Use Case 1 - Herbarium - Exercise Sheet

This use case is not graded, but instead is used for your notes.

# Exercise 1: Biodiversity Data Mobilization Planning

|  |  |  |  |
| --- | --- | --- | --- |
| **People and institutions** | **Affiliations** | **Stakeholder** | **Roles** |
| The University of White Plains |  |  |  |
| The Plant Biology Department |  |  |  |
| Professor of Plant Systematics |  |  |  |
| The Departmental Admin |  |  |  |
| Faculty and students |  |  |  |
| Two retired botanists who regularly volunteer |  |  |  |
| University’s central web-team |  |  |  |
| The Head of the Plant Biology Department |  |  |  |
| Hosted servers |  |  |  |
|  |  |  |  |
| **Rationale for your assignment** |
|  |

|  |  |  |
| --- | --- | --- |
| **Stage** | **Task** | **Role** |
| Initiating |  |  |
| Planning |  |  |
| Monitoring |  |  |
| Executing |  |  |
| Closing |  |  |
|  |  |  |
| **Rationale for your assignment** |
|  |

# Exercise 2: Biodiversity Data Capture

|  |
| --- |
| **Describe the actions you took to transcribe the herbarium specimens** |
|  |

# Exercise 3: Biodiversity Data Quality and Standardization

## 3a: Data validation check: data cleaning

|  |
| --- |
| **Description of the suggested data cleaning procedures** |
| **Column** | **Hint** | **Error** | **Type** | **Subtype** |
| **Country col.** | Check for blanks |  | Technical |  |
| **YE** | 2016 is the current year |  | Technical |  |
| **countryCode** | Code should include only letters |  | Technical |  |
| **MO** | JAN is 01, FEB is 02 and so on |  | Technical |  |
| **taxonRank** | Check for blanks |  | Technical |  |
| **Elevation** | 8872 is the elevation of Mt. Everest |  | Technical |  |
| **phylum** | Maybe a new Phylum of Plantae? |  | Technical |  |
| **coordinate Uncertainty** | Uncertainty in meters |  | Technical |  |

#

|  |
| --- |
| **Did you notice other types of errors?** |
|  |

## 3b. Data validation check: other data management tools

|  |
| --- |
| **Description of the suggested data cleaning procedures** |
| **Column** | **Hint** | **Error, solution and lessons learned** | **Type** | **Tools used** |
| **Full Name**  | Are the names valid? |  | Nomenclatural |  |
| **lat/lon** | Are all the coordinates consistent and in decimal format?  |  | Format |  |
| **lat/lon** | Are all the occurrences taking place in Guatemala? |  | Geographic |  |
| **eventDate** | Does the column exist ?  |  | Format |  |

##

## 3c. OpenRefine

|  |
| --- |
| **Compare/contrast data cleaning in OpenRefine vs Excel?** |
|  |

#

# Exercise 4: Biodiversity Data Publishing

|  |
| --- |
| **List the “Core” you used to publish the data? Why did you choose this core?** |
|  |
| **Describe the suggested data publishing procedure** |
|  |
| **Link to the result published dataset** |
|  |