen (Voss 2018).

Integration of museum and institutional collections has become the focus for the Smithsonian Museum and the NPS. In an Esri Presentation, Cole (2016) discusses how the Smithsonian Museum plans to integrate their databases with GIS and the long-term benefits. Data on location, depth, environment, department, geologic age, biasing, and spatial analysis would be accessed (Cole 2016; Lieberman, Bruce S. and Kimmig 2018).

GIS combined with museum and institutional records provides a quality method for assessing the completeness of the fossil record (Lieberman, Bruce S. and Kimmig 2018). One such exploration of combining museum collections with field collections and outreach was performed by Pham (2015). Pham (2015) worked with paleontologists from the Page Museum at the La Brea Tar Pits to develop queries within Microsoft SQL to pull information from the museum collections. To display these queries, Pham (2015) created a web application that would allow visitors to interact with these queries. This web application allows visitors to explore the tar pits and associated fossil assemblages. Due to the sophistication of the queries, problems with the application's ability to properly function resulted in complications with the application's ability to respond to visitor interaction.

Burrows (2016) builds upon Pham's (2015) geodatabase design for her thesis by developing a field geodatabase to inventory archeological artifacts collected in the field for the Isreali Archaeology Consortium. This geodatabase was designed for field collection and inventorying archeological artifacts found within the site. This geodatabase and Pham's (2015) field geodatabase both use occurrence data and polygons to identify sites. Burrows (2016) focuses on the importance of ontology to define the differences in order to query the artifacts, whereas Pham (2015) focuses on relationships. While Pham (2015) identified the problem with her field geodatabase and its accompanying SQL queries as an issue of over specification and defining relationships, resulting in an overly complicated design, Burrows (2016) had the problem of being too specific with the ontology, resulting in problems with overly complicated queries.

2.3. Petroleum County Geology and Paleontology

Reports pertaining to federal paleontological resources within the NPS, BLM, and the Utah Geological Survey (UGS) begin with an introduction of the PRPA, which discusses the scientific and outreach programs, benefits to paleontological resource management, curation, and illegal collection. Each report documents the geology and paleontology found within the park to provide an overview for the area of effect. The Potential Fossil Yield Classification (PFYC), used by the BLM, classifies geology as a formation’s potential to produce fossils. To provide the geological and paleontological context necessary for the geodatabase, this section provides a literature review of Petroleum County and its geology and paleontology.

Petroleum County has an extent of 1,675 square miles and is located within the Missouri Plateau in Central Montana (Figure 2) (Lindahl 1993). The geomorphology of the county consists of rolling hills with low to moderate slopes and steep, incised valleys along the Missouri and Musselshell Rivers bordering the county to the North and East (Lindahl 1993). Elevation for Petroleum County ranges between 2,250 to 4,000 feet and it has a mean annual precipitation of 13 inches (Lindahl 1993).

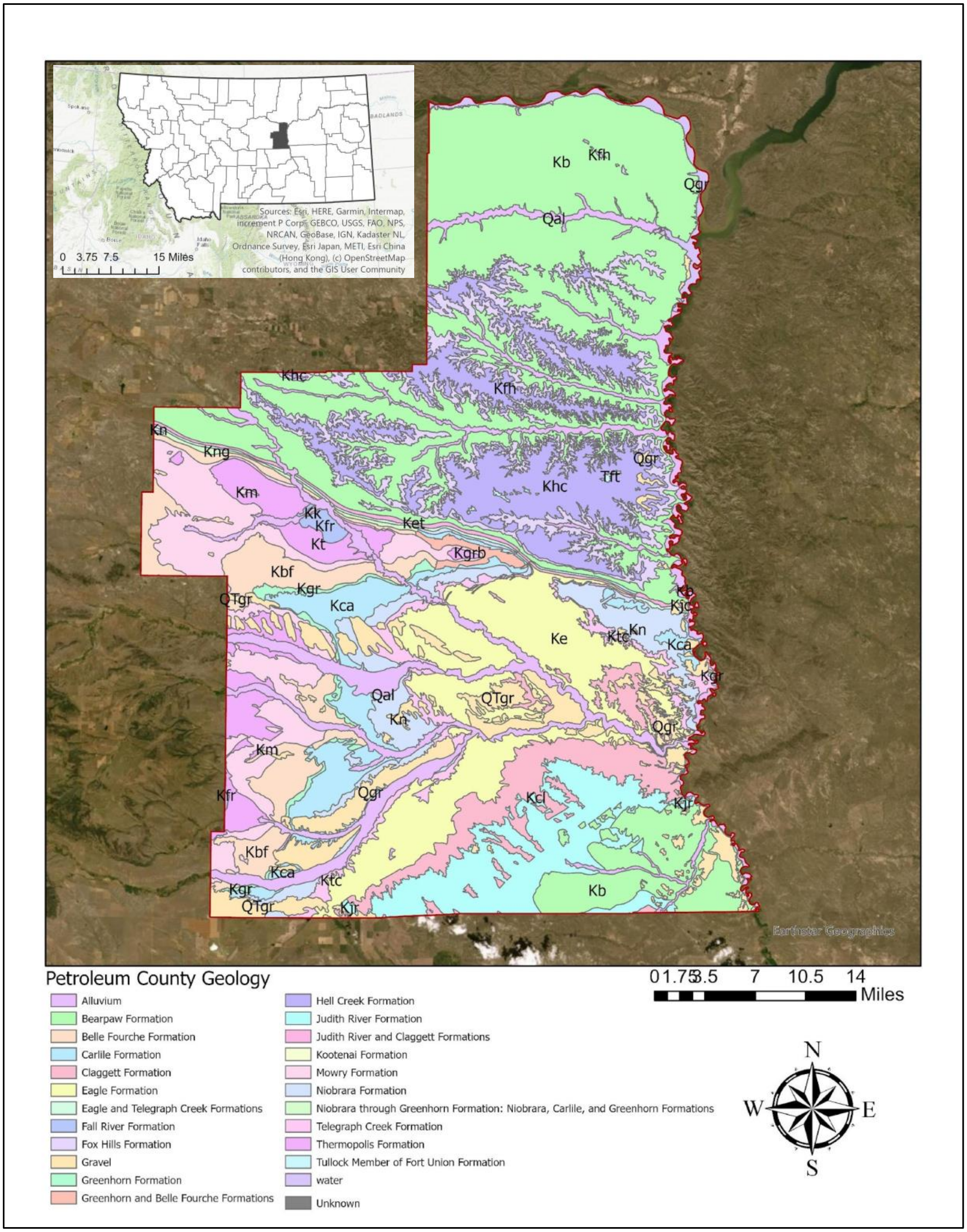


Figure 2 Petroleum County, Montana

3.2 Geology

The Cat Creek Anticline is the most prominent geological structure within the county (Lindahl 1993). The Cat Creek Anticline is made up of a series of asymmetrical domes trending westward and divides the county into two separate and distinct geological provinces (Lindahl 1993). North of the Anticline are the Bearpaw Shale and the Hell Creek Formations (Lindahl 1993). South of the Cat Creek Anticline extends the Bearpaw Formation, which dips in a gentle eastward direction into the Flat Willow Anticline (Lindahl 1993). The Eagle, Claggett, and Judith River Formations define the southern part of the Cat Creek Anticline.

Marine shales deposited during the Cretaceous represent the oldest geological units within the county and are known as the Colorado Group (Lindahl 1993). These fissile shales are dark grey to black in color and are easily prone to mechanical weathering (Lindahl 1993). The shales are often accompanied by calcareous sandstones grading into sandy or sometimes shaly limestone (Lindahl 1993). Fossils are present but not common within these formations.

Overlaying the Colorado Group is the Eagle Formation, which consists of buff to tan fine-grained sandstones (Lindahl 1993). According to Vuke et al. (2007), the Eagle Formation throughout Montana is grey to greyish brown, thickly bedded sandstone, and sandy shale with interbedded coal. Vuke et al. (2007) describes a member of the formation as the Virgelle Member, which is light grey or brown sandstone that coarsens upward with thin carbonaceous beds at the uppermost part of the member.

These sandstones vary in thickness and in some places around the state can be as much as 150 m thick; shale partings are common in certain areas (Vuke et al. 2007). The depositional environment for the Eagle Formation are coastal plains, off shore bars, tidal flats, deltas, and marine shoreface (Lindahl 1993; Vuke et al. 2007).

Overlaying the Eagle Formation is the Claggett shale and sandstone. The Claggett Formation consists of marine shales and sandstones that are dark grey in color and fissile. The shale is dark grey to grey in color that weathers brown (Vuke et al. 2007), while the sandstones are calcareous, thinly bedded, and fine-grained (Lindahl 1993). Calcareous concretions are found in the lower section of the formation (Vuke et al. 2007). The Claggett sandstones are not as highly resistant as the Eagle Formation, and unlike the towering Eagle ridges they are found in gradual slopes, small ridges, and valleys and can have a maximum thickness of 170 m (Vuke et al. 2007).

Overlaying the Claggett shales and sandstones is the Judith River Formation. The Judith River was deposited in a brackish, nearshore marine and delta complex and consists of alternating sequences of shale and sandstone beds (Lindahl 1993; Vuke et al. 2007). The shales are carbonaceous and silty and black to grey in color (Vuke et al. 2007). The sandstone sequences are highly prominent, discontinuous, vary in lithology (Lindahl 1993), and can range in thicknesses up to 305 m in parts of the state (Vuke et al. 2007). The sandstones within Petroleum County are fine to coarse-grained and range in color from pale greens to browns (Lindahl 1993). Due to the resistivity of the sandstone beds, the Judith River Formations often form ridges and also some carbonaceous beds (Lindahl 1993).

The Bearpaw formation overlays the Judith River Formation and consists of dark grey to black marine shales with several zones of calcareous concretions (Lindahl 1993; Vuke et al. 2007). Bentonite beds are commonly found within the formation and range in color from pale yellow or buff (Lindahl 1993). Large iron and calcium concretions from bands within the formation are highly fossiliferous (Lindahl 1993). Shale present within the Bearpaw Formation is easily weatherable and erodible (Lindahl 1993).

The Fox Hill overlays the Bearpaw Formation and has three members: Colgate, Timber Lake, and Trial City Members (Vuke et al. 2007). The lithology of the Fox Hills Formation consists of yellow orange to grey non-calcareous sandstone that is fine to medium-grained in texture (Vuke et al. 2007). The lower section of the formation contains interbedded sandstone, siltstone, and shale with a concretion zone (Vuke et al. 2007). The Colgate member directly underlies the Hell Creek Formation. This member is a white sandstone that has a thickness of a meter or greater (personal communication with Dr. Hastings). The Fox Hills Formation consists of marginal marine deposits that can have a thickness between 30 to 45 m (Vuke et al. 2007).

The Hell Creek Formation overlays the Fox Hills Formation and is made up of sandstones, shales, mudstones, ironstones, and bentonite (Stein and Triebold 2013). The Hell Creek Formation consists of light grey bentonitic claystone with alternating grey or brown sandstones with carbonaceous shale (Vuke et al. 2007). The sandstone is highly prominent and can be either calcareous to non-calcareous (Lindahl 1993). The sandstones are fine grained and resistant to erosion (Lindahl 1993). The shales are soft and fissile in nature (Lindahl 1993). Their colors can range from grey to pale green to brown (Lindahl 1993). The Hell Creek Formation can be highly resistant in areas, causing the formation of steep ridges and cliffs (Lindahl 1993). The ridges are capped by the exposure of the lower section of the Hell Creek Formation within the study area (personal communication with Dr. Hastings).

The strata for the entirety of the Hell Creek Formation consists of fine-grained fluvial and floodplain deposits and channel sequences, ranging from 40 to 170 meters thick and originating from the Williston Basin, North Dakota and Southern Canada (Fastovsky and Bercovici 2016; Jackson and Varricchio 2016). Fluvial channeling originated from the Rocky Mountains to the west have a lithology of sandstones, lignite, silt stones, and claystone with floodplain deposits exhibiting laminated silts, mudstones, and paleosol development (Jackson and Varricchio 2016; Fastovsky and Bercovici 2016). The Hell Creek Formation represents deposits from the margin of the Cretaceous Seaway and represents the final eastwardly regression (Lindahl 1993).

Overlying the Hell Creek Formation is the Tullock Member of the Fort Union Group. The Tullock Member consists of yellow interbedded sandstone and grey-black shale inside thin beds of coal within the layer (Vuke et al. 2007). The Tullock member has a maximum thickness of 180 m and was deposited as an alluvial plain (Vuke et al. 2007). Some glacial erratics are present within the area but speculation through transportation from the Missouri River may have caused their deposits (Lindahl 1993). Modern floodplain deposits from the Holocene are present.

3.2 Notable Paleontological Finds

The majority of the finds within Petroleum County have little to no documentation accompanying the specimens. The purpose of this section is to describe some of the documented finds within the county. The most notable find is documented by Stein and Triebold (2013), which describes a theropod skeleton discovered in 2002. In their chapter, Stein and Triebold (2013) describe the discovery of a skeleton nicknamed "Sir William" and attempt to identify the therapod taxonomically. As of this publication there is no current taxonomic classification given to this theropod.

The documentation provided by the Smithsonian National Museum of Natural History contained three vertebrate specimens and 21 invertebrate specimens. These species included ammonites, fish scales, and Pteranodons discovered within the Mowry Shale. Ammonites were discovered in the Mosby member of the Colorado Group and were documented and described by Cobban (1953).

Hadrosaurs and torosaurs were discovered within the Hell Creek formation in the neighboring county of Garfield. Wosik, Goodwin, and Evans (2013) describe the cranial morphology of the *Edmontosaurus annectens* from nesting size to adulthood. Sullivan, Boere, and Lucas (2005) compared the *Torosaurus utahensis* with similar species and with the occurrences of the *Torosaurus latus* within the Lance Formation in Montana and South Dakota. Sullivan, Boere, and Lucas (2005) suggest the *Torosaurus utahensis* and *Torosaurus latus* are distinct species based on unique features within the squamosal and postorbital horncore.

According to Jackson and Varricchio (2016), over the past 30 years Montana has yielded fossil dinosaur eggs and embryos of the theropod *Troodon formossus*, and its Two Medicine foundation yielded eggshells from the *Maisasaura peeblesorum*. The Judith River Formation has yielded information on the eggs, embryos and juvenile examples of *Hypacrosaurus stebingeri* (Jackson and Varricchio 2016). Turtle eggs have also been recorded within the Judith Formation (Jackson and Varricchio 2016).

3.2 Summary

In 2009, the passage of the PRPA laid the guidelines for defining the processes of collection, management, and inventorying of the fossil resources found within public lands. Researching these resources has been in part an effort by museums, institutions, NPS, BLM, and researchers to better understand past life on earth and educate visitors through outreach programs and exhibits. This is important and frequently documented within each inventory report.

Documentation outlined in past paleontological inventories reference the PRPA, its importance to conservation and why the information was protected from publication. Each inventory outlines the geology of the region, the fossils associated with the geology, and the age of the formations before it documents the process of the survey and its results. This process is necessary for describing the site and how the park's or monument's resources were documented using geodatabases.

Chapter 3 Methods

This chapter describes the methodology used during the construction of the geodatabases for the use of site management, assessment, and inventory of paleontological resources. It focuses on Petroleum County, Montana as a case study to evaluate the usage of museum collections and the Potential Fossil Yield Classification (PFYC) in GIS geodatabases for site assessment, inventorying, and analysis. Section 3.1 provides an overview of the methodology necessary to construct the geodatabases. Section 3.2 breaks down the data processing required prior to the construction of the geodatabases. Section 3.3 describes the creation of the geodatabases for which the accessioned paleontological records from museum collections and resulting field collections are to be documented per the guidelines of the BLM and NPS, and to the request of the Science Museum of Minnesota, for whom this case study was designed.

3.1. Methodology Overview

Construction of the geodatabases resulted from reviewing paleontological resource assessments and inventorying on public land. It also resulted from conversations with GIS and the NPS and BLM paleontologists who are using GIS and geodatabases to perform these assessments. Documentation on the construction of these geodatabases is limited and contains information protected by the PRPA of 2009 from the Freedom of Information Act. While geodatabases are being used within paleontological resource assessment, they are designed for field assessment only and have not yet been combined directly with museum collections.

Museum collections are used for paleontological resource assessment and inventorying of public lands but have yet to be incorporated directly into GIS. Digitization of museum

<table>
<tr><td>Data Collection</td><td>Collected DEM, Geology, PFYC, Paleontological Collections, Land Cover, Hydrology, Roads, Public Lands and County Fossil data was gathered from iDigBio and GBIF
Requested information pertaining to specific formations from museums, institutions as well as the BLM for additional information
Queried the data down to the specific county related to the case study</td></tr>
<tr><td>Data Process</td><td>Compiled the fossil information into a singular csv file already designed to the requirements of the BLM, NPS, and specified for the paleontologist involved in the case study
Reviewed the tabular information for duplicates and multiple records
Imported the fossil data into ArcGIS Pro to visually review missing or combined records
Mosaicked the DEM
Converted and projected the shapefiles into NAD 1983 StatePlane Montana FIPS 2500 (Meters)</td></tr>
<tr><td>Creation of Geodatabases</td><td>Created two geodatabases to house collections and field feature classes
Subtypes and Domains were created for paleontology and geology
Domains were set for: PFYC, Formation, and Fossil Context
Used a domain for lithology generated by the USGS
Museum Collections Geodatabase:
Created feature datasets for Montana Paleontology, Petroleum County, and Field Collection
Defined relationships between paleontology and geology
Performed a many-to-many join on the geology and the PFYC feature classes
Created a toolbox to contain the models that utilize the paleontology and geology layers
Set each feature class to allow for editing and image attachments
Field Geodatabase:
Created datasets for field collection of fossils, locality, and geology
Based these off the BLM and NPS schema and designed them for the case study's application for future work
Set feature datasets to allow editing and image attachments
Defined relationships between fossils and geology
Created a one-to-many join between fossils and locality</td></tr>
<tr><td>Topology</td><td>Topology was created for both geodatabases
Rules dictated that the field collection feature class must be inside of a locality and geologic formation
A locality must contain at least one fossil site
Geology must not have any gaps
Collections fossil data must be contained within geologic formations</td></tr>
<tr><td>Model</td><td>Models were created using ModelBuilder
Models focus on querying the Museum Collections geodatabase
Results are found within the field geodatabase so they can be brought into the field for future works
Parameters are defined within the models to allow paleontologists to define their criteria for how the model should be run</td></tr>
</table>

	● Results are automatically displayed on the map for immediate review
Visualization Through Tasks	● Visualization and interaction with the geodatabases used Tasks in ArcGIS Pro ● Tasks guide the paleontologist through each step of the data entry process and modeling ● Tasks simplify the process and allows non-GIS users to perform data entry and analysis

Figure 3 Methodology Workflow

collections are growing with the advent of online databases such as iDigBio and GBIF. Incorporating collections partnering with iDigBio and GBIF into GIS has been slowly explored in studies such as Graham et al. (2004), which explains the use of GIS to incorporate these repositories with biodiversity studies. The development of iDigBio and GBIF simplified the process of gathering information pertaining to the case study. To fill in missing data, requests for information were sent to museums and institutions housing collections that focus on public lands with fossiliferous formations that are protected by the PRPA of 2009.

Understanding these challenges was necessary in the development of the methodology to construct this project since it identifies the need for such a project. The methodology is broken down in Figure 3. Each step of the methodology is needed to use paleontological collections' information to perform the various analyses, while still allowing for the geodatabases to be set up within ArcCollector for future use in field application and assessment. The following paragraphs break down each step of the methodology used to construct the geodatabases.

The first step began with defining the locality for the case study. Petroleum County, Montana was chosen as the study area for this project through communication with Dr. Hastings, Paleontologist for the Science Museum of Minnesota. The Science Museum of Minnesota was chosen due to the author's former connection and familiarity with the department. For the benefit of the style of the geodatabase, it was decided that the active sites within the county would best fit this geodatabase's purpose for site analysis.

The next step was to determine data accessibility for the proposed site. The purpose of the project is to show the benefits of combining museum collections' information pertaining to paleontological collections within a geodatabase. After data was collected, validation of the data was necessary to determine whether the downloaded information was acceptable for use. This was performed through visual assessment within the tables while data was being compiled and through visual inspection in Esri ArcGIS Pro. Upon the revelation that data was not present, information was pulled from The Paleobiology Database to confirm data absence. Confirmation of the possibility that data may be restricted came from personal communications with the NPS.

While the purpose of iDigBio and GBIF are to standardize biological collections to a global standard, allowing citizens and scientists around the world access to curated collections held to scientific and bioinformatic standards, they are only available for sites that are not protected by laws or are actively being studied by scientists. When it comes to open-access databases such as iDigBio and GBIF, complications arise when information is protected by the Freedom of Information Act. Fossils found within public lands are protected by the PRPA and access to this information requires permission from institutions. A request for information was sent out to DOI institutions and was accepted. A list of specimens and participating institutions can be found in Appendix 1.

A total of 1,043 specimens were compiled from iDigBio, GBIF, and participating institutions. 1,001 of the specimens compiled have already been digitized and projected in NAD 1983. 42 of the specimens are considered to be legacy data since they were collected prior to GPS. These 42 specimens will be used for comparison and analysis upon completion of the geodatabases. The final list of participating museums and institutions providing paleontological data are: iDigBio, GBIF, the BLM upon request, the Smithsonian National Museum of Natural History (NMNH), the Field Museum, and the Science Museum of Minnesota. A specimen count can be found within Appendix A, where the names of participating museums and institutions are listed along with the number of specimens obtained from each museum.

Data validation of the digital elevation model, geology, roads, and hydrology were performed through Esri ArcGIS Pro. The DEM tiles were mosaicked to create a single layer then projected into NAD 1983 StatePlane Montana FIPS 2500 (Meters) to match the roads and hydrology layers. The geology layer was trimmed to fit within the county boundary layer and projected into the same projection as the road and hydrology layers.

The third step was the development of the Entity-Relation diagram, the geodatabase schema, the geodatabases, and replication. Design of the geodatabase Entity-Relation diagram and schema used Lucidcharts and Microsoft Excel. Creation of the geodatabase and its accompanying model will use Esri's ArcGIS Pro 2.5.2 and Modelbuilder. Creation followed the NPS and BLM geodatabase schema, the Potential Fossil Yield Classification defined by the BLM, Pham (2015) and Burrows (2016), and was adjusted to meet the criteria developed to meet the specifications of this project. This schema was broken down for a dual geodatabase system where one would support accessioned and curated data meant for collections, and the other would be designed for field accessioned data which would contain additional fields related to field collection processes.

The fourth step was to define the topology of the geodatabases. This process followed the NPS topology rules as defined in the Grand Canyon National Park's geodatabase schema provided by their GIS Program Manager Mark Nebel, as described in Figure 3, which will be discussed in more detail later on in this section.

The fifth step incorporated a series of models constructed in ModelBuilder within Esri ArcGIS Pro. These models allow the paleontologist to use queries that will reflect their hypotheses. These queries display information relating to the query and allow the paleontologist to further refine the parameters to develop a corresponding map depicting the results. These models are contained within the collection's geodatabase, so the information modeled has been properly accessioned into the museum collections.

Displaying and interacting with the geodatabases uses Tasks within ArcGIS Pro to guide the paleontologist through each step, from data entry to modeling. Tasks by Esri ArcGIS Pro are designed to guide non-GIS users while maintaining data integrity. Developing Tasks designed to guide the paleontologist in entering both field and collections information allows the paleontologist to enter data within specified feature classes while maintaining data integrity. The models can also be run through Tasks, allowing a paleontologist who may not be familiar with GIS to easily perform multiple analyses and easily review the results.

The final step was testing the geodatabases and models through Tasks to make sure topology rules, data entry, and models worked. Using the legacy data allowed tests of whether the editing feature within the feature classes would accept data entry using both location and georeferencing to create new features. The models were also run through Tasks and created various results from simple querying of fossils and geology to complex modeling.

3.2. Data Acquisition and Description

The gathering of data for the creation of the file geodatabases needed to contain information pertaining to paleontological resources and site assessment of public lands within

Petroleum County can be found in Table 1 in Appendix A. Two types of data were gathered for the construction of the geodatabases: spatial and tabular data. The tabular, or aspatial data, collected for this project relates to paleontological information for the State of Montana. This information was queried and downloaded from open source databases, iDigBio and the GBIF, which work with member institutions by opening their collections to the public for research. However, due to the Paleontological Resource Preservation Act (PRPA) of 2009, not all paleontological information is available to the public.

Since PRPA protects paleontological resources on public lands under the Statute 16 U.S.C. § 470aaa 1-11, requests for information were sent and accepted by the BLM, the Field Museum, the Smithsonian National Museum of Natural History, the Museum of the Rockies, and the Science Museum of Minnesota. A total number of 1,001 specimens were compiled between GBIF, iDigBio, and the participating museums that have accepted the request.

All biological and paleontological data uploaded into iDigBio and GBIF are converted into WGS84 for international use. Since WGS84 is not an ideal projection for spatial analysis, the downloaded data necessary for this project were projected into the NAD 1983 StatePlane Montana FIPS 2500 (Meters) to match the other layers.

The paleontological information provided by iDigBio and GBIF contain information on paleobiology and both vertebrate and invertebrate specimens. The total specimen count for iDigBio was 684 and the total specimen count for GBIF was 141. Information downloaded from these sites focused on taxonomy, location, geologic formation and age to provide the most pertinent information within the collections relating to this study.

Since the museum collections' information is spatial in nature, spatial information relating to rivers, roads, land cover, Federal lands, elevation, and imagery were downloaded from the Montana State Library, the USGS, and Landsat At A Glance. This information will be used to identify ease of access and environmental risks to current sites that can impact paleontological resources within Petroleum County. Appendix B breaks down data type, intended purpose, and availability of each dataset and has been updated to reflect the current datasets in use.

3.3. Data Preparation

Collection of data was the first step of determining what information was necessary for the geodatabase. Preparation of the information to determine data fitness and quality is important for the construction of the geodatabase since the geodatabase will be the backbone of any future work. Success of the geodatabase depends on the quality of data collected prior to importation.

Preparation of the tabular data was performed within Microsoft Excel, where the downloaded tables from iDigBio and GBIF were cleaned and reviewed for data quality and quantity. After cleaning the tables, they were uploaded into Esri ArcGIS Pro to assess the museum collections for data fitness and accuracy. Data pertaining to Petroleum County’s roadways, hydrology, and public lands were also evaluated for fitness and usability. This section breaks down this process by examining how it was performed and also the reasons for which each decision was made.

3.3.1. Discovery of Protected Records

Preparation of the museum data was necessary prior to its importation into ArcGIS Pro since each spreadsheet was organized differently. The GBIF data required the least amount of preparation when it came to organization and cleanup. GBIF required delimiting of the xlsx worksheet into a comma separated value (CSV). Evaluation of data integrity and missing fields was performed through reviewing the table and importing the table into Esri ArcGIS Pro. The

XY Table to Point Tool was used to create a feature class called GBIF Montana Fossils, project it into NAD 1983 StatePlane Montana FIPS 2500 (Meters) and map it within the Petroleum County Polygon to review occurrence count and placement. Data missing from the GBIF table data were ‘county’ and ‘geology’, so the Extract tool was used to retrieve Petroleum County occurrence information from the GBIF Montana Fossils Feature Class.

Preparation of the museum data was necessary prior to its importation into ArcGIS Pro since each spreadsheet was organized differently. The GBIF data required the least amount of preparation when it came to organization and cleanup. GBIF required delimiting of the xlsx worksheet into a comma separated value (CSV). Evaluation of data integrity and missing fields was performed through reviewing the table and importing the table into Esri ArcGIS Pro. The XY Table to Point Tool was used to create a feature class called GBIF Montana Fossils, project it into NAD 1983 StatePlane Montana FIPS 2500 (Meters) and map it within the Petroleum County Polygon to review occurrence count and placement. Data missing from the GBIF table data were ‘county’ and ‘geology’, so the Extract tool was used to retrieve Petroleum County occurrence information from the GBIF Montana Fossils Feature Class.

3.3.2. Petroleum County Spatial data

Processing of the spatial data describing Petroleum County, Montana required processing vector and raster datasets. The preparation of both vector and raster datasets was performed using the software, Esri ArcGIS Pro 2.5.2. The vector datasets (such as counties roads, hydrology, public lands, and geology) were processed using the Extract tool to make sure the accompanying attribute information was preserved within the boundary of the county polygon.

Once the vector data relating to hydrology, public lands, and roads was finished, the raster layers were analyzed. These raster layers were digital elevation models (DEMs) and land classes. The DEMs were downloaded as tiles and mosaicked together to create an elevation model of the county. After mosaicking the DEM's, they were projected into NAD 1983 StatePlane Montana FIPS 2500 (Meters) to maintain the same projection.

After the mosaicking of the DEMS, land class needed to be narrowed down to show only land class information for Petroleum County. This process required using the Raster Clip Tool so the data would only show for Petroleum County instead of the entire State.

3.4. Geodatabase Design and Construction

After processing the data, developing a geodatabase schema (Figure 6) was necessary before the construction of the geodatabase. Construction of the geodatabase schema followed the schema provided by the NPS, BLM, Pham (2015), Burrow (2016), and Murphy et al. (2015), and was developed through personal communication with the paleontologist at the Science Museum of Minnesota who will be using the geodatabase. From the schema, the geodatabases were designed.

Since there is little discussion on actual geodatabase construction outside of theses and a passing article, communications with the NPS, the BLM, and research into Esri Documentation Help Desk were used to define geodatabase type. Two file geodatabases were constructed: one to house the accessioned paleontological information gathered from museums specifically pertaining to Petroleum County, and the other designed for field excavation and surveying (Figure 3). These geodatabases were given names to reflect each geodatabase's function. The Museum Collections geodatabase was designed to hold accessioned data gathered from various museums and online biodiversity databases listed in Appendix A, and it contains models to assist with data interpolation.

The second geodatabase was named and designed to record and collect field data. This geodatabase allows the user to enter in information pertaining to field excavation sites, such as the locality of each site and the geology of the sites. The field geodatabase also houses the results for the models so they can be taken in the field for surveying and assessment purposes. Uploading the field geodatabase to ArcCollector is a future works development and will be discussed in a later chapter.

Investigation into Feature Classes, Subtypes and Domains, Relationship Classes, and Topology was performed to aid in maintaining data integrity and accuracy as data is entered into the Field Collection geodatabase and the Collection geodatabases. The creation of feature classes house data pertaining to each dataset that is imported into the geodatabase. Relationship classes define how these feature classes relate or interact with each other. Subtypes allow for feature classes to be categorized (Esri 2020c) while domains are created to define values that are specific to that field (Esri 2020a). Topology applies rules that maintain data integrity between each feature class and can be used to identify and fix errors (Esri 2020e).

Once the composition of the geodatabase was defined, the user needed to interface with the geodatabases and its contents. The interface needed to be easy to use and able to accommodate non-GIS users. Tasks were selected for this project due to their ability to walk the user step-by-step all the way through the process of data entry to modeling without the need to know how ArcGIS works.

Once the geodatabase was understood, construction of the geodatabase schema could begin. Construction of this geodatabase required directly contacting the NPS and reviewing the theses, federal reports, and journal articles that identify each geodatabases' strengths, weaknesses, and overall usefulness. The best features of each thesis were combined with the

NPS and BLM geodatabases to define relationships, identify attributes that relate to the topic, and define domains that will work for this project.

3.4.1. Feature Classes

A feature class is a group of similar objects that have a shared set of attributes or characteristics while having a similar structure or spatial representation (Esri 2020b; Yeung and Hall 2007). The feature classes developed for this project focus on museum collections, the region for the case study, Petroleum County, and field collection.

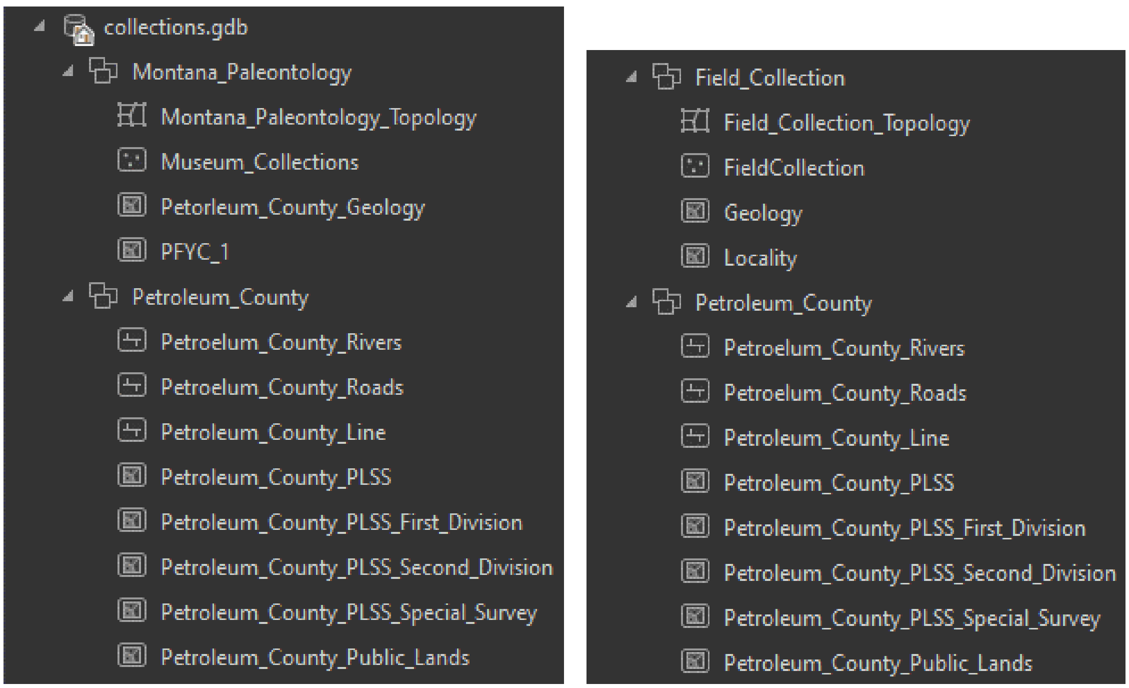


Figure 4 Feature Datasets for the Museum Collections and Field geodatabase

Figure 4 pictures the feature datasets and their accompanying feature classes. Each feature dataset is tailored to the specific feature dataset. For example, under the Museum Collections geodatabase is the feature dataset Montana Paleontology, which contains fossils and geology.

This set of feature classes contain the paleontological information gathered and compiled from museum collections as well as the PFYC and geology layers which were joined together to create Petroleum County Geology. The Museum Collections feature class is a point feature class that contains curated paleontological information that has been prepared, identified, and accessioned into a collection and is available for scientific research. The Petroleum County Geology feature class is a polygon feature class that contains geologic information pertaining to Petroleum County's geology. This layer was joined together with the PFYC polygon feature class so that when queried, the information pertaining to the formations and associated classification would be available. Petroleum County contains feature classes that pertain to Petroleum County, such as the line feature classes of roads, PLSS, and hydrology, and Public Lands, which are polygon feature classes.

Within the Field Collection geodatabase are the feature datasets, Field Collection, and the Petroleum County feature dataset. The Field Collection feature dataset is made up of a point feature class named Field Collection, and the polygon feature classes Geology and Locality. These feature classes have been designed to follow the standards of both the NPS and BLM and are customized to the paleontologist's preferences. These feature classes were designed this way so they would match permitting regulations and reporting to make documenting easier on the paleontologist. Field data are identified and documented in the field but require cleaning and further identification before being entered into the museum's collection. This feature class has additional fields to allow for this process.

The Geology feature class allows the paleontologist to record the geology of the site such as formation, lithology, PFYC, Fossil Context, and structure. This feature class was added since the last recorded geological survey was performed by Lindahl in 1993 and smaller outcrops may not be noted within the USGS geology layer. The addition of the Fossil Context allows for the paleontologist to record whether the fossil was discovered within the proper geologic formation or has eroded out. This field is also found within the field collection and locality feature classes.

Locality is a polygon feature class that allows the paleontologist to mark out the boundary of the site. Following the BLM schema, this feature class allows the paleontologist to document the site. This feature class was designed to record the site for excavation and surveying. This feature class also follows the BLM standards to make the following documentation easier.

Photo documentation is an integral part of surveying and excavation. Within the feature datasets for the geodatabases, the option for allowing images to be collected was selected. Thus, when the field geodatabase is uploaded into ArcCollector images can be attached to each attribute.

3.4.2. Relationships

Defining which attributes are needed for each feature class and defining their interaction with each other requires an understanding of the relationship between each dataset. A relationship can be defined as either simple or complex. A simple relationship is defined to have a one-to-one relationship between entities. Where a complex relationship is defined as having a one-to-many or many-to-many relationship between entities. These relationships are diagrammed through an entity-relation diagram found in **Error! Reference source not found.**.

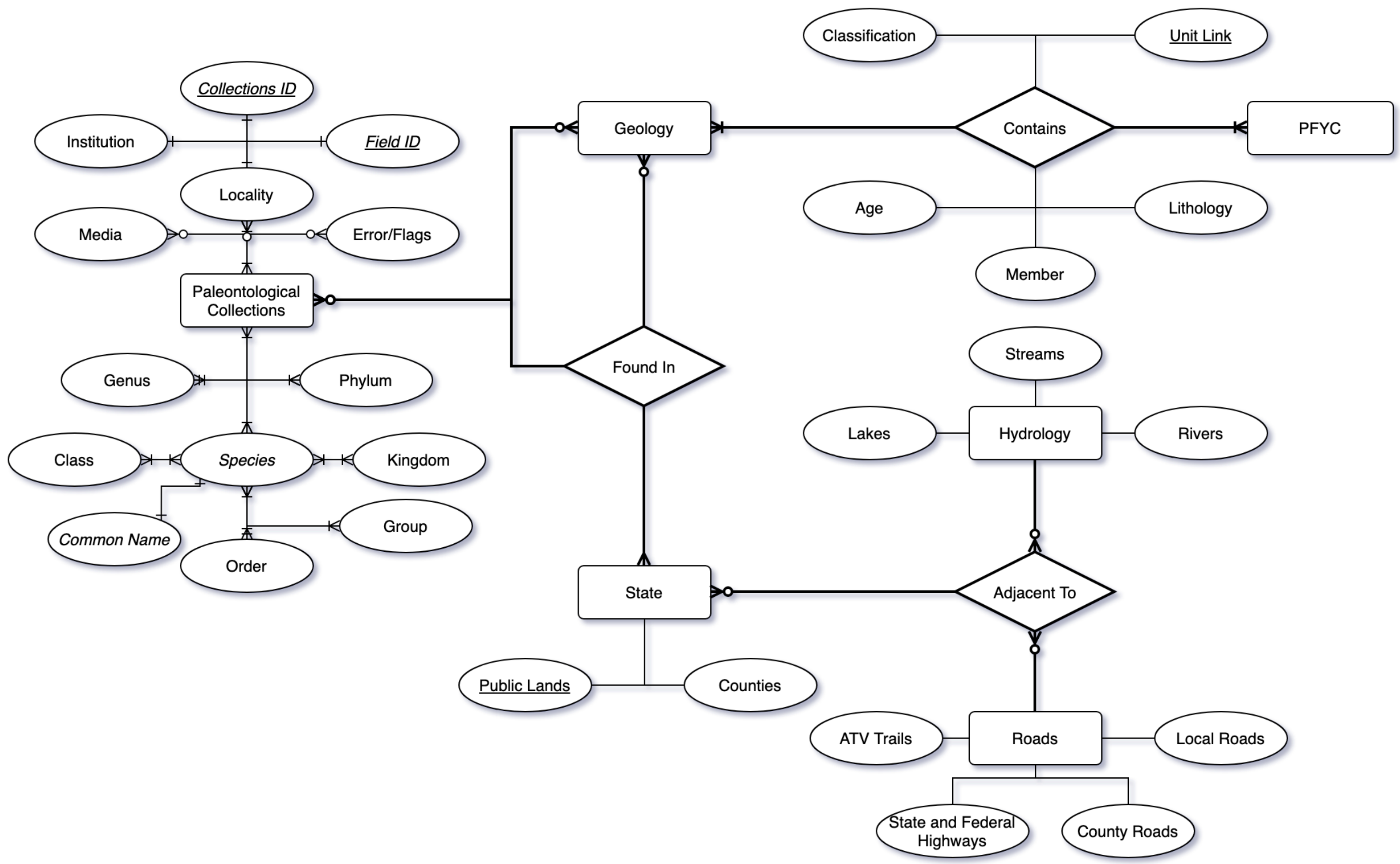


Figure 5 ER Diagram

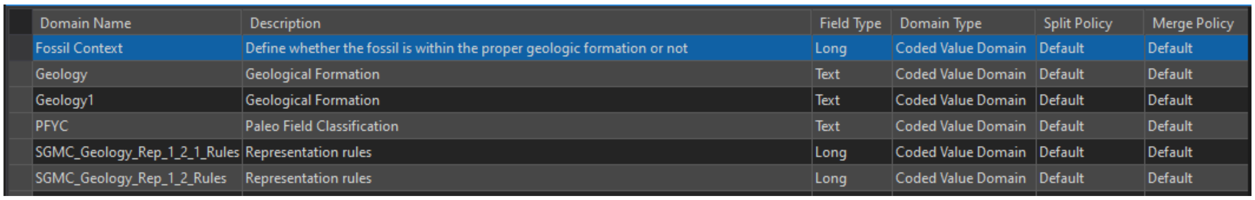
The E-R diagram depicted in **Error! Reference source not found.** maps out the attribute relationships to each dataset and their relationships to each other. Within Paleontological Collections, a species is given a one-to-one relationship with common names but given a zero-to-many description within Geology. The Paleontological Collections dataset was given a zero-to-many relationship in the Geology feature class. Not all geologic formations contain fossils, but all fossils must be found within a geological formation. Rivers and roads are designated as adjacent to the state of Montana since roads and rivers continue outside the boundaries of the state and its counties.

While each feature class is related to each other, defining relationship classes for this project resulted in a simpler result. This choice came after conversations with professors and GIS professionals within the field and studying the results of Pham (2015). Complex relationships had made querying complex, resulting in problems when trying to extract information from the database, and this also created problems with field applications as recounted by Pham (2015) in the experience with her app.

For this project, relationships and relationship classes were simplified and kept between geology, fossils, and locality. Relationships were defined for the geology, field collections, and collections feature classes since geology and fossils have a defined relationship. In the case of the geology and the PFYC feature classes, a relationship class and a simple join were created. With the join between the PFYC and geology, the queried table will return geological information while providing the PFYC classification value for that formation. For the Field Collection geodatabase, all three feature classes (field collection, geology, and locality) were joined using a one-to-many. Each feature class is related by Field ID so the information pertains to the same record.

3.4.3. Subtypes and Domains

Subtypes are used to characterize the feature class (Esri 2020c) and were used on the hydrology layer to identify rivers and streams. There was an attempt to perform a subtype classification for roads, but due to the remote nature of the county this was impossible to create.



Domain Name	Description	Field Type	Domain Type	Split Policy	Merge Policy
Fossil Context	Define whether the fossil is within the proper geologic formation or not	Long	Coded Value Domain	Default	Default
Geology	Geological Formation	Text	Coded Value Domain	Default	Default
Geology1	Geological Formation	Text	Coded Value Domain	Default	Default
PFYC	Paleo Field Classification	Text	Coded Value Domain	Default	Default
SGMC_Geology_Rep_1_2_1_Rules	Representation rules	Long	Coded Value Domain	Default	Default
SGMC_Geology_Rep_1_2_Rules	Representation rules	Long	Coded Value Domain	Default	Default

Figure 6 Entering Domains

Domains are stored values that restrict input values within columns to a certain coded value (Esri 2020a; Yeung and Hall 2007). Domains were created for this project to simplify and unify data entry for geology, lithology, fossil context, and PFYC. Found within Figure 6 are the domains created for both Collections and Field geodatabases. These domains are used to describe whether fossils are found “in context,” meaning whether the fossils were discovered within the formation associated with the fossils or whether the fossils discovered were eroded out. Allow the paleontologist to define the fossil probability of the geologic formation in question, and describe the formations using the USGS lithology descriptions. Each domain is given a coded value and a description (Figure 7) and assigned to a field. When used, a drop-down list is created and will depict only the assigned values created specifically for a particular field.

Code	Description
0	Out of Context
1	In Situ

a

Code	Description
1	Very Low Potential
2	Low Potential
3	Moderate Potential
4	High Potential
5	Very High Potential
U	Unkown

b

Code	Description
Qal	Alluvium
Kb	Bearpaw Formation
Kbf	Belle Fourche Formation
Kca	Carlile Formation
Kcl	Claggett Formation
Ke	Eagle Formation
Kfr	Fall River Formation
Kfh	Fox Hills Formation
Qgr	Gravel
QTgr	Gravel
Kgr	Greenhorn Formation
Khc	Hell Creek Formation
Kjr	Judith River Formation
Kk	Kootenai Formation
Km	Mowry Formation
Kn	Niobrara Formation
Ktc	Telegraph Creek Formation
Kt	Thermopolis Formation
Tft	Tullock Member/Fort Union Formation
W	Water

c

In a, b, and c represent some of the domains and their coded values created for this project. A represents the coded values for Fossil Context domain. This is to allow the paleontologist to define the value for the field upon their discovery of the fossil.
B represents the coded value for the PFYC domain. This allows the paleontologist to define the potential fossil productivity of the formation.
C represents the values for the geological formation. This coded value uses the formation unit name to define the formation. This was created to keep the uniformity and to simplify data entry for the paleontologist.
The one that is not represented is the USGS lithology description. This is because this domain and its coded values were left alone for the purpose of simplicity. This domain is coded to describe the lithology of the formation according to the USGS standards.

Figure 7 Domain Values

Each domain was given a name and a description of what each domain contains, and each domain contains a coded value and a description. Field type defines the coded value as either an integer or a text value which can be seen in Figure 7 and can be either given either a numerical value or a text value as is shown in Figure 7a and Figure 7c. Accompanying each coded value is a description describing each value. For this project domains were created to display whether fossils found during a survey or an excavation were found within or outlying the formation, classifying each formation's ability to produce fossils, describe the formation, and describe its lithology.

Fossil Context used long integers to define the codec value for the domain while the PFYC and the USGS Geology domains uses text values to define their coded values. Using a long integer to define the coded values for Figure 7a, uses binary to describe whether the fossil was found within the formation or has eroded out. A value of zero is given for the domain "out of context" and a value of one is given for "in context". Figure 7b are the coded values for the PFYC and were created to follow the BLM classification of potential fossil bearing formations. Each value is given a value one to five with a "U" being designated as Unknown. The SGMC is the USGS (Figure 7c) description for lithology which accompanied the geology feature class which was downloaded from their site. This domain was kept provide a unified system for describing the regions lithology. This domain was applied to all of the geology feature classes to maintain consistency with lithology descriptions.

Figure 8 PFYC

Figure 8 presents a visual representation of the PFYC. Ranks are given to each formation by the probability of the formation producing fossils. One represents that there is an exceptionally low potential of finding fossils within the geologic formation being surveyed. Two represents that there is a low potential for fossils being present within the formation that is being surveyed. Three is given to formations that have a moderate potential for bearing fossils. Four has a high potential of being a fossil bearing formation. Five is to be given to the formations that have the highest potential of producing fossils, where U is to be given to unknown formations (Liggett 2020). For site assessments and management, a classification of four and five have the greatest chances to produce the greatest chances to provide multiple field seasons.

Field Name	Alias	Data Type	✓ Allow NULL	Domain	Default	Length
Shape_Area	Shape_Area	Double	✓			
Site_ID	Site_ID	Text	✓			255
Site_Description	Site_Description	Text	✓			255
Formation	Formation	Text	✓	Geology1	Alluvium	255
Member	Member	Text	✓			255
Lithology	Lithology	Long	✓	SGMC_Geology_Rep_1_2_Rules	Sedimentary, carbonate	
PFYC	PFYC	Text	✓	PFYC	Unkown	255
Fossil_Context	Fossil_Context	Long	✓	Fossil Context	In Situ	

Figure 9 Setting Domains

Once the domains have been created, they need to be assigned to the specific attribute field. Figure 9 shows these domains being set to the Field geodatabase's Geology feature class. Domains were created and assigned after the table has been created or imported. Setting the domains is dependent upon their coded value and Data Type. If the field has been assigned as a long integer only coded values that are defined as long integers will appear in the domain field. In the case of Lithology, the data type was easily changed to match the coded value of the SGMC domain.

Default allows the user to define a value from the domains to be set as a default value for the attribute field instead of displaying a NULL value. An example of this can be seen in Figure 9, which displays the default values assigned using Figure 7's listed domain values.

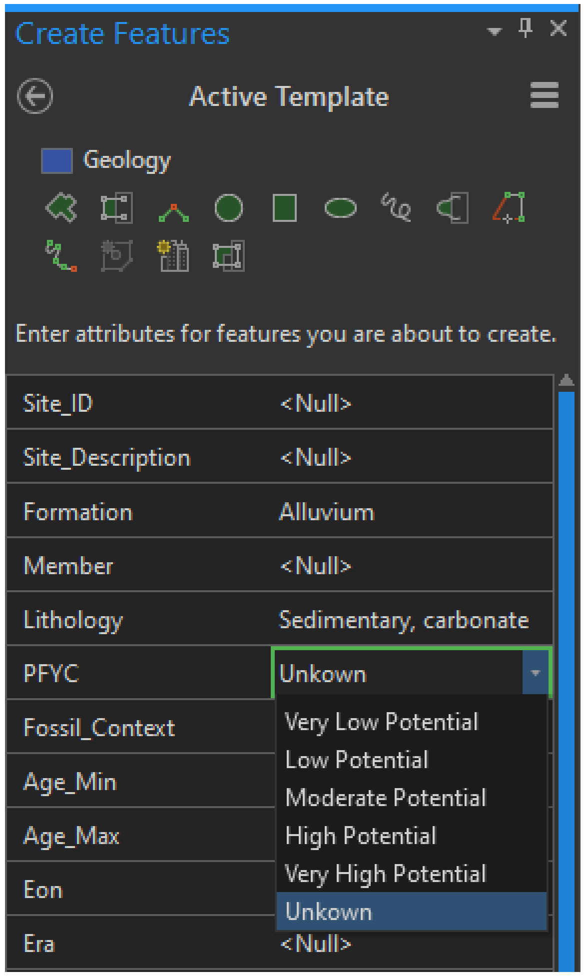


Figure 10 Displaying Domains

Once the domains were assigned to the appropriate fields, testing the fields (Figure 10) using the Editor tool was performed to make sure each assigned domain was displaying the correct information. When the field for an attribute containing a domain is selected a list of values drops down specific to that attribute. In the case of Figure 10 the paleontologist can select the PFYC value to assign the Field geodatabase's Geology feature class. Since the Default value for the field was assigned as Unknown, this value is the default value for the field and will be shown instead of NULL if nothing is selected. This can also be seen for the attributes Formation, Lithology, and Fossil Context which were also assigned default values.

3.4.4. Topology

Within ArcGIS Pro, topology is used to define, perform, or constrain integrity rules and features that share geometry, while also supporting editing of these features (Esri 2020f, 2020e; Yeung and Hall 2007). Integrity rules can be applied to a particular feature class such as the Geology feature class in both geodatabases where there must not be gaps between polygons (Figure 9), or they can be applied to multiple feature classes to maintain coinciding geometry (Yeung and Hall 2007).

Topology rules were created to enforce integrity rules that maintain data integrity while allowing the editing of features for the Montana Paleontology and Field Collection feature datasets. These rules enforce specific editable feature classes: the museum collections, the field geodatabase's field collection, geology, and locality in order to maintain data integrity as information is entered or updated.

Museum Collections Geodatabase Topology Rules

Feature Class 1	Subtype 1	Rule	Feature Class 2
Collections_Fossils_1		Must Be Properly Inside (Point-Area)	Petroleum_County_Geology_1
Petroleum_County_Geology_1		Must Not Have Gaps (Area)	

a

Field Geodatabase Topology Rules

Feature Class 1	Subtype 1	Rule	Feature Class 2	Subtype 2
FieldCollection		Must Be Properly Inside (Point-Area)	Locality	
FieldCollection		Must Be Properly Inside (Point-Area)	Geology	
Locality		Contains Point (Area-Point)	FieldCollection	
Geology		Must Not Have Gaps (Area)		

b

Figure 11 Topology Rules

Figure 11 depicts each feature class with their associated rules and are broken down by their geodatabase. Figure 10a looks at the collections and geology feature classes that are stored within the collection's geodatabase. All fossils must be inside the county's geology layer to make sure that no fossils are recorded outside the county line. In a visual check of the USGS geology layer, the petroleum county geology feature class was given a topology rule that each geology polygon must not contain any gaps between polygons. This visual check of the layer made sure that there were no problems with the layer.

Figure 10b shows the rules created for the field geodatabase topology rules that were designed to enforce data collection. Each point and polygon layer were defined to follow the BLM and the NPS schema and created for data entry during surveying or excavations. All fossils found or collected are to contain within a locality and each locality must contain at least a single fossil to allow the creation of the polygon.

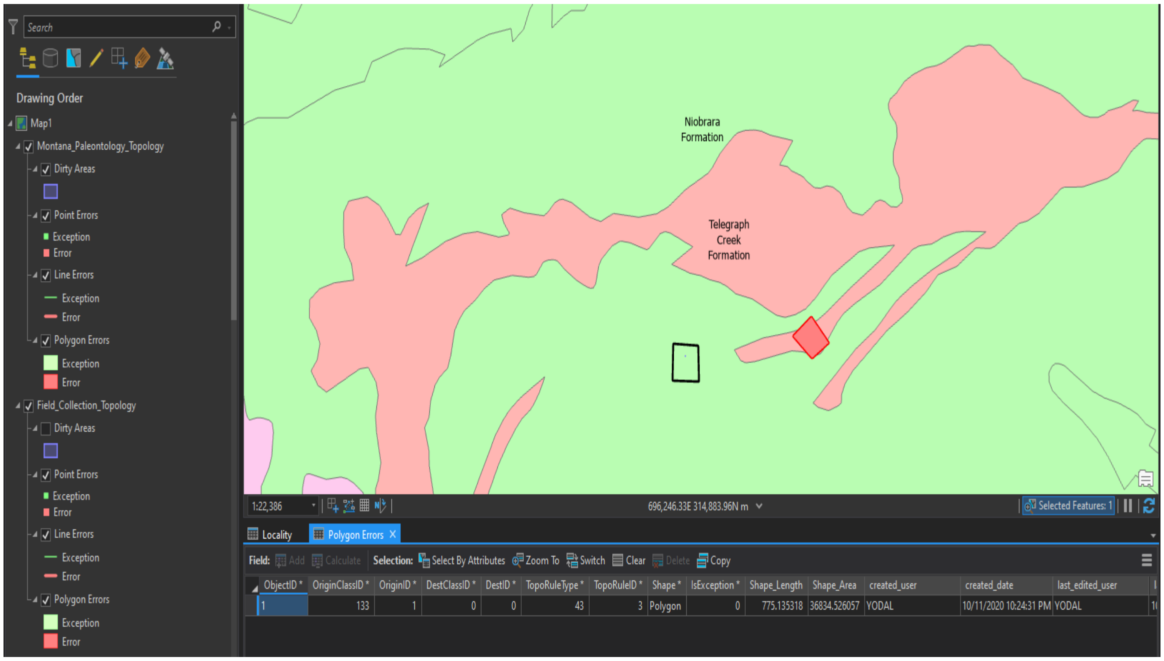


Figure 12 Topology Projected

Figure 12 shows the results of successful (black outlined polygon) and failed (red polygon) creations of a locality. Failure will result with the topology result within the topology layer for each geodatabase identifying errors and then notifying the user. All fossils must be recorded within a geology polygon in order to enforce accurate recording.

3.1.1. *Summary*

This section broke down the creation of each of the geodatabases, from describing the creation of feature datasets to defining topology rules to enforce data integrity. Two geodatabases were created: Museum Collections and Field Collections. The Museum Collections geodatabase was designed to house museum information and related geological and county information to assist with site management, assessment, and inventorying. The Field Collections geodatabase was designed for information collected in the field and was designed with the idea of uploading into ArcCollector to allow for portability. However, field data is often changed upon returning to the lab as the specimens are prepared. Finalization of data comes through accessioning within museum collections. This was why the Museum Collections geodatabase was created to replicate the information entered within museum paleontological collections. To assist the paleontologist in data manipulation of the accessioned data, models was constructed and added to the geodatabase. These models are discussed in the next section.

.

3.5. Models

Performing paleontological resource inventorying requires reviewing a collection's specimens along with its accompanying records, published documentation, field notes, and geologic maps. This in order to verify the information within the collection, the documentation of the region and field location, and to begin an inventory of the potential area of interest. This section discusses the use of ModelBuilder to create a set of custom tools to assist with querying and analysis of museum collections for the purpose of assisting inventorying, assessment, and management practices. The following sections break down the construction of each model and their purpose.

3.5.1. Construction

Modelbuilder within ArcGIS Pro offers different ways to perform programming by constructing a workflow that consists of a variety of components and connectors to create complex tasks or a custom tool (Allen 2011). Model components are what create the model and define how the model will operate (Allen 2011; Esri 2020d). They are represented by colored ovals and rectangles using connectors to denote direction of the workflow.

Variables are used either to contain data or store values (Allen 2011; Esri 2020d). Variables defined for these models are: the Montana Paleontology feature class, the Petroleum County feature, land cover, and the DEM found within the collection's geodatabase. The Museum Collections geodatabase contains the accessioned paleontological information and geologic information stored in the Montana Paleontology feature dataset. The Petroleum County feature dataset contains feature classes, roads, and hydrology which are often used to determine accessibility to sites and any potential erodibility of the regions. Land cover plays a large role in the discovery of paleontological resources since paleontological resources can only be found within regions subjected to erosion. The ability to classify the region's land cover to suit the paleontologist's need made the addition of this information important. The DEM was used to extract Aspect and Slope to help identify factors which can affect erosion, slope stability, or access to potential areas of interest.

Models use four types of connectors to define interaction between each component, although only one type was used for this project. Connectors are determined by: the type of data, the environment for which it is being used, the precondition which determines when the rest of the model will be run, and the feedback which allows the derived variable to be entered as an input (Allen 2011; Esri 2020d). For this project data input and parameters were used to allow for data entry to be performed.

Parameters allow for interaction with the models and are identified with the letter "P" next to the variable. Parameters also allow for new variables to be introduced and overwritten into the model without having to reconstruct a new model (Allen 2011). When the models are run, a window will appear allowing the paleontologist to define certain variables or perform analysis multiple times without having to recreate the model each time. The results are stored in the Field Collection geodatabase so the newly generated results can accompany the paleontologist in the field.

Models created for this project range from querying to performing analysis and were pulled into Task Designer for the ease of accessibility. The reasoning behind this decision is to allow a paleontologist with minimal GIS experience to perform analysis using museum collections to assess sites or regions that may benefit from further exploration, site management, or inventorying of paleontological resources. Centralizing the models with the data entry process keeps every process used by the paleontologist within a single, easy-to-use interface that is tailored to their needs without their needing to be proficient with GIS.

3.5.2. Defining Paleontology and Geology of Petroleum County

Integration of museum collections within GIS has been made easier with the development of spatial databases such as iDigBio and the GBIF. Studies into species distributions and ranges using museum paleontological collection have been performed by Rode (2005) and Stigall and Lieberman (2006) while Graham et al. (2004) explored during their development. Pham (2015) explored the combination of GIS and museum records through the use of Microsoft SQL Server while also focusing on field data. However, she noted that the complexity of the queries and their relationships made it difficult.

Defining Paleontology and Geology models allows for the ability to query attributes within the Museum Collections feature class and the Petroleum County Geology feature class, creating a unique way of interacting with paleontological collections. Intended to assist the paleontologist in the office for site assessment, inventorying, and management practices, the model is designed to act alone or in conjunction with other models to aid in the decision-making process.

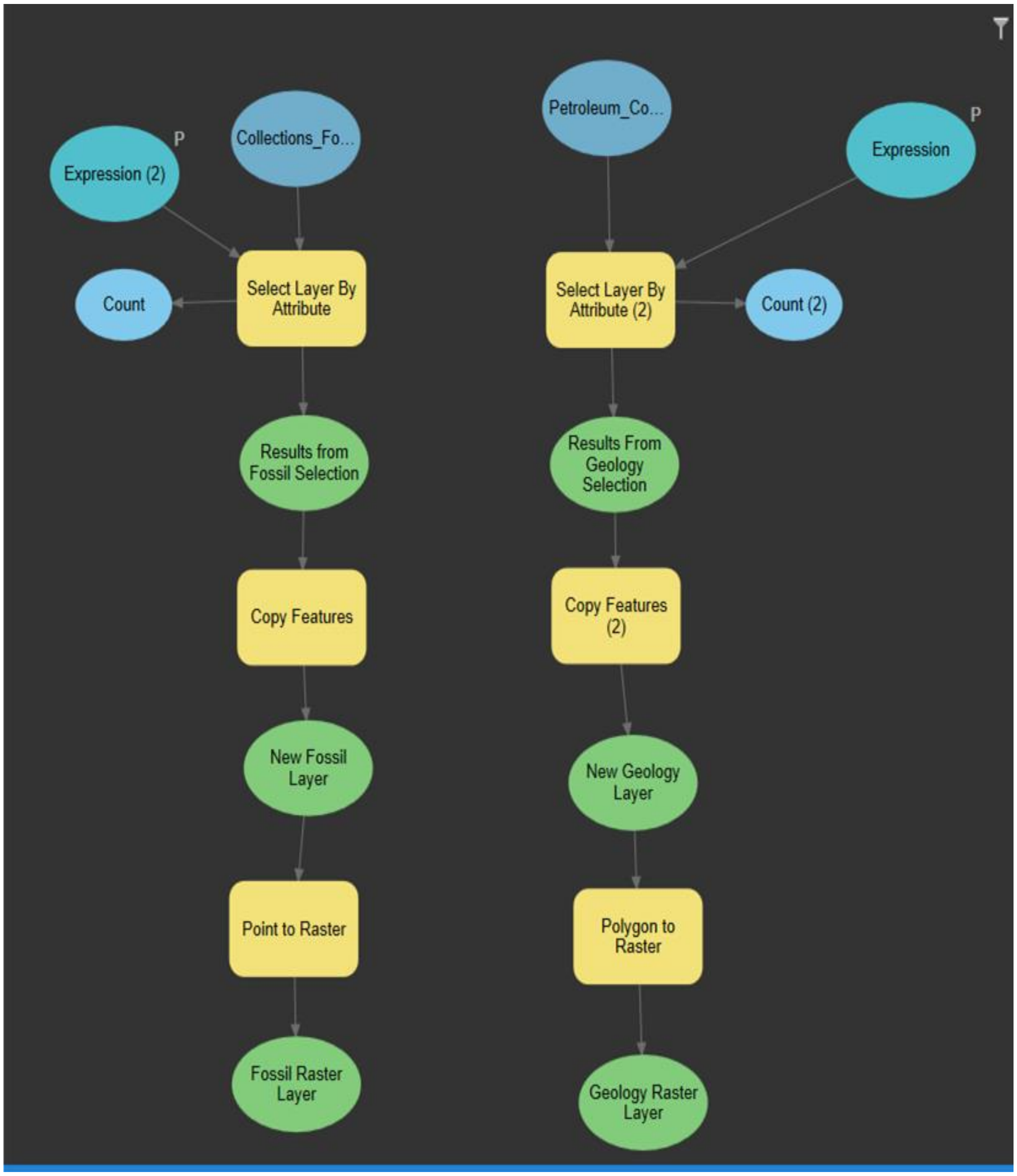


Figure 13 Define Paleontology and Geology

Shown in Figure 13 is the workflow constructed for this model. The Museum Collections and Petroleum County Geology feature classes were used as the primary variables to define the intent of the model. Parameters for this model allow for the Select By Attribute tool to be queried and selected to fit the paleontologist's inquiry. The selected results are copied and saved into a new layer within the Field Collection geodatabase and converted into raster for use within the suitability model, *High Potential Fossil Suitability* model.

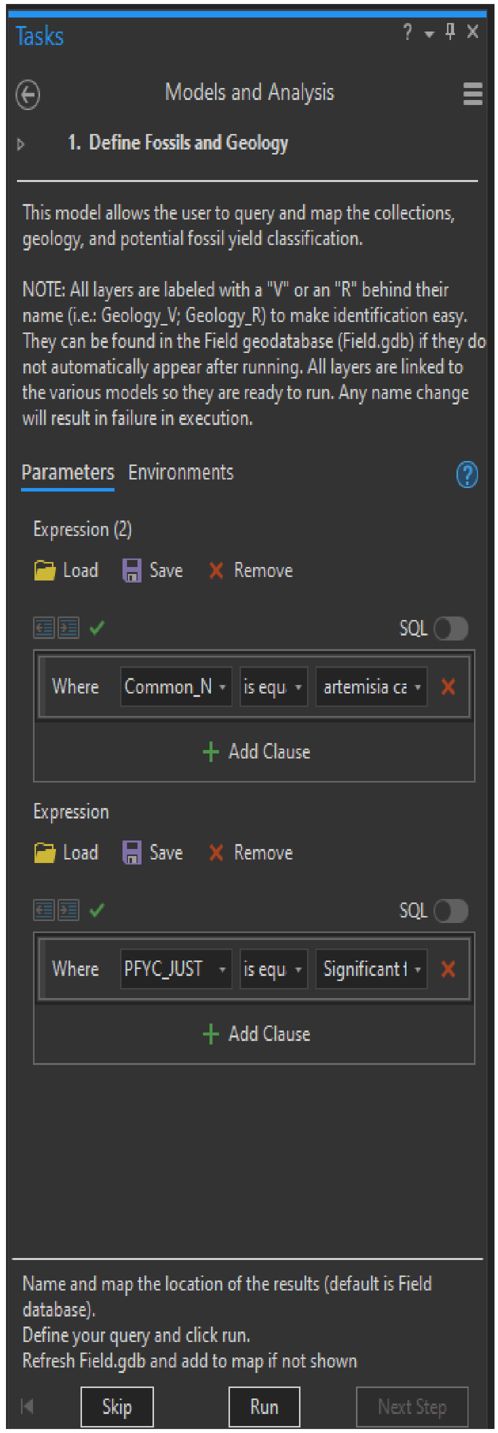


Figure 14 Define Paleontology and Geology Task Interface

Figure 14 is the interface the paleontologist sees when accessing Tasks. A description of the model informs the paleontologist about the model, where to find the results and which variable is being queried. A where clause expression allows the paleontologist to query the attribute table from the Museum Collections and the Petroleum County Geology feature classes. The paleontologist has the option to save and load queries they wish to create, simplifying the process. Running the query will result in the creation of a layer that will display the results, and which is stored in the Field Collection geodatabase for use in the field.

3.5.3. Define Roads and Hydrology Buffer

A site's accessibility from the road and the determination of whether is is endangered by the erosional processes of downcutting rivers were constructed into a model to allow for assessment of the region's accessibility to transportation and the identifying of endangered sites due to the region's geomorphology. Using the *Buffering Roads and Hydrology* model allows the paleontologist to determine accessibility to locations, identify regions that are affected by over collection or illegal collecting, and assess regions around rivers prone to downcutting into the bedrock and exposing outcrops.

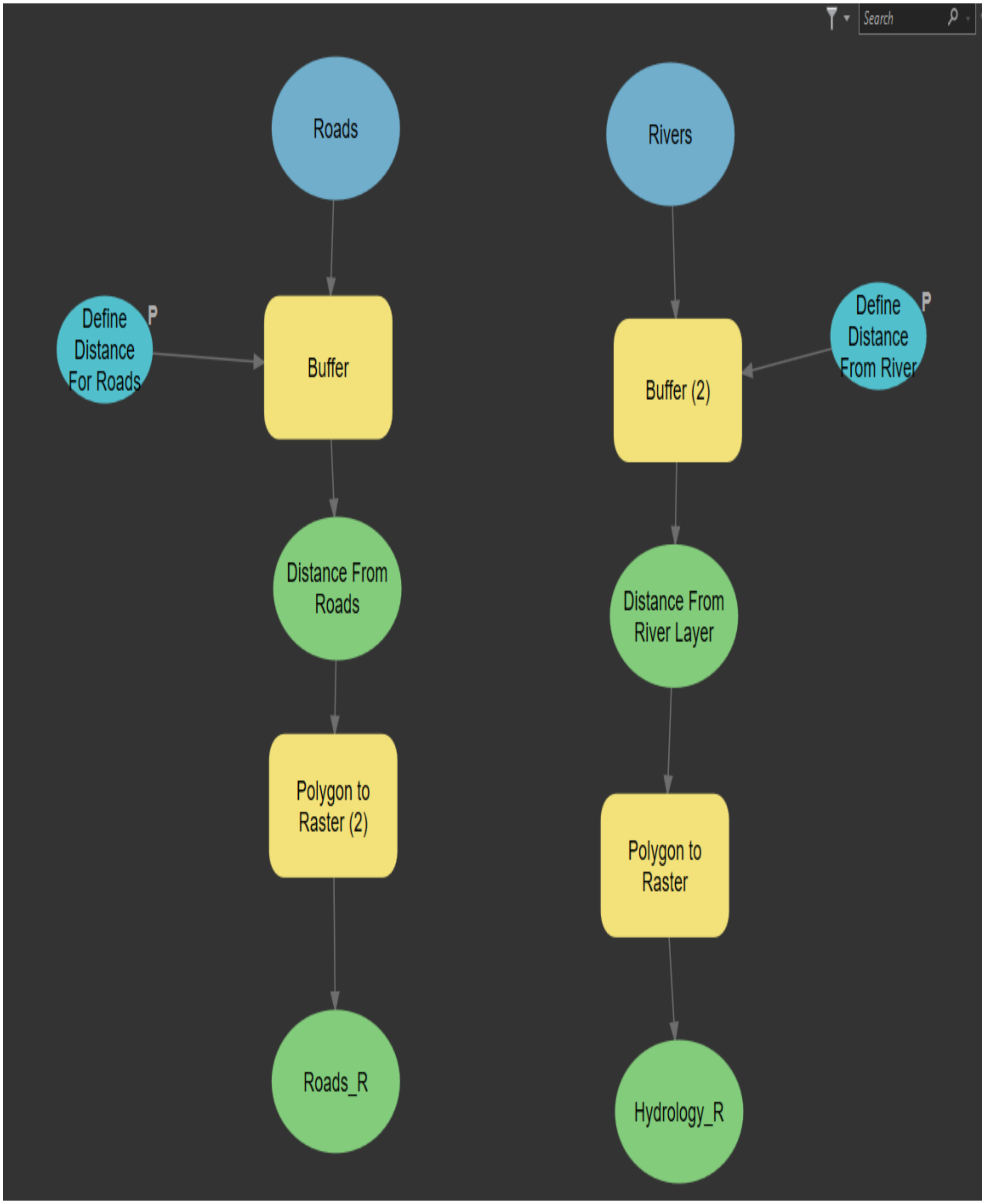


Figure 15 Roads and Hydrology

Figure 15 allows the paleontologist to define a buffer zone around the Roads or Hydrology feature classes found within the Museum Collections geodatabase. Parameters allow the paleontologist to define the distance that they want to investigate. The results of the buffer are stored in the Field Collections geodatabase so they can be taken in the field.

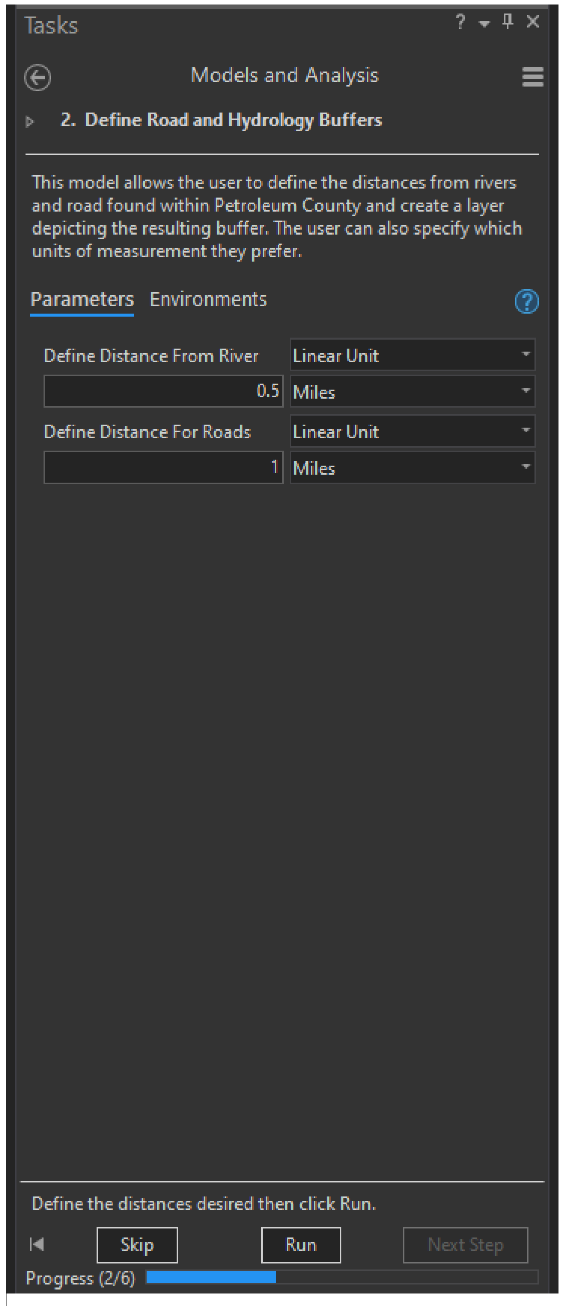


Figure 16 Define Roads and Hydrology Task Interface

Figure 16 depicts the interface the paleontologist would use to perform this process. The interface allows the paleontologist to define the units of measurement and the distance for their inquiry. The model, like all the other models, can either be run separate or in conjunction with the other models.

3.5.4. Define and Reclass Aspect and Slope

Two versions of the Aspect Slope models were created for this project with the intention that one would be used for defining the slope direction and percentage and the other one would be used primarily for reclassification of the results if the suitability model was to be used. The model uses the DEM of the region and defines the Aspect, or the direction each slope is facing, and the Slope of the terrain. These newly created layers are added together using the Raster Calculator to create a single layer that defines both the direction and the degree of slope.

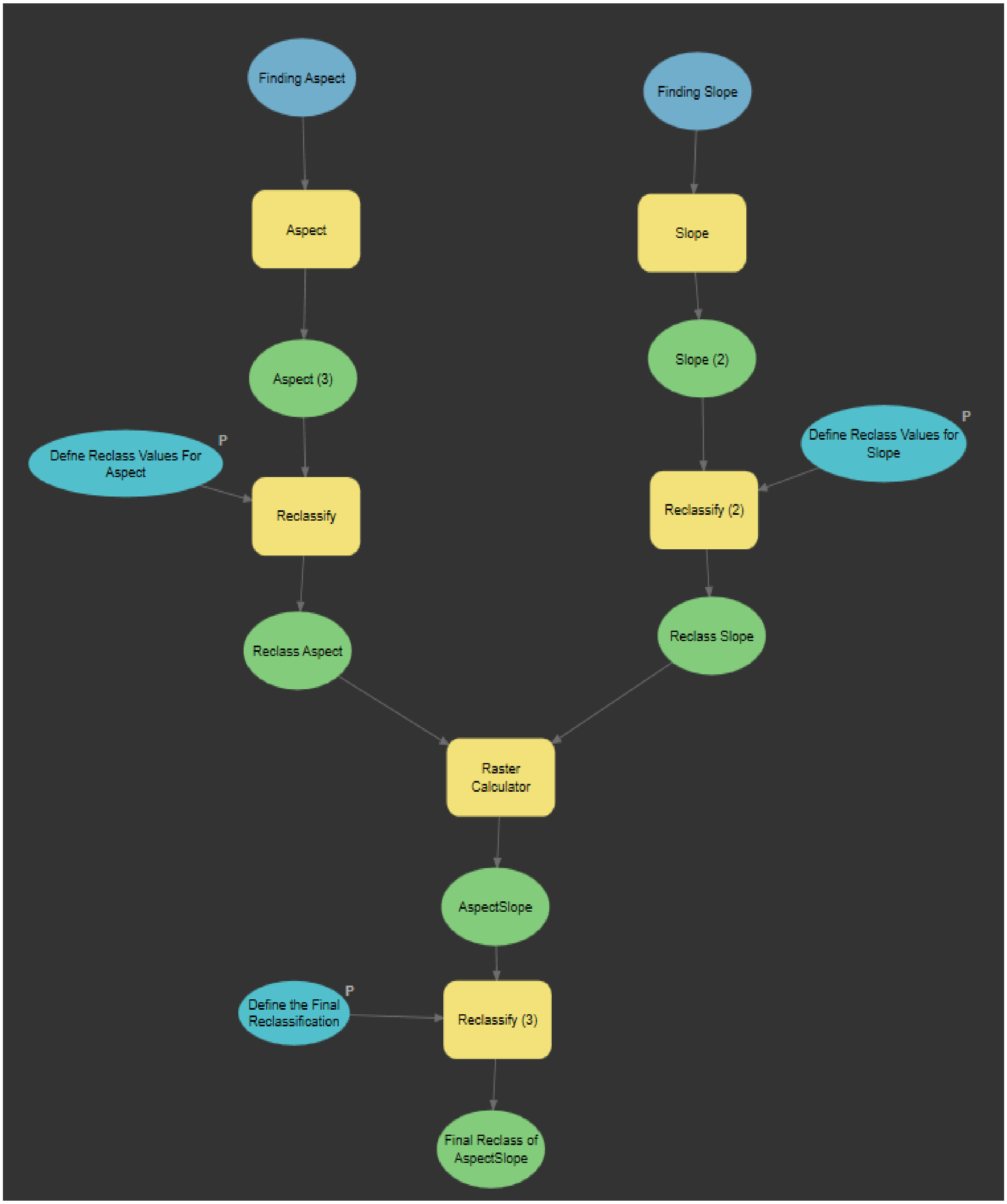


Figure 17 Aspect and Slope

To define Aspect and Slope (Figure 17) a DEM was used to create two new layers: Aspect and Slope. These layers allow the paleontologist to define whether the slope will be measured as either percent rise or degree (Figure 18a). After using Raster Calculator to add the two layers together to create a single layer, the new layer is saved to the Field Collections geodatabase to be used for analysis.

If the suitability model *High Potential Area* is used, then the second version of the Aspect Slope model should be used. Figure 18 continues from the newly created AspectSlope layer and allows the paleontologist to perform reclassification of the layer using the interface seen in Figure 18b. Reclassification allows for each cell value to be reclassified to a specific value, in this project's case either a 1 for "suitable" or 0 for "not suitable". By having the paleontologist assign the unique values of the layer with either a "1" or a "0", the newly created layer will be depicted with each cell value defined by the reclassification identifying suitable and not suitable regions.

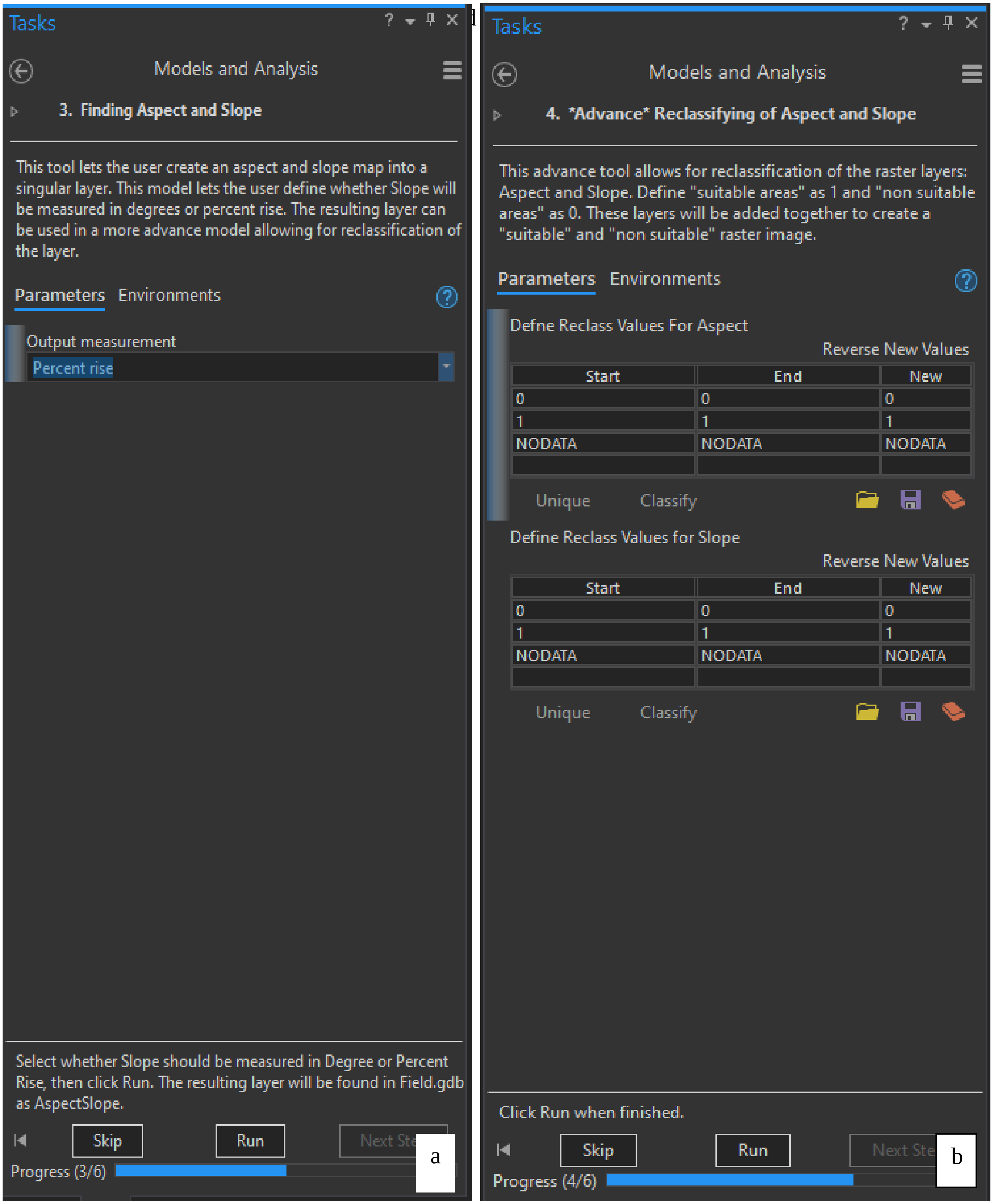


Figure 18 Defining and Reclassifying Aspect and Slope

3.5.5. Reclassifying Land Cover

Land cover defines the vegetation over a region and provides a classification for each cell that is related to the vegetation found within the area. Dense vegetation makes it impossible to discover fossils. Reclassifying these values to a range of 1 to 5 can simplify the identification of regions most likely to be suitable for the discovery of fossils. The resulting reclassified layer undergoes a second reclassification to define values 1 through 5 as either a 1 or a 0 in order to fit the classification of 1 being “suitable” and 0 as “not suitable” for identifying fossils.

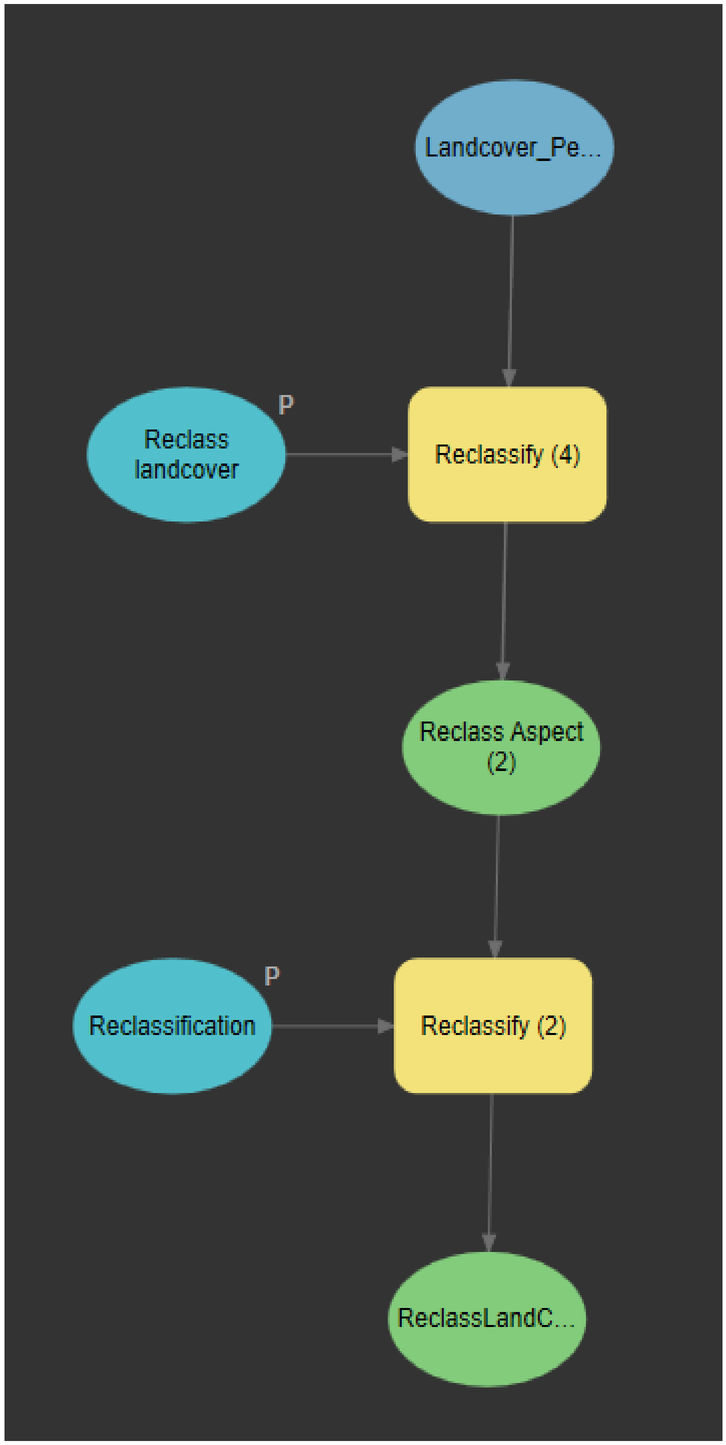


Figure 19 Landcover Reclassification

Figure 19 shows the workflow designed for the creation of the land cover reclassification model. The land cover layer allows the paleontologist to take the existing land cover layer and reclassify the cell values to their preference. For this project, each of the land cover types have been reclassified using a one to five system to identify regions from unsuitable to highly suitable.

Allowing the paleontologist to classify and reclassify the land class layer, means they can define regions to their own classification. Thus, it allows them to have more power over the final product and its use within the *High Potential Fossil Suitability* model.

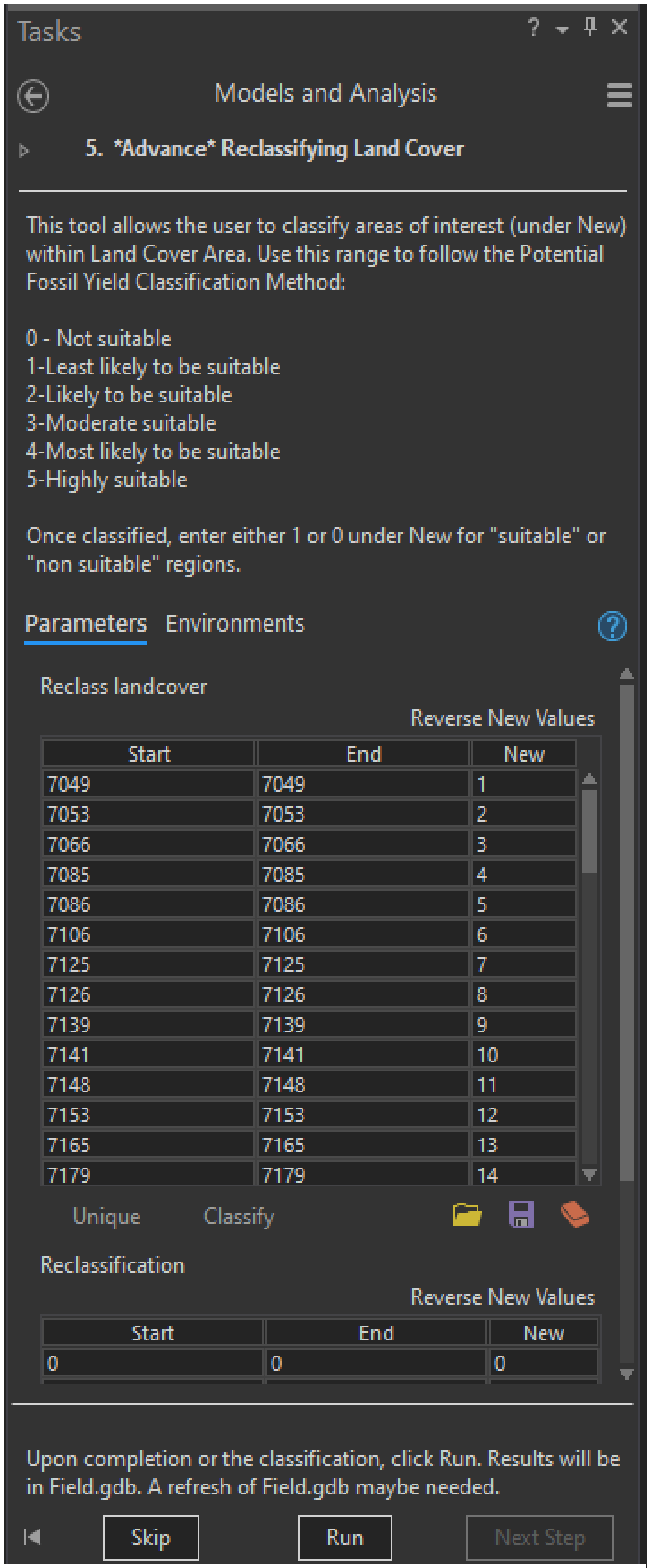


Figure 20 Land Cover Reclassification in Tasks

There are two parameters assigned within the model to allow the paleontologist to perform both of the reclassification processes needed to prepare the layer for the final model. The first parameter, identified as “reclass land cover” in Figure 20, is present as the first set of values within Figure 20. This process has already been reclassified for the paleontologist using the values between zero and five, as stated in Figure 20, and was listed earlier in this section. If the paleontologist wishes to change the reclassification values instead of using the preset values, they are free to adjust the reclassification to their preference.

The last parameter depicted in Figure 20 is the second reclassification of the land cover layer. This step is necessary if the *High Potential Fossil Suitability* model is going to be used. This reclassification takes the previous five reclassifications of each land cover feature and reclassifies them again using the same “suitable” and “not suitable” values that were used with the previous model.

The *Define Paleontology and Geology*, the *Define Roads and Hydrology Buffers*, the *Define and Reclassify Aspect and Slope,* and *Reclassifying Land Cover* models can be run either by themselves or in conjunction with each other. However, when adding running with the next model, *High Potential Fossil Suitability Model,* these models should be run back-to-back as their resulting raster datasets can be found within the *High Potential Fossil Suitability Model.*

3.5.6. High Potential Fossil Suitability Model

Using the results from the previous models, the *High Potential Fossil Suitability Model* uses reclassification and weighted sums to insure cell values remain consistent between each layer and are weighted against each other to identify regions that have the greatest potential for finding fossils.

Figure 21 High Potential Fossil Suitability Model

Figure 21 depicts the results from the previous models with the added Public Lands feature class. Each of these variables undergo reclassification to assign “suitable” cell values as a “1” and “not suitable” cell values as “0”. Since the Land Cover and the AspectSlope models end with the second reclassification, these layers are left alone.

A Weighted Sum is used against the reclassified layers as well as Land Cover and Aspect and Slope. A weight of 3 given to geology, 2 to land cover, and 1 for the other layers. This weighted value gives the highest value the most weight so those regions have a higher probability to identify potential regions suited to the criteria given earlier. The resulting model is saved to the Field Collection geodatabase where the resulting analysis can be taken into the field work.

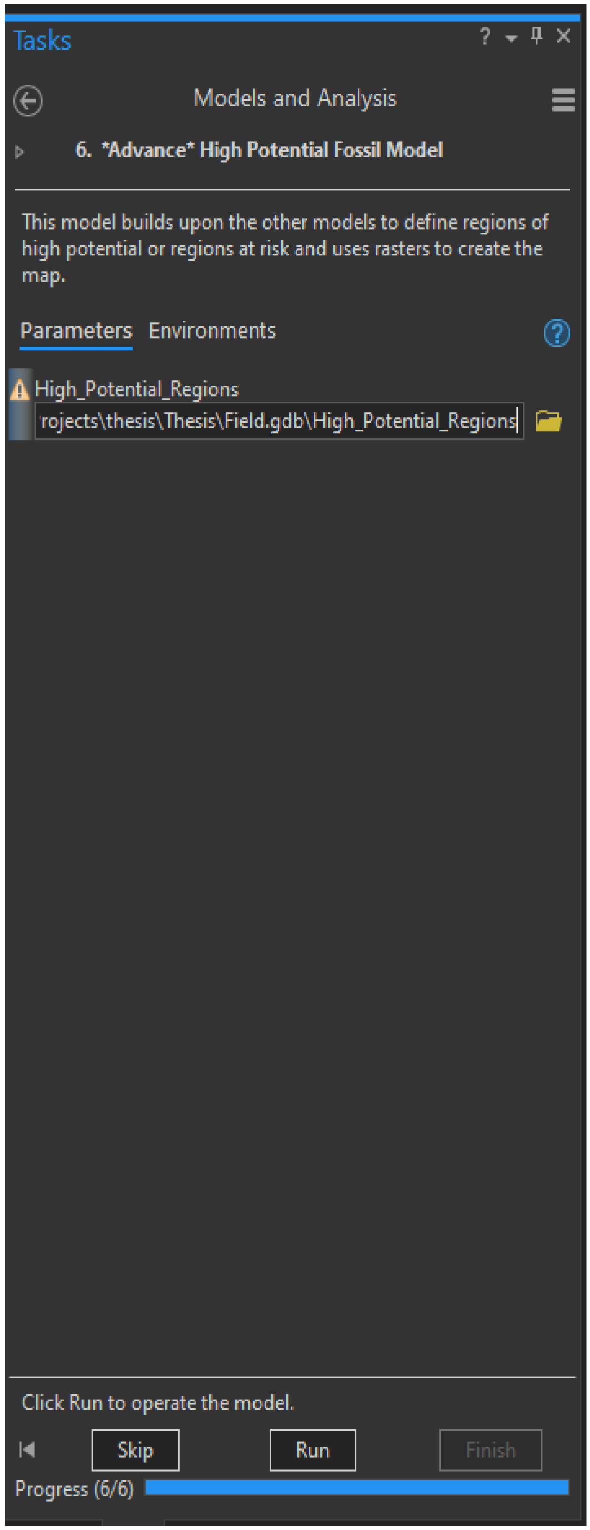


Figure 22 Saving the model

While there are no step for Tasks as the information entered into the model is reliant on the other models discussed within this section, the option to save the final results are depicted in Figure 22.

3.5.7. Summary

This section covered the models that were created as customized tools for paleontologists to use in conjunction with museum collections. From performing simple queries to running complex models, the purpose of these models is to assist the paleontologist with site assessment and management. These models have the ability to be updated using highly curated data collected from museum collections and offer the ability for the paleontologist to update the paleontological feature class through the use of Tasks, which will be discussed in a separate section. The following section will discuss how the parameters allow the models to be querriable and showcase the window created using ArcGIS Pro's Tasks Designer to create the interface, allowing for easy data entry and use of custom models created from the Museum Collections geodatabase.

3.6. Tasks

This section describes the interaction of the geodatabases and models through the use of Tasks within ArcGIS Pro. Tasks provides a unique interface with the data by allowing tasks to be assigned to direct the paleontologist through each step of the process. This gives paleontologists who are not familiar with GIS a simple way to perform tasks that are often left for GIS professionals. For this project tasks were designed for data entry, modeling for collections, field collection, and the geology and locality feature classes associated with the Field geodatabase.

3.6.1. Task Design Editor

Two separate task lists were created for data entry and modeling: editing and updating Field and Collections Information and Running Models. Within these two task lists are editable feature classes and models for the paleontologist to use to update information as well as perform analysis based on the information. Models are kept up to date since any new information being entered reflects changes and developments as new areas are being studied. Displayed in Figure 23 are the various editable steps that allow the paleontologist to add information pertaining to a survey or an excavation. The last pane called Models and Analysis provides an example of an interface to one of the models described in the previous section, allowing the paleontologist the ability to manipulate data and project the results on the accompanying map. Each of the models use an interface tailored to each model pictured in the previous section. This gives the option to run each model back-to-back or run each model separately to allow the paleontologist different levels of complexity.

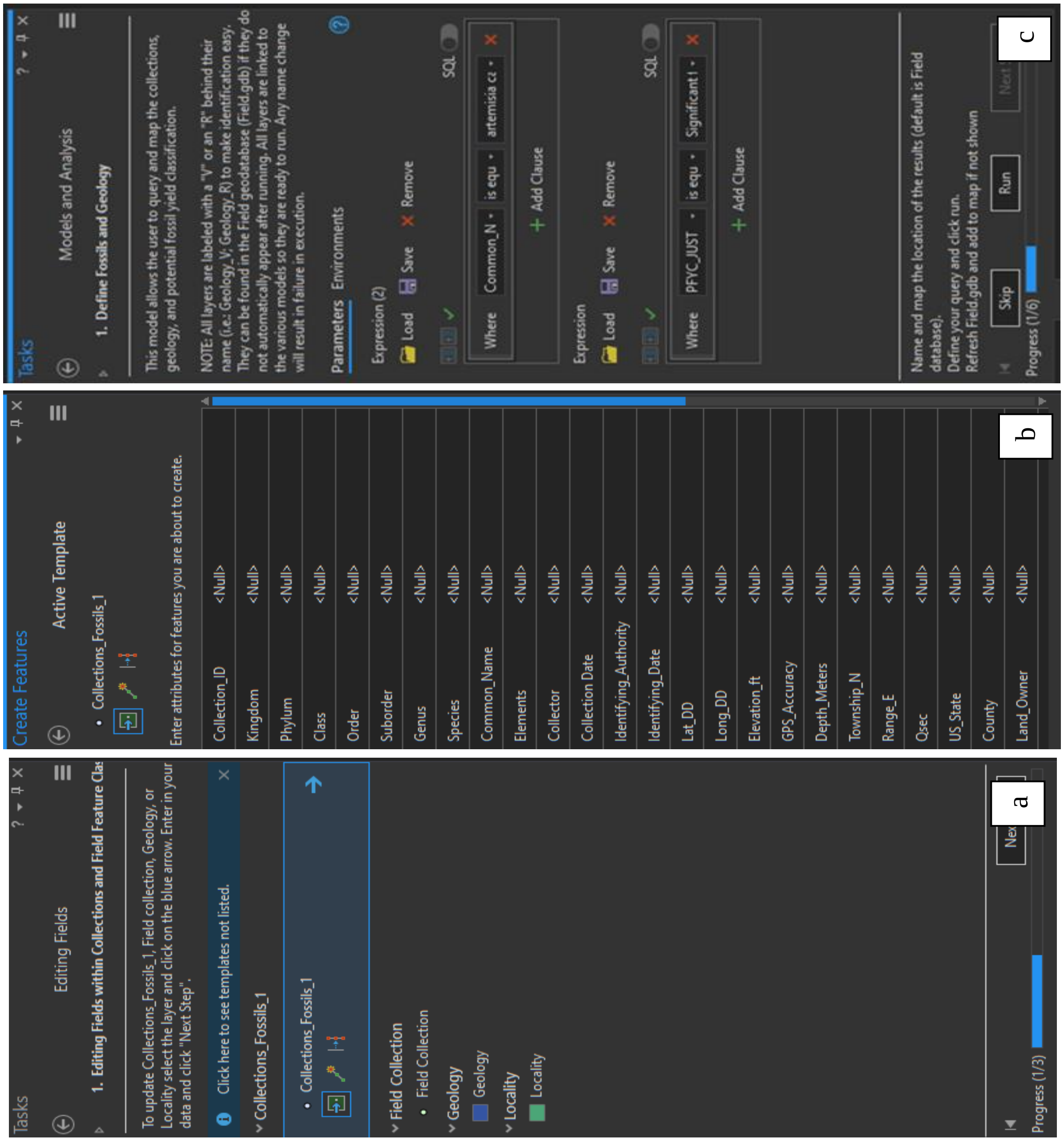


Figure 23 Tasks

3.6.2. Integrating Topology into Tasks

Topology rules described in the Topology section are entered into the task and run against the newly created dataset. If the information entered follows the rules, the topology rules will validate it and allow the paleontologist to continue. If there is any violation when the datum was entered the tool will prevent the continuation of data entry until the erroneous datum is corrected.

Creation of each task was performed using Tasks Designer. Within Tasks Designer, the recording of tools allows a tool to be linked with a task, allowing for easy access to a tool that is needed for a particular action. Tools selected for these Tasks were Editor and the models within the toolbox found within the Museum Collections geodatabase. Within the designer were directions, Actions, Viewability, and Editability as seen in Figure 24 and Figure 25.

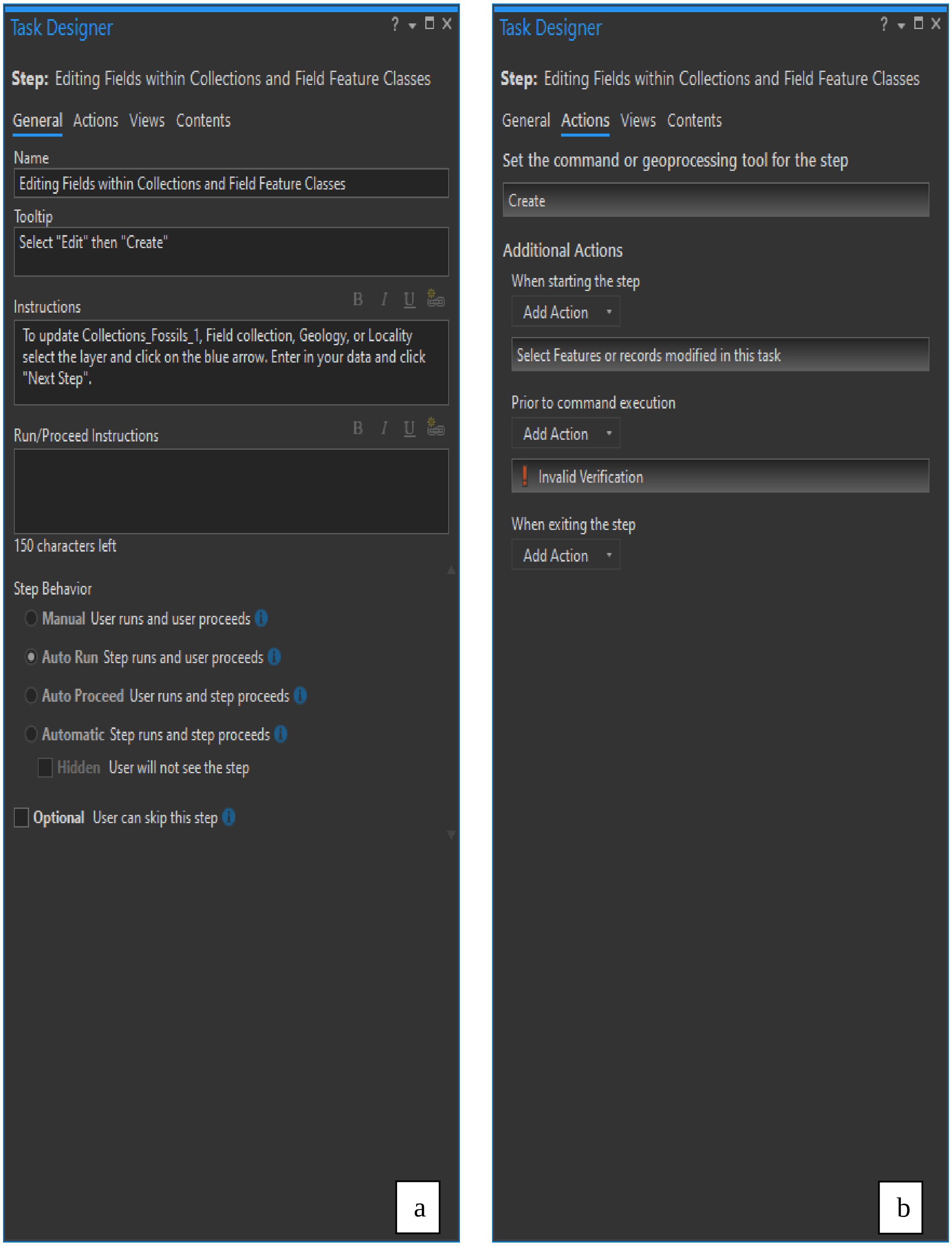


Figure 24 Task Template

Within Task Designer the task creator can determine which type of Action is required for the tool. The option to record allows the creator to record which command or tool will be selected for their purpose. Once the Actions and Contents have been selected the result will look something similar to Figure 24. Figure 24a shows the interface that the paleontologist will encounter first, which allows the paleontologist to choose which layer they want to edit first. Selecting the layer to be edited will take the paleontologist to Figure 24 a and b, depending on which set of tasks the paleontologist selects. Once completed the edited feature will be evaluated through topology, which has been designated as a hidden task. This can be seen as a choice in Figure 24a.

.

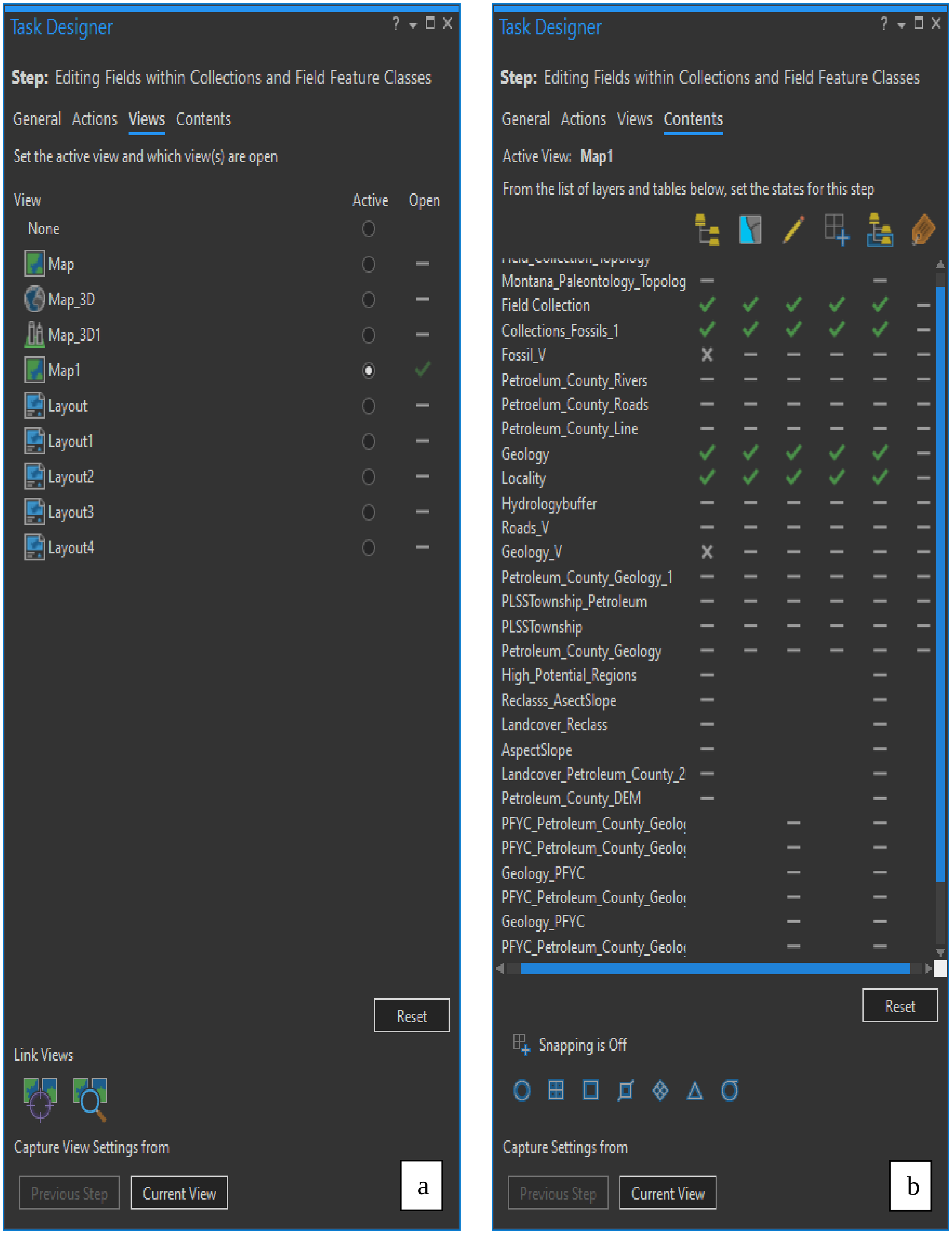


Figure 25 Task Designer

A hidden task allows a tool to be run without input or interaction from the paleontologist and run in the background once the step has been reached. In the case of the topology task, the task runs the tool Validate Topology which has access to the topology rules created for the Field geodatabase. As seen in Figure 24a additional actions, expressions, and commands can be set up to further define the interaction with the task. These commands or expressions allow for the task to be customized and either modify or create the selected task. Upon completion of these actions, the next step are views and contents.

As seen in Figure 25, the creator of the task can designate the window to view the results of the task as well as designate which layers are editable or not. Figure 25b shows the selection of which layers are selected as editable and viewable with the designation of a check mark. Layers that are not to be editable are either left alone or given an x to prevent the layer from being visible or editable within tasks.

Chapter 4 Results

This chapter describes the results of the geodatabases development and implementation of Tasks using the Science Museum of Minnesota's paleontological investigation of Petroleum County, Montana as a case study. Section 4.1 discusses the results of the geodatabases' construction and Section 4.2 discusses the application of the geodatabases to the case study, Petroleum County, Montana.

4.1. Geodatabase Design

A primary key is given to a field that holds a unique identifier within that field and is never duplicated (Yeung and Hall 2007). This key is used to define joins or relationships between datasets and are often depicted in Entity-Relations Diagrams (ERD). The following section will discuss the joins and relationships established between each feature class created for the dual geodatabase system. Section 4.1.1 will review the joins and relationships established for the Museum Collections Geodatabase and Section 4.1.2 will review the Field Collection geodatabases joins. Section 4.1.3 and 4.1.4 will review the results of each geodatabase.

4.1.1. Museum Collections Geodatabase ERD

An ERD was created to demonstrate the relationships found within the Montana Paleontology feature dataset found within the Museum Collections geodatabase. The ERD below describes the relationships between Paleontological Collections, Petroleum County Geology, and the PFYC through a series of joins which describe the relationships between each of the primary keys assigned to each table.

The Montana Paleontology feature class is a point feature class which contains accessioned data from museums and institutions pertaining to Petroleum County, Montana.

Petroleum County Geology feature class is a polygon feature class which contains data from the USGS *The State Geologic Map Compilation (SGMC) geodatabase Conterminous Geology of the United States* geodatabase which was trimmed down for the purpose of this study. The PFYC is a polygon feature class that was provided by the BLM and is made up of a compilation of geologic maps that were rated by each formation's ability to produce paleontological resources.

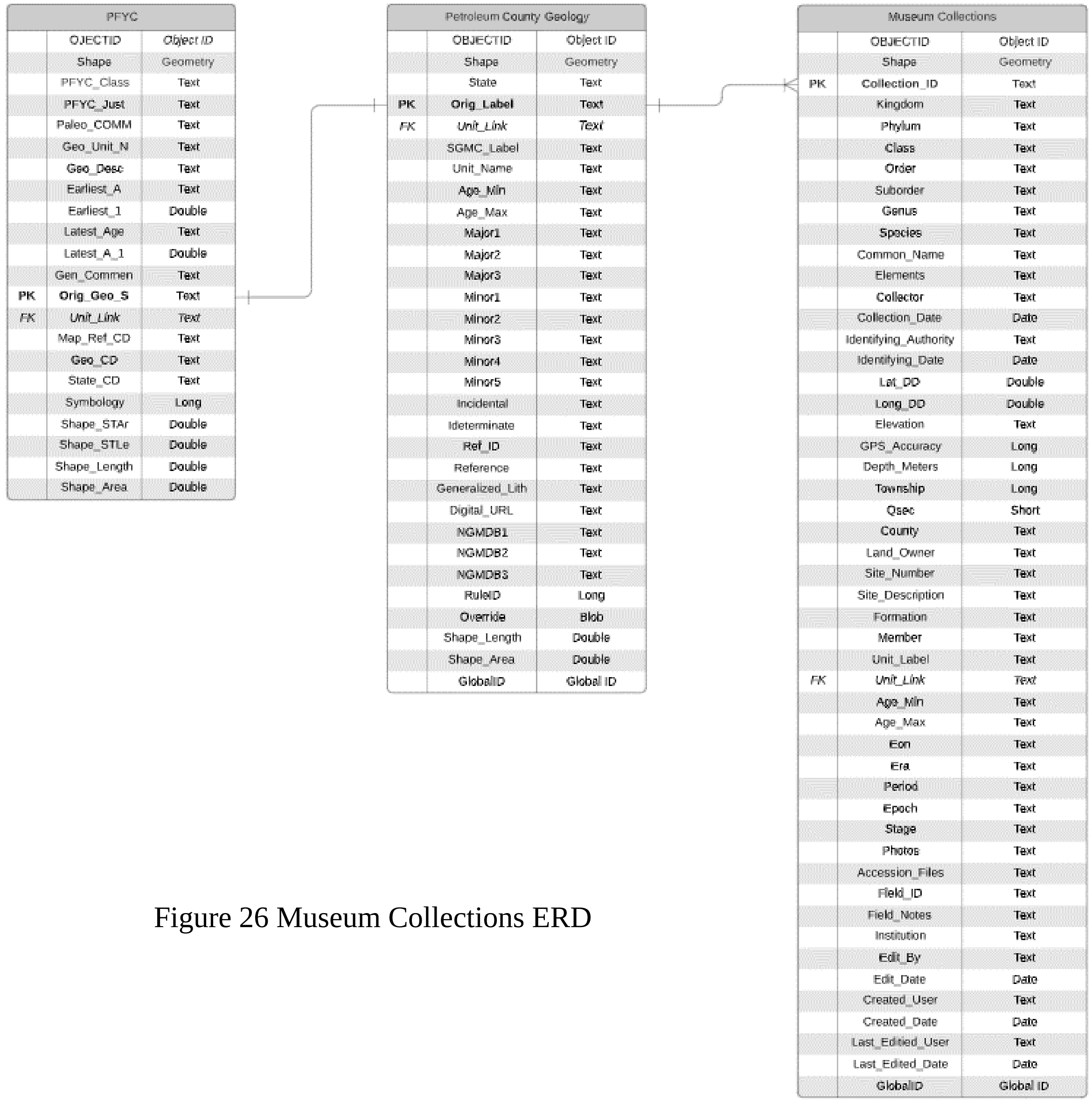


Figure 26 Museum Collections ERD

Museum collections require each collection to contain a unique identifier, which was used as the primary key for the Museum Collections feature class (Figure 26). In the case of the Science Museum of Minnesota's accessioning process, each paleontological specimen begins with a "P" to denote the paleontological department, the year the collection is accessioned, the next accessioned number for which paleontological specimens were accessioned that year, and the number of specimens cataloged within that accession (personal communication with Dr. Hastings, 2020). This primary key is given a many-to-one relationship with Petroleum County Geology (Figure 26) since many collections can come from a single geologic formation.

A relationship class between Petroleum County Geology and PFYC (Figure 26) was given a one-to-one relationship between Orig_Label and Orig_Geo_S. This join allowed for the Montana geologic formations to become classified with the potential fossil classification values: U is given to geologic formations where there is no evidence to support whether or not fossils are present; 1 is used for geologic formations that have a very low potential of containing fossils; 2 is used on geologic formations where the presence of fossils is very low; 3 is given to formations that have moderate potential for being fossiliferous; 4 is given to formations that have a high potential for producing fossils; and 5 is given to geologic formations that have an overabundance of fossils present.

4.1.2. Field Collection Geodatabase ERD

Creation of the Field Collection geodatabase was based on preferences given by the Science Museum of Minnesota paleontologist along with the BLM and NPS geodatabase schemas. Three feature classes were created within the Field Collection Feature Dataset: Field Collection, Locality, and Geology, in order to assist with field data collection and surveying.

An Entity-Relation Diagram was created to show the relationship between each attribute and which fields were used for the joins. The primary key that was used to join each of these feature classes together was the site identification number. A site identification number is given to each new site or locality and links all related data collected to a particular site. For this purpose, the Site_ID field was defined as the primary key.

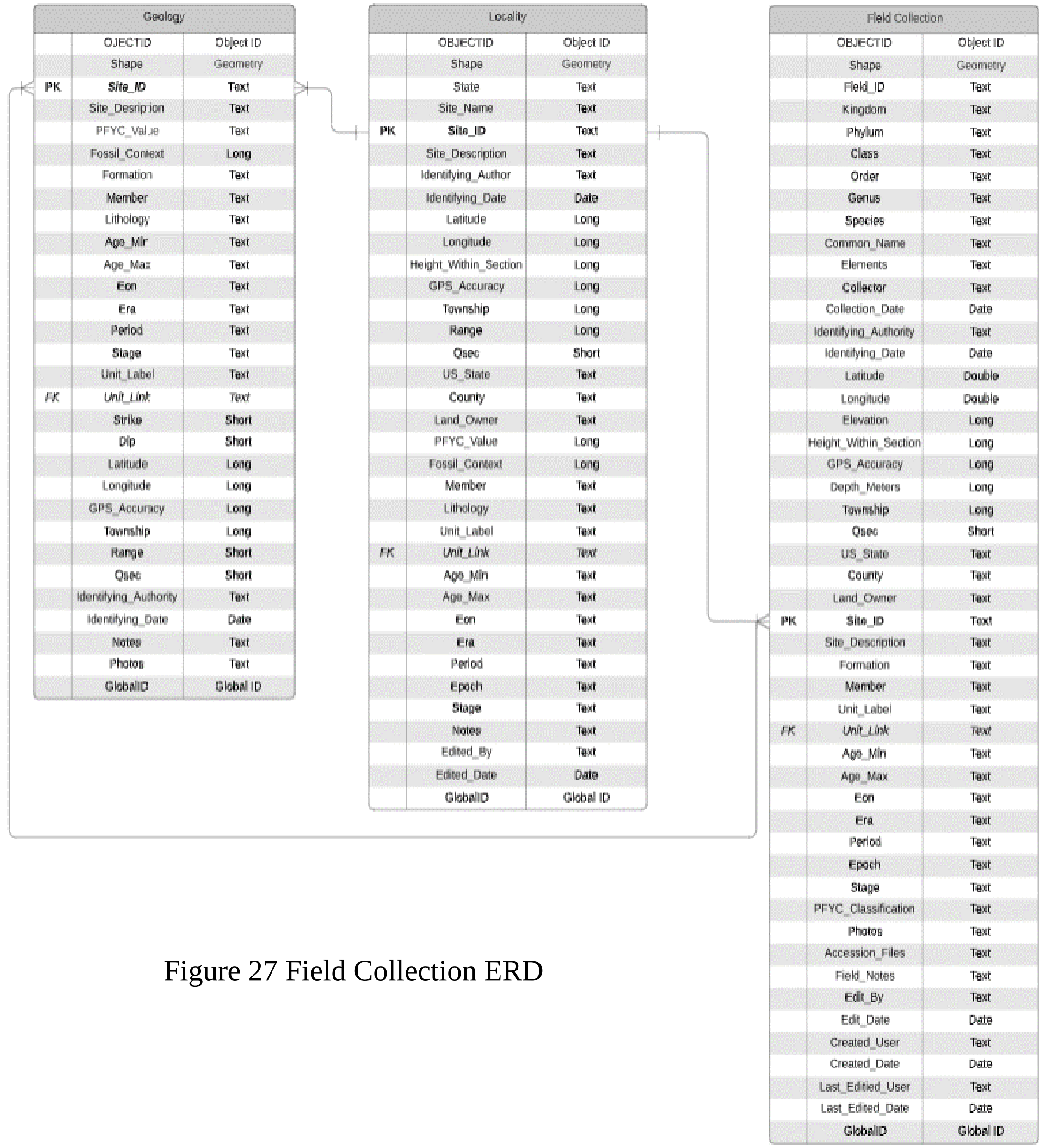


Figure 27 Field Collection ERD

.

The Field Collection feature class, diagrammed in Figure 27, breaks down the fields, data types, and relationships between Field Collections, Locality, and Geology. While Field Collection is designed for fossil collecting, Locality and Geology were designed to be multipurpose. The Field Collection feature class follows a similar design as the collections feature class found within Montana Paleontology Feature Dataset in the Museum Collection's geodatabase. Slight differences are present within Field Collection to accommodate for initial field analysis. Certain fields were altered or deleted to accommodate these differences

A Site Identification number is given to each survey, boring, or excavation performed in the field. For this purpose, Site_ID was used as the primary key for each of the Field Collection geodatabase's feature classes. When identifying potential locations for fossils, a location is defined by the geology, the geologic formation's potential, site sensitivity, abundance, and ability to produce paleontological resources. Since a single locality can contain either a single fossil or many fossils for it to be worth further investigation, a one-to-many join was created between Locality and Field Collections.

Geology plays a large part in producing fossils since specific environmental, mineral, and depositional conditions are required during the deposition and burial of the plant or animal. Paleontological resources found within geologic formations can range from no fossils to an abundance of fossils. If this case study focused on a region where this was the case, a zero-to-many join would be given. However, due to Petroleum County's geology, a join of one-to-many was created between Locality and Field Collection feature classes. This was based upon previous research and the PFYC classification of the region. A location must contain at least one geologic formation within each location, but a location can cover multiple geologic formations as each fossil is discovered and cataloged.

Geology and Field Collection feature classes were given a many-to-many relationship. The Geology feature class for the Field Collections geodatabase was created for the purposes of surveying and excavation, so it is to be expected that only potential fossil-bearing geological formations will be entered into this feature class. Therefore, each formation can contain one or more fossils and multiple fossils can be found in different formations.

4.2. Museum Collections Geodatabase

A total of 1,043 specimens were compiled, cleaned for importation into the new spreadsheet, imported into ArcGIS Pro, and projected into the appropriate projection. Once the test of the data was successful the feature class was created. 846 of the specimens were left out of the importation into the feature class so they could be used for testing the Editing Fields Tasks. Out of the 846 records, 42 of those specimens required georeferencing due to the age of the record. This provided the opportunity to test the Tasks' ability for georeferencing.

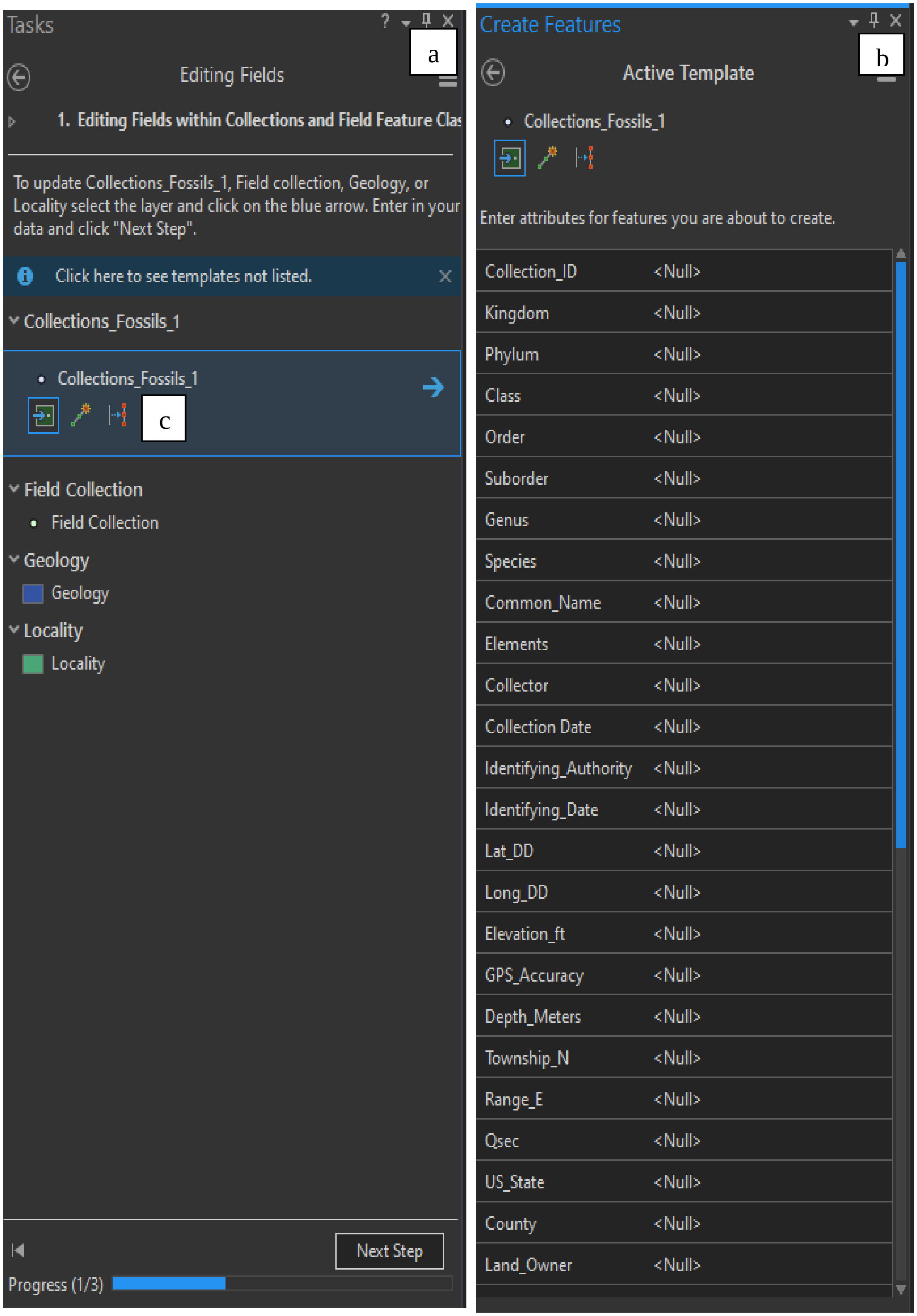


Figure 28 Tasks Edit Fields

Each entry was entered through the Tasks interface created for this project and the entries are depicted in Figure 28a and b. The first interface to be presented is shown in Figure 28a, which allows the paleontologist to select which feature class to edit.

For this project, the Museum Collections geodatabase is the primary file geodatabase which stores museum records and was the one selected for the 804 records. Either a double click or selection of the blue arrow will open up the editing field for the feature class. Figure 28b shows the interface presented to the paleontologist for the collections feature class found within the Museum Collections geodatabase. When selected, each feature class will display their attributes created from the geodatabase schema seen in Figure 28. Each field was filled out for each of the 804 records that contained coordinate information. 42 were missing legacy data that required georeferencing and this georeferencing was performed with the Create-A-Point tool shown in Figure 28c. The Create-A-Point tool allowed for the data to be entered into the same pane as seen in Figure 28b. Upon completion a point was created and had to be placed using the field information given at the time. Since documentation was not provided to perform a more accurate georeferencing of the legacy data, future work in obtaining the documentation is necessary to correct these 42 legacy points.

4.2.1. Data Integrity

Georeferencing requires an extensive amount of time to manually plot each record correctly. The information provided within these records only contain so much due to either the PRPA of 2009 16 U.S.C. § 470aaa 1-11 or lack of information attached to the record. These 42 legacy records were not the only problem noted within the records provided. After further review of each record it was noted that there was a broad range of error with the accompanying spatial information.

Data accuracy for the museum paleontological information is charted in Figure 29 to depict the problem confronting mapping museum collections. The furthest bars on the right indicate NULL values within the dataset that are associated with records missing any correction. The second largest accuracy problem is presented on value 836, which represents 168 individual records of *neogastroplites mulleri* whose coordinate information has been rounded to four decimal places. Review of the spatial information shows that latitude and longitude for each record range from being rounded between four to six decimal points.

Rounded coordinate information also poses a problem for performing analysis or accurately mapping field or museum data. When coordinate information is rounded, this presents a problem when performing analysis or accurately depicting information related to each site. This can mean the difference between centimeters to kilometers from the actual site where the fossil was collected.

Sites such as iDigBio and GBIF have participating museums and institutions link their sites' Darwin Core with the museums' or institutions' databases. This requires participating museums and institutions to follow the standards of the TDWG, or the ABCD, to maintain the data integrity that has become standard for biodiversity informatics while also maintaining their compliance with their country's laws on the preservation of biodiversity or paleontological specimens. For the United States this law would be the PRPA of 2009.

Unfortunately, there are no standards for the accuracy of spatial information in terms of extent to prevent the problems associated with those 168 records or the NULL values represented in Figure 29. This is but one of the issues being addressed by the TDWG and the ABCD that are still in development and should be standardized within the field of paleontology as geodatabases and GIS become more popular.

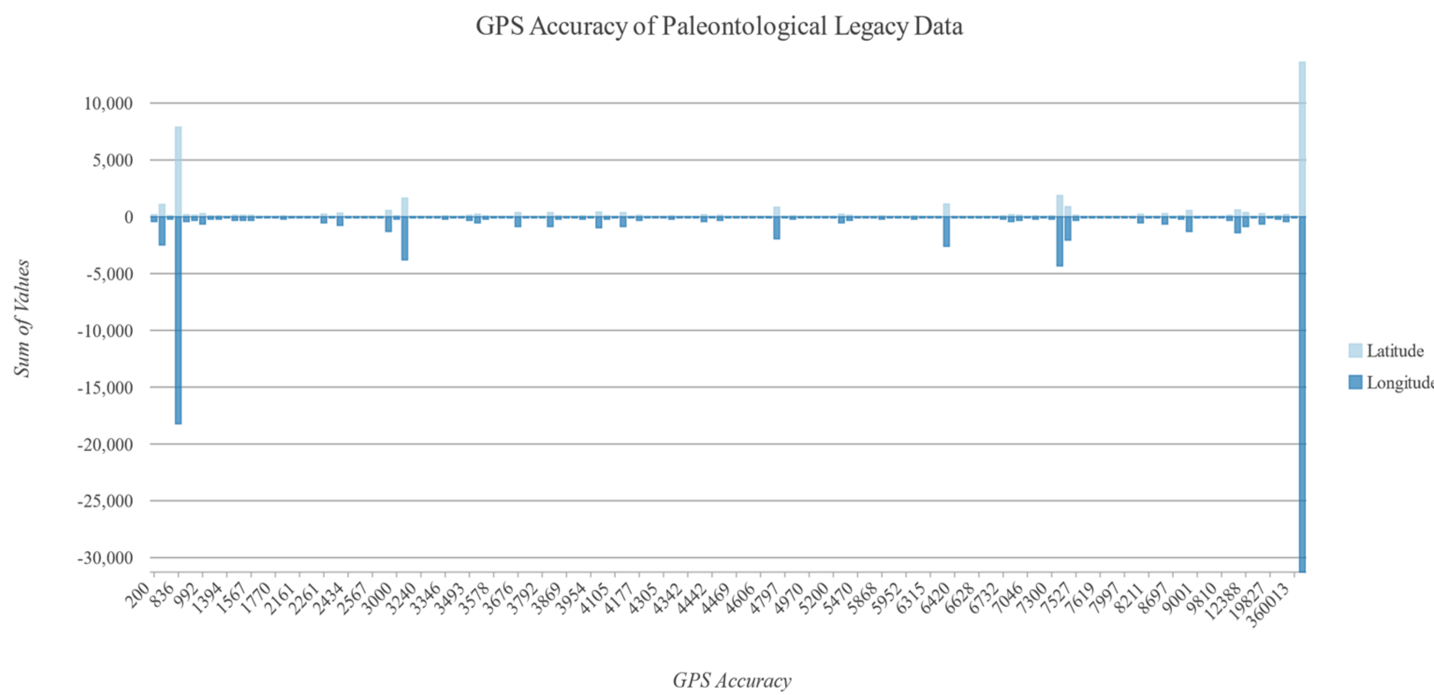


Figure 29 GPS Accuracy

4.2.2. Models

Investigating new sites or making plans for reinvestigating older sites often requires hours of detective work. With the creation of the collections feature class, integrating museum collections into these investigations can provide a unique perspective. For this purpose, a toolbox with customized tools created using Modelbuilder was used to assist with applications for paleontologists who are not familiar with GIS. The construction of the models was broken down in Chapter 3 to showcase the process.

Presented within the toolbox are a series of tools ranging from simple querying to complex suitability modeling. Easily accessible through Tasks, these models can be run separately or as a series. This can be seen in Figure 30 as well as Figure 31, 32, and 33, and also

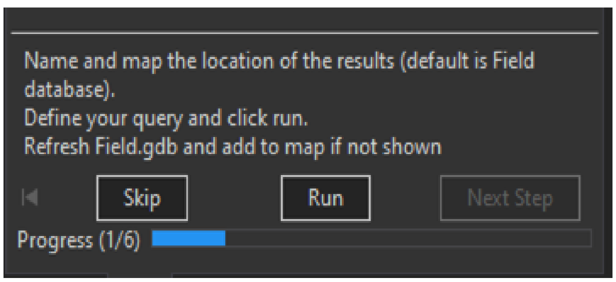


Figure 30 Navigating Between Steps

in Chapter 3, which shows each model's interface. At the bottom of each Task are the options to "Skip", "Run", or "Next Step," as well as a progress bar to tell the paleontologist which step they are on.

Complex models require classification of each cell to create the suitability model for the High Potential Fossil model. Figure 31 is a reclassification of the Land Cover for Petroleum County following Oheim (2007) weight scale, which favors a granting a weight of 5 to bare earth with the highest potential of finding fossils and to regions where the potential of finding open

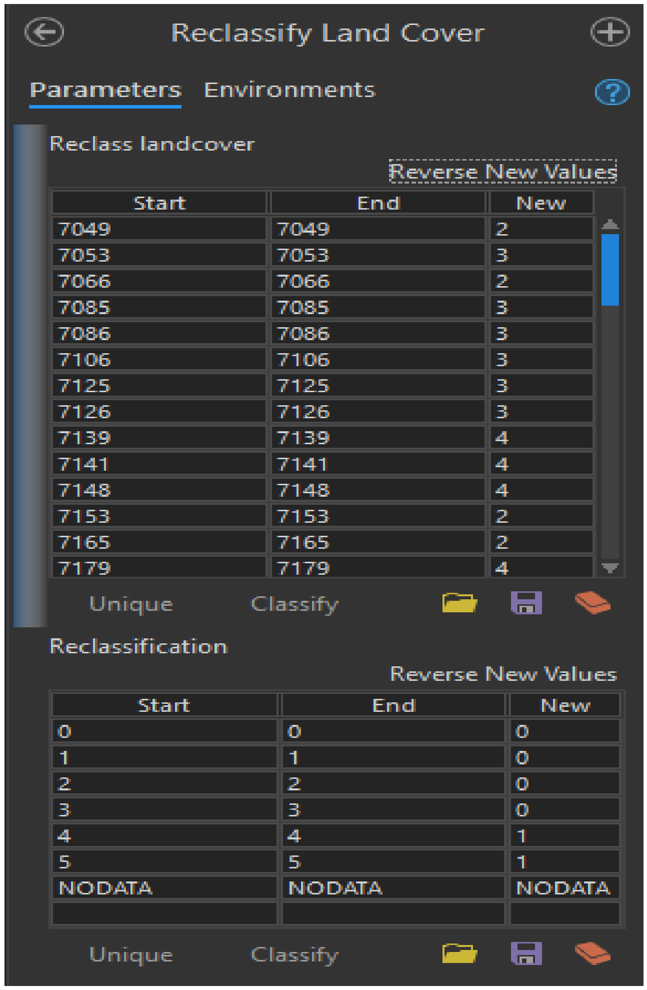


Figure 31 Reclassification Values

ground is present. A value of 1 is given to places such as developed areas or waterways with a 2 to 4 is given to vegetation coverage with low to moderate potential for finding fossils, as it is less likely or there is minimal exposure.

For the complex model that combines the results of all the models into a single model there is the High Potential Fossil Suitability model presented back in Figure 21. This model uses the same weighted value system as the reclassified land cover model for weighting each input value and combining each raster into a single raster image. Figure 32 depicts the weighted sum values applied to the model.

.

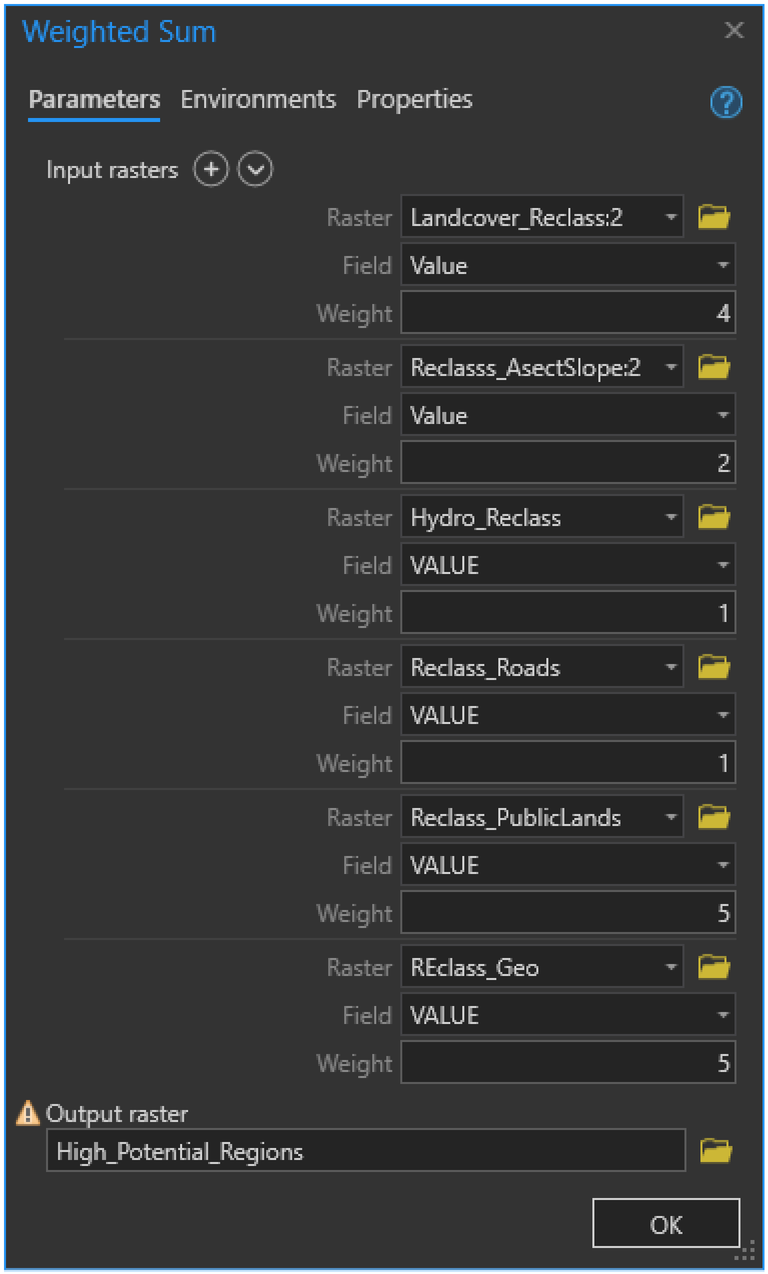


Figure 32 Weighted Sums

Figure 32 depicts the weighted values defined for the *High Potential Fossil Suitability* model. Classification of the values is based on Oheim (2007), where geology and land class were weighted higher than roads and hydrology. A value of five is given to the highest weighted value of the layers that have the greatest importance to fossil exploration. For the case study, the highest values were given to Public Lands and Geology and the lowest values to the road and hydrology layers. Permits for the case study focus on public lands for the Hell Creek Formation, so geology and Public Lands were weighted at five with land cover being weighed at four. Land cover describes the coverage of the region and was reclassified to highlight regions with the highest potential for finding fossils.

Depending on the results of the previous model, a new map is easily created. This saves time, since the traditional model will take hours to reconstruct and run even as variables are changed. The interactive nature of these models and their ability to build upon one another and add the result to the next model or run separately as individual models make this Task feature unique and extremely adaptive to the paleontologist's queried hypothesis.

4.3. Field Collection Geodatabase

The main focus of this chapter is the Field Collections geodatabase housing the museum paleontological records, but a note about the Field Collection geodatabase should be made about the basic functions this geodatabase is required to do when brought into the field.

4.3.1. Field Collection Feature Class

The Field Collections feature class allows for the paleontologist to record each fossil discovered during an excavation. This feature class has similar attributes to the Montana Paleontology feature class found within the Museum Collections geodatabase. Changes within the attributes were designed so that the Field Collection feature class would be used for the initial documentation of the fossil.

Photo documentation of the site is one of the processes for documentation within the field. This feature class has been constructed to allow for photo documentation and attachment of accessioned files. This will allow for the record of the fossil and its accompanying documentation to stay in one place from collection to accessioning.

Topology rules created for this feature class dictate that all fossils must be contained within a geology polygon. As each fossil is required to have location and geologic information entered, these feature class topology rules are primarily used to assist in accidental data entry or plotting.

4.3.2. Locality Feature Class

Localities are represented as polygons which require the user to define the locality by drawing a polygon. Topology enforces that each location should contain at least one fossil as well as not overlap. Test runs of this feature depict an outlined polygon when accepted by the topology rules. If the polygon accidently overlaps with another polygon, they will turn red, preventing the user from being able to save the polygon unless this is corrected. This result can be seen in Figure 10, where one site is accepted and the other is red. While not visible, the accepted polygon contains a single, recorded fossil. The fossil was turned off to protect the site.

Similar to the Field Collection feature class, Locality also allows for the attachment of photo documentation and accessioned files. This is to maintain each record associated within the locality and the fossil sites contained therein.

4.3.3. Geology Feature Class

Petroleum County is found on the western side of Montana and is the focus of the Science Museum of Minnesota's research into the Hell Creek Formation. The last recorded geologic survey was written by Lindahl (1993) presenting a Soils Survey of the county. Quadrangle maps created by the USGS were last updated in 2007 and can be found in Vuke et al. (2007) publication on Montana's Geology.

The maps constructed for the USGS geology layer have a scale ranging between 1:50,000 to 1:1,000,000 and the PFYC geology shape file is a combination of differently scaled quadrangle maps. These maps were a factor in creating the relationship class for the Petroleum County Geology feature class within the Museum Collections geodatabase. The problem with scale regarding geology was one of the discussions that involved the NPS. Geologic maps for fine scales are often constructed during surveys performed for the region. A Geology feature class was created for the Field Collections geodatabase to allow for surveying.

The Geology feature class is a polygon feature class that records the results of surveys and allows the paleontologist to create a new polygon to map the extent of an outcrop. This will allow for the creation of a local scale geology map of the case study. As with the Field Collections and Locality feature classes, photo documentation and any documentation associated with the geologic survey can be attached to the record.

4.3.4. Summary

1,043 records containing information on Petroleum County's paleontological resources were created in the collections feature class for the purposes of integrating museum collections in order to assist site planning, assessment, and management. Ten percent of the data was withheld to test the Tasks "Editing Field" feature where 42 out of 846 records required georeferencing within the Museum Collections geodatabase's feature class. Georeferencing the 42 records was limited to the locality information provided for the individual records, along with their accompanying publications.

Tasks also utilizes the Museum Collections geodatabase toolbox containing customized, interactive models that allow the paleontologist to easily query against the collections, geology and the Petroleum County feature classes and datasets. These queries allow the paleontologist to define geology, which fossils they would like to investigate, and to create buffer zones around roads and rivers. Advanced models allow the paleontologist to create complex analysis with minimal knowledge of GIS while still using the necessary steps often used to perform suitability analysis for locating new sites or investigating older sites.

The field geodatabase was designed primarily for the field and constructed with future application with ArcCollector in mind and allows for the attachment of documents and photographs. These feature datasets allow for the uploading of images, surveying, or recording excavations. Topology rules were created to maintain data integrity, so no accidental entries are recorded into the feature classes.

Chapter 5 Conclusion

The dual geodatabase system was constructed to determine whether combining museum collections and the PFYC within GIS was beneficial for paleontological resource management, assessment, and inventorying. This chapter discusses the significance of this work, the strengths and weaknesses of this process, and describes future directions to better enable this project to assist in field surveying and collection. Section 5.1 describes the strengths and weaknesses of these geodatabases in their application for Petroleum County, Montana. Section 5.2 discusses the significance of this project to paleontological resource management. Section 5.3 discusses future applications for this project and section 5.4 concludes with a project summary.

5.1. Strengths and Weaknesses

The purpose of the dual geodatabase system was to integrate museum collections and PFYC with field data to provide paleontologists with a greater understanding of the paleontological resources within a specific region. This project was a case study focusing on Petroleum County, Montana, and was used to assess the viability and usability of the dual geodatabase structure for this purpose. The geodatabases were combined with accessible interfaces designed using tasks, which allow paleontologists to query information pertaining to a region's geology and to use legacy data gathered from various museums and institutions through online data repositories with permission from the BLM. These tasks were also designed to allow modeling data to assist the paleontologist in refining their hypothesis and/or proposed area of interest. Section 5.1.1 discusses the strengths of this dual geodatabase design and section 5.1.2 discusses the weakness of this approach.

5.1.1. Strength of the dual geodatabase system

Information held within museum collections relating to paleontological specimens contains a wealth of information which can benefit GIS applications. At present, analyses and studies focusing on museum collections using GIS are limited and slow to gain attention. Petroleum County, Montana was used as a case study to show the benefits of incorporating GIS using this method and presented a new method to incorporate the process of referencing museum collections with field data. Paleontological information pertaining to Petroleum County, Montana was gathered from open data repositories with permission from the BLM. One geodatabase was constructed to house the museum collection data along with the region's geological and paleontological classification system developed by the BLM. In the case of the Field geodatabase, field data was provided by the BLM and the Science Museum of Minnesota to test the field geodatabase data entry process as well as to assess the overall performance of how museum collections, field collections, and the PFYC can aid in site management practices.

Results of the dual geodatabase system were encouraging since the geodatabases and their models were updated as new information was entered. Interfaces using Tasks allow data to be entered easily without the paleontologist needing to know GIS. The process of using Tasks made data entry easy and accurately enforced the rules upon entry, thus preventing the need to correct errors later on. Tasks were also set up to perform modeling, and this was also successful and incorporated new data as it was entered.

Models offer a novel way to explore and interact with museum collections. Museum collections are static in nature and performing the necessary tasks for inventorying and the assessment of paleontological resources is time consuming. Models created for the Museum Collections geodatabase were problematic at times due to a scripting error. This issue was easily fixed by resetting the queries back to the original loaded query. These models pull from the museum information gathered on Petroleum County and allow the paleontologist to query based on the museum records and the region's geology. This process is unique to this study and was designed to be repeatable, saving paleontologists time. As datum is entered into the Museum Collections geodatabase the models will incorporate the information, thus allowing the development of increasingly accurate results to queries.

Records with excavated fossils can be entered into the field geodatabase until the fossil is ready for accessioning into the museum's collection. This feature is unique to this method as the dual geodatabase system allows for the paleontologist to maintain the fossils and their records within a single database, as well as attach documentation to the records as they are processed for the museum's collection. Once a record has been prepared for the museum's catalog the data can be transferred from the field geodatabase into the collections, where the record can be displayed and queried while maintaining both the record and the documentation attached to it.

Lastly, the dual geodatabase system was designed for portability. These geodatabases were meant to incorporate ArcCollector for future use by allowing the addition of documents, having editable fields accompany the domains, and allowing for the drawing of new features. The ability to bring this system into the field and perform data entry in the field information while referencing museum information were intentional parts of the geodatabases' design. As technology evolves, the ability to bring museum collections and their records into the field should become a necessary function

5.1.2. Weaknesses of the dual geodatabase system

As collection management policies change to reflect diversifying fields and standardize spatial data standards, it is necessary to reexamine legacy data and data collection standards. The

Petroleum County case study exemplifies this. Museum data compiled for Petroleum County only contains records with spatial references and the list is by no means exhaustive. Not only is this compilation of paleontological records not complete, but a large portion of the historic museum information was collected prior to the advent of GPS. The process of defining the records' coordinates is assumed to use georeferencing from field data, notes, and maps. However, there are inherent errors in these records.

Records collected during the era of GPS collection are not uniform with their spatial extents. Some of the data was intentionally rounded for iDigBio and GBIF databases, while other records suffered from rounding during collection and post-processing. In the case of the records of the 168 ammonites, it is evident that they were rounded for the sake of simplicity. These records contained the same numbers for latitude and longitude. Each value was rounded to four decimal places and these all were identical. When mapped, these 168 records were projected into a singular point containing each of the 168 records, instead of 168 individual points with individual records for each location. Failure to generate individual points for each record due to the rounding issue hinders the models. These records will need to be fixed to help balance out models or other analyses. As museums digitize their collections, care is needed for the collections with more than one specimen recorded within the same location.

While rounding is an issue with this dataset, there are other issues involving the projection of fossil data. Some fossils were found within a lake or outside of the formations they were recorded in. These issues could be attributed to projection problems, poor data collection, or poor georeferencing of legacy data. iDigBio and GBIF convert their spatial data to WGS84, which needs to be converted in order to have any meaningful analysis. The BLM dataset also went through a projection conversion. Collection managers also expressed concern about the accuracy of the legacy data, hence the questioning of the georeferencing of the legacy dataset.

Incomplete datasets that include null values and inappropriate projections limit the efficiency of the dataset and its accompanying geodatabases. This limits the ability to query information accurately and affects the ability of performing analysis. In the case of Petroleum County, the dataset contains these problems but also offers the unique ability for correction. Research into the Hell Creek area within Petroleum County offers paleontologists the opportunity to validate, update, and build upon museum paleontological information, and this benefit on the dataset will be ongoing. Future field validation and georeferencing of the legacy datasets will increase the geodatabases' ability to assist the paleontologist in the evaluation of sites.

5.2. Significance

In the case of the BLM and NPS collections, all fossils found on public lands are property of the United States, whether they reside within onsite collection facilities or are housed in museums and institutions across the country. Traditional methods of performing site assessments, inventorying, and management involve investigating each specimen collected in a region, comparing it to past documentation and field investigations, referencing geologic formations, and comparing past records with the collection records. This process involves referencing each specimen and investigating each specimen's locality and the collection where it originated. The NPS and BLM paleontological inventorying process requires an extensive amount of time to track each collection.

A geodatabase system that houses museum collections' information and contains an accompanying field collection geodatabase may help better organize these collections. This is beneficial whether used on a large-scale basis for federal inventorying of paleontological resources or for museums to aid in their management practices. These geodatabases can be used to analyze museum collections to better understand the paleontological resources within a potential area of interest or within a region. This allows for the implementation of better site management practices. Whether it is used in the office or in the field, the dual geodatabase system allows museum collections to become portable. Compared with the PFYC, it can bring a new perspective to management practices.

Not only can the dual geodatabase serve as a separate tool, it can be used to accompany traditional inventorying processes or used in the field. These geodatabases and their models can be used in conjunction with current management practices either for reference or for performing field analysis to determine the potential success of an area of interest. Performing traditional inventorying with physical collections and doing reviews of the geodatabase will ensure data is accurate and complete, verifying both repositories while new information is added or changed.

Data integrity rules were established to follow museum data standardization practices for digitizing collections, as well as to follow the permitting guidelines of the NPS and BLM while using domains to follow the USGS descriptions. Attribute tables were designed to match fields found within the museum collections database, topology upholds the rules created to enforce data integrity, and domains simplify data entry. This ensures data entered within the geodatabases will be the same as the museums' databases and will simplify reporting back to the NPS and BLM. Topology rules will ensure that accidental recordings will not occur, thus saving the paleontologist the need to correct the datum later. Having fields with specific terms designated for specific fields, such as "Formation", will only allow for the paleontologist to choose from a list of formations or "Lithology." A list of descriptors allows for uniform data entry and will ensure that nothing is missed when queried. This simplifies data entry processes and can help limit redundancies. Not only do the data integrity rules allow paleontologists to keep track of the fossil and its records from field collection to accessioning into museum collections, but it ensures that information is entered accurately. This makes it easier for reporting on permitted collections as the field geodatabase is designed for this process. Information collected within the database houses some of the same information requested by the BLM and NPS when information on associated permits are entered after field collection.

5.3. Future Work

5.3.1. ArcCollector

Portability was a priority while designing and executing this project. Workflows and field collection methods are increasingly moving to mobile devices; thus, these geodatabases were designed with ArcCollector in mind. The NPS and BLM have developed applications for performing field surveying and documentation using ArcCollector to inventory paleontological resources found within public lands. However, the inventorying of museum collections is still performed via the traditional method and is not incorporated into ArcCollector maps. The purpose of the dual geodatabase system is to allow museum information and field information to be displayed simultaneously while in the field. Offering paleontologists in the field the use of museum collections alongside field collections with up-to-date information can also assist with on-site management practices. Along with the PFYC, paleontologists can reference the geodatabases with their notes to develop more accurate on-site information. As information is collected from field surveys or excavations using ArcCollector, the geodatabases will store the information and allow paleontologists to view information in both the office and the field.

5.3.2. Project Continuation

Due to the nature of the data needed for this case study, the geodatabases were not designed to pull directly from museum repositories. With the dual geodatabase system, future work connecting museum collections and geodatabases will provide direct access to museum repositories. This system could work with the entirety of the museum's paleontological collection instead of just Petroleum County, Montana. These geodatabases were also set up with global IDs to allow for the option to sync these geodatabases with a geoserver. Once established they can be set up to pull information from each other to reduce the amount of data entry needed.

Further development of this project is essential if it is to benefit the field of paleontology and further work is necessary to ensure the success of the dual geodatabase system. For these reasons, teaming with a museum will ensure the continued use and development of this project. Since the case study was provided by the Science Museum of Minnesota the project will be turned over to them upon completion. Their paleontologist will be the primary user for the dual geodatabase system and will assist in field testing the project to help identify and fix any problems with the system. He has also agreed to allow the author of this document to stay on the project and assist in its maintenance so that development can continue and benefit the paleontology department.

5.4. Summary

Past documentation on how GIS was used to assist with paleontological resource management and inventorying has been assembled slowly, and also offers little information about how other organizations and individuals developed their workflows. This may be due to the slow nature of digitizing museum collections. The goal of this thesis was to create a dual geodatabase system that would incorporate museum data, field data, and the PFYC in order to assist paleontologists with site management, inventory, and assessment.

The case study, which focused on Petroleum County, Montana, involved the creation of the geodatabases and the assessment of whether this application of museum collections and the PFYC would be suitable for such an undertaking. The results suggest that further research into this application of museum collections is well worth pursuing. The dual geodatabase design allows for paleontologists to maintain a complete chain-of-custody method with the site and its fossils, from field data collection to accession into the museum's collection. The design of the databases allows for easy data entry, topology rules enforce data integrity, and domains make sure nomenclature remains uniform. While a field test is still needed, the entry of field data within the Field Collections geodatabase was successful.

Legacy data has proven to work reasonably well with the geodatabase and models. While there are some limitations that come with legacy data, the designed Tasks allow the paleontologist to manually enter or correct this information when necessary. As more information is added, verified, or corrected, the legacy data's effect on the models should reduce. While working with legacy data at the beginning of the project may be more of a hindrance than an assistance, the fact remains that legacy data offers a unique opportunity to explore a region such as Petroleum County, Montana. Using this database can aid in applying management projects to the region and the database can grow alongside the investigation of the geology and paleontology of the county.

Understanding the impact of the PRPA of 2009 on museum curation standards is necessary to keep in mind with respect to the construction of these geodatabases. While it may be common knowledge within the field of paleontology, it is necessary for the students pursuing similar projects to understand the difference between absence of information and protected information. With digitization of collections becoming standardized and utilizing sites such as iDigBio and GBIF, it is doubly important to understand this distinction as well as other restrictions that may be in place.

Overall, this project shows the importance of digitizing museum collections and integrating them with GIS and shows the benefits to paleontological resource management, assessment, and inventorying. Being able to interact with museum collections and compare information with field data and the PFYC can significantly aid site management and inventorying practices and offers a unique opportunity to expand museum collections beyond traditional applications.

References

Allen, David W. 2011. *Getting To Know ArcGIS Modelbuilder*. Redlands: Esri Press.

American Alliance of Museums. 2012. "Developing a Collections Management Policy." Arlington. http://www.aam-us.org/docs/default-source/continuum/developing-a-cmp-final.pdf?sfvrsn=4.

Berendsohn, W. G., A. Güntsch, N. Hoffmann, A. Kohlbecker, K. Luther, and A. Müller. 2011. "Biodiversity Information Platforms: From Standards to Interoperability." *ZooKeys* 150. Pensoft Publishers: 71–87. doi:10.3897/zookeys.150.2166.

Bonnie, Robert. 2015. "Part IV Department of Labor." *Federal Register*. Vol. 80.

Breithaupt, Brent H. 2016. "Passport-In-Time Microvertebrate Fossil Project At The University of Wyoming Geological Museum: Late Cretaceous Paleontological Resources In The Public Eye." In *Proceedings of the 6th Fossil Resource Conference*, 107–12.

Burrows, Cynthia. 2016. "Developing an Archaeological Specific Geodatabase to Chronicle Historical Perspectives at Bethsaida, Israel." University of Southern California.

Canhos, V P, S Souza, R Giovanni, and D A L Canhos. 2004. "Global Biodiversity Informatics: Setting The Scene For A 'New World' of Ecological Modeling." *Biodiversity Informatics* 1: 1–13. http://www.bdt.org.br/bin21/wks95/chm_doc.html.

Cobban, W. A. 1953. "Cenomanian Ammonite Fauna from the Mosby Sandstone of Central Montana." *Geological Survey Professional Paper* 243-D: 1–33.

Cole, Daniel G. 2016. "Incorporating Museum Collections into GIS Research." In *Esri Proceedings*. Washington, D. C.: Esri. https://proceedings.esri.com/library/userconf/fed16/papers/fed_126.pdf.

Cronin, Keith. 2014. "A Bone to Pick: The Paleontological Resources Preservation Act and Its Effect on Commercial Paleontology." *Albany Government Law Review*, 2013–14.

"Download." 2020. Accessed July 7. https://www.gbif.org/occurrence/download/0016669-200613084148143.

Esri. 2020a. "Domains and Subtypes—ArcGIS Insights | Documentation." *Esri Online*. https://doc.arcgis.com/en/insights/latest/get-started/domains-and-subtypes.htm.

Esri. 2020b. "Feature Class Basics—ArcGIS Help | ArcGIS Desktop." *Esri Online*. https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/feature-class-basics.htm.

Esri. 2020c. "Introduction to Subtypes—ArcGIS Pro | Documentation." *Esri Online*. https://pro.arcgis.com/en/pro-app/help/data/geodatabases/overview/an-overview-of-subtypes.htm.

Esri. 2020d. “Model Elements—Help | ArcGIS for Desktop.” *Esri Online*. https://desktop.arcgis.com/en/arcmap/10.3/analyze/modelbuilder/model-elements.htm.

Esri. 2020e. “Topology in ArcGIS—ArcGIS Pro | Documentation.” *Esri Online*. https://pro.arcgis.com/en/pro-app/help/data/topologies/topology-in-arcgis.htm.

Esri. 2020f. “What Is Topology?—ArcGIS Pro | Documentation.” *Esri Online*. https://pro.arcgis.com/en/pro-app/help/data/topologies/an-overview-of-topology-in-arcgis.htm.

Fastovsky, David E, and Antoine Bercovici. 2016. “The Hell Creek Formation and Its Contribution to the CretaceousePaleogene Extinction: A Short Primer.” *Cretaceous Research* 57: 368–90. doi:10.1016/j.cretres.2015.07.007.

Gavin, Daniel G., Matthew C. Fitzpatrick, Paul F. Gugger, Katy D. Heath, Francisco Rodríguez-Sánchez, Solomon Z. Dobrowski, Arndt Hampe, et al. 2014. “Climate Refugia: Joint Inference from Fossil Records, Species Distribution Models and Phylogeography.” *New Phytologist* 204 (1): 37–54. doi:10.1111/nph.12929.

Graham, Catherine H., Simon Ferrier, Falk Huettman, Craig Moritz, and A. Townsend Peterson. 2004. “New Developments in Museum-Based Informatics and Applications in Biodiversity Analysis.” *Trends in Ecology and Evolution* 19 (9): 497–503. doi:10.1016/j.tree.2004.07.006.

Jackson, Frankie D., and David J. Varricchio. 2016. “Fossil Egg and Eggshells from the Upper Cretaceous Hell Creek Formation, Montana.” *Journal of Vertebrate Paleontology* 36 (5): 1–15. doi:10.1080/02724634.2016.1185432.

Kirkland, James I, Donald D Deblieux, Martha Hayden, and Grant Willis. 2006. “Utah Geological Survey: A Valuable Partner in the Management of Federal Fossil Resources.” *New Mexico Museum of Natural History and Science Bulletin* 34: 77–80.

Lazerwitz, David J. 1994. “Bones of Contention: The Regulation of Paleontological Resources on The Federal Public Lands.” *Indiana Law Journal* 69 (2). Bloomington: 601–36. https://www.repository.law.indiana.edu/iljAvailableat:https://www.repository.law.indiana.edu/ilj/vol69/iss2/8.

Lieberman, Bruce S. and Kimmig, Julien. 2018. “Museums, Paleontology, and a Biodiversity Science – Based Approach.” *Special Paper of the Geological Society of America* 535 (September). Lawrence: 335–48. doi:10.1130/2018.2535(22)CITATION.

Liggett, Greg. 2020. “U . S . Department of the Interior Bureau of Land Management Potential Fossil Yield Classification (PFYC) for the Geologic Formations of Montana , North Dakota , and South Dakota.” Billings.

Liggett, Gregory A., S Terry Childs, Nicholas A Famoso, H. Gregory McDonald, Alan L Titus, Elizabeth Varner, and Cameron L Liggett. 2019. “From Public Lands to Museums: The Foundation of U.S. Paleontology, the Early History of Federal Public Lands and Museums, and the Developing Role of the U.S. Department of the Interior.” In *Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making*, edited by G. D. Rosenberg and R. M. Clary, Special Pa, Geological:311–34. Geological Society of America. doi:10.1130/2018.2535(21).

Lindahl, John A. 1993. “Soil Survey of Petroleum County, Montana.” Billings.

Marshall, C. R., S. Finnegan, E. C. Clites, P. A. Holroyd, N. Bonuso, C. Cortez, E. Davis, et al. 2018. “Quantifying the Dark Data in Museum Fossil Collections as Palaeontology Undergoes a Second Digital Revolution.” *Biology Letters* 14 (9). doi:10.1098/rsbl.2018.0431.

Murphey, Paul C, Georgia E Knauss, Lanny H Fisk, Thomas A Deméré, and Robert E Reynolds. 2019. “San Diego Society of Natural History BEST PRACTICES IN MITIGATION PALEONTOLOGY.”

Murphy, Paul C., Kate Zubin-Stathopoulos, Courtney Richards, and Meredith A. Fontana. 2015. “Paleontological Overview of the Royal Gorge Field Office Planning Area, Colorado.” Denver. https://www.researchgate.net/publication/297944941.

Nebel, Mark. 2020. “Grand Canyon National Park Geodatabase Schema.” Flagstaff.

Oheim, Kathryn B. 2007. “Fossil Site Prediction Using Geographic Information Systems (GIS) and Suitability Analysis: The Two Medicine Formation, MT, a Test Case.” *Palaeogeography, Palaeoclimatology, Palaeoecology* 251 (3–4): 354–65. doi:10.1016/j.palaeo.2007.04.005.

Pham, Kacey Johnson. 2015. “GIS Data Curation and Web Map Application for La Brea Tar Pits Fossil Occurrences in Los Angeles, California.” *ProQuest Dissertations and Theses*. https://login.pallas2.tcl.sc.edu/login?url=https://search.proquest.com/docview/1767205694?accountid=13965%0Ahttp://resolver.ebscohost.com/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQuest+Dissertations+%26+Theses+Global&rft_v.

PRPA. 2009. *Paleontological Resources Preservation Act of 2009*. Vol. 2009.

Rode, Alycia L. Stigall. 2005. “The Application of Geographic Information Systems to Paleobiogeography: Implications for the Study of Invasions and Mass Extinctions.” *The Paleontological Society Papers* 11: 77–88. doi:10.1017/s108933260000125x.

Rowley, Jennifer. 2002. “Using Case Studies in Research.” *Management Research News* 25: 16–27.

Santucci, Vincent L, and Alison L Koch. 2003. “Paleontological Resource Monitoring Strategies.” *Park Science*.

Santucci, Vincent L, and Justin S. Tweet. 2020. “Grand Canyon National Park Centennial Paleontological Resource Inventory (Non-Sensitive Version).” Fort Collins.

Sax, Joseph L. 2016. “Implementing The Public Trust In Paleontological Resources.” In *Proceedings of the 6th Fossil Resource Conference*, edited by Vincent L Santucci and Lindsay Mcclelland, 173–77. Lakewood: National Park Service Geological Resource Division.

Stein, Walter W., and Michael Triebold. 2013. “Preliminary Analysis of a Sub-Adult Tyrannosaurid Skeleton from the Judith River Formation of Petroleum County, Montana.” In *Tyrannosaurid Paleobiology*, edited by Michael J. Parrish, Ralph E. Molnar, Philip J. Currie, and Eva B. Koppelhus, 55–77. Bloomington: Indiana University Press.

Stigall, Alycia L., and Bruce S. Lieberman. 2006. “Quantitative Palaeobiogeography: GIS, Phylogenetic Biogeographical Analysis, and Conservation Insights.” *Journal of Biogeography* 33 (12): 2051–60. doi:10.1111/j.1365-2699.2006.01585.x.

Sullivan, Robert M., Arjan C. Boere, and Spencer G. Lucas. 2005. “Redescription of the Ceratopsid Dinosaur Torosaurus Utahensis (Gilmore, 1946) and a Revision of the Genus.” *Journal of Paleontology* 79 (3): 564–82. doi:10.1666/0022-3360(2005)079<0564:rotcdt>2.0.co;2.

Voss, John-Henry. 2018. “Museum Collections Management with Google Earth and GIS.” Texas Tech University.

Vuke, Susan M., Karen W. Porter, Jeffery D. Lonn, and David A. Lopez. 2007. “Geologic Map of Montana” Geologic M (1). Billings: Montana Bureau of Mines and Geology: 1–68. doi:10.16309/j.cnki.issn.1007-1776.2003.03.004.

Wosik, Mateusz, Mark B. Goodwin, and David C. Evans. 2013. “Nestling-Sized Hardosaurine Cranial Material from the Hell Creek Formation of Northeastern Montana, USA, with an Analysis of Cranial Ontogeny in Edmontosaurus Annectens.” *PaleoBios* 36: 1–18. doi:10.11436/mssj.15.250.

Yeung, Albert K. W., and G. Brent Hall. 2007. *Spatial Database Systems*. Dordrecht: Springer.

Appendix A Fossil Specimens

This Appendix lists the provided museum records gathered from open-sourced databases and with permission by the BLM. For more information on the BLM specimen list, please reach out to Gregory Liggett.

Table 1 Specimen Count

Museums	Specimen Count
University of Montana Herbarium	395
American Museum of Natural History	88
University of Kansas	69
University of Montana	62
Smithsonian National Museum of Natural History	19
The New York Botanical Garden	9
Royal Ontario Museum	7
University of Michigan Museum of Zoology	5
Colorado State University	4
University of Alabama	4
id	4
University of New Mexico	3
University of Puget Sound	3
The University of Washington The Burke Museum	3
Oregon State University	2
University of Alberta Museums	2
University of Minnesota Bell Museum	1
Boise State University Snake River Plains	1
Ohio State University	1
Angelo State Natural History Collections	1
Yale Peabody Museum	1
Total	**684**

A list of specimens from iDigBio (2020).

Table 2 Specimen Count

Museums	Specimen Count	
American Museum of Natural History	87	
University of Kansas	5	
Paleobiology Database	48	
Yale Peabody Museum	1	
Total	**141**	

A list of specimens downloaded from GBIF (2020).

Table 3 Specimen Count

Museums	Specimen Count	
Museum of the Rockies	1	
Science Museum of Minnesota	90	
The Field Museum	18	
The Smithsonian National Museum of Natural History	24	
The Bureau of Land Management	85	
Total	218	

A list of specimens provided from participating museums and federal collections.

Appendix B Data

Data
Data necessary for this project requires data listed in Table 1. The following table breaks down and describes the necessary datum needed to create this project.

Table 4 Datasets Used

Data	Type	Format	Source
Fossil Occurrence Data	Fossils	Tabular	iDigBio, GBIF, NMNH, FM, SMM, BLM, MOR
Montana Land Cover	Land Cover	Geodatabase	University of Montana umt.edu
Public Lands	Montana Public Lands	Polygon	Montana Cadastral
The State Geologic Map Compilation (SGMC) geodatabase Conterminous Geology of the United States	Geology	Geodatabase	USGS
Major Lakes and Streams for Montana	Hydrology	Polyline	Montana State Library geoinfo.msl. mt.gov
USGS 13 arc-second 1 x 1 degree	Elevation	Raster	USGS The National Map Viewer.nationalmap.gov
State and Counties of Montana Boundary	Counties	Polygon	Montana State Library geoinfo.msl. mt.gov
Montana Transportation Framework	Roads	Polyline	Montana State Library geoinfo.msl. mt.gov
Public Land Survey System (PLSS)	Township and Range	Polyline	Montana Cadrastral
Montana Roads from TIGER/Line Pile	Roads	Polyline	USGS Sciencebase.gov

Cartographic Boundary Files	States	Polygon	United States Census Bureau www.census.gov