**Assignment 2**

Chloe Lee-Hone

CST8215 – Introduction to Database

Section 362

Professor Alemeseged Legesse

Algonquin School of Advanced Technology

05/12/2021

Table of Contents

[Introduction 4](#_Toc89551995)

[1. Abstract 4](#_Toc89551996)

[2. Business Rules 4](#_Toc89551997)

[3. Logical Entity-Relationship Diagram (ERD) 6](#_Toc89551998)

[4. Dataset origin: 7](#_Toc89551999)

[5. Metadata 8](#_Toc89552000)

[Referenced cited: 10](#_Toc89552001)

[Reference cited in Queries file and DDL file: 10](#_Toc89552002)

# Introduction

This project simulates the creation of a database for the United-States National Park Service (NPS). This document includes an abstract, business rules for each entity, a logical ERD, a brief outline of the changes brought to the original datasets, and a summary table documenting the metadata of all entities in the National-Parks database. The last page of this document contains the references cited and the dataset sources.

# 1. Abstract

The U.S. National Park Service (NPS) keeps records of the animal species, the trails, the state location of every national park, as well as every federal employee working for the Parks Service. Each state may contain zero or many parks, however a park may overstep state borders, and therefore be located in several states. Similarly, since many of the same animal species are found throughout North America, several parks may contain the same species. For the scope of this assignment, the species entity will represent a list of the animal species that can be found in each park, but will not include the individual animals in each one.

Each federal National Park Service employee has a specific role, and is part of a larger department under the umbrella of the NPS. In order to compartmentalize their work, every employee works for one specific national park. Lastly, each department has one headquarter located in one state.

This database contains 8 entities, two of which are associative.

# 2. Business Rules

**Park**

1. A park is distinguished by a unique four-letter code.
2. A park contains one or many walking trails.
3. A park may contain zero, one or many species.

**Trail**

1. Each trail is distinguished by a unique 8-number code.
2. A trail is located within one park.
3. A trail belongs to only one state.

**Species**

1. Each species is distinguished by a unique six-number code.
2. A species may be found in zero, one or many parks.

**State**

1. A state has a unique two-letter abbreviation code that distinguishes it.
2. Each state may contain zero, one or many parks.

**Employee**

1. Each employee has a unique four-number employee code.
2. Each employee works for one department within the National Park Service (NPS).
3. Every employee is assigned to work on one specific national park.
4. An employee does not have to reside in the same state as the park they work on.

**Department**

1. Each department has a unique department code composed of one letter and two numbers.
2. Each department has a main office located in one state.
3. A department may have zero, one or many employees.

**Park\_Species (associative)**

1. Each species may be present in zero, one or many parks.
2. A park may contain zero, one or many species.

**State\_Park (associative)**

1. A park may be located in one state, or may span several states.
2. A state may contain zero, one or many parks.

# 3. Logical Entity-Relationship Diagram (ERD)

# 4. Dataset origin:

In this section, I will discuss the original datasets and the modifications I brought to them in the context of this project.

I generated the data in the State table based on information on YourDictionnary [1]. The State-Park table, as well as the Park-Species tables, were generated based on the information in both those tables, as they are associative entities.

I created the data in the Employee table, though the role names come from the National Park Service website [2]. Similarly, for the Department table, I drew all but two of the department names, the “Animal Protection and Education Unit” and the “National Water and Geological Survey Department”, from the National Parks Service website [3].

The Trail table was modified from a Kaggle dataset published by a user name Jane [4]. For this project, I used a sample of this table’s data, as it originally contained 3313 rows [4]. I brought the following changes to the national\_park\_trails table [4]:

* I reduced the number of columns. Removed columns: area\_name, state\_name, country\_name, geoloc, popularity, length, elevation, difficulty\_rating, route\_type, visitor\_usage, average\_rating, num\_reviews, features, activities, and units [4].
* Changed “name” column to “trail\_name” [4].
* Added a park\_code as a foreign key.
* Removed values with no corresponding park code.

The Park table was modified from a dataset published by the National Park Service on Kaggle [5]. I brought the following changes to it:

* Changed the primary key from “Park Name” to “parc\_code” [5].
* Put everything in lower caps.
* Removed the State column.

The Species table was modified from a dataset published by the National Park Service on Kaggle [5]. For the scope of this project, I used a sample of this dataset, as it originally contained 119,248 rows of data [5].

* I changed “Park Name” to ”park\_name” [5].
* I changed the species’ ID to be exclusively numeric.
* I put everything in lower caps.
* Any space between names was changed to an underscore “\_”.
* I removed the columns Park Name, Family, Category, Record Status, Nativeness, Occurrence, Abundance, Seasonality, and Conservation Status [5].

# 5. Metadata

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table** | **Attribute Name**  character varying | **Data Type**  character varying | **Max Length**  integer | **Description** | **Source** |
| Department | dept\_id | character varying | 5 | Department primary key | Entered by user |
| dept\_name | character varying | 60 | Department name | Defined by the National Parks Service [3] and user |
| dept\_location | character | 2 | Department headquarter location | Entered by user |
| Employee | employee\_id | character | 5 | Employee primary key | Entered by user |
| fname | character varying | 30 | Employee's first name | Entered by user |
| lname | character varying | 50 | Employee's last name | Entered by user |
| address | character varying | 50 | Employee's address | Entered by user |
| postal\_code | character | 5 | Employee's postal code | Entered by user |
| phone\_number | character | 12 | Employee's phone number | Entered by user |
| email | character varying | 100 | Employee's email | Entered by user |
| job | character varying | 50 | Employee's job title | Entered by user |
| dept\_id | character | 5 | Employee foreign key | Entered by user |
| park\_id | character | 4 | Employee foreign key | Entered by user [5] |
| Trail | trail\_id | character | 8 | Trail primary key | Entered by user [4] |
| trail\_name | character varying | 50 | Trail name | Defined by the National Parks Service [4] |
| city\_name | character varying | 30 | City closest to trail | Entered by user [4] |
| park\_id | character | 4 | Trail foreign key | Entered by user [4] |
| Species | species\_id | character | 6 | Species primary key | Entered by user [5] |
| species\_order | character varying | 30 | Species order | Entered by user [5] |
| scientific\_name | character varying | 30 | Species scientific name | Entered by user [5] |
| common\_name | character varying | 50 | Species common name | Entered by user [5] |
| Park\_Species | species\_id | character | 6 | Park\_Species primary foreign key | Entered by user [5] |
| park\_id | character | 4 | Park\_Species primary foreign key | Entered by user [5] |
| Park | park\_id | character | 4 | Park primary key | Entered by user [5] |
| park\_name | character varying | 50 | Park name | Defined by the National Parks Service [5] |
| acres | bigint |  | Park's acres | Entered by user [5] |
| latitude | numeric |  | Park's latitude | Entered by user [5] |
| longitude | numeric |  | Park's longitude | Entered by user [5] |
| State | state\_id | character | 2 | State primary key | Entered by user [1] |
| state\_name | character varying | 20 | State name | Entered by user [1] |
| State\_Park | park\_id | character | 4 | State\_Park primary foreign key | Entered by user [5] |
| state\_id | character | 2 | State\_Park primary foreign key | Entered by user [1] |

# Referenced cited:

[1] J. Gunner, “List of All 50 US State Abbreviations,” *Your Dictionary*. <https://abbreviations.yourdictionary.com/articles/state-abbrev.html> (accessed Nov. 18, 2021).

[2] National Park Service, “Careers in the National Park Service (U.S. National Park Service),” *National Park Service*. <https://www.nps.gov/aboutus/nps-careers.htm> (accessed Nov. 18, 2021).

[3] National Park Service, “Organizational Structure of the National Park Service (U.S. National Park Service),” *National Park Service*. <https://www.nps.gov/aboutus/organizational-structure.htm> (accessed Nov. 18, 2021).

[4] Jane, “National Park Trails. Every trail in the National Parks Service gathered from alltrails.com,” *Kaggle*, Sep. 2019, Accessed: Apr. 11, 2021. [Online]. Available: <https://www.kaggle.com/planejane/national-park-trails>

[5] National Park Service, “Biodiversity in National Parks Plant and animal species found in the American national park system,” *Kaggle*, Jan. 2017, Accessed: Apr. 11, 2021. [Online]. Available: <https://www.kaggle.com/nationalparkservice/park-biodiversity>

# Reference cited in Queries file and DDL file:

[1] W. H. Osgood, Biographical Memoir of Clinton Hart Merriam 1855-1942, vol. XXIV. National Academy of Sciences of the United States of America, 1944.