The Thesis Committee for Laura Marie Brenskelle
Certifies that this is the approved version of the following thesis:

The Use of Modern Digital Technology to Store and Serve Biodiversity Data For Research and Educational Purposes

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Timothy B. Rowe

Christopher J. Bell

Unmil Karadkar

Matthew Brown

The Use of Modern Digital Technology to Store and Serve Biodiversity Data For Research and Educational Purposes

by

Laura Marie Brenskelle, B.A.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Geological Sciences

The University of Texas at Austin

December 2015

Dedication

To Fred and June Underwood, my grandparents, who instilled in me a deep love of nature, which has fueled my passions in graduate school.

Acknowledgements

Thank you to my supervising committee, my parents, Ann Molineux, Angie Thompson, Benn Breeden, Adam Marsh, Colleen Dawes, Joshua Lively, Lana Jones, the Landis family, and my dog Chance for all of their support in completing my Master's work.

Abstract

The Use of Modern Digital Technology to Store and Serve Biodiversity Data For Research and Educational Purposes

Laura Marie Brenskelle, M.S.Geo.Sci.

The University of Texas at Austin, 2015

Supervisor: Timothy B. Rowe

Herein I describe two different projects I completed during the course of my Master's at The University of Texas at Austin. These projects broadly focused on the application of technology to maintain scientific data for research and education. The first chapter is a case study of a website I developed as part of a group project in a graduate database management course. Our group took a module from proprietary instructional software developed in the 1990s, and moved it into an online format with a MySQL database on the backend. In chapter one, I provide the appropriate documentation for this project to be expanded in the future. The second chapter describes a project where I interviewed collection managers of natural history collections about their database practices. These practices have implications for the downstream use of these data for research, education, and conservation. As technology inevitably advances, this thesis will serve as a historical snapshot of modern practices, and today, it can provide a starting point of how to further the emerging discipline of biodiversity informatics.

Table of Contents

List of Tables

List of Figures

Chapter 1: Migrating an Instructional Module from Proprietary Software to a MySQL-Based Website

ABSTRACT

The Age of Dinosaurs is an undergraduate paleontology course taught at The University of Texas at Austin that utilizes instructional software created by Kyoko Kishi in the early 1990s. The students in this course use the software as a study tool to prepare for exams and laboratory quizzes. However, due to its age and the advancement of technology, the software is no longer functional on all operating systems. In 2014, my colleagues and I developed a website to replace one of the instructional modules on the software as a project for a database management graduate class. We aimed to migrate the functionalities of the software to a website in order to make access to the instructional resources easier for all students in the course, and to allow the content in the resource to be updated to reflect the current understanding of evolutionary relationships among dinosaurs.

INTRODUCTION

During the early 1990s, the affordability of personal computers and compact discs revolutionized the mobility of software. Efforts were made to utilize this technology for educational purposes but at the time, it was not clear how quality digital content should be developed or how it should be introduced into the classroom. In 1992, a project funded by the National Science Foundation (USE-9156073) at The University of Texas at

Austin began to develop multimedia courseware for a popular undergraduate paleontology course called The Age of Dinosaurs, taught by Dr. Timothy Rowe. The project aimed to develop courseware that augmented and extended the course content, while also advancing the development of digital media. The project was directed by Rowe, and was carried out by students led by Kyoko Kishi, a Master's student in the UT Department of Geological Sciences (Kishi 1995).

Due to the growing popularity and availability of personal computers and the ability to inexpensively produce inexpensive compact discs, it was decided to create a CD that could be distributed to students with course textbooks (Kishi 1995). *The Age of Dinosaurs* CD-ROM consists of nine separate modules – Plate Tectonics, Systematics, Archosaur Anatomy, Vertebrate Cladogram, Vertebrate Cladogram Taxon Drill, Dinosaur Cladogram, Dinosaur Drill Instructor, Dinosaur Flashcards, and Collecting a Fossil.

The modules were built in Adobe Authorware Professional version 2.0, and skeletal drawings were prepared using Adobe Illustrator. Initially, the software was compatible with both Apple Macintosh and PC computers, but after OS 9, Apple no longer supported older Authorware applications. Fortunately, the skeletal drawings are still compatible with current versions of Adobe Illustrator. Today, *The Age of Dinosaurs* CD-ROM can be run on most Windows computers, but students with Mac computers must install a Windows Emulator in order to use the program. The students are provided an instructional PDF on how to install this emulator, but anecdotally, the undergraduates frequently run into problems while trying to open the CD-ROM on their computers, and require significant help from the teaching assistants.

This courseware has its merits and was a significant advance for instructional technology at the time. In 2014, as part of a database management course at The University of Texas School of Information, my colleagues and I migrated one of the instructional modules from this courseware to a Web-based platform. The goal was not only to move the module online so it is readily available to all students in the Age of Dinosaurs course, but also to create a platform that easily allows course content to be updated when necessary.

Migrating the different components of *The Age of* Dinosaurs CD-ROM to the web is one way to prolong the longevity of this program, and to make accessing the educational content easier for undergraduate students. In 2005, the Systematics component was moved to a website developed by Eleanor TenBarge under Kyoko Kishi's supervision. This module can be viewed at http://www.geo.utexas.edu/courses/302d/Systematics/systematics.swf. This page uses Adobe Flash file format to allow an essentially unchanged version of the Systematics module to be accessed online. Stimulated by this approach, my colleagues and I planned to transfer another module from *The Age of Dinosaurs* program to a web-based platform that could be easily opened by students using a diverse array of computer operating systems.

METHODS

In the spring of 2014, Melissa Brown, Jin Gao, Katherine Carter, Mallorie Sayre, and I created an online interface for the Flashcards module from *The Age of Dinosaurs* CD-ROM for our Database Management course. Using the PHP and MySQL skills gained from this class, we developed a website that is available to any student with internet access. The website was used as a study tool by undergraduates taking the Age of Dinosaurs class in the Fall 2014 semester.

Our INF 385M course used MariaDB, an open-source, freely available version of MySQL. Because MariaDB is open-source and currently utilized by large corporations around the world, its longevity over the next 5-10 years should be ensured. Even if a technological shift does occur in the future, its open-source nature means that data can be easily extracted from MariaDB. The website is currently hosted on a University of Texas School of Information server, and can be viewed at http://corvette.ischool.utexas.edu/dino. In the short term, Mr. Stan Gunn has archival copies of the Dino code stored on the corvette server at the School of Information. Additionally, the PHP code for the website is included in the appendices of this thesis, and has been uploaded to GitHub at https://github.com/utdinos/Dinos. We believe that using an open source database management system and providing documentation on how the website was developed will improve the possibility that it could be built upon or improved in the future. If the course material were to be revised, a person with knowledge of MySQL could simply delete our data and repopulate the fields with updated information.

In order to create the online Dino modules, we first developed a schema for the MySQL database on the back-end. Relational databases are comprised of a set of related tables, and the schema is the structure of how the data is stored within the database. To know what tables and fields were necessary, we needed to understand what information was presented on the old CD-ROM software. All group members had access to the CD-ROM and assessed the Dinosaur Flashcards module. Because the course instructor was going to continue to also use the software, we had to ensure that the information presented on the website was consistent with that on the CD-ROM in order to avoid confusion among the undergraduate students. Benn Breeden, a Master's student of Dr. Timothy Rowe, helped the group parse out the synapomorphies (shared, derived characters that define a clade) associated with each of the dinosaur taxa presented on the CD-ROM. Adobe Illustrator images were archived by the original authors of the CD-ROM, who provided them to us for this project. Our group created a MySQL schema for representing these data and images to serve as the back-end of our dynamic web application (see Figure 1.1).

The arrows with the forked ends on one side indicate a one-to-many relationship. In the figure, PK stands for ‘primary key’, which indicates a unique numerical field that is used to identify an individual entry in that table. FK stands for ‘foreign key’, which indicates where a primary key from one table has been used in another table in order to link information. For example, the two foreign keys in the table ‘taxonchar’ link each taxon to its synapomorphies.

Melissa Brown and Mallorie Sayre were responsible for the Bootstrap CSS (Cascading Style Sheets) and JavaScript used to create the Web display, and the home page for the online study guide. The remaining pages were developed by Katherine Carter (http://corvette.ischool.utexas.edu/dino/Dinos_TaxonFC.php), Laura Brenskelle (Dinos_List.php), and Jin Gao (Dinos_Search.php). Due to the nature of the Database Management course requirements and our group's coding abilities, some of the functionalities of these pages changed between The Age of Dinosaurs CD-ROM and the Dino website. The Results section describes the development and functionality of each page on the Dino website, as well as distinctions in behavior between the website and the CD-ROM.

RESULTS

INDEX.HTML

The home page is mobile responsive, and will change its resolution to fit the device that is used to access the page. It contains a navigation header with links to the different pages of the website, a brief description of the website and its purpose, descriptions of the functionalities of the Search, Flashcards, and Quiz pages, and three buttons that link to each page. Figures 1.2a and 1.2b display the top and bottom portions of this page.

DINOS_TAXONFC.PHP

The Dinos_TaxonFC.php page (Figure 1.3a), designed by Katherine Carter, mimics the experience of using physical flashcards. Upon opening, the page displays images of a group of dinosaurs along with the question, “What is the monophyletic group these dinosaurs all belong to?” Because these are flashcards, the user can check his or her answer by clicking the green ‘Answer’ button, which is implemented as a JavaScript toggle button. This results in the display shown in Figure 1.3b, while avoiding the time and computationally intensive process of re-loading the entire page.

This page queries the MySQL database to request a random set of dinosaurs that share the same dinotaxon_id (see Figure 1.1 for clarification), and then displays the dinosaur images associated with this id in the database. The interface prompts the user to identify the smallest monophyletic group that describes these dinosaurs. To bring up the next question, the user clicks the gray ‘Next Question’ button, which reloads the page. As a result of the SQL query being randomly generated, some times the reloaded page may be identical to the current page. The PHP script that generates the web page for Dinos_TaxonFC.php can be found in Appendix B.

DINOS_LIST.PHP

Dinos_List.php (Figure 1.4a), designed by Laura Brenskelle, enables students to test their ability to identify a dinosaur group based on visual representations of its skeletal reconstruction. The page works similarly to Dinos_TaxonFC.php, and returns a random dinosaur from the SQL database. The user is asked to answer the question: “What is the smallest monophyletic group that this dinosaur belongs to?” as a free text entry. When in doubt, the user also has an option to click the green ‘Need a Hint?’ button, which toggles the synapomorphy for the group (Figure 1.4b).

Figure 1.4c depicts the result display when a user provides and incorrect answer. Because the code works by comparing the user's answer to the string entered in the ‘prime’ field of the dinotaxon table, it is very important that the submitted answer be correctly spelled and appropriately capitalized. While the technology behind our software is mature enough to accept spelling variants, this specificity serves a purpose; it re-enforces the importance of spelling and capitalization of taxon names to the students, who are expected to know how to write these names properly on exams in the Age of Dinosaurs class. For this reason, the error message does not indicate how close a student’s answer is to the correct answer (i.e., whether the first letter needs to be capitalized or whether there is a small spelling error).

Figure 1.4d illustrates the resulting page when the user enters the correct answer (spelled and capitalized properly). In addition to a textual message, the page includes the genus of that dinosaur, which becomes important when utilizing the Dinos_Search.php page. The page also allows the user to move on to a new question. The user is not given a new question until they answer the current question correctly. This feature was intended to let students answer as many times as possible in order to get the right answer. The PHP source code for Dinos_List.php is available in Appendix C.

DINOS_SEARCH.PHP

The search page, Dinos_Search.php (Figure 1.5a), designed by Jin Gao, was a new feature (i.e., one that does not have an analogous functionality in earlier versions of the software) that our group developed in order to provide users with a complete view of all the dinosaurs available in our database. To facilitate browsing the list of dinosaurs, this page presents users with a dropdown list of all genera in the database. Students may then view the skeletal reconstruction images associated with a genus, the smallest monophyletic taxon they should know for the course (the 'prime' field stored in the database), and the synapomorphies of that taxon (Figure 1.5b). The PHP source code for the web page Dinos_List.php is included in Appendix D.

CONCLUSIONS AND FUTURE DIRECTIONS

We migrated the Flashcards module from *The Age of Dinosaurs* CD-ROM to a web format. There are a few differences between the website and the Flashcard component of the 1995 software. In general, the CD-ROM offers the user more options and control over their experience. However, our website lacks certain functionalities due to time constraints, because this was a group assignment for a course, and because our group members lacked the programming background to create these features within the available time frame. By documenting the development, the back-end MySQL database, and coding behind the Dino webpage, we expect to enable future developers to expand and improve our project.

Before accessing the actual flashcard function on the CD-ROM, the user may optionally choose to focus their study on particular dinosaur taxa. In addition, the user may choose between three different levels of difficulty. The easiest – 'novice' – allows a user to receive hints and does not impose a time limit. The second level of difficulty – 'fledgling' – does not provide hints, and it enforces a 60 second time limit on the user for providing the correct answer. The most difficult level – 'veteran' – shows the user disarticulated skeletons, does not provide hints, and also enforces a 60 second time limit. We have maintained the hint functionality in our website via the 'Need a Hint?' toggle button. We did not have access to disarticulated versions of the dinosaur skeleton images, and given more time, the three levels of difficulty would be technically easy to implement. This change would also necessitate changes to the database schema, for example, the addition of an additional field to the dinosaur table for pointing to the images of the disarticulated skeletons. The idea of enforcing a time limit was quickly ruled out in our group discussions due to the fact that programming for this was outside the scope of our Database Management course and outside of our personal web design knowledge.

The CD-ROM also lets users choose whether or not the program will accept the genus name of a given dinosaur as a correct answer. With an eye toward future implementation of this feature, we have included the genus names as a field in our MySQL database.

The fact that our website does not indicate if a user's answer is improperly spelled or capitalized may seem like a pedagogical misstep. Given more technical knowledge and time, our group would have liked to address this by creating two different prompts to inform the user if their answer is incorrect or if it is misspelled. This could potentially be solved if the page recorded commonly misspelled user responses, and then querying this list to compare the user's answer to the commonly misspelled responses.

Authorware formats, like the one used to create the original software, do not lend themselves to easy upgrades by future programming teams, thus stifling continuing updates to course content. Scientific advancements in dinosaurian phylogenetics in the past 20 years are not represented in the Age of Dinosaurs course because the class design relies heavily on the students' use of the CD-ROM to prepare for exams. In contrast, the database-based website uses open standards such as HTTP and SQL, popular programming languages such as PHP and JavaScript, and freely available infrastructure such as the Apache webserver and the MySQL database engine. This would make future updates to the website technically easier, but appropriate funds would still be necessary to compensate programmers.

In the future, it will be necessary to move the website to a different server. If the Jackson School of Geosciences or The University of Texas at Austin have particular servers for hosting instructional technology used in courses, that would be an ideal place to move this website. The freely available code, the well-documented schema, and the .jpg image files detailed here and saved on GitHub will facilitate replication of our web site.

This project served as a case study in exploring the potential for transforming proprietary educational software to a web-based format, which can be accessed and updated easily. Although the technology we employed is freely available, sustaining such development involves personnel and infrastructure, which are not trivial and must be considered in the context of long-term use of academic software. However, the technical risk of attempting such a project is low, and the potential advantages are numerous. Though a formal survey was never completed, anecdotal evidence through interactions with students enrolled in the Fall 2014 Age of Dinosaurs course indicates that students prefer this format because it is simple to use and does not require installing new software to their computers. Furthermore, the benefits of attempting such a project are enhanced when proper documentation of the development and technology is created. The documentation provides an important resource for future work and improvement as technology inevitably advances.

FIGURES

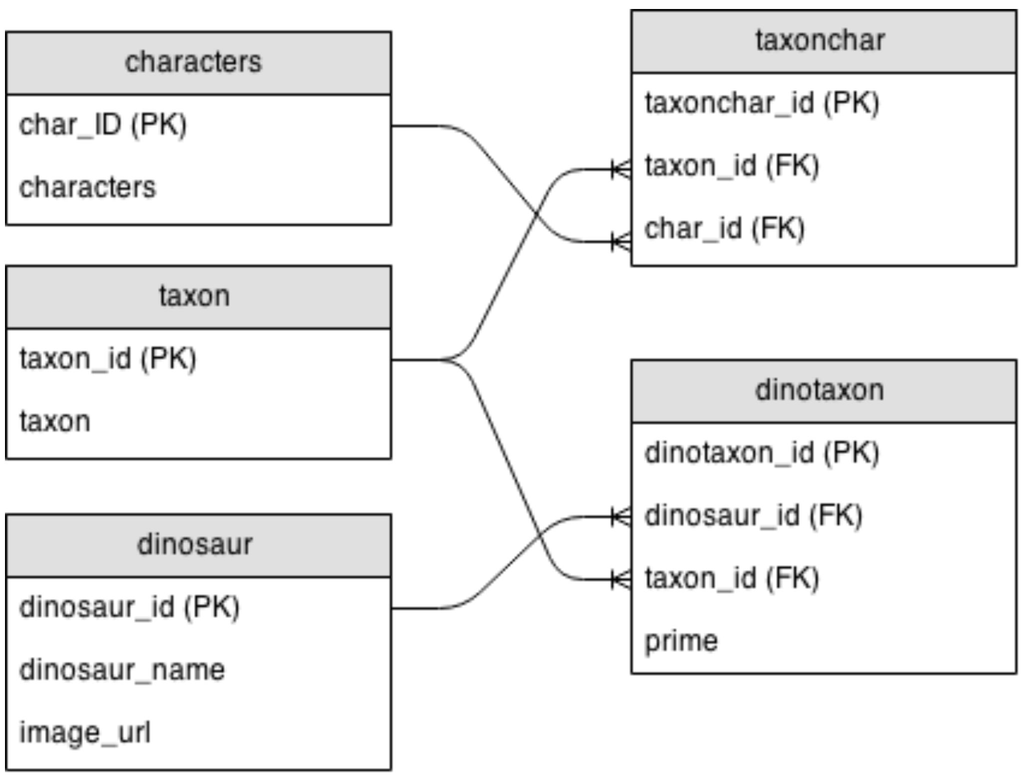

Figure 1.1 – The MySQL database schema operating on the back-end of the Dino webpage.

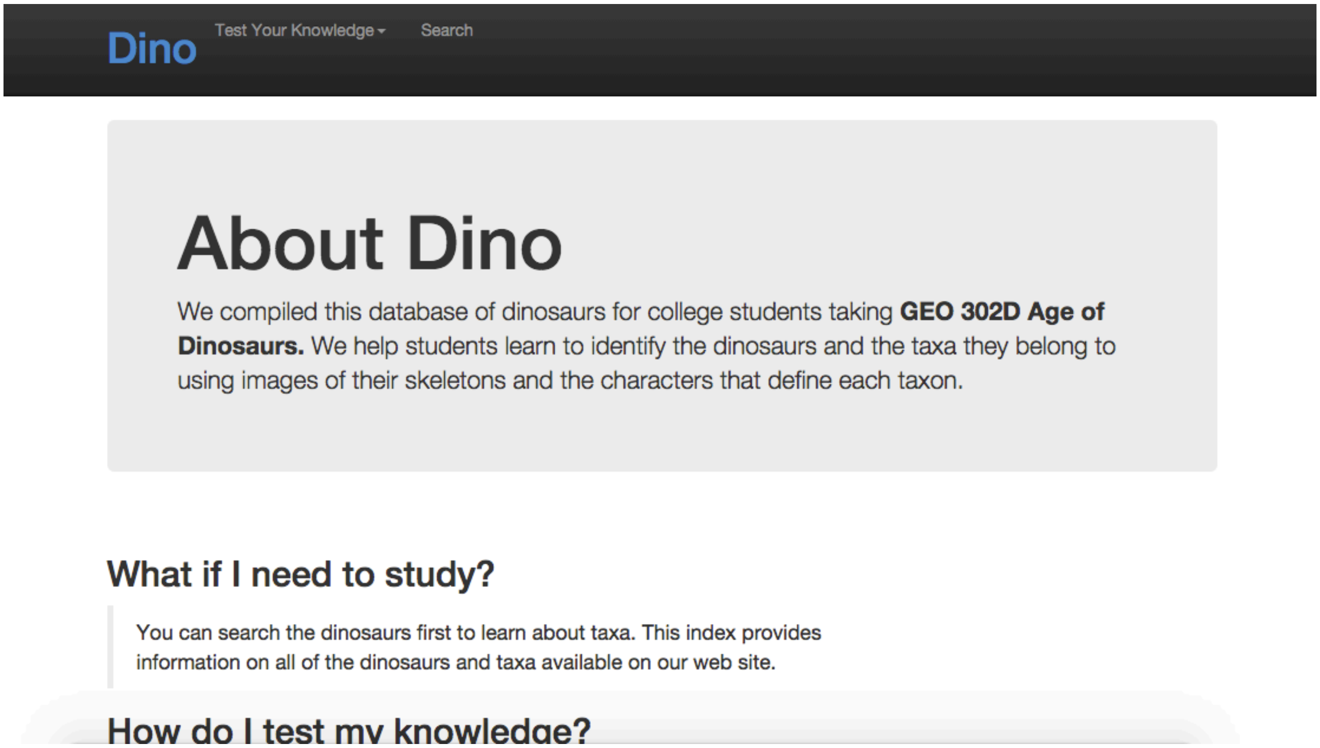

Figure 1.2a: This is the top portion of the Dino website homepage.

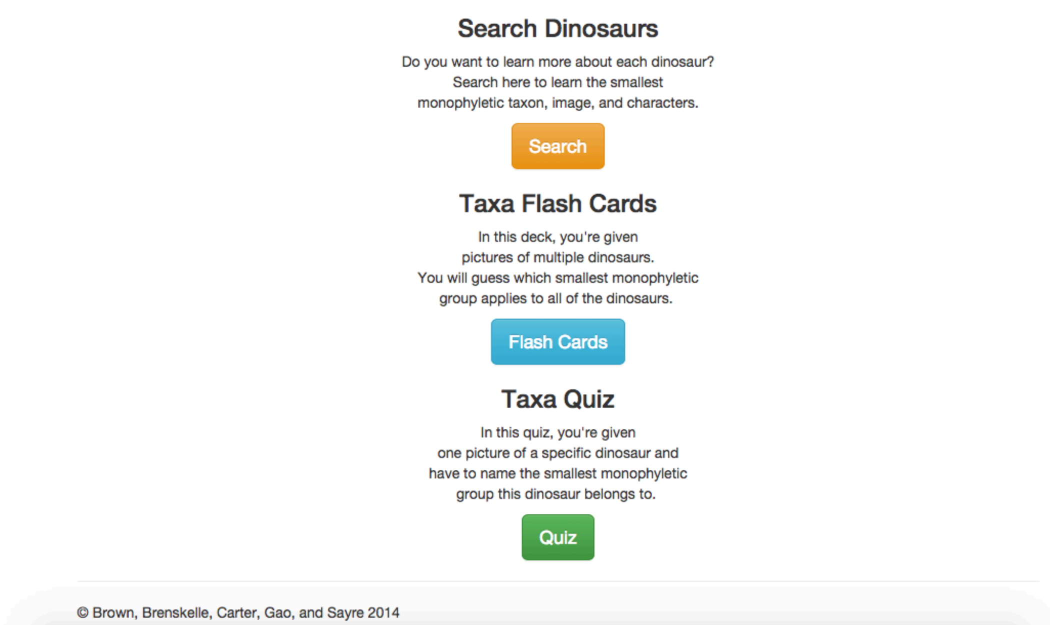

Figure 1.2b: This is the bottom portion of the Dino website homepage.

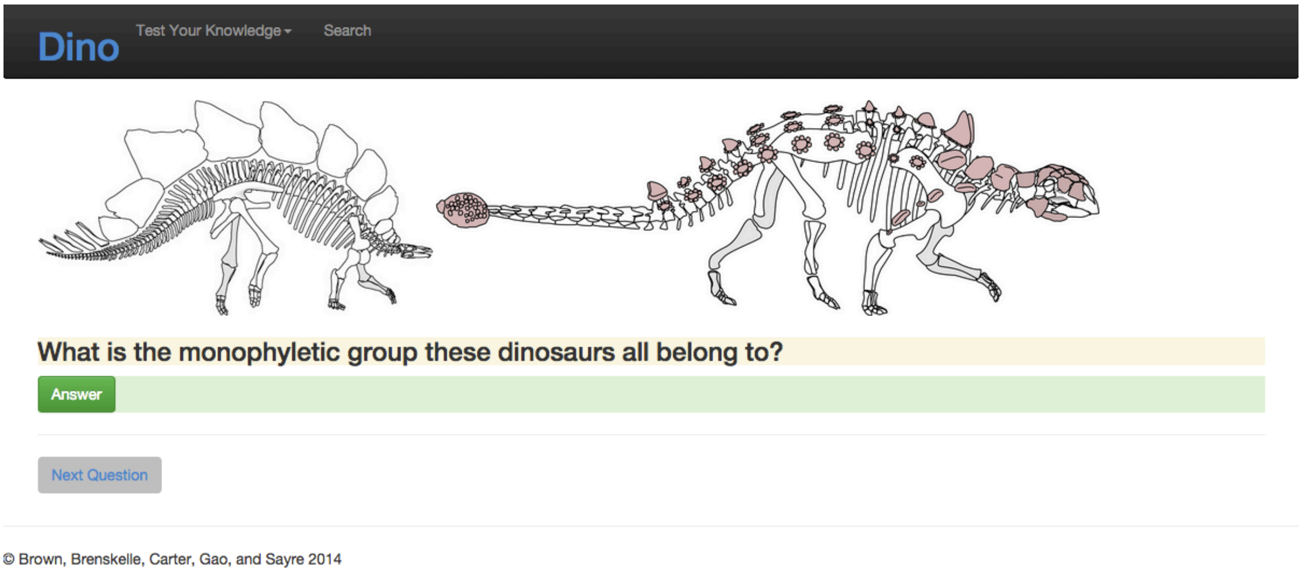

Figure 1.3a: This shows the flashcards web application.

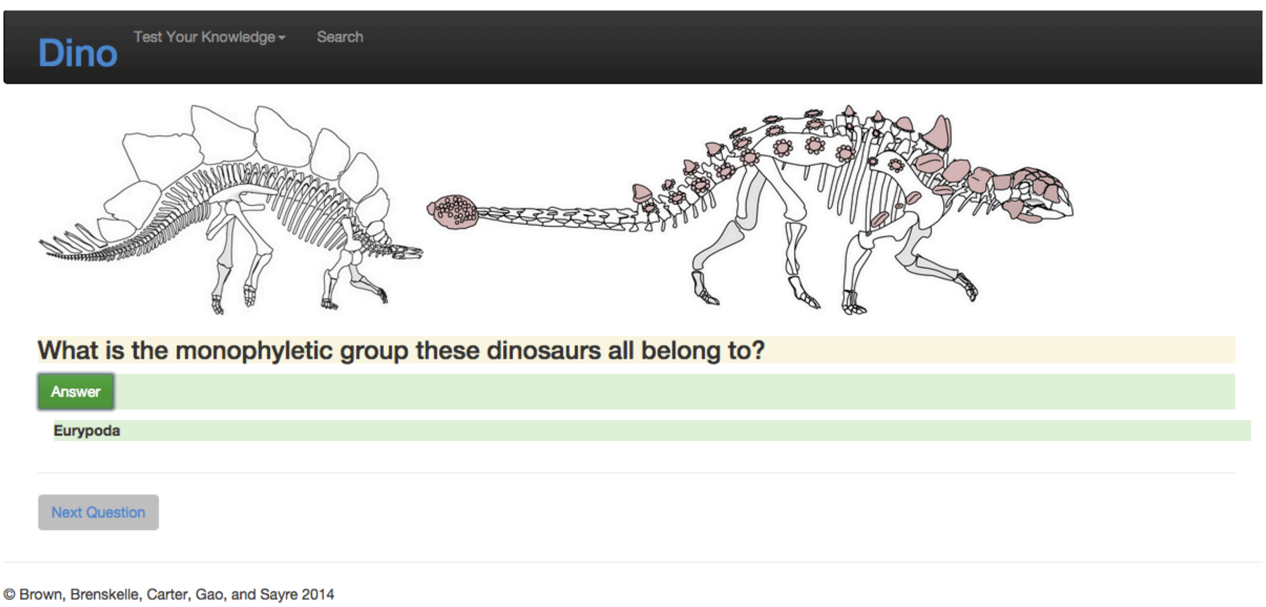

Figure 1.3b: This is the flashcards web application after the green 'Answer' button has been toggled.

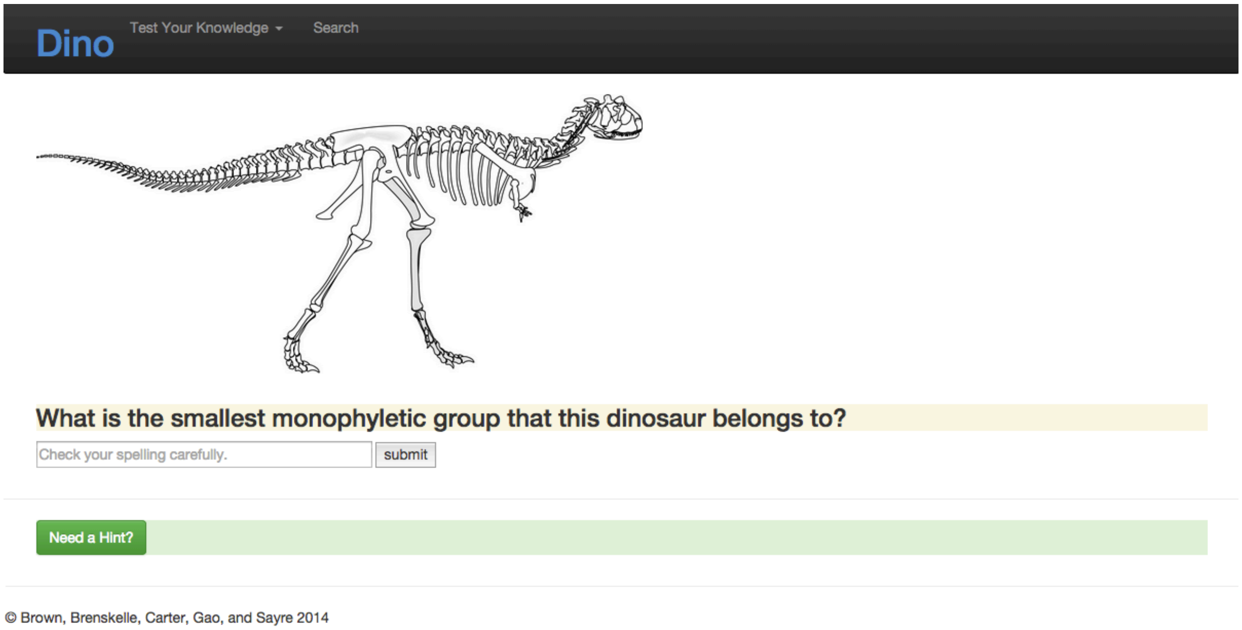

Figure 1.4a: This shows the quiz web application.

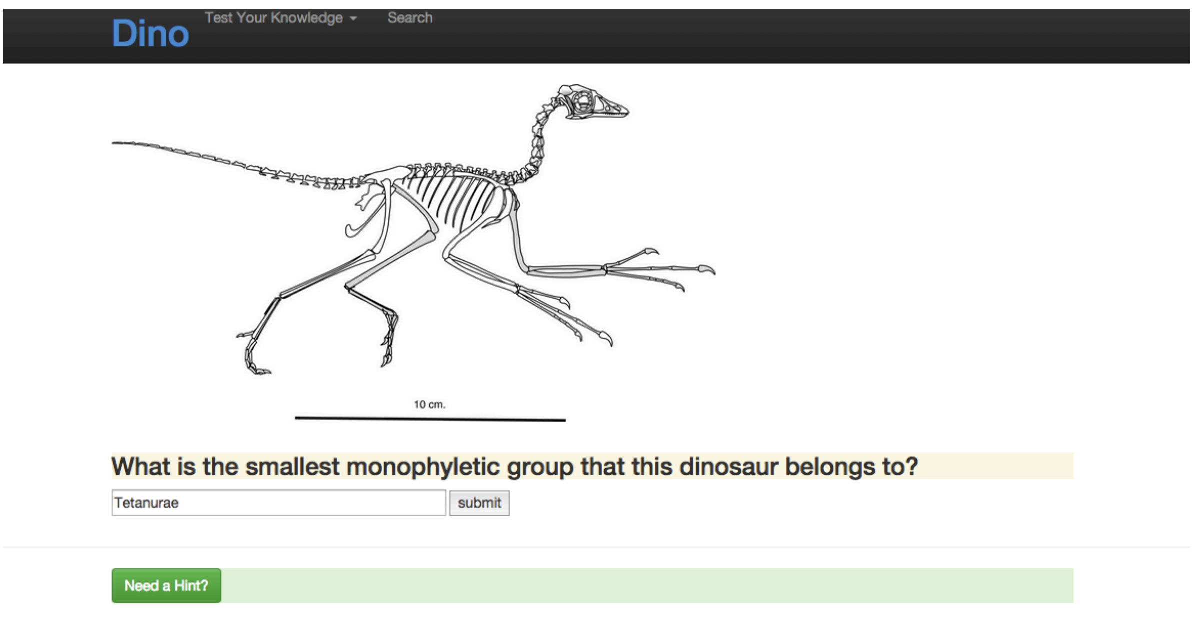

Figure 1.4b: View of the quiz web application with an answer typed into the text box.

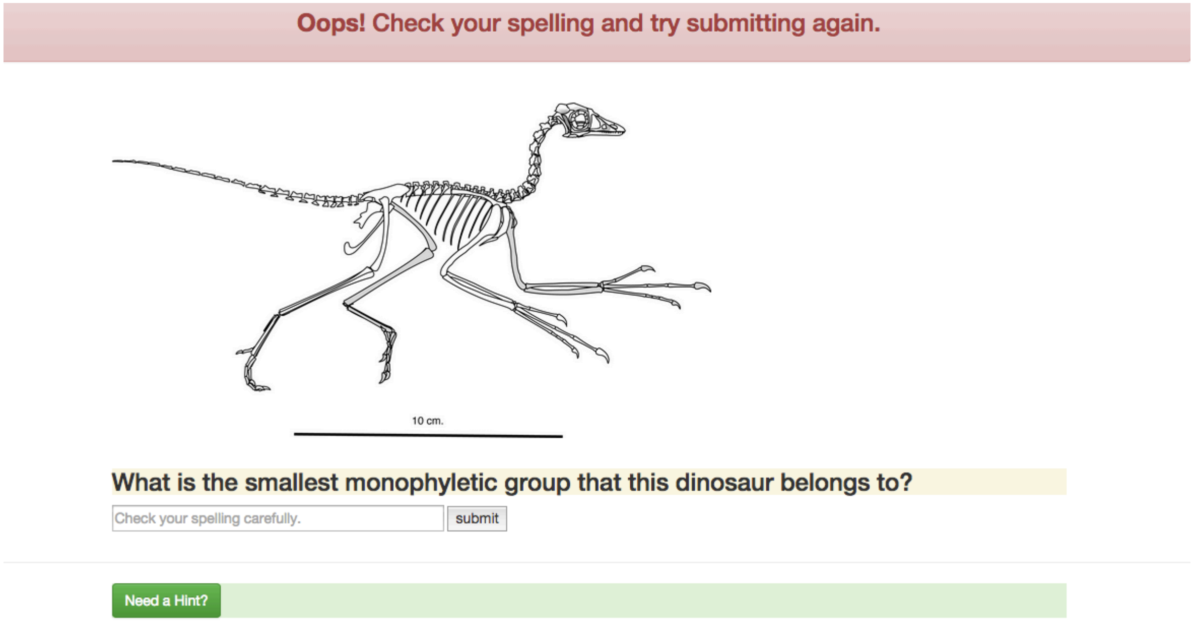

Figure 1.4c: View of the flashcards web application if a wrong or incorrectly spelled answer is submitted.

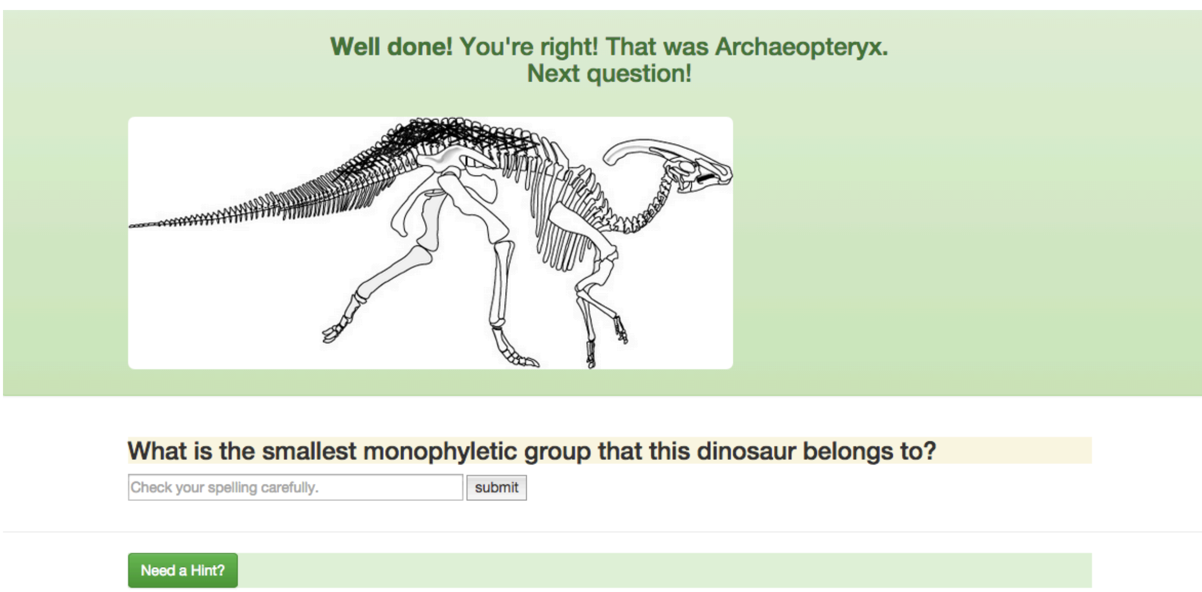

Figure 1.4d: View of the flashcards web application if a correct answer is submitted.

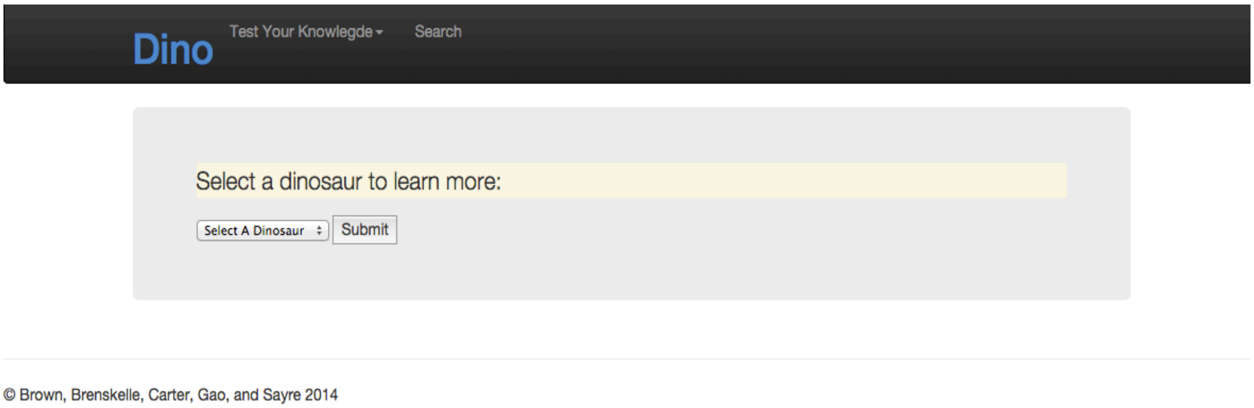

Figure 1.5a: This shows the search page.

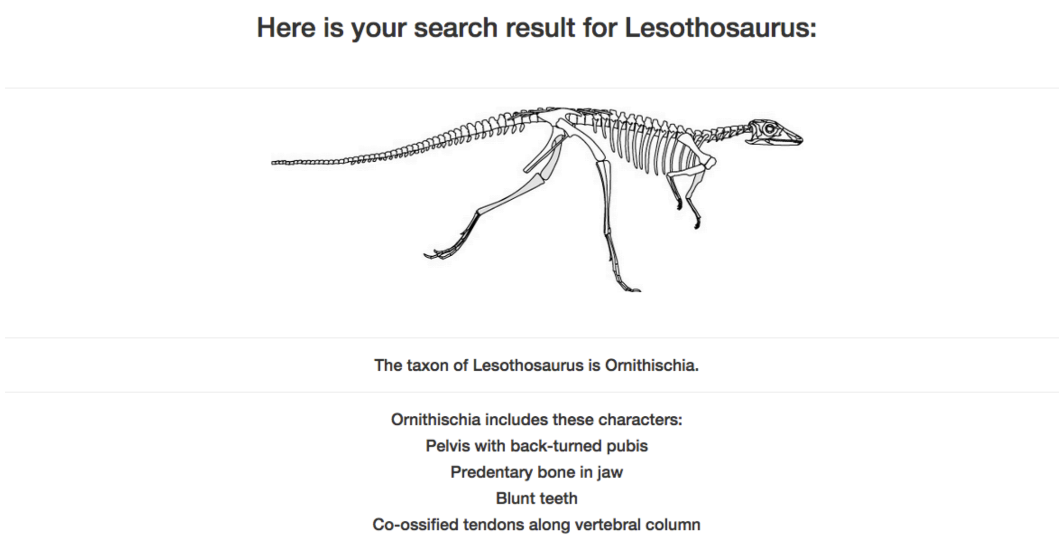

Figure 1.5b: View the search page once *Lesothosaurus* was chosen from the dropdown list.

Chapter 2: An Assessment of Database Practices of Natural History Collections

Abstract

Current national and global scientific initiatives aim to digitize natural history collections in order to build online resources that will provide scientists, educators, policymakers, and the public in general with access to the vast stores of information housed in museums. The data sources for these websites are individual collections, and the management of these digital data falls on collection managers. For this study, I was interested in exploring how well collections databases are serving these purposes, and how much natural history collection managers know about database management. The database practices and protocols at individual collections have larger implications when their data are sent to data aggregators and shared via data portal websites like iDigBio or GBIF. Overall, I found that the majority of collection managers lack formal training in database management, and this impacts the searchability, usability, and accuracy of data shared by these collections. This study recognizes a professional gap in modern biological sciences that is not being filled by curators, collection managers, or researchers, but could be filled by the professionalization of biodiversity informatics as a discipline.

Introduction

Biodiversity informatics, an emerging field that applies information technology to enhance the study of the organismal biology and ecology of our planet, has the potential to play an essential role in answering important questions about the Earth's biota during a time of dramatic change (Thomas et al. 2004; Peterson et al. 2010). Some estimate that half of all species could be extinct by the end of this century (Pimm et al. 1995; Jenkins 2003). Others warn that unless changes are made to alleviate environmental and human pressures that are driving extinction, we will be in the midst of the sixth mass extinction in only a few generations (Barnosky et al. 2011). Digitizing data from natural history collections across the globe will create a massive dataset of approximately 2-3 billion historical species-occurrence records, representing scientific efforts to collect and understand Earth's diversity dating back to the 16^{th} century (Krishtalka & Humphrey 2000; Cowan 1969). This would undoubtedly be a valuable resource for all disciplines of the biological sciences, and if these data were integrated with information from other domains, such as geology, climatology, and demography, then they would have an increased value for informing current and future conservation and policy decisions (Peterson et al. 2010; Graham et al. 2004).

At present, the museum community is still in the throes of digitizing the vast stores of information in natural history collections. Despite the fact that these data form the foundation of everything that biodiversity informatics has the potential to accomplish, the reality of how these data are made available is sometimes ignored. The attitudes and biases that collection managers bring to data curation have implications for the accuracy and usability of these data downstream. Researchers bring their own biases when they use these data to address research questions without proper knowledge of where these data come from and what limitations they may have. Furthermore, as my results show, there are other systemic issues that constrain collection managers' abilities to take on the massive task of digitizing a complete collection. The purpose of my study is to augment understanding of database practices and other challenges accompanying digitization of natural history collections. The goal is to identify steps that can be taken to improve institutional practices and data interoperability, and to inform researchers who are intent on using this data.

When some people in the natural history museum community use the word 'digitization', they are speaking strictly about the imaging process. However, in the context of this paper, I use 'digitization' to describe the goal of modern efforts to make any and all specimen data available online for user consumption. Computerization of natural history collections began in the 1970s with the purpose of converting previously handwritten catalogues into electronic formats. Computerization is differentiated from digitization because the objectives of these two processes are different. Initial computerization of collections data was done to improve the collection manager's ability to search for and locate specimens (Graham et al. 2004; Sunderland 2013). Before electronic catalogues, searching natural history collections was a much more arduous task. Prior to computerization, Ellis Yochelson (1969) of the U.S. Geological Survey stated that, if you wanted to search for fossil data in the Washington mega-fossil collection based on age or geographic location, this might take half an hour. If you knew an exact locality number, your search would be cut down to only two or three minutes.

Certain vital information was left out of these early efforts due to technological gaps, and biodiversity informatics is still trying to find ways to address these problems today. In 1977, the Museum of Vertebrate Zoology (MVZ) at UC Berkeley was awarded a National Science Foundation grant to computerize their collections. Due to the inability of 1970s-era computers to handle images, it was during this time that photographs, which were a required part of Joseph Grinnell's strict collection protocols, stopped being considered a systematic part of the MVZ collection (Sunderland 2013). Today, the importance of primary records and archives are again recognized as significant parts of natural history collections, and efforts are being made through projects like Notes From Nature (http://www.notesfromnature.org/) and The Field Notebook Project at the Smithsonian (http://www.mnh.si.edu/rc/fieldbooks/) to digitize these resources and further supplement data from natural history collections.

Given that the intent of computerization was collection management-based and not research-oriented, and that consequently some data were left out of early electronic catalogs, modern natural history collection databases are undergoing a shift in purpose. The first web-accessible databases for natural history collection information, ORNIS, MaNIS, FishNET, and HerpNET, were taxonomically limited, and their intended purpose was to serve specimen data in the form of an online catalog (National Science Foundation). Today, the ideal database would capture every iota of information about a specimen in order to allow these data to be shared online and analyzed by researchers from diverse disciplines. Increased computing power has allowed 'big data' to become not just a trendy topic, but a reality that now faces every domain from the humanities to the sciences, and the biodiversity community is no exception.

Portals like Integrated Digitized Biocollections (iDigBio) and the Global Biodiversity Information Facility (GBIF) serve natural history collections data from around the United States and the globe, respectively. These sites allow users to search collections from diverse geographic locations in one virtual space. However, as I discovered, non-standardized database management systems and data entry practices are prevalent in the natural history museum community. This has implications for data quality and, in turn, data usability. While some data clean-up can be done post hoc, there are some problems, documented herein, related to training and standardization that could be addressed to improve the reliability of data from natural history collections as they are databased.

METHODS

Sample

The study population was natural history collection managers who work with a database on a regular basis to record and store specimen data. All of the collections at which those collection managers work are associated with large research universities in the United States and Canada. I recruited subjects by sending an email to collection managers at institutions of this kind. Due to travel limitations and a number of unanswered recruitment emails, this study uses a convenience, not systematic, sample. There were 19 subjects, representing 16 different collections. Participation was voluntary, and subjects did not receive compensation for their contribution to the study. All subject names, contact information, and employing institutions are confidential.

The decision to make the identities of subjects confidential was made in order to allow them the ability to speak freely without fear of repercussions. Although the names, employing institutions, and professional reputations of individuals may provide extra weight to the opinions and practices discussed in this study, the honesty of subjects was a top priority. Providing subject confidentiality allows subjects to speak freely about collection practices and institutional choices. I am confident that the absence of identifying information will not keep other museum professionals from recognizing similarities and differences between the cases examined here and their own collections.

Data Collection

I designed the interview questions and questionnaire to inform my research questions. Table 2.1 for how the interview questions relate to the research questions, and the questionnaire can be viewed in Appendix E. The questionnaires were intended to collect general information about the subjects, their educational backgrounds and experiences, and the collections where they work. The interview questions were meant to warrant responses about how work involving the database and specimen data curation is done at each collection.

I contacted each collection manager in advance to schedule data collection at his or her convenience. Collection visits took place over a span of six months – the first collection visit occurred on August 26, 2014, and the final one on January 8, 2015. Data collection took place in the offices, laboratories, or collections space of the participants, wherever the computer they most frequently use to access the database was kept. I described the purpose of the study and the nature of the observations, interview, and questionnaire, and gave the participant a consent form to read and sign. Once I received the signed consent form, an audio recording commenced.

The data collection process started with the subject performing a set of common tasks in the collection database. I took notes while watching the subject complete these tasks, and the subjects were asked to describe what they were doing as they went along to enhance the usefulness of the audio recording. In cases where I was unfamiliar with the particular database management software, I then repeated the database tasks. Following observations, the subjects were interviewed about how their database is used to record different kinds of data and how they feel about its overall usability. Finally, participants filled out a questionnaire detailing their educational background and relevant training. On average, each collection visit took approximately one hour. The observation and interview portions of the visits were audio recorded on a mobile device for transcription and analysis.

Data Management

In accordance with Institutional Review Board protocols, data from this study will only be kept for two years after collection. After that period, all physical documentation will be destroyed, and all digital files will be erased. The audio recordings are for research purposes only, and have only been accessible to me. All collected data, whether they are from audio, questionnaires, or interviews, were coded, and no personally identifiable information may be disclosed to third parties. Participants in the study have been assigned randomized ID numbers.

Data Analysis

I transcribed the audio from the collection visits, and the three modes of data collection – observations, interviews, and questionnaires – were combined to generate a comprehensive view of the training and database use of collection managers. The transcriptions were used for content analysis to identify themes in the data (Krippendorff 1980). I employed open coding methods, by which the list of codes characterizing each theme was not pre-determined, but was compiled during the data analysis process. The interview consisted of seeding questions, which were meant to get the conversations between the subjects and I started, but in most of the 19 cases, these topics mentioned in these questions were not the only things we discussed. If a subject thought a particular, unmentioned topic was important, then we talked about that as well. Additionally, I asked for any clarification or follow-up questions I felt necessary. Although the interviews did include certain questions, conversations were allowed to flow naturally, and usually, this resulted in more detailed responses from the subjects. The relationship between the interview seeding questions to specific research questions is presented in Table 2.1a and 2.1b.

Research Questions	Interview Questions
RQ1. Collection managers in the 21st century not only need content knowledge about their collections, but they are also effectively database managers. To what degree are natural history collection managers that interact with a database regularly able to effectively manage it?	• What database management software does your collection use? • In your opinion, how easy is this software to use? • What technical support is available for your database management software? Have you ever used this technical support, and if so, did it help you to resolve the problem you were experiencing? • Have you received any specific training to work with this database management system? • Does your database utilize controlled vocabularies? What are some examples of this? • How many staff and how many volunteers modify and enter data into your collection's database? • Within the database, how do you deal with specimens that may consist of multiple parts or pieces? • Do you always have the data required? If not, what do you do to fill informational gaps?

Table 2.1a: Research question one and the associated interview questions.

RQ2. Because databases play a key role in the organization and maintenance of natural history collections, how will the emergence of new forms of digital data impact these collections? How are the caretakers of collections preparing for these changes?	• Are there any fields or functionalities you wish your database management software had? • Are data about specimen loans recorded in the database? • If researchers analyze a specimen in your collection, are they required to return data from that analysis to you? If so, where and how are these data stored? • Is there a bibliography or record of scientific papers based on specimens from your collection? • What workflows have been developed by collections staff to allow digital data to be incorporated into the collection? • What are the future plans for your institution's collections, both physical and digital? • How much does funding impact these future plans? • Are there any other possible complications that might hinder these plans?

Table 2.1b: Research question two and the associated interview questions.

RESULTS

Participant and Collection Demographics

Nineteen natural history museum professionals from 16 diverse collections participated in this study. Among the 16 collections, seven use FileMaker to manage their data; two use Specify 6; two use Microsoft Access; two use Microsoft Excel; one uses Specify 5; one uses 4th Dimension; and one uses both Excel and FileMaker. On average, these database management systems have been in place for 11 years, and this ranged between 2 years to 22 years.

The sample also consists of all major types of natural history collections, including nine modern zoological, four paleontological, two botanical, and one genetic collection. Table 2.2 presents a further breakdown of the natural history disciplines represented in this study. Using the numbers reported in the questionnaires, these 16 collections contain an estimated 9,200,000 specimens. The average operating budget for each collection is approximately $8,000 per year. However, this ranged from $2,000 to $21,000, and these numbers reflect institutional support, not other funds from endowments or external grants.

Most of the participants self-reported their job title to be 'collection manager,' but there was some diversity in the responses. Table 2.3 shows the job titles reported by all 19 participants on the questionnaires. On average, each subject has been in his or her current position for 8 years, but this ranged from a minimum of 1 year to a maximum of 39 years.

NATURAL HISTORY DISCIPLINE	NUMBER OF COLLECTIONS SURVEYED
Vertebrate paleontology	1
Invertebrate paleontology	2
Paleobotany/micropaleontology	1
Total paleontological collections	4
Vascular plants/mycology/phycology/lichenology	1
Bryology	1
Total botanical collections	**2**
Mammalogy	1
Herpetology	2
Ornithology	1
Ichthyology	2
Invertebrate zoology	1
Entomology	1
Tetrapods	1
Total modern zoological collections	**9**
Tissues	1
Total genetic collections	**1**
TOTAL COLLECTIONS SURVEYED	**16**

Table 2.2: A full description of the different natural history collection types included in this study.

JOB TITLE	NUMBER OF RESPONSES
Collection manager	9
Assistant or associate collection manager	2
Curator	1
Curator/collection manager	1
Curatorial assistant or associate	3
Assistant curator	2
Research assistant	1
TOTAL	19

Table 2.3: Self-reported job titles of all 19 subjects.

Ease of Database Use

The majority of participants (16 out of 19) stated that their collection database management system was easy to use. However, almost every one of those 16 subjects went on to reveal that perhaps the database was not as easy to use as they originally thought. They explained that the database is not easy for beginners to use, but becomes easy "once you know it". R105, whose collection utilizes Specify 6, said:

> For basic curatorial needs, I think it's easy. Getting into deeper database, I don't know… There's a big learning curve. It depends on the task at hand, but for basic curatorial needs, pretty easy, pretty user-friendly.

Subjects that work with FileMaker or Microsoft Access often found its flexibility, specifically the ability to edit the fields or schema, to be one of the reasons the software is easy to use. Two of the three subjects who said their databases are not easy to work with use Microsoft Excel. One reported that the difficulty was because the database is not a relational format, which required data that were updated in one location of the database to be replicated elsewhere in order to maintain internal consistency. The other, whose collection uses both Excel and FileMaker to manage data, complained that the database has too many separate parts and different paperwork associated with each part. That participant, R115, prefers to enter data into Microsoft Excel, and then upload it to FileMaker on a monthly basis.

Cataloging Philosophies and Standardization

There was an array of personal views about the role of the cataloger. One subject, R105, explained that tweaking the data as you enter them is against the law of how cataloging is done, and that information should always be entered verbatim to how it was received from the donor. The only exception to this was if the donor made an obvious spelling error. Alternatively, some collections have internal data control, usually in the form of a Wiki page or a binder that defines the database fields and describes the standards for how those fields should be entered.

In some cases, the participants knew their data were not standardized, and that this sometimes made it difficult to search even within their own single-source database. As R101 explained:

> There's definitely more variety in that than is ideal for search purposes because it kind of depends on who catalogs it and who enters it, how various things are phrased... There's some variety, but mostly it fits under conventions, and you can usually find stuff by doing more open searches.

Database Training

Eight of the 19 subjects were self-taught, and two of them reported relying on database knowledge from a preceding job or coursework. If any training did occur, participants reported that the previous collection manager most commonly gave that training. In those cases, training usually lasted one day or even only a matter of hours. They were able to ask questions about how the database works during that time span, and beyond that, they were left to figure out how to solve any future problems on their own. Two other subjects reported receiving similar database training from a curator, and temporary or permanent information technologists trained two participants. Only one participant stated that he or she took a Microsoft Access course offered by the university associated with the collection.

R119 described the experience of trying to understand the collection database:

> I was told, "Oh, here's the login password. Why don't you look around and see what kind of files are there that the previous collection manager left you?" I've been trying to figure out where things were on there, and what sort of documentation might've been on her Access database, but she had a weird way of backing up her files. It was somewhat logical for certain things, but not logical for others... It took me six months to deconstruct her Access database to get her flat file back out.

Additionally, the collection managers are usually not the only ones in the collection doing data entry. The majority of subjects said they allow curators, graduate students, undergraduates, volunteers, work-study students, and other part-time staff members to enter data directly into the database. In some cases, other people are allowed to indirectly enter data by first cataloging specimens into an Excel spreadsheet, which is passed on to the collection manager and then vetted to the best of their ability.

The interviews revealed a spectrum of opinions on who should be allowed to enter data, and after how much training. At one end of the spectrum were participants like R105 who said: “Entering the new data is probably the easiest of tasks so that's handed down to the volunteer.” Alternatively, R108 said: “Usually I do not let my volunteers do the entry unless they've been with me for many years, and they've gone through the system.” R113 also said: “We don't want too many people mucking in our database. It's a very scary thing.” However, some participants’ opinions were more moderate, particularly those working in collections in which students and volunteers indirectly enter data into spreadsheets. In one case, different database training levels were coordinated with different access levels to the database. R114 explained:

> It depends on the access level you have. We do have separate access levels. I have full access, [the curator], and IT has full access. We also have student and volunteer accounts. The students can modify most fields. They cannot delete records. I think they can create records. They just can't delete any because that's not something we can undo. And then the volunteers are find only. That doesn't necessarily mean... There are some volunteers with student access. If somebody's been intro'd into the database, they get student access. If not, all you can do is search, not edit.

Technical Support

Technical support was unavailable in many collections. In most cases, participants devised their own methods for backing up digital assets from their collection. Two subjects reported that they include funds for technical support into their grant applications. However, I visited one institution that has its own internal IT department. This IT department works closely with the collection managers, and actually developed a script to standardize collection data before sending them to various portals or data aggregators. If the script returns an error about data from a specific collection, the IT staff informs the collection manager so they can attempt to fix the problem. Additionally, if collection managers at this institution wish to change their database schema, they must notify the IT department so appropriate changes can be made to the script.

Multiple subjects discussed frustrations about working with student or temporary IT staff. R117 said: "The guy that built my database is gone now. He was a student. I don't have anyone here anymore who understands my database, or how it works. So that's kind of a pain." In another case, a collection decided to stop doing georeferencing after the collection manager realized that the temporary IT staff member who had recently left had been double-checking the georeferences. R112 shared:

> She only left two to two-and-a-half years ago. It's taken me a full round of a couple years of how we're going to deal with this, which is why I had to drop geolocations because I realized she was checking them, and I didn't have time to check them. I don't think we should be doing this anymore until we have somebody to check them because I think we're causing more problems than good.

Specimen Loans, Publications, and Data Return

Most of the collections surveyed do not require researchers to give data back after a specimen is borrowed or destructively sampled. They do, however, request a reprint or a copy of the publication. Five subjects said they would prefer to have physical copies of the publication, and four said they would prefer receiving a PDF of the publication from the researcher. Six participants said they incorporate these publications either in the remarks field of the database or a specific citations field. One collection, which uses Specify 6, attaches publications or associated data to the specimen record in the database.

Five collections were uninterested in returned data of any kind unless the researcher changed taxonomic identification of the specimen.

Only two collections did not keep specimen loan records in their databases; instead, they were kept in a separate spreadsheet or in paper files. Of the remaining collections, nine have a dedicated museum registrar who tracks loans in a separate database, and seven of these also note the loan status of specimens within their own collection database.

Although most collections told researchers they wanted a copy of any resulting publications, this request is commonly ignored. It usually falls on the collection manager to find resulting publications and data from other sources, like Google Scholar or GenBank. In at least one case, the collection manager sends follow-up emails to researchers who borrowed specimens within the last few years to ask for copies of any publications. Thirteen of the 16 collections keep an archive – whether digital, physical, or both – of publications that relate to their specimens. The remaining three collections that do not currently have these records stated that this was on their 'to-do' lists.

Additionally, some participants mentioned the expansive archives associated with their collections. R112 gave an example of how these can prove helpful:

> What most people don't realize is that once your database is good enough, it actually corrects the collection. It corrects old labels, old mappings. We were able to database and image all our historical specimens... In the very beginning when they started being collected from 1894 to 1920. Now that we've done that, we can now say who John Davidson was and who Jack Davidson was. And who the other John Davidson was because we have two John Davidsons. They're very different people. We can say who Miss Abby is, and she had a daughter that was also called Miss Abby. We can now look up death records and figure out which Abby was it… It's as simple as being able to use your database to actually see the

> history of people and where they were at any point in time. That's what most people don't see in their database as a huge advantage.

Domain-Specific Digitization

Two participants, one from a mammalogy collection and one from a bryology collection, expressed that imaging their entire holdings would not be a productive use of resources. R108 said: “Dry bryophytes do not look that attractive. It’s just a clump of gray stuff… Text data is still more important.” R114 said: “One rodent in a drawer looks like another rodent… If we were to digitize anything, it would be tags.”

R112’s attitude toward digitization can be summarized by the following:

> For vasculars, if the whole place burns down, the images will be useful - label information, handwriting of people, the specimen itself, all kinds of information is in that picture. For bryophytes? I just don't think so. Lichens, not a lot of information off that little thing that looks like dirt. The label, possibly, the handwriting, the history, you're getting an inventory of that. Algae are very similar as well. It's so hard to identify algae off a picture.

Funding

Funding is a vital factor affecting the future capabilities of natural history collections. Eighteen of the 19 participants said funding impacts the future plans for their collection in some way. Specifically, funds are necessary to hire student workers, collaborators, and other staff to help with backlog or special projects. Substantial external grants are required to cover large undertakings, such as rehousing of specimens or digitization. Four collections had experienced budget cuts in recent years.

Incomplete or Inaccurate Databases

Due to the size of collections and the funding troubles faced by these institutions, it can take a long time for the entire holdings of a collection to be computerized, much less digitized. R108 described this problem and how it impacts the users of the collection data:

> Even though we've been databasing for over 30 years now, 70% of the collection is in our database. But that 30% is still a large number, but until we get the funding, that collection is going to be waiting. Quite often, people will look at our database, "You guys don't have such & such thing, it must be rare." Have you ever really talked to us and come and see? That is an extremely common species, and you're calling it rare? "You've only got 5 in there." That's because only 5 are in the database, but we've got maybe 500 yet to be databased. So, you do not call that rare. Quite frequently, people look & they do not really look into it in detail. Especially conservation data centers, they have this list of such-and-such species is rare, but we said, "No, it's not rare."

Furthermore, collection managers are aware that all of the information entered into these databases is not always correct. R112 explained: "I put it in, a lot of it's wrong. It gets out there. It gets corrected, and also, you'll start to see patterns. You can't see a pattern when you look at 150 cabinets, especially when the doors are closed. You can definitely see a pattern very quickly here."

Database Schemas

When I began this project, there was a third research question that I wanted to address: how do the structures of different natural history collection database schemas vary, and how well do those schemas match the information of the science they are mapping? However, my methods of surveying collection management professionals were insufficient to answer this question. This would require an in-depth look at different natural history database schemas used in each of these collections, and evaluating the schemas for their ability to fit data from different natural history domains. The prevalence of 'homegrown' natural history collection databases, which are created in programs like Microsoft Access and FileMaker, further complicate answering this question. Out of the 16 collections represented in this study, no two databases were identical, even among the collections that used Specify 6. Across the same discipline, the databases were also not the same. In part, this can be explained by the fact that a database is a tool. The customization of a database must suit the needs of the people who interact with it on a regular basis. However, there were some common themes that emerged in this study that may warrant future research concentrating on these problems.

Some collection managers felt they needed a database that is capable of handling parent/child or hierarchical relationships to convey the fact that there are multiple specimens of different species present on one physical object. For example, in invertebrate paleontology collections, it is common to find one slab of rock with several different species of organisms preserved on it. The micropaleontology collection faced a complicated problem in the cataloging of coal balls and their associated slides (Figure 2.1).

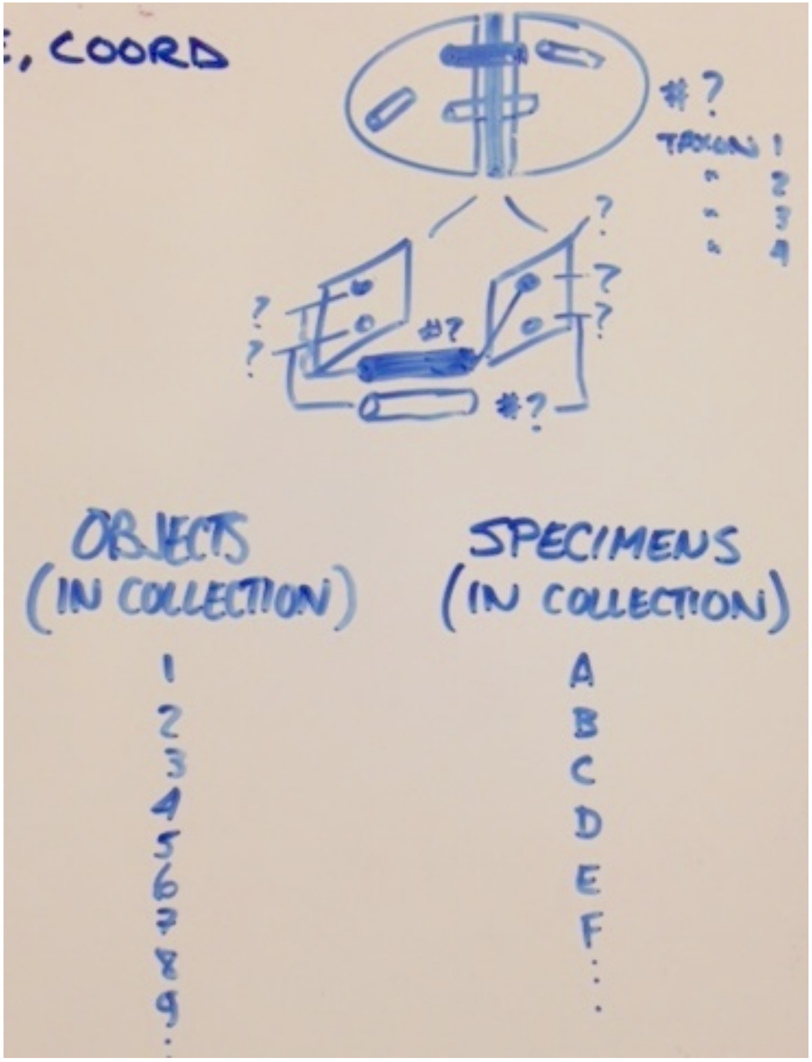

Figure 2.1 An image summarizing the problems of cataloging coal balls in a collection. Many specimens representing different species can be found on one slide.

Coal balls are permineralized plant material. By segmenting and making microscope slides of coal balls, researchers can identify different taxa of plants that existed in a distinct geographic and temporal area. These slides usually contain grains of pollen from different plant species. If a collection manager wanted to keep track of which types of pollen could be seen on which slides, they would need for their database to be able to record hierarchical relationships of taxa and the physical entity, which in this case is the slide. To maintain information on the place from which these slides came in the original coal ball would add another layer of hierarchical complexity to the database. This leads to some interesting cataloging problems in collections that maintain coal balls. R119 was struggling with these problems while trying to move the collection's Microsoft Excel data management system to a relational database. Is a catalog number representative of a physical object, or a distinct specimen with a single species identity? Depending on which option is chosen, there are implications for how these data are accurately stored in a database, how specimen numbers are relayed in publications, and even how loans are done. Sharing these data in a portal where different collections have taken a different approach would impact what search results look like to a user.

Future research should be done to look at problems such as these in detail and develop best practices for how to catalog and manage data associated with more complex types of natural history specimens. If collections come up with their own unique solutions to these problems, it will obfuscate searchability when data from different collections are mixed in a portal. It should also be noted that if a collection has already chosen a method for how they deal with these specimens in their databases, changing all of this to meet a new standard could be a large undertaking, especially if these changes impact a large number of specimens. Perhaps, instead of asking collections to unify the way complex specimens are cataloged, biodiversity informaticians could establish standards within Darwin Core or an ontology that would help untangle the data from these specimens at the portal-level rather than the collection-level.

DISCUSSION

We are in a transitionary period for natural history collection databases. A quote from R103 introduces this idea well:

> There are two schools of thoughts on databases - one of which is that a database is nothing but a big file cabinet, and it doesn't matter what information is there. All you want to do is be able to retrieve the object and know where it is and that it's in good shape. That's kind of the baseline museum database. Then there's the school of thought that you want to have all the words in those journals or whatever's in your file cabinet actually available in the database so you can look and see if there's any additional information… I've always looked at databases as a discovery tool. It's something that you can use to find things that aren't necessarily apparent.

Originally, natural history collection databases were intended for internal use only. Their purpose was to store basic information about specimens, and to facilitate locating specimens in the physical collection. With the emergence of digitization initiatives and data portals, these internal databases have been repurposed. ORNIS, MaNIS, FishNET, and HerpNET, NSF-funded projects to develop taxon-specific data portals, all began between 1999-2004. VertNet, which eventually replaced the four initial taxon-specific data aggregators, and iDigBio both began in 2011. The average age of the collection databases in this study was eight years; most of these systems predate the emergence of the most widely used data portals in the United States (National Science Foundation 2015). These initiatives, as R103 pointed out, are aimed at creating collaborative databases that can be used for discovery and analyses by researchers, policymakers, and the public. However, by putting together multiple institutional databases intended for collection management purposes, can we actually determine reliable, accurate patterns about national or global biodiversity?

Research Question One: To what degree are natural history collection managers able to effectively manage their collection's database?

Initially, this research question was worded: "To what degree do natural history collection managers that interact with a database regularly understand its structure and how to manage it effectively?" However, the data collected over the course of this study are more suitable to answer the question listed above. The database knowledge or training of a natural history collection professional is not the only factor that affects the quality of the data in a collection database. In fact, as shown by the data collected in this study, there are a number of different factors that may inhibit a collection manager's ability to enter a complete data set that accurately reflects every specimen in his or her care.

Natural history collection management professionals are almost all trained as biologists or paleontologists. In my sample, 12 of the 19 subjects have a Bachelor of Science degree, and five subjects have attained a PhD. Their degree specialties range from biology, botany, geology, and museum studies. However, little attention is paid to the fact that collection managers often have no formal training in database management systems. In most cases, they are left to learn on the job. If any training is offered, it is brief, and usually taught by someone who also has no formal database management training.

To add to the challenges presented by a lack of training in database management, participants in this study, on average, oversee six students, volunteers, or additional staff members. Only six of the 16 collections surveyed do not allow anyone but the collection manager or trained staff members to enter specimen data into the database. Attitudes about how much training is necessary for students or volunteers to become 'qualified' to catalog specimens vary, but 13 out of the 16 collections allow students or volunteers to directly enter data into the collection database.

Disparate ideas about how cataloging is done – is it verbatim to the donor, or do you make every effort to standardize it as it is entered? – further complicate the community's capability to standardize specimen data. If all collections across the globe do not have the same cataloging philosophy, the value of compiling collaborative databases diminishes. Although there is some worth to storing verbatim fields, all data must be standardized before they are sent off to various portals in order to be useful. If there is too much variety in the content of a field to be adequately searched in a single collection-wide database, incorporating these data into a portal where data is being sourced across multiple collections will only further reduce searchability.

If the goal is indeed to build research-quality data portals, formalized database training should be made mandatory for all collection management professionals in order to make the most of governmental and institutional investments. In this study, collection managers seem to prefer databases that give them the freedom to easily add or edit fields. If collection managers received thorough database training, they would be better prepared to make changes to schemas in the most efficient ways possible. This training should include a homogeneous cataloging philosophy, which would set a community standard for how to enter data. Collection managers should also be trained on how to train others to use the database. This would make them better equipped to pass along their knowledge to volunteers, undergraduates, and graduate students working in collections. This would also standardize the amount of training received prior to giving someone access to enter or edit data in the database.

Digitization of natural history collections creates valuable digital assets that require a lot of money and time to generate. Since 2011, the National Science Foundation's Advancing Digitization of Biodiversity Collections program has invested over $57,000,000 in grants to create Thematic Collection Networks (TCNs), collaborative cross-institutional networks that aim to digitize their collections in order to enable future research (National Science Foundation 2015). Therefore, properly backing up these digital assets that have been created through governmental support and taxpayer dollars should be a priority for any institution that is participating in digitization. However, this is frequently not the case. Collections professionals must often figure out how to back-up data on their own, and if they choose to contract out their data storage, this must be funded by external grants.

In cases where temporary IT staff or computer or information science students are hired to address the database needs of a collection, this can create a cyclical problem. Inevitably, the temporary person or student leaves the institution, and again, technical support falls to the collection manager. However, in this case, the person who understands the database is now gone. If any documentation is left behind to describe the methods employed to create the database, it may be inadequate or too difficult for a non-specialist to understand. This leaves the collection manager to deconstruct what has been done, or if the decision is made to wait until the next grant to address the problem, then the newcomer must spend time figuring out how the current database works.

The single institution studied with a dedicated museum IT department provides a model of how data curation and management can work differently. That institutional structure removes data cleaning and data sharing from the myriad of job responsibilities piled on to collection managers at other institutions. Moreover, these tasks are now delegated to IT professionals who have the expertise to navigate the technical aspects of data sharing via the Integrated Publishing Toolkit (IPT), a web application for publishing data on the internet.

In the future, it would be ideal for collection managers and IT professionals to have knowledge of each other's domains to ensure that natural history collections data are maintained accurately. However, creating an environment in which these two groups can work together on a long-term basis is another example of how natural history collections can be dutifully managed. If an institution can afford to hire in-house, permanent IT staff, these professionals can become invaluable in the digitization and maintenance of specimen data. This kind of specialization within institutions allows collection managers to focus more on managing the physical collections and less on trying to grapple with the intricacies of database management, digital asset storage, and the process of sending data to aggregators. As biodiversity informatics becomes an established field, graduate programs could train people to perform these job responsibilities.

Funding is an important issue that impacts almost every other previously discussed factor. Without appropriate funding, natural history collections cannot afford to purchase appropriate resources or to hire well-equipped staff to help with the backlog or the collection's IT needs of the collection. The one participant who said funding does not impact future plans of his or her collection has no intention of undergoing any large digitization or backlog projects, and this collection is located within the institution that has a dedicated IT staff. R115 explained how external grants allow collections to take on new projects and expand beyond their traditional operations: "We have enough annual funding to operate museum/library style. We can ship things out, bring things in. We have supplies to maintain the collection, but we do not have funding to do anything outside of that."

Without proper training, technical support, and funding, digitization and maintenance of these digital assets are insurmountable tasks. If the desired output of data portals is to provide researchers, policymakers, and educators with a comprehensive view of biodiversity as it is represented in natural history collections, these needs must be addressed. Collection managers and the students, volunteers, and staff who help enter data are the sources of all data seen on any online portal. If appropriate training and funding for the sources of these data are overlooked, then collection managers cannot be faulted for sharing incomplete, inaccurate data.

Research Question Two: Because databases play a key role in the organization and maintenance of natural history collections, how will the emergence of new forms of digital data impact these collections? How are the caretakers of collections preparing for these changes?

With the open-access movement, academic publishing companies, such as Elsevier B.V. and Nature Publishing Group are creating data repositories that allow scientists to get credit for data they have created (Shaklee et al. 2015; Nature.com). Given the importance of collection use and subsequent publications, natural history museums could also enforce stricter data policies to ensure that they are aware every time their specimens are used for scientific or educational purposes. Specimens are the underpinning of most research in the biological and paleontological communities. Developing new methods to track the data and consequent research value tied to natural history specimens could help improve the funding situation for the institutions that house those collections.

A few collection managers reported that they have created a Google Scholar account for the collection itself, and some others are planning to do this in the future. This feature on Google Scholar was originally intended to allow researchers to track their publications and citations, but these institutions are using it to track publications and citations where their specimens are being referenced. This, in turn, can help museum administration get a better sense of the amount of research their collections have aided. Although most institutions maintain archives of field notes, publications, permits, and other documents related to the specimen holdings, making this information available in a more easily accessible format could prove beneficial to other user groups beyond researchers. In R112's case, archives were used to augment historical knowledge of how the collection was built, and subsequently, collecting information, dates, and geolocations in the database were confirmed by these primary resources. Specifically in cases where the database information seems wrong, such as in R112's example where collectors have similar or even identical names, archives can prove to be a valuable resource to avoid mistakenly modifying these data.

Although archiving and linking today's diverse formats of derivative analytical data might be too much for a collection manager to handle on top of many other job responsibilities, archiving scientific data associated with specimens should become an expressed priority for institutions. Granted, this relies on researchers to properly curate their data so that they can prove useful to someone with no background knowledge of a given project. The open-access movement and the emergence of methods that allow researchers to gain credit for being a data producer, such as *Scientific Data* (http://www.nature.com/sdata/) and *DataLink* (http://www.datalink.com/), may help motivate researchers to share their data if this credit that can be used to their benefit for tenure or future grants. By tracking loan and publication information, perhaps advances in data sharing will show that the impact of natural history museums goes beyond a list of scholarly articles.

Collection managers, like R114 and R108, are keenly aware of what kinds of data are relevant to the scientific community in their domain. While macroscopic images of dry bryophytes or a drawer full of rodent skins may not have research value, there are other ways of imaging that can be used by different natural history disciplines to serve their users' needs. As R114 mentioned, digitization of specimen tags might be valuable. In the case of bryophytes or other groups for which diagnostic information exists at a microscopic level, microscopic imaging techniques may help supplement text data that is available online. These microscopic images would allow researchers to better search these collections and improve their ability to confirm taxonomic identifications. Imaging techniques should be domain-specific to serve the needs of specialized researchers. iDigBio has focused heavily on imaging techniques for different disciplines via workshops and working groups, but there is still room for new, creative approaches for lesser represented taxa such as bryophytes.

The word 'digitization' can have varying meanings to different stakeholders. Who are the intended users? This is an important question to ask as a collection is being imaged or text data are being shared. What types of information are valuable for domain-specific researchers? What kind of specimen imagery is useful in a K-12 classroom, undergraduate course, or to a domain specialist who is executing research? Although digitization of natural history collections can serve educational purposes, educators do not have time to share all 2-3 billion specimens (Krishtalka & Humphrey 2000) in collections worldwide with their classes. If digitization efforts are aiming to serve an educational role, the needs of those users are different than those of specialized scientists.

CONCLUSIONS AND FUTURE DIRECTIONS

For Biodiversity Informatics

Through the course of this project, the needs of the natural history museum community and of biodiversity informatics more broadly have become more defined. It is clear that data needs to become more standardized to facilitate searchability. It is clear that there are certain technological tools, such as citation tracking or databases that can handle complex hierarchical specimen relationships, that need to be built to serve the collections community. Collection managers cannot and should not have to bear the burden of digitization and managing the subsequent digital assets on their own. They are generally specialists in their taxonomic field, but this training does not include data management. A collection manager's specialized knowledge of a domain has value in other aspects of collection management, and I do not think the answer to the problems detailed herein is to replace collection managers with information scientists. Instead, I think it is important that the natural history community cultivate a new breed of professional that has knowledge in both data management and biology.

Given the amount of time and money being spent to develop these online resources and the tremendous potential they hold, the problems addressed in this study are not going to disappear. Biodiversity informatics should become an established discipline where students are trained in biological, computer, and information sciences. With this background, biodiversity informaticians could help biological collections manage their data, train collections staff on proper data entry procedures, train researchers to properly use informatics tools, construct new tools to advance the kinds of analyses and visualizations that can be derived from these data, and generally use technology and their expertise to streamline the data life cycle, making the whole process easier for all stakeholders. A large, collaborating community of biodiversity informaticians could address larger, community-level problems, such as creating taxonomic synonymy databases or creating better data standards and ontologies.

Despite this apparent need, at present, there are no universities in the U.S. that offer biodiversity informatics degree programs. This problem expands beyond the biological sciences. As data across the sciences becomes increasingly openly available and the reuse of these data in new research projects becomes commonplace, the need for professionals who have both domain-specific scientific knowledge and knowledge of information technology will undoubtedly increase. Data management cannot be passed down to research scientists or information technologists without a background in a given scientific area because in order to properly manage and store any kind of scientific data, knowledge of what these data are and how they can be reused is vital. An understanding of best practices of how to manage data is also essential. However, in most cases, these best practices do not yet exist due to the lack of specialized professionals who understand both the science and information management.

As of January 18, 2011, all National Science Foundation grant applications must include a data management plan, which details how resulting data and samples will be archived and disseminated for reuse (National Science Foundation 2014). The importance establishing biodiversity informatics as an established discipline is only one facet of this larger challenge facing modern science. If the goal is to create a scientific environment where data are freely shared and reused in new studies, then the development of new scientific informatics programs should become a priority in this country in order to train professionals who have the proper skills to address the issues associated with this broader goal.

For Downstream Users of Collections Data

Many users of biodiversity data portals may wrongly assume that it is an accurate, comprehensive representation of every applicable specimen within the geographic or taxonomic sphere of that database. This can be a treacherous assumption if a researcher publishes an analysis based on these data sets, or if a conservation decision is based on data from one of these sites. Collections data can be incomplete, representing only a small percentage of the actual holdings. Collections data can also be inaccurate, especially if collection managers upload their data to portals in hopes of receiving “free curation”. Beyond this, digitization itself is inherently biased by institutional collecting efforts (Soberón et al. 1996; Reynolds 1998; Peterson et al. 1998; Wilkes et al. 1999). Not only that, but as shown in this study, digitization projects are generally only funded by external grants. This adds more biases to biodiversity data sets because only institutions awarded grants in countries where digitization initiatives are supported are able to contribute their data. Assuming that any online database constitutes a complete inventory for any taxonomic group held in museum collections around the world shows a lack of understanding in how these databases are populated in the first place, and what kinds of biases are introduced during this process.

Although the objective is eventually to create freely available biodiversity data portals that offer a thorough view of natural history collection holdings, we have not yet reached this goal. For researchers aiming to mine these portals for publication-worthy data sets, it is important to note that these data represent only a fraction of specimens in collections, and that each field – whether it houses the taxonomic identification, its collection date, or location – is subject to change if new information emerges. For policymakers, conservation decisions based upon incomplete data could potentially be damaging. Different groups of users of natural history collections data must understand the context in which these data are made available because this has important implications for how they can be applied.

To some, the term ‘biodiversity informatics’ refers to the use of biodiversity data for research. This is undoubtedly a central purpose for the development of these online portals. It is important for biologists, ecologists, paleontologists, and researchers from any discipline that concerns itself with the study of the Earth’s biodiversity to become familiarized with the data resources and analysis tools that can be used to their advantage. Developing new tools to aid the analysis and visualization of biodiversity data would be a key role of biodiversity informaticians, and in turn, these professionals could help train scientists and students how to use these tools to further their research goals. Indeed, building new tools to serve the research and collections communities is a valid research pursuit on its own.

With the digitization of natural history collections and other ecological and observational data, the biological community has clear needs that are not being met by curators, collection managers, or research scientists. Given all of the money, time, and resources that have been put into digitization thus far, it is important that the community stays committed to fulfilling this goal of creating vast online resources about the changes to the Earth's biota through time and space in the best and most accurate way possible. These resources could reveal vital information about the present, past, and future of the Earth's fauna and flora that could guide policy decisions to slow current biodiversity loss. The development of biodiversity informatics as a discipline would professionalize research and career opportunities that would help overcome current challenges, and lead the way to creating the best data resources possible.

Appendices

APPENDIX A: Code for index.html

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <meta name="description" content="">
    <meta name="author" content="">
    <link rel="shortcut icon" href="#">

                <title>Home for Dino</title>

         <!-- Bootstrap theme -->
        <link rel="stylesheet" type="text/css" href="Bootstrap/css/bootstrap.css">
     <!-- Bootstrap responsive theme -->
    <link href="Bootstrap/css/bootstrap-theme.min.css" rel="stylesheet">
    <!-- Custom styles for this template -->
    <link href="Bootstrap/css/bootstrap-theme.css" rel="stylesheet">
        </head>
<!-- Helps to read page but may cause problems depending on web browser -->
        <body role= "document">

    <!-- Standard navigation bar, toogle-->
     <div class="navbar navbar-inverse">
       <div class="container">
         <div class="navbar-header">
           <button type="button" class="navbar-toggle" data-toggle="collapse" data-
target=".navbar-collapse">
             <span class="sr-only">Toggle navigation</span>
             <span class="icon-bar"></span>
             <span class="icon-bar"></span>
             <span class="icon-bar"></span>
           </button>
             <!-- need to link to homepage-->
           <a class="active" href="index.html"><h1>Dino</h1></a>
         </div>
         <div class="navbar-collapse collapse">
           <ul class="nav navbar-nav">
```

APPENDIX A: Continued

```
        <li class="dropdown">
             <!-- need to link to homepage section-->
           <a href="#" class="dropdown-toggle" data-toggle="dropdown">Test Your
Knowledge<b class="caret"></b></a>
           <ul class="dropdown-menu">
             <!-- need to link to Flash Cards webpage-->
           <li><a href="Dinos_TaxonFC.php">Flash Cards</a></li>
             <!-- need to link to Quiz webpage-->
             <li><a href="Dinos_List.php">Quiz</a</li>
             </li>
             </ul>
            <!-- need to link to Search webpage-->
            <li><a href="Dinos_Search.php">Search</a></li>
          </ul>
        </div><!--/.nav-collapse -->
      </div>
     </div> <!-- end navigation bar-->


        <div class="container theme-showcase" role="main">

               <!-- About Us area -->
                        <div class="jumbotron">
               <h1>About Dino</h1>
               <p class= "lead"> We compiled this database of dinosaurs for college
students taking <strong>GEO
302D Age of Dinosaurs.</strong> We help
students learn to identify the dinosaurs and the taxa they belong to using images of their
skeletons and the characters
that define each taxon.</p>
               </div>
               <br>

                      <!--Description of Resources -->
                      <div class="row">
                      <div class="col-xs-9">
        <div class="text-left">
                      <h2>What if I need to study?</h2>
                               <blockquote>You can search the dinosaurs first to learn about
taxa. This index
provides information on all of the dinosaurs and
```

APPENDIX A: Continued

```
taxa available on our web site.</blockquote>
                <p class="divider"></p>
               <h2>How do I test my knowledge?</h2>
               <blockquote> First, you can view the <strong>Flash
Cards</strong> of the different
which will show you either one dinosaur or a group of dinosaurs. Then, you guess the
taxon. Once you think you have the
answer, click the answer button to see if you're correct.
Once you think you are an expert, try the <strong>Quiz</strong>! you need to type the
answer. Don't worry; we provide
hints!</blockquote>
            </div>
        </div>
        </div> <!-- close theme -->
<hr>
<!-- Resources -->
<div class="row">
<div class="text-center">

     <div class="col-lg-4">
                  <h3>Search Dinosaurs</h3>
                  <p>Do you want to learn more about each
dinosaur?<br>Search here to learn the
smallest<br> monophyletic taxon, image, and characters.</p>
                  <form>
                <input type="button"
value="Search"onClick="window.location.href='Dinos_Search.php'" class="btn btn-lg
btn-warning">
               </form>
              </div>
           <div class="col-lg-4">
                  <h3>Taxa Flash Cards</h3>
                  <p>In this deck, you're given<br/>pictures of multiple
dinosaurs.<br>
You will guess which smallest monophyletic<br>group applies to all of the dinosaurs.
<br>
                  <form>
                <input type="button" value="Flash
Cards"onClick="window.location.href='Dinos_TaxonFC.php'" class="btn btn-lg btn-
info">
               </form>
```

APPENDIX A: Continued

```
                        </div>

        <div class="col-lg-4">
                                        <h3>Taxa Quiz</h3>
                                         <p>In this quiz, you're given<br>one picture of a specific
dinosaur and<br/>
have to name the smallest monophyletic<br>group this dinosaur belongs to. <br>
                                        <form>
                                   <input type="button"
value="Quiz"onClick="window.location.href='Dinos_List.php'" class="btn btn-lg btn-
success">
                                  </form>
                                </div>



                </div>
        </div>
        </body>

    <!-- Bootstrap core JavaScript
    ================================================== -->
    <!-- Placed at the end of the document so the pages load faster -->
    <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js"></script>
    <script src="Bootstrap/js/bootstrap.min.js"></script>
        </body>
                <hr>
                        <div class="footer">
                                <p>&copy; Brenskelle, Brown, Carter, Gao, and Sayre
2014</p>
                        </div>
                </div>
</html>
```

APPENDIX B: Code for Dinos_TaxonFC.php

```
 <!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <meta name="description" content="">
    <meta name="author" content="">
    <link rel="shortcut icon" href="#">

              <title>Dino Flash Cards</title>

        <!-- Bootstrap theme -->
       <link rel="stylesheet" type="text/css" href="Bootstrap/css/bootstrap.css">
     <!-- Bootstrap responsive theme -->
    <link href="Bootstrap/css/bootstrap-theme.min.css" rel="stylesheet">
    <!-- Custom styles for this template -->
    <link href="Bootstrap/css/theme.css" rel="stylesheet">
    <!-- jQuery Toogle which I use to first hide and then display an element: taxon -->
    <style type="text/css">
    #toggle1{
display:none;
}
</style>
<script src="http://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js">
</script>
<script>
$(document).ready(function(){
  $("#1").click(function(){
    $("#toggle1").toggle();
    });
   });
</script>

       </head>

<!-- Helps to read page but may cause problems depending on web browser -->
       <body role= "document">

    <!-- Bootstrap CSS navigation bar, toogle-->
     <div class="navbar navbar-inverse">
```

APPENDIX B: Continued

```
        <div class="container">
          <div class="navbar-header">
            <button type="button" class="navbar-toggle" data-toggle="collapse" data-
target=".navbar-collapse">
              <span class="sr-only">Toggle navigation</span>
              <span class="icon-bar"></span>
              <span class="icon-bar"></span>
              <span class="icon-bar"></span>
            </button>
              <!--link to homepage-->
            <a class="active" href="index.html"><h1>Dino</h1></a>
          </div>
          <div class="navbar-collapse collapse">
            <ul class="nav navbar-nav">
               <li class="dropdown">
              <a href="#" class="dropdown-toggle" data-toggle="dropdown">Test Your
Knowledge<b class="caret"></b></a>
              <ul class="dropdown-menu">
               <!--link to Flash Cards-->
                <li><a href="Dinos_TaxonFC.php">Flash Cards</a></li>
                 <!--link to Quiz-->
                <li><a href="Dinos_List.php">Quiz</a></li>
                </li>
                </ul>
                 <!--link to Search-->
               <li><a href="Dinos_Search.php">Search</a></li>
            </ul>
          </div><!--/.nav-collapse -->
        </div>
      </div> <!-- end navigation bar-->

<!-- In order to access the information, we have the login and opening of the connection -
->
<?php
$host = "localhost";
$user = "dino";
$password = "dinosaur";
$database = "dino";
$link = mysqli_connect($host, $user, $password, $database);
```

APPENDIX B: Continued

```
//Randomize the information pulled but make sure it is all associated; some is hidden for
verifying the correct answer
$rand = RAND(1,25);

//only relevant data pulled from the database; Laura determined which ones apply to this
question
$input = array("1", "4", "7", "12", "9", "13", "17", "21", "25");
$rand_key = array_rand($input);
$randtax = $input[$rand_key];

//our retrieved data from our tables
//since this website will be used in the future, certain changes to what is displayed for the
flashcards such as hints or dinosaur names could be included
$searchq = "SELECT dinotaxon.dinotaxon_id, dinosaur.dinosaur_id,
dinosaur.dinosaur_name, dinosaur.image_url, taxon.taxon_id, taxon.taxon,
dinotaxon.prime FROM dinotaxon, dinosaur,
taxon WHERE dinosaur.dinosaur_id=dinotaxon.dinosaur_id AND
dinotaxon.taxon_id=taxon.taxon_id AND taxon.taxon_id LIKE (?) GROUP BY
dinosaur.dinosaur_id ORDER BY RAND()";
$listresult = mysqli_prepare($link, $searchq);
// Below command binds variables for the parameter markers in the SQL statement that
was passed to above command
//corresponding variable type string so use letter s
mysqli_stmt_bind_param($listresult, 's', $randtax);
mysqli_stmt_bind_result($listresult, $dinotaxon_id, $dinosaur_id, $dinosaur_name,
$image_url, $taxon_id, $taxon, $prime);
mysqli_stmt_execute($listresult);
// Once it is prepared and determined to be correct, then it can be executed
//The above protects this page from mysqli injections
?>
<!--The php and html are seperated to ensure that even if a problem arises with the query,
something appears on the page-->
<!--In addition, it is easier to read the code and find errors when the two are separated.-->
<div class="row">
   <div class="container">
                  <?php
                  while (mysqli_stmt_fetch($listresult)){
                  //echo $dinosaur_name, $dinosaur_id ; //this is for our group reference to
make sure they are pulling correctly; users will not see the id
                  //This step is the first
                  ?>
```

APPENDIX B: Continued

```
			<!--Image of the dinos which corresponds with the taxon-->
			<img height='200' width='auto'
src='http://corvette.ischool.utexas.edu/dino/<?=$image_url;?>'>
			<?php
			}
			?>
   </div>
</div>

<div class="row">
   <div class="container">
		<h3 class="bg-warning"> What is the monophyletic group these dinosaurs all
belong to?</h3>
				<p class="divider"></p>
			<!-- jQuery Toogle which I use to first hide and then display an element:
taxon -->
				<p class="bg-success"><button id="1" class="btn btn-
success">Answer</button></p>
					<div id="toggle1" class="container">
					<!--Taxon of the dinos which corresponds with the
image(s) above-->
					<p class="bg-success"><strong><?php echo
$taxon;?></strong></p>
					</div>
				<p class="divider"></p>
				<hr>
				<!--refreshes page so that a new question is displayed; better
choice that running the same query twice in two different places in one file-->
	<p><a
href="http://corvette.ischool.utexas.edu/dino/Dinos_TaxonFC.php"><button
type="button" class="btn btn-sucess" action="#">Next Question</button></a><p>
   </div>
</div>
<p class="divider"></p>

<?php
mysqli_stmt_close($listresult);
mysqli_close($link);
?>
```

APPENDIX B: Continued

```
</body>

   <!-- Bootstrap core JavaScript
   ================================================== -->
   <!-- Placed at the end of the document so the pages load faster -->
   <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js"></script>
   <script src="Bootstrap/js/bootstrap.min.js"></script>
         </body>
                  <hr>
                           <div class="footer">

                                    <p>&copy; Brown, Brenskelle, Carter, Gao, and Sayre
2014</p>
                           </div>
</html>
```

APPENDIX C: Code for Dinos_List.php

```
    <!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <meta name="description" content="">
    <meta name="author" content="">
    <link rel="shortcut icon" href="#">

            <title>Dino Quiz</title>

       <!-- Bootstrap theme -->
      <link rel="stylesheet" type="text/css"
href="Bootstrap/css/bootstrap.css">
     <!-- Bootstrap responsive theme -->
    <link href="Bootstrap/css/bootstrap-theme.min.css" rel="stylesheet">
    <!-- Custom styles for this template -->
    <link href="Bootstrap/css/theme.css" rel="stylesheet">
    <style type="text/css">

    #toggle1{
display:none;
}
#toggle2{
display:none;
}
</style>
<script
src="http://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js">
</script>
<!-- Below, enclosed in the script tags, is the correct Javascript
function to make the buttons toggle. -->
<script>
$(document).ready(function(){
        $( "#1" ).click(function() {
                $('#toggle1').toggle();
        });
        $( "#2" ).click(function() {
                $('#toggle2').toggle();
        });
```

APPENDIX C: Continued

```
});
</script>
       </head>
<!-- Helps to read page but may cause problems depending on web browser
-->
       <body role= "document">

    <!-- Standard navigation bar, toogle-->
     <div class="navbar navbar-inverse">
       <div class="container">
         <div class="navbar-header">
           <button type="button" class="navbar-toggle"
data-toggle="collapse" data-target=".navbar-collapse">
             <span class="sr-only">Toggle navigation</span>
             <span class="icon-bar"></span>
             <span class="icon-bar"></span>
             <span class="icon-bar"></span>
           </button>
             <!-- need to link to homepage-->
           <a class="active"
href="http://corvette.ischool.utexas.edu/dino/"><h1>Dino</h1></a>
         </div>
         <div class="navbar-collapse collapse">
           <ul class="nav navbar-nav">
              <li class="dropdown">
             <a href="#" class="dropdown-toggle"
data-toggle="dropdown">Test Your Knowledge <b class="caret"></b></a>
             <ul class="dropdown-menu">
               <li><a href="Dinos_TaxonFC.php">Flash Cards</a></li>
               <li><a href="Dinos_List.php">Quiz</a></li>
               </li>
               </ul>
             <li><a href="Dinos_Search.php">Search</a></li>
           </ul>
         </div><!--/.nav-collapse -->
       </div>
      </div> <!-- end navigation bar-->

<?php
// Let's connect to the dino database...
$host = "localhost";
```

APPENDIX C: Continued

```
$user = "dino";
$password = "dinosaur";
$database = "dino";
$link = mysqli_connect($host, $user, $password, $database);

/* If the user has already submitted an answer, then one of two things
will happen. They either got the answer right or wrong. This is what
happens for each instance. */
if (isset($_GET['taxonguess'])) {
  $taxonguess = $_GET['taxonguess'];
  $taxonanswer = $_GET['taxonanswer'];
/* Below is a query which pulls the taxon_id from the previous page if
there was one, and checks the user's answer for the taxon name against
the right value. */
  $guessq = "SELECT dinotaxon.dinotaxon_id, taxon.taxon,
dinosaur.dinosaur_id, dinosaur.dinosaur_name, dinosaur.image_url,
taxon.taxon_id,
dinotaxon.prime,
characters.characters FROM dinotaxon, dinosaur, taxon, taxonchar,
characters WHERE dinosaur.dinosaur_id=dinotaxon.dinosaur_id AND
dinotaxon.taxon_id=taxon.taxon_id AND taxon.taxon_id=taxonchar.taxon_id
AND taxonchar.char_id=characters.char_id AND
taxon.taxon_id='$taxonanswer' AND
dinotaxon.prime LIKE 'Y' ORDER BY RAND() LIMIT 1";
  $guessresult = mysqli_query($link, $guessq);
  $guess = mysqli_fetch_array($guessresult);
  $taxon = $guess['taxon'];
  if ($taxonguess == "$taxon") {
    echo "<div class='alert alert-success'><center><p>
      <strong><h3>Well done!</strong> You're right! That was
$guess[dinosaur_name].<br/>Next question!</h3></center><br/>";
    $searchq = "SELECT dinotaxon.dinotaxon_id, dinosaur.dinosaur_id,
dinosaur.dinosaur_name, dinosaur.image_url, taxon.taxon_id,
dinotaxon.prime,
characters.characters FROM dinotaxon, dinosaur, taxon, taxonchar,
characters WHERE dinosaur.dinosaur_id=dinotaxon.dinosaur_id AND
dinotaxon.taxon_id=taxon.taxon_id AND taxon.taxon_id=taxonchar.taxon_id
AND taxonchar.char_id=characters.char_id AND dinotaxon.prime LIKE 'Y'
ORDER BY
RAND() LIMIT 1";
   $listresult = mysqli_query($link, $searchq);
```

APPENDIX C: Continued

```
  $row = mysqli_fetch_array($listresult);
/* If the user's answer was correct, the query above pulls another
random record for a new question and the BLOCK tag below loads the page
with this new record. */
    echo <<<BLOCK
    <div class="row">
            <div class="container">
                <p><img
src="http://corvette.ischool.utexas.edu/dino/$row[image_url]"
width="590px"
height="auto" class="img-rounded"><p>
           </div>
     </div>
</div>

<div class="row">
            <div class="container">
                <form name="input" action="Dinos_List.php"
method="get">
                <h3 class="bg-warning"> What is the smallest
monophyletic group that this dinosaur belongs to?</h3>
                <input type="text" size="45" placeholder="Check
your spelling carefully." name="taxonguess">
                <input type="hidden" value="$row[taxon_id]"
name="taxonanswer">
                <input type="submit" value="submit"></p>
                <p class="divider"></p>
           </div>
     </div>
</div>
</form>

<p class="divider"></p>
<hr>
<div class="row">
<div class= "container">
<div class="row">
</div>
<div class="row">
<div class="container">
  <p class="bg-success"><button id="1" class="btn btn-success">Need a
```

APPENDIX C: Continued

```
Hint?</button></p>
                    <div id="toggle1" class="toggleItem">
                    <p
class="bg-success"><strong>$row[characters]</strong></p>
             </div>
       </div>
</div>

<div class="row">
BLOCK;
        }
/* Now, if the user's answer is wrong, we reload the page using the
query up top where we pulled the same record from the database that was displayed
on the previous page. This prompts the user to try again. Ultimately, they cannot
advance to a new question if they don't get the right answer. */
     else {
     echo "<center><div class='alert alert-danger' class='text-
center'><h3><strong>Oops!</strong> Check your spelling and try submitting
again.</div></h3></center><br>";
     echo <<<BLOCK
     <div class="row">
               <div class="container">
                    <p><img
src="http://corvette.ischool.utexas.edu/dino/$guess[image_url]"
width="590px"
height="auto" class="img-rounded"><p>
             </div>
       </div>
</div>

<div class="row">
               <div class="container">
                    <form name="input" action="Dinos_List.php"
method="get">
                    <h3 class="bg-warning"> What is the smallest
monophyletic group that this dinosaur belongs to?</h3>
                    <input type="text" size="45" placeholder="Check
your spelling carefully." name="taxonguess">
                    <input type="hidden" value="$guess[taxon_id]"
name="taxonanswer">
                    <input type="submit" value="submit">
```

APPENDIX C: Continued

```
                    <p class="divider"></p>
                </div>
        </div>
</div>

</form>
<p class="divider"></p>
<hr>
<div class="row">
<div class= "container">
<div class="row">
<div class= "container">
  <p class="bg-success"><button id="1" class="btn btn-success">Need a
Hint?</button></p>
                    <div id="toggle1" class="toggleItem">
                    <p
class="bg-success"><strong>$guess[characters]</strong></p>
                </div>
        </div>
</div>


<div class="row">
BLOCK;
            }
   }
   else {

/* Finally, if no answer from a previous page was submitted, the page
loads normally for the first time, pulling a random record.
Below is the query for the random record. It is the same as the one at
the top that loads if you get a correct answer. */
$searchq = "SELECT dinotaxon.dinotaxon_id, dinosaur.dinosaur_id,
dinosaur.dinosaur_name, dinosaur.image_url, taxon.taxon_id,
dinotaxon.prime,
characters.characters FROM dinotaxon, dinosaur, taxon, taxonchar,
characters WHERE dinosaur.dinosaur_id=dinotaxon.dinosaur_id AND
dinotaxon.taxon_id=taxon.taxon_id AND taxon.taxon_id=taxonchar.taxon_id
AND taxonchar.char_id=characters.char_id AND dinotaxon.prime LIKE 'Y'
ORDER BY
```

APPENDIX C: Continued

```
RAND() LIMIT 1";

$listresult = mysqli_query($link, $searchq);
$row = mysqli_fetch_array($listresult);

echo <<<BLOCK

<div class="row">
          <div class="container">

               <p><img
src="http://corvette.ischool.utexas.edu/dino/$row[image_url]"
width="590px"
height="auto" class="img-rounded"><p>
          </div>
     </div>
</div>

<div class="row">
          <div class="container">
               <form name="input" action="Dinos_List.php"
method="get">
               <h3 class="bg-warning"> What is the smallest
monophyletic group that this dinosaur belongs to?</h3>
               <input type="text" size="45" placeholder="Check
your spelling carefully." name="taxonguess">
               <input type="hidden" value="$row[taxon_id]"
name="taxonanswer">
               <input type="submit" value="submit">
               <p class="divider"></p>
          </div>
     </div>
</div>
</form>
<p class="divider"></p>
<hr>
<div class="row">
<div class= "container">
<div class="row">
<div class= "container">
</div>
</div>
```

APPENDIX C: Continued

```
<div class="row">
<div class="container">
<!-- This is the coding for the toggle button, where row[characters] is
the hint. -->
                    <p class="bg-success"><button id="1" class="btn
btn-success">Need a Hint?</button></p>
                    <div id="toggle1" class="toggleItem">
                    <p
class="bg-success"><strong>$row[characters]</strong></p>
             </div>

</div>
<div class="row">
BLOCK;
}
mysqli_close($link);
?>

   <!-- Bootstrap core JavaScript
   ================================================== -->
   <!-- Placed at the end of the document so the pages load faster -->
   <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js"></script>
   <script src="Bootstrap/js/bootstrap.min.js"></script>
             <hr>
                    <div class="footer">
                           <p>&copy; Brown, Brenskelle, Carter,
Gao, and Sayre   2014</p>
<div class="container"></div>
             </div>
      </div>
</body>
</html>
```

APPENDIX D: Code for Dinos_Search.php

```
        <!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <meta name="description" content="">
    <meta name="author" content="">
    <link rel="shortcut icon" href="#">

                <title>Home for Dino Web Site</title>

         <!-- Bootstrap theme -->
        <link rel="stylesheet" type="text/css" href="Bootstrap/css/bootstrap.css">
     <!-- Bootstrap responsive theme -->
    <link href="Bootstrap/css/bootstrap-theme.min.css" rel="stylesheet">
    <!-- Custom styles for this template -->
    <link href="Bootstrap/css/theme.css" rel="stylesheet">
        </head>
        <!-- Helps to read page but may cause problems depending on web browser -->
        <body role= "document">

 <!-- Standard navigation bar, toogle-->
     <div class="navbar navbar-inverse">
       <div class="container">
         <div class="navbar-header">
          <button type="button" class="navbar-toggle" data-toggle="collapse" data-
target=".navbar-collapse">
            <span class="sr-only">Toggle navigation</span>
            <span class="icon-bar"></span>
            <span class="icon-bar"></span>
            <span class="icon-bar"></span>
          </button>
            <!-- need to link to homepage-->
          <a class="active" href="index.html"><h1>Dino</h1></a>
        </div>
        <div class="navbar-collapse collapse">
          <ul class="nav navbar-nav">
              <!-- need to link to homepage section-->
             <li class="dropdown">
               <!-- need to link to homepage section-->
```

APPENDIX D: Continued

```
        <a href="#" class="dropdown-toggle" data-toggle="dropdown">Test Your
Knowlegde<b class="caret"></b></a>
            <ul class="dropdown-menu">
<!-- need to link to Taxa webpage-->
              <li><a href="Dinos_TaxonFC.php">Flash Cards</a></li>
              <!-- need to link to Dinosaurs webpage-->
              <li><a href="Dinos_List.php">Quiz</a></li>
              </li>
              </ul>
             <!-- need to link to Search webpage-->
             <li><a href="Dinos_Search.php">Search</a></li>
           </ul>
         </div><!--/.nav-collapse -->
       </div>
     </div> <!-- end navigation bar-->



<div class="container theme-showcase" role="main">

              <!-- Search area -->
                     <!-- Create dropdown list form -->
                     <div class="jumbotron">
<p class="bg-warning">Select a dinosaur to learn more:<p>
<form name="dinosaurs" method="get" action="Dinos_Search.php">
<select name="dinosaur">
  <option value="Select A Dinosaur">Select A Dinosaur</option>
  <option value="Allosaurus">Allosaurus</option>
  <option value="Archaeopteryx">Archaeopteryx</option>
  <option value="Brachiosaurus">Brachiosaurus</option>
  <option value="Carnotaurus">Carnotaurus</option>
  <option value="Dilophosaurus">Dilophosaurus</option>
  <option value="Euoplocephalus">Euoplocephlus</option>
  <option value="Gallimimus">Gallimimus</option>
  <option value="Homalocephale">Homalocephale</option>
  <option value="Hypsilophodon">Hypsilophodon</option>
  <option value="Ichthyornis">Ichthyornis</option>
  <option value="Lesothosaurus">Lesothosaurus</option>
  <option value="Oviraptor">Oviraptor</option>
  <option value="Parasaurolophus">Parasaurolophus</option>
  <option value="Plateosaurus">Plateosaurus</option>
```

APPENDIX D: Continued

```
<option value="Psittacosaurus">Psittacosaurus</option>
  <option value="Scelidosaurus">Scelidosaurus</option>
  <option value="Stegosaurus">Stegosaurus</option>
  <option value="Syntarsus">Syntarsus</option>
  <option value="Triceratops">Triceratops</option>
  <option value="Tyrannosaurus">Tyrannosaurus</option>
  <option value="Velociraptor">Velociraptor</option>
</select>
<input type="submit" name="submit" value="Submit">
  </form>
 </div>
</div>

 <!-- PHP for pulling search onto this page -->

<?php
$host = "localhost";
$user = "dino";
$password = "dinosaur";
$database = "dino";
$link = mysqli_connect($host, $user, $password, $database);

//Conditionals set for nothing selected

if (isset($_GET['dinosaur'])) {
  $dinosaur= $_GET['dinosaur'];
  $dinosaur = preg_replace("/[^ 0-9a-zA-Z]+/", "", $dinosaur);
  // Use the dinosaur name to locate the dinosaur id in the dinosaur table
  $select = "SELECT dinosaur_id FROM dinosaur WHERE dinosaur_name =
'$dinosaur'";
  $dinosaur_id = mysqli_query($link, $select) or die("Error: ".mysqli_error($link));
  $return = mysqli_fetch_array($dinosaur_id);
  if (empty($return)) {
        echo "<div class='alert alert-danger' class='text-center'><strong>Oops!</strong>
Please select one dinosaur.</div>";
  }
  else {
  // Use the dinosaur_id to find the relationship of this particular dinosaur, its belonging
taxon and the taxon's characters
    $image_url = "SELECT image_url FROM dinosaur WHERE dinosaur_name =
'$dinosaur'";
```

APPENDIX D: Continued

```
            $image = mysqli_query($link, $image_url);
            $display = mysqli_fetch_array($image);
     $searchq = "SELECT dinosaur_name, taxon, characters FROM dinosaur, taxon,
dinotaxon, taxonchar, characters
             WHERE dinosaur.dinosaur_id = '$return[dinosaur_id]'
                                 AND dinosaur.dinosaur_id = dinotaxon.dinosaur_id
                                 AND taxonchar.taxon_id = dinotaxon.taxon_id
                                 AND characters.char_id = taxonchar.char_id
                                 AND taxonchar.taxon_id = taxon.taxon_id
                                 AND dinotaxon.prime = 'Y'";
            $listresult = mysqli_query($link, $searchq) or die ("Error:
".mysqli_error($link));
            $result = mysqli_fetch_array($listresult);

     $char = "SELECT characters FROM characters, taxonchar, dinosaur, dinotaxon,
taxon
                    WHERE dinosaur.dinosaur_id = '$return[dinosaur_id]'
                                        AND dinosaur.dinosaur_id = dinotaxon.dinosaur_id
              AND dinotaxon.taxon_id = taxon.taxon_id
                                        AND taxon.taxon_id = taxonchar.taxon_id
                                        AND taxonchar.char_id = characters.char_id
                                        AND dinotaxon.prime = 'Y'
                                        GROUP BY characters.char_id";
            $list_char = mysqli_query($link, $char);
     // Print the search result
     echo "<h2 class='text-center'>Here is your search result for $dinosaur: </h4><br>";
            echo "<hr>";
            echo "<h4 class='text-center'><img src = '$display[image_url]' width='590px'
height='auto'></h4><br>";
            echo "<hr>";
            echo "<h4 class='text-center'>The taxon of $result[dinosaur_name] is
$result[taxon].</h4>";
            echo "<hr>";
            echo "<h4 class='text-center'>$result[taxon] includes these characters: </h4>";
            // This statement was used to display multiple characters in separate lines
            while ($result_char = mysqli_fetch_array($list_char)) {
                echo "<h4 class='text-center'>$result_char[characters]<br/></h4>";
           }
}
}
mysqli_close($link);
```

APPENDIX D: Continued

```
?>

</body>

   <!-- Bootstrap core JavaScript
   ================================================== -->
   <!-- Placed at the end of the document so the pages load faster -->
   <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.0/jquery.min.js"></script>
   <script src="Bootstrap/js/bootstrap.min.js"></script>
        </body>
                <hr>

                        <div class="footer">
                                <p>&copy; Brown, Brenskelle, Carter, Gao, and Sayre
2014</p>
                        </div>
                </div>
</html>
```

APPENDIX E: Subject Questionnaire

Please fill out the following.

Name:

Email address:

Employing institution:

Job title:

Please answer the following questions with a short answer.

1. What is your educational background?

2. What job-specific/professional training have you received? Please circle all that apply.

   Taken coursework Attended community-specific workshops

   Attended software-specific workshops Institutional-specific training

   Other - please describe:

3. Did you receive a certification from completing any of kind of training?

4. What was the duration of this specific training?

5. How long have you been employed in your current position?

6. How large is your staff that you in particular oversee?

APPENDIX E: Continued

7. How many specimens does your collection contain?

8. What is the annual operating budget of your collection?

9. How many users does your collection have per year?

10. Which database management software does your collection use?

11. Are there any current plans to change the software? If so, in what timeframe?

12. How long has your natural history collection been utilizing the database software currently in place?

APPENDIX F: Observation and interview seeding questions for the collection visits

Observation Tasks:

During a collections visit, the participant will be asked to complete the following tasks:

- create a new record in the database,
- query the database to return a particular specimen,
- edit an existing record,
- query the database to return all specimens from a single locality,
- create a record for a new loan.

Interview Seeding Questions:

The following questions will be asked during the interview portion of the study.

1. In your opinion, how easy is this database management software to use?
2. What technical support is available for this database software?
    a. Have you ever utilized this technical support? If so, did the technical support help resolve the problems you were experiencing?
3. Have you received any specific training to work with this database management system?
4. How many staff and how many volunteers modify and enter data in your collection's database? What are the different ways these people use the database?
5. Does your institution use controlled vocabularies in your database?
    a. What are some examples of this?
6. Within the database, how do you deal with specimens that may consist of multiple parts or pieces?
7. Are data about specimen loans also recorded in the collections database?
8. Are there any fields or functionalities that you wish were your database management software had?
9. Do you always have the data required? If not, what, if anything, do you do to fill informational gaps?
10. If researchers analyze a specimen in your collection, are they required to return data from that analysis to you? If so, where and how is that data stored?
11. Is there a bibliography or record of scientific papers based on specimens housed in your collection?
12. What workflows have been developed by collections staff to allow digital data to be incorporated into the collection?
13. What are the future plans for your institution's collections, both physical and digital?
14. How much does funding impact these future plans?
15. Are there any other possible complications that might hinder these plans?

References

"(SGER) A Predictive Data Network for Fish Diversity: The FISHNET Prototype System." National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=9985737.

"ABI Development: Collaborative Research: VertNet, a New Model for Biodiversity Networks". National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=1062193.

"Chapter II - Proposal Preparation Instructions." National Science Foundation. December 26, 2014. Accessed November 23, 2015. http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp.

Barnosky, Matzke, Tomiya, Wogan, Swartz, Quental, Marshall, McGuire, Lindsey, Maguire, Mersey, and Ferrer. "Has the Earth's Sixth Mass Extinction Already Arrived?" *Nature* 471, no. 7336 (2011): 51-7.

Cowan, R. "The National Collections as Biological Standards". *The Proceedings of the Biological Society of Washington* 82 (1969): 611-17.

"Development of an Integrated Network for Distributed Databases of Mammal Specimen Data." National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=0108161.

"Digitization HUB: A Collections Digitization Framework for the 21st Century." National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=1115210.

Graham, Ferrier, Huettman, Moritz, Peterson. "New Developments in Museum-Based Informatics and Applications in Biodiversity Informatics." *Trends in Ecology & Evolution* 19, no. 9 (2004): 497-503.

Jenkins, M. "Prospects for Biodiversity". *Science* 302, no. 5648 (2003): 1175-77.

Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, 1980.

Krishtalka and Humphrey. "Can Natural History Museums Capture the Future?" *Bioscience* 50, no. 7 (2000): 611-17.

Kyoko, K. “Design, Development, and Implementation of Interactive Multimedia Software for an Introductory Geology Course”. M.A. thesis, The University of Texas at Austin, 1995.

“ORNIS: A Community Effort to Build an Integrated, Distributed, Enriched, and Error-checked ORNithological Information System.” National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=0345448.

Peterson, Navarro-Sigüenza, and Benítez-Díaz. “The Need for Continued Scientific Collecting; a Geographic Analysis of Mexican Bird Specimens.” *Ibis* 140, no. 2 (2008): 288-94.

Peterson, Knapp, Guralnick, Soberón, and Holder. “The Big Questions for Biodiversity Informatics.” *Systematics and Biodiversity* 8, no. 2 (2010): 159-68.

Pimm, Russell, Gittleman, and Brooks. “The Future of Biodiversity”. *Science* 296, no. 5222 (1995): 347-50.

“Principles.” Nature.com. Accessed May 13, 2015. http://www.nature.com/sdata/about/principles.

Reynolds, J. H. “Data Management Fundamentals.” In *WCMC Handbooks on Biodiversity Information Management*. Vol. 7. Cambridge: World Conservation Monitoring Centre, 1998.

Rowe, Kishi, Merck, Jr., and Colbert. 1998. *The Age of Dinosaurs*. Educational Interactive Multimedia on CD-ROM for Macintosh and PC computers. Third Edition. W. H. Freeman & Co.

Shaklee, Motter, and Zudilova-Seinstra. “DataLink Tool Helps Researchers Search, Cite, and Write About Data.” March 3, 2015. Accessed May 13, 2015. http://www.elsevier.com/connect/datalink-tool-helps-researchers-search-cite-and-write-about-data.

Soberón, Llorente, and Benítez. “An International View of National Biological Surveys.” *Annals of the Missouri Botanical Garden* 83, no. 4 (1996): 562-73.

Sunderland, M. E. “Computerizing Natural History Collections.” *Endeavour* 37, no. 3 (2013): 150-61.

"The HerpNET Community Informatics Project: Development of a Distributed Information Network of North American Herpetological Databases (HerpNET)." National Science Foundation. Accessed May 11, 2015. http://www.nsf.gov/awardsearch/showAward?AWD_ID=0132303.

Thomas, Cameron, Green, Bakkenes, Beaumont, Collingham, Erasmus, Ferreira de Siqueira, Grainger, Hannah, Hughes, Huntley, van Jaarsveld, Midgely, Miles, Ortega-Huerta, Peterson, Phillips, and Williams. "Extinction Risk From Climate Change." *Nature* 427, no. 6970 (2004): 145-8.

"What Has Been Funded (Recent Awards Made Through This Program, with Abstracts)." National Science Foundation. Accessed November 23, 2015. http://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=6895&#;results.

Wilkie, Cassis, and Gray. "Quality control in invertebrate biodiversity data compilations." *The other 99%: the conservation and biodiversity of invertebrates* (1999): 147-53.

Yochelson, E. "Fossils – The How and Why of Collecting and Storing". *The Proceedings of the Biological Society of Washington* 82 (1969): 585-601.