

Report:

Data Project #2: Format Ocean Tracking Data and Report Data Center

File: otnunit_aat_datacenter_attributes_8a94_cefd_f8a3.csv

1. Datacenter_abstract consisted of data which wasn't readable except for one column, the remaining rows are also changed with the appropriate data.
2. Data center is derivative attribute of datacenter name. So, it can be deleted and ignored
3. Datacenter_licence has the same case just like datacenter_abstract, similar steps are followed.
4. The columns datacenter_distribution_statement, datacenter_date_modified, time_coverage_start and time_coverage have null in all fields so they can be ignored.
5. datacenter_pi_organization is taken as a primary key in the table datacenter_Unique_PrimaryKey. In the table datacenter_Normalised_ForeignKey, datacentre_pi_organization acts a foreign key in order to normalise the tables.

Output File: datacenter_Normalised_ForeignKey.csv; datacenter_Unique_PrimaryKey.csv;

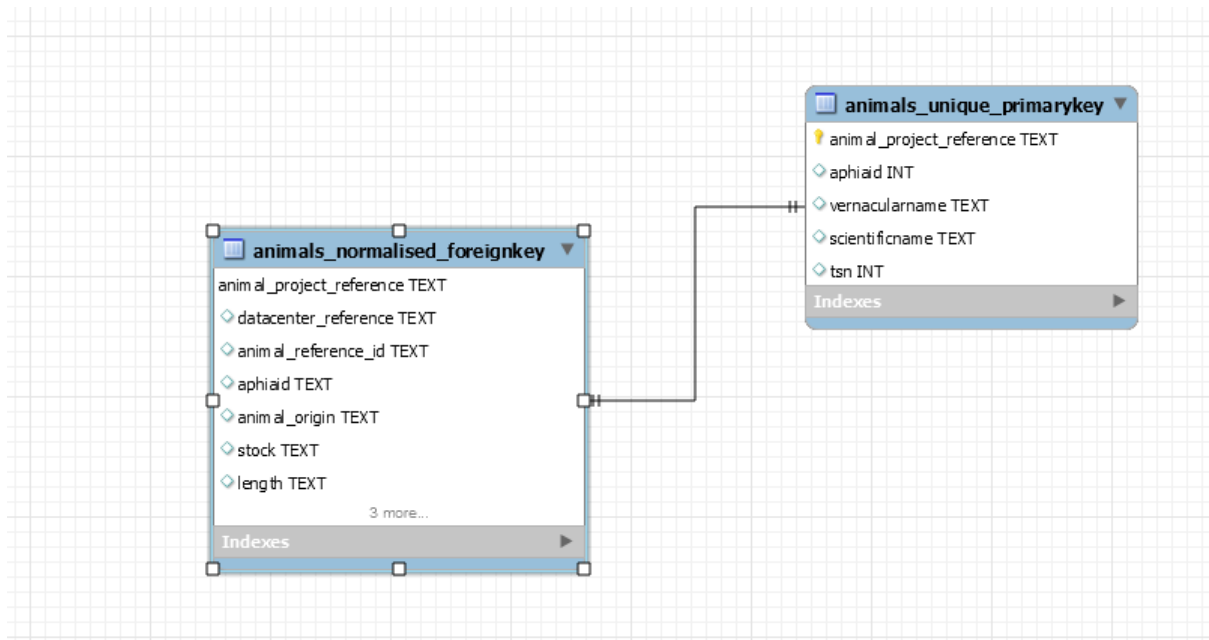
ER Diagram:

File: otnunit_aat_animals_8dc3_4d15_c278.csv

1. The attribute animal_guid is derived attribute of animal_reference_id which can be ignored.
2. Taxorank column have null in all fields, it can be deleted.
3. Age, Life_stage and sex column mentioned here has fewer values which can be ignored as it doesn't make difference in thousands of records.
4. Animal_project_reference is taken as a primary key in the table Animals_Unique_PrimaryKey. In the table Animals_Normalised_ForeignKey, animal_project_reference acts a foreign key in order to normalise the tables.

Output File: Animals_Normalised_ForeignKey.csv; Animals_Unique_PrimaryKey.csv;

ER Diagram:



File: otnunit_aat_detections_9062_5923_1394.csv

1. detection_guid is derived attribute from detection_id. So, it can be deleted
2. detection_transmittername is derived attribute of transmitter_id. So, it can be deleted
3. receiver_log_id has null in all fields so they can be ignored.
4. deployment_id is partially derived from detection_project_reference and detection_serial_number.
5. Sensor_data and sensor_data_units have fewer records compared to thousands which are even inappropriate and can be ignored and deleted
6. Detection_quality isn't accurate as only some records reflected "found receiver" so can be ignored.
7. All the records in depth, uncertainty_in_latitude, uncertainty_in_longitude, depth_data_source, uncertainty_in_depth, other_position_data, dataset_equality are null values so it can be deleted.
8. detection_reference_type has the same value "ANIMAL" for all the records so it can be mentioned in the documentation and ignored in the table.

Output File: otnunit_aat_detections_cleaned.csv;

File: otnunit_aat_manmade_platform_0735_7c9f_329c.csv

1. platform_type has same value for all the columns so it can be ignored and can be referred separately in the documentation.
2. Platform_guid is a derived attribute from the combination of data_center_reference and platform_reference_id.
3. Platform_reference_id and platform_name share the same values in all the rows. So, either of them can be ignored, and only one can be considered.

Output File: otnunit_aat_manmade_platform_cleaned.csv;

File: otnunit_aat_project_attributes_f29c_fb21_23a3.csv

1. Project_keyword, project_licence and project datum have similar values in all the columns, so it can be taken in another table by taking data_center as primary key in that table.
2. Data_reference is used as a primary key in the table Project_attribute_Unique_PrimaryKey in order to normalise it and acting as foreign key in Project_attribute_Normalised_ForeignKey which is in normalised form.
3. Project_references, project_distribution_statement, project_data_modified, project_linestring, geospatial_vertical_positive, time_coverage_start, time_coverage_end and project_doi have empty fields in all the records so can be removed from Project_attribute_Normalised_ForeignKey to further normalise it.
4. geospatial_vertical_min in the normalised table consist of fewer values like '0'. So, that can be deleted too.

Output File: Project_attribute_Normalised_ForeignKey; Project_attribute_Unique_PrimaryKey.csv;

File: otnunit_aat_receivers_c595_05f4_68b2.csv

1. expected_receiver_life and deployed_by has all the records as null. So, the entire column can be ignored.
2. Deployment_guid is derived attribute from combination of deployment_project_reference, datacenter_reference and deployment_id. So, it can be deleted as it's a derived attribute.
3. Array_name and deployment_project_reference share the same values so either of them can be ignored.
4. Receiver_reference_type can be set default value as "manmadePlatform", deleting the column can save space.
5. Receiver_reference_id is again the derived attribute of deployment_id which we can get rid of.
6. frequencies_monitored and receiver_coding_scheme can be deleted as they have all null valued for further normalisation.
7. Receiver_model and receiver_serial_number can be further derived by deployment_id. However, we can delete those columns.

Output File: otnunit_aat_receivers_cleaned.csv;

File: otnunit_aat_recover_offload_details_4b23_f002_f89a.csv

1. Recovery_guid is the derived attribute of recovery_id and deployment_id. It can be ignored.
2. offload_datetime_utc has few record set compared to thousands and remaining are null, we can get rid of it as it's negligible.
3. Columns clock_synchronized and recovered have All-Null values, ignored.
4. Note that recovery_id and deployment_id are different after certain thousands of records though they seem similar at start.

Output File: otnunit_aat_recover_offload_details_cleaned.csv;

File: otnunit_aat_tag_releases_b793_03e7_a230.csv

1. release_guid is the combination of datacentre_reference and tag_device_id, derived attribute, can be deleted.
2. Transmitter_id is derived from integer part of tag_device_id and even transmittername. So, these columns can be ignored.

3. Tag_device_id column has few records in different format, all the records have changed to prefix "A69-1303-"
4. Tag_coding_system is also a derived attribute, we can ignore it.
5. transmitter_type, tag_frequency and tag_programming_id are all-null fields which has to be deleted.

Output File: otnunit_aat_tag_releases_cleaned.csv;