27 January 2006

Task Description: Data Integrity Check and Finalization

Click to return to Database Documentation

The following queries can and should be run throughout the trapping season. It is most important that someone checks the data for the following conditions at the end of the season before the maps are produced for the annual planning meetings.

Check both the SITE_LOCATIONS and SITE_LOCATIONS_CTY tables for instances where the PLACED column = 'Y' and there are no records in either the PLACEMENT or OMITTED_SITES tables:

```
SELECT count(*)
FROM site_locations_cty s
WHERE (placed ='Y') and
(NOT EXISTS (SELECT *
FROM placement_cty p
WHERE (s.state = p.state and s.county = p.county AND s.block_id = p.block_id and s.site = p.site))
AND NOT EXISTS (SELECT *
From omitted_sites_cty o
WHERE (s.state = o.state and s.county = o.county and
s.block_id = o.block_id and s.site = o.site)));
```

Check both the SITE_LOCATIONS and SITE_LOCATIONS_CTY tables for instances where the PLACED column is NULL even though there is a record in either the PLACEMENT or OMITTED_SITES tables:

```
SELECT *
FROM site_locations_cty s
WHERE (placed is null) AND
(EXISTS (SELECT *
FROM placement_cty p
WHERE (s.state = p.state AND s.county = p.county AND s.block_id = p.block_id and s.site = p.site))
OR EXISTS (SELECT * FROM omitted_sites_cty o
WHERE (s.state = o.state AND s.county = o.county AND s.block_id = o.block_id AND s.site = o.site)))
ORDER BY county,block_id, site;
```

In the quad-based PLACEMENT and OMITTED_SITES tables, check the county column for a value of zero. This occurs when records for 'random' traps are inserted. The value of 0 is included for the county column. The GIS can be used to determine the correct county value based on the trap location (UTM coordinates). **NOTE: This should be completed prior to preparation of the NAPIS files because the data are organized by county**.

```
SELECT COUNT(*),state
FROM placement WHERE county = 0
GROUP BY state;
```

Compare the values in various columns between the SITE_LOCATIONS and PLACEMENT or SITE_LOCATIONS and OMITTED_SITES tables (also compare county data):

Check for mismatched county codes (quad-based only):

```
SELECT count(*),state
FROM placement P
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (P.quad = S.quad AND P.site = S.site AND P.state = S.state)
AND (P.county <> S.county)))
GROUP BY state;
```

```
SELECT count(*),state
FROM omitted_sites O
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (O.quad = S.quad AND O.site = S.site AND O.state = S.state)
 AND (O.county <> S.county)))
GROUP BY state;
```

Check for mismatched quad codes (county-based only):

```
SELECT COUNT(*),state
FROM placement_cty P
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (P.state = S.state AND P.county = S.county AND P.block_id = S.block_id
AND P.site = S.site) AND
```

```
(P.quad <> S.quad)))
GROUP BY state;

SELECT COUNT(*),state
FROM omitted_sites_cty O
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (O.state = S.state AND O.county = S.county AND O.block_id = S.block_id
AND O.site = S.site) AND
(O.quad <> S.quad)))
GROUP BY state;
```

Check for mismatched grid types:

```
SELECT COUNT(*), state
FROM placement P
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (P.quad = S.quad AND P.site = S.site AND P.state = S.state)
AND (UPPER(P.grid) <> UPPER(S.grid))))
GROUP BY state;

SELECT COUNT(*), state
FROM omitted_sites O
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (O.quad = S.quad AND O.site = S.site AND O.state = S.state)
AND (UPPER(O.grid) <> UPPER(S.grid))))
GROUP BY state;

SELECT count(*),state
FROM placement_cty P
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (P.state = S.state AND P.county = S.county AND P.block_id = S.block_id
AND P.site = S.site) AND (UPPER(P.grid) <> UPPER(S.grid))))
GROUP BY state;

SELECT COUNT(*),state
FROM omitted_sites_cty O
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (O.state = S.state AND O.county = S.county AND O.block_id = S.block_id
AND O.site = S.site) AND (UPPER(O.grid) <> UPPER(S.grid))))
GROUP BY state;
```

Check for mismatched grid node coordinates:

```
SELECT COUNT(*), state
FROM placement P
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (P.quad = S.quad AND P.site = S.site AND P.state = S.state) AND
(P.gridnode_utme <> S.utm_east OR P.gridnode_utmn <> S.utm_north)))
GROUP BY state;

SELECT COUNT(*), state
FROM omitted_sites O
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (O.quad = S.quad AND O.site = S.site AND O.state = S.state)
AND (O.utm_east <> S.utm_east OR O.utm_north <> S.utm_north)))
GROUP BY state;

SELECT COUNT(*), state
FROM placement_cty P
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (P.state = S.state AND P.county = S.county AND P.block_id = S.block_id
AND P.site = S.site) AND
(P.gridnode_utme <> S.utm_east OR P.gridnode_utmn <> S.utm_north)))
GROUP BY state;

SELECT COUNT(*), state
FROM omitted_sites_cty O
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (O.state = S.state AND O.county = S.county AND O.block_id = S.block_id
AND O.site = S.site) AND (O.utm_east <> S.utm_east OR O.utm_north <> S.utm_north)))
GROUP BY state;
```

Check for mismatched project areas:

```
SELECT COUNT(*), state
FROM placement P
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (P.quad = S.quad AND P.site = S.site AND P.state = S.state)
AND (UPPER(P.project) <> UPPER(S.project))))
GROUP BY state;

SELECT COUNT(*), state
FROM omitted_sites O
WHERE (EXISTS (SELECT * FROM site_locations S
WHERE (O.quad = S.quad AND O.site = S.site AND O.state = S.state)
```

```
AND (UPPER(O.project) <> UPPER(S.project))))
GROUP BY state;

SELECT COUNT(*), state
FROM placement_cty P
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (P.state = S.state AND P.county = S.county AND P.block_id = S.block_id
AND P.site = S.site) AND (UPPER(P.project) <> UPPER(S.project))))
GROUP BY state;

SELECT COUNT(*), state
FROM omitted_sites_cty O
WHERE (EXISTS (SELECT * FROM site_locations_cty S
WHERE (O.state = S.state AND O.county = S.county AND O.block_id = S.block_id
AND O.site = S.site) AND (UPPER(O.project) <> UPPER(S.project))))
GROUP BY state;
```

Check for mismatched quad codes between placement_cty and inspection_cty tables (at MSU):

```
SELECT COUNT(*) FROM mi_inspection_cty I
WHERE EXISTS (SELECT * FROM mi_placement_cty p
WHERE (i.state = p.state AND i.county = p.county
AND i.block_id = p.block_id AND i.site = p.site )
AND (i.quad <> p.quad));
```

Review the QC reports for each agency to check for problems that may have been overlooked, especially the 'Outside Target' report and the 'Inspection Distance' report.

A liaison in each agency should review the QC reports for data accuracy throughout the trapping season. It is especially important that all problems have been identified (and corrected if necessary) before the maps are produced for the planning meetings in the fall.

Scroll through the 'Outside Target' reports periodically and check some of the sites that are more than 1,000 meters beyond the edge of the target circle. ArcIMS or ArcView are very useful for this task. If the site appears to be within the target circle of another site, check with the agency liaison to determine whether the site was labeled with an incorrect quad and/or site number (county, block, or site for county-based trapping areas). In some cases, two sites may have been switched. You would find each trap site located within the other's target circle.

Another report to check closely is the 'Inspection Distance' report in which inspection records are listed if they are reported to be more than 150 meters from their respective trap placement locations. Sites are often included on this report when the records are submitted with incorrect site labels (quad and/or site in quad-based areas and county, block, and/or site in county-based areas). However, it is also possible that the inspection records are correct but the trap PLACEMENT records were labeled incorrectly. Both the inspection and placement records should be checked.

You will also need to look for a large number of records (say, in the hundreds) from an agency where there is a difference of approximately 200 – 300 meters between the UTM Northing of the inspection record and the UTM Northing of the placement record. This could occur if there is a difference in datums when the GPS records were downloaded. You would need to contact the agency liaison and ask that they check to be sure that the datum was set to NAD27 in all hardware and software used.