

Rank and Distribution Data Consistency User Guide for the Field and the Function Behind It

September 10, 2018 (G. Davis)

The Distribution Data and Rank Validation Function (DDF) compares data within a national or subnational element tracking record to check internal consistency. Certain ranks require certain distribution data to be present in the record; other combinations of values in these fields are inconsistent or not valid. [Rules for distribution data and rank consistency](#) followed by the function are described later in this document.

The function has been supplemented by the addition of a calculated field in Biotics 5 called “Rank and Distribution Data Consistency”. It appears in National and Subnational Element Tracking records for species under the distribution data grid. In the data model the field is called “DDF” and exists in the tables TAXON_NATIONAL and TAXON_SUBNATIONAL.

Biotics provides warnings and error messages when a record is saved with values in the national or subnational rank, distribution data grid, or hybrid indicator fields that are internally inconsistent or potentially so.

Definitions of Values

INVALID – ANIMAL RECORD MISSING POPULATION VALUE

For animal records only, the Population value is null for one or more rows in the distribution data grid. This field is required for each row entered for all animal records.

A record with this value in Rank and Distribution Data Consistency cannot be saved.

INVALID DATA COMBO

The combination of distribution data values, rounded rank (national or subnational), and/or the hybrid indicator value are internally inconsistent. See the [Business Rules](#) section of this document for a description of the rules of internal data consistency the function follows.

A record with this value in Rank and Distribution Data Consistency cannot be saved.

INVALID DATA COMBO - HYBRID WITH RANK NOT SNA

The ‘Hybrid without conservation value’ (hybrid_ind) field in the national or subnational tracking record is checked but the rank is not ‘SNA’ or ‘NNA’.

Note that the function currently accepts as valid a rank other than ‘NA’ if the Nname, Sname, Gname, or *global* hybrid indicator fields (taxon_global.g_hybrid_ind and taxon_global.d_hybrid_id) indicate that the taxon is a hybrid, as long as the ‘Hybrid without conservation value’ field in the national or subnational tracking record is not checked. It may be beneficial to do an independent review of these records.

A record with this value in Rank and Distribution Data Consistency cannot be saved.

INVALID EST IDENTIFIER NO_DATA_FOUND

The identifier for the given ENT or EST does not match any identifier in the ENT or EST (element_national or element_subnational) table.

A record with this value in Rank and Distribution Data Consistency cannot be saved.

INVALID - NO DISTRIBUTION DATA FOUND

There are no distribution data recorded in the ENT or EST (i.e., no taxon_natl_dist records for the ENT or no taxon_subnatl_dist records for the EST).

Temporarily, a record with this value in Rank and Distribution Data Consistency can be saved, however, trying to save a record without distribution data will result in an error in the long term.

INVALID RANK SYNTAX

Rounded rank (element_national.rounded_n_rank or element_subnational.rounded_s_rank) is 'INVALID.' The record cannot be assessed for rank and distribution data consistency. Note that the function will only check that the rounded rank field equals 'INVALID'. It is possible to have a rounded rank that is valid but out of sync with the rank field (e.g., by updating rank with sql instead of through the Biotics interface without also correctly updating rounded rank) in which case the DDF value will be determined using the rounded rank and may give an incorrect result when reviewed against the actual rank.

A record with this value in Rank and Distribution Data Consistency cannot be saved.

MULTIPLE RANKS - UNABLE TO EVALUATE

The rounded rank field contains a comma, which indicates more than one rank is applied to this element in this location (e.g., breeding and nonbreeding ranks are different). The function does not evaluate these records for valid combinations.

A record with this value in Rank and Distribution Data Consistency produces no warning message and can be saved.

NOTHING VALID OR INVALID FOUND

None of the combinations of distribution data row, rank, and the hybrid indicator value in this EST exists in the valid, valid but review suggested, or the invalid matrices (d_dist_valid_matrix, d_dist_valid_but_matrix, d_dist_invalid_matrix). For example, if a plant record has a value in the Population field (which is not visible in the user interface for plants), an otherwise valid combination will produce this value (this error is not tested by the function – [see Appendix B](#) – and should be identified using a separate QC).

A record with this value in Rank and Distribution Data Consistency produces no warning message and can be saved.

VALID DATA COMBO

The combination of at least one distribution data row, rank, and the hybrid indicator value in this record is considered valid.

A record with this value in Rank and Distribution Data Consistency produces no warning message and can be saved.

VALID DATA COMBO – BUT REVIEW SUGGESTED

The combination of distribution data row, rank, and the hybrid indicator value in this national or subnational tracking record is considered valid but review is suggested. Data in the record may be valid, but the combination so rarely applicable that it may be recorded in error. Comments in Distribution Data Comments should help explain the recorded values if they are recorded intentionally ([see Appendix A for more info](#)).

A record with this value in Rank and Distribution Data Consistency produces a warning message but can be saved once the values are reviewed and confirmed.

Notes about Hybrids

If the element has been assigned a rank of 'NA' because it is a hybrid, 'Hybrid Without Conservation Value' should be checked under Reason Rank "Not Applicable" in the national or subnational tracking record (taxon_national.hybrid_ind or taxon_subnational.hybrid_ind = Y). Without another reason in the data supporting a rank of 'NA' the function will return 'invalid' for the data combination *even if the element is flagged as a hybrid in the global tracking record* (taxon_global.d_hybrid_id = 1).

If a hybrid has 'Hybrid Without Conservation Value' checked and a rank of 'NA', the function does go on to confirm that the distribution data is also internally consistent. In addition, if at some point in the future we decide that the function should allow ranks other than 'NA' for hybrids without conservation value, the function can be modified at a high level to accept these and records for hybrids as well as for non-hybrids will still be analyzed for internally consistent distribution data and the 'NA' rank requirement ('NA' may be required independent of whether the taxon is a hybrid) due to the way the 'invalid' matrix (d_dist_invalid_matrix) is constructed.

Using the field in queries

To find records with each kind of data combinations, run the following query in the Biotics Query Builder or another SQL processor. In the WHERE part of the query, you can edit the value selected for DDF to be any value the function returns. For example, changing it to 'MULTIPLE RANKS – UNABLE TO EVALUATE' selects records with multiple ranks (usually for migrating species) for manual review. Changing it to 'NOTHING VALID OR INVALID FOUND' selects EST records with other combinations not evaluated by the function. [Appendix B](#) contains information about checks of the data the function does not perform.

```
/*
Query for plant and animal element_subnational records to find
inconsistent or invalid distribution data and rank using
Rank and Distribution Data Consistency (TAXON_SUBNATIONAL.DDF)

Records with the following TAXON_SUBNATIONAL.DDF values (reported in the
rank_dist_check column) can be selected:
```

```

INVALID - ANIMAL RECORD MISSING POPULATION VALUE
INVALID DATA COMBO
INVALID DATA COMBO - HYBRID WITH RANK NOT SNA
INVALID EST IDENTIFIER NO_DATA_FOUND
INVALID - NO DISTRIBUTION DATA FOUND
INVALID RANK SYNTAX
VALID DATA COMBO - BUT REVIEW SUGGESTED
MULTIPLE RANKS - UNABLE TO EVALUATE
NOTHING VALID OR INVALID FOUND
VALID DATA COMBO

```

Distribution data are reported in 2 columns (keep whichever you prefer):
 DISTRIBUTION_DATA: distribution data with field names concatenated in one column
 DIST_DATA_COND: distribution data without field names concatenated in one column
 (via delimlist)

AUTHOR: G. Davis

DATE: 12/22/14

*/

```

SELECT est.element_subnational_id,
nation.iso_nation_cd nation,
subnation.subnation_code subnation,
name_cat.name_type_desc name_type,
sname.scientific_name sname,
ts.ddf rank_dist_check,
est.s_rank srank,
est.rounded_s_rank rnd_srank,

```

/* DISTRIBUTION DATA */

```

DelimList('SELECT ' 'ORG: ' || substr(d_origin.origin_desc,1,1) ||
'; REG: ' || substr(d_reg.regularity_desc,1,1) ||
'; CNF: ' || substr(d_conf.dist_confidence_cd,1,3) ||
'; PRS: ' || substr(d_pres.curr_presence_absence_desc,1,1) || (case when
s_dist.d_population_id is not null then
'; POP: ' || substr(d_pop.population_desc,1,1) else null end)
FROM taxon_subnatl_dist s_dist, d_curr_presence_absence d_pres,
d_dist_confidence d_conf, d_origin, d_population d_pop, d_regularity d_reg
WHERE s_dist.d_curr_presence_absence_id = d_pres.d_curr_presence_absence_id (+)
and s_dist.d_dist_confidence_id = d_conf.d_dist_confidence_id (+)
and s_dist.d_origin_id = d_origin.d_origin_id (+)
and s_dist.d_population_id = d_pop.d_population_id (+)
and s_dist.d_regularity_id = d_reg.d_regularity_id (+)
and s_dist.element_subnational_id = ' || est.element_subnational_id, ' xx ') AS
distribution_data,

```

/* DISTRIBUTION DATA CONDENSED - NO FIELD NAMES */

```

DelimList('SELECT ' '|| substr(d_origin.origin_desc,1,1) ||
'. ' || substr(d_reg.regularity_desc,1,1) ||
'. ' || substr(d_conf.dist_confidence_cd,1,3) ||
'. ' || substr(d_pres.curr_presence_absence_desc,1,1) || (case when
s_dist.d_population_id is not null then
'. ' || substr(d_pop.population_desc,1,1) else null end)
FROM taxon_subnatl_dist s_dist, d_curr_presence_absence d_pres,
d_dist_confidence d_conf, d_origin, d_population d_pop, d_regularity d_reg
WHERE s_dist.d_curr_presence_absence_id = d_pres.d_curr_presence_absence_id (+)
and s_dist.d_dist_confidence_id = d_conf.d_dist_confidence_id (+)

```

```

and s_dist.d_origin_id = d_origin.d_origin_id (+)
and s_dist.d_population_id = d_pop.d_population_id (+)
and s_dist.d_regularity_id = d_reg.d_regularity_id (+)
and s_dist.element_subnational_id = ' || est.element_subnational_id, ' xx ' ) AS
dist_data_cond,

est.s_distribution_com est_dist_cmt,
ts.hybrid_ind s_hyb_wo_cons_val,
d_hybrid.hybrid_desc g_hybrid,
d_maint.maintained_by_status_cd maint

FROM element_subnational est,
element_national ent,
element_global egt,
taxon_global tg,
scientific_name gname,
scientific_name sname,
nation,
subnation,
taxon_subnational ts,
d_name_category name_cat,
d_hybrid,
d_maintained_by_status d_maint

WHERE est.element_national_id = ent.element_national_id
and ent.element_global_id = egt.element_global_id
and egt.gname_id = gname.scientific_name_id
and gname.d_name_category_id = name_cat.d_name_category_id
and egt.element_global_id = tg.element_global_id
and tg.d_hybrid_id = d_hybrid.d_hybrid_id (+)
and est.sname_id = sname.scientific_name_id
and est.subnation_id = subnation.subnation_id
and ent.nation_id = nation.nation_id
and est.element_subnational_id = ts.element_subnational_id
and est.d_maintained_by_status_id = d_maint.d_maintained_by_status_id
and egt.inactive_ind = 'N'

and name_cat.d_class_framework_id in (1, 3) /* plant, animal */

and ts.DDF like 'INVALID%' /* change this to any TAXON_SUBNATIONAL.DDF value */

order by
subnation.subnation_code,
name_cat.name_type_desc,
ts.ddf,
name_cat.name_type_desc,
sname.scientific_name

```

The Function Behind the Field – How it Works

This section describes how the function works for subnational element tracking records, but the same rules and process apply to national element tracking records.

The function **VALIDATE_EST_DIST** and its dependent functions **VALID_DIST** and **INVALID_DIST** use data from the following fields:

Field	Biotics Table and Column	Values
Element subnational tracking ID	element_subnational.element_subnational_id	ID
Rounded Subnational Rank	element_subnational.rounded_s_rank	SNR, SNA, SU, SX, SH, S1, S2, S3, S4, S5 (plus migrant modifiers on all ranks except SNA)
Hybrid Without Conservation Status (subnational)	taxon_subnational.hybrid_ind	Y, N
Name Type (zoological or botanical)	d_name_category.name_type_cd (from subnational scientific name)	A, P
Subnational distribution data ID	taxon_subnatl_dist.element_subnational_id	ID
Origin	taxon_subnatl_dist.d_origin_id	Native (1), Exotic (2), Unknown/Undetermined (3)
Regularity	taxon_subnatl_dist.d_regularity_id	Regularly occurring (1), Accidental/Nonregular (2), Unknown/Undetermined (3)
Distribution Confidence	taxon_subnatl_dist.d_dist_confidence_id	Confident (1); Reported but unconfirmed (2); Reported but doubtful (3); Reported but false (4); Potential (5); Potential, but false report exists (6); Never was there (7)
Current Presence Absence	taxon_subnatl_dist.d_curr_presence_absence_id	Present (1), Absent (2), Unknown/Undetermined (3)
Population	taxon_subnatl_dist.d_population_id	Year-round (1), Breeding (2), Non-breeding (3), Transient (4), Unknown (5)

For each Element Subnational Tracking record (EST) the function checks the internal consistency of the data in the fields above and compares data combinations from the record against rows in the following three domain tables to determine whether data in the record is valid or invalid.

D_DIST_VALID_MATRIX
D_DIST_VALID_BUT_MATRIX
D_DIST_INVALID_MATRIX

These tables have been copied into the Excel file DDF_matrices_2009_05_18.xlsx to ease viewing them, however, reviewing the matrices is not necessary to use the function.

The sequence of checks performed by the function is:

1. Check that EST ID is valid --- if not, stop: **INVALID EST IDENTIFIER NO_DATA_FOUND**
2. Check that the Srank is valid – if not, stop: **INVALID RANK SYNTAX**

3. Check that where 'Hybrid without conservation value' is checked in the *EST* (taxon_subnational.hybrid_ind = 'Y') Srank is 'SNA' --- if not, stop: **INVALID DATA COMBO - HYBRID WITH RANK NOT SNA**
4. Check for any distribution data rows in the grid --- if not, stop: **INVALID - NO DISTRIBUTION DATA FOUND**
5. Check for multiple ranks in Srank (rank field has comma) --- if so, stop: **MULTIPLE RANKS - UNABLE TO EVALUATE**
6. Check for a value in Population for animal records --- if Population is null, stop: **INVALID - ANIMAL RECORD MISSING POPULATION VALUE**
7. Check for the combination of name type, rounded Srank, hybrid indicator, and *any* distribution row in this record in the valid matrix – if so, stop: **VALID DATA COMBO**
8. Check for this combination in the valid_but matrix --- if so, stop: **VALID DATA COMBO – BUT REVIEW SUGGESTED**
9. Check for this combination in the invalid matrix --- if so, stop: **INVALID DATA COMBO**
10. If there is no value assigned up to this point: **NOTHING VALID OR INVALID FOUND**

In records where there are multiple rows in the distribution data grid, as long as one of the rows combines with other data in the record (such as rank) to make a valid combination, the record is identified as valid. If there are other invalid distribution data rows they will not be identified by the function.

Records may have more than one invalid combination, but the function reports the first invalid issue it detects. For example, if a record has an invalid rank for a hybrid, other data in the record will not be checked. Records found to be invalid or valid with review suggested should be reviewed completely to detect any secondary errors. Alternatively, once the first errors found are corrected, the function can be queried a second time to detect any remaining issues and to make sure new errors have not been introduced.

The function does not evaluate records where there are multiple ranks in the Srank field, as in the case of some migrating animals; these records are identified by the function but must be reviewed manually.

Distribution Data and Rank Consistency Business Rules
followed by the Distribution Data Function for EST (same rules apply to ENT)

General

Every EST record must have at least one row in the distribution data grid.

If the Srank is a number rank, there must be one row in the distribution data grid with Native (1), Regular (1), Confident (1), Present (1). **Exception** ([see Appendix A: VALID DATA COMBO BUT REVIEW SUGGESTED](#)): it is valid but flagged for review if there is a number rank but Origin is Unknown (3).

Distribution data must be internally consistent regardless of rank (e.g., combining Reported but Unconfirmed with Currently Present in the same row is invalid, regardless of whether the rank is 'SNA'). This is also true for hybrids (distribution data must be internally consistent even if the EST identifies the element as a hybrid without conservation value and the rank is 'SNA').

Origin

If in the only distribution row the **Origin is Exotic** (2), rank must be SNA.

If in the only distribution row the **Origin is Unknown** (3), rank should be SNR or SU. The rank SNA is discouraged because an element that has not been determined to be exotic may turn out to be a conservation target and the rank SNA may be interpreted to mean that it has been determined not to be. A number rank is allowed in rare cases where it is critical to indicate rarity in the event the element is found to be native. For these cases, distribution comments should explain the situation. In these cases (number rank or SNA without another reason that rank is SNA), the record will be flagged for verification ([see Appendix A: VALID DATA COMBO BUT REVIEW SUGGESTED](#)).

Regularity

If in the only distribution row the **Regularity is Accidental** (2), rank must be SNA.

If **Regularity = Accidental** (2) and **Confidence = Confident** (1) then **Current Presence Absence** can be any value; rank must be SNA

Recording **Regularity = Unknown** (3) is not recommended for most situations. If the rest of the distribution data suggest a rank is applicable and is internally valid the record will be flagged for verification ([see Appendix A: VALID DATA COMBO BUT REVIEW SUGGESTED](#)).

Distribution Confidence

If **Distribution Confidence** is *not* **Confident** (2-7) then **Current Presence/Absence cannot be Present** (1). If Distribution Confidence is **Reported but unconfirmed** (2), rank must be SNA or SU. If Distribution Confidence is anything *other than* **Confident** (1) or **Reported but unconfirmed** (2), rank must be SNA.

If **Distribution Confidence** is **Reported but doubtful** (3) then **Current Presence/Absence** must be **Unknown** (3). Reported but ‘doubtful’ implies there is uncertainty about the current presence; if an element is confirmed to be absent then, not only should Current Presence/Absence be changed to Absent, but ‘Reported but doubtful’ also should be changed to ‘Reported but false’ or ‘Reported but unconfirmed’ (if the original report could not be definitively rejected).

If Distribution Confidence is **Reported but false** (4) then **Current Presence/Absence CAN be Unknown or Absent** because there may be a false report AND someone has determined that the element is currently absent in which case it can be argued that recording ‘Absent’ is more logical. But in this case, there should be an additional row where Distribution Confidence = Never was there (7) and Current Presence/Absence = Absent (2). Note that the function will not return ‘INVALID’ if this row is not present for such a case. The rank must be **SNA**.

If **Distribution Confidence** is **Potential** (5) or **Potential but false report exists** (6) then **Current Presence/Absence CAN be Unknown** (3) or **Absent** (2) because the element may be known to be absent with potential to migrate there due to proximity to an existing occurrence and/or suitable habitat. The rank must be **SNA**.

If Distribution Confidence is **Never was there** (7) then **Current Presence/Absence** must be **Absent** (2) and rank must be **SNA**

Current Presence Absence

If the rank is **SH**, then there must be a distribution row with Native (1), Regular (1), Confident (1), **Unknown** (3). If there is a distribution row with Native, Regular, Confident, and Unknown, the rank can be SH, SNR, or SU.

If the rank is **SX**, then there must be a distribution row with Native (1), Regular (1), Confident (1), **Absent** (2). If there is a distribution row with Native, Regular, Confident, and Absent, the rank should be SX. In rare cases SU can be entered but these should be confirmed so the function will flag these as valid with suggested review ([see Appendix A: VALID DATA COMBO BUT REVIEW SUGGESTED](#)).

Populations (zoological elements only)

For all zoological elements (“animals”), Population must have a value (i.e., must not be null) for all rows entered in the distribution data grid. The Biotics system allows entry and saving of data without a value in Population because botanical elements (“plants”) do not require data in this field (nor can a value be entered for botanical elements in the Biotics interface), however, it is required for all animals so the function checks for this.

Where a single Srank ends in **B**, **N**, or **M**, there must be a corresponding row in the distribution table where Population is **Breeding** (2), **Non-breeding** (3), or **Transient** (4), respectively. Likewise, if there is a distribution row where Population is Breeding (2), Non-breeding (3), or Transient (4) and there is no row where Population is Year-round (1) and the Srank is not SNA, then the Srank must have a B, N, or M modifier as appropriate.

Note about multiple ranks: Records where Srank contains multiple ranks (separated by a comma) are not analyzed by the function. For these records, the function will return the value, 'MULTIPLE RANKS – UNABLE TO EVALUATE.'

Rank Not Applicable (SNA)

Where the element is flagged as a '**hybrid without conservation value**' at the subnational level (taxon_subnational.hybrid_ind = Y), the Srank must be **SNA**. If it is not, distribution data are *not* checked by the function.

If distribution data suggest a **rank is applicable** (Native – Regular – Confident), the Srank must not be SNA. **Exception** (see [Appendix A: VALID DATA COMBO BUT REVIEW SUGGESTED](#)): if a breeding (2) or nonbreeding (3) population is Native, Regular, Confident but there are no definable EOs, SNA is valid. Since these cases are rare, the combination is called valid with review suggested; distribution comments explaining the details of the situation will speed review.

If distribution data indicate a **rank is not applicable** [Origin Exotic (2), Regularity Accidental (2), or Distribution Confidence is not Confident (1) or Reported but Unconfirmed (2)], then rank must be SNA. It cannot be SNR or SU. If Distribution Confidence is Reported but Unconfirmed (2), rank may be SNA or SU.

If factors determining rankability are uncertain (**Origin** or **Regularity** are **Unknown** (3)), then rank can be SNR, SU, or SNA. However, records for non-hybrids where the Srank is **SNA** and **Confidence is Confident** (1) will be flagged for review/verification (even though Biotics Help suggests this is the appropriate rank) because the recorded distribution data does not indicate conclusively that a rank is not applicable. Distribution comments explaining the details of the situation will speed review. This includes cases where Origin and/or Regularity are Unknown and Current Presence Absence is Unknown and rank is SNA.

APPENDIX A

SUMMARY of cases where function = VALID DATA COMBO - BUT REVIEW SUGGESTED

The following cases will need manual review to verify that the entry is correct. Entering Distribution Comments is recommended for the exceptional, allowable cases so they can be quickly reviewed and accepted.

- Breeding and/or nonbreeding populations of the element are Native (1), Regular (1), and Confident (1) in the subnation but the rank is SNA (any Current Presence/Absence, Breeding or Nonbreeding population): These populations should have an SNA rank only if the taxon occurs often in jurisdiction but not at a regular location so there are *no definable EOs*. The entry of comments in the 'Distribution Comments' field is encouraged to facilitate quick review. For example, see the Pine Siskin in North Carolina example under case #3 in the [Protocol for Assignment of NatureServe Conservation Status and Distribution Status for Long Distance Migrants](#) document.
- **Origin Unknown – Number Rank:** When the origin of the element cannot yet be determined to be native but is not known conclusively to be exotic either, a number rank can be entered in very limited circumstances such as when an indication of rarity is critical to convey the fact that a taxon of possible conservation value is rare in a location where it is possibly but not known for certain to be native (e.g., native nearby). Distribution comments for these cases will facilitate quick review.
- **Origin Unknown – Rank SNA** – If an element is not confirmed to be exotic, it could eventually be found to be a conservation target, therefore ranking it SNA rank is not recommended (SNR or SU are suggested). However, because Biotics Help (under the task [Enter Distribution Data for Species Elements](#), Additional guidelines, section #1) does say that the NA rank should be applied if Origin is *other than* Native this combination is valid but should be reviewed to insure it has not been recorded in error. Distribution comments for these cases are encouraged to allow quick elimination of these records as errors.
- **Regularity Unknown** – Biotics Help (under the task [Enter Distribution Data for Species Elements](#), Additional guidelines, section #2) recommends against creating tracking records for situations where the element cannot at least be confirmed as an accidental. It is possible that the regularity of the element in the jurisdiction cannot be confirmed as regular so the Unknown value is technically valid. However, review of most records where Regularity has been recorded as Unknown has shown that this value is usually recorded in error. Therefore, this entry will always be flagged as one that should be confirmed. Entry of Distribution Comments where this value has been verified will speed future review.
- **If Current Presence/Absence is 'Absent' (with Native, Regular, Confident) rank should be SX**, however, SU may be valid but records should be reviewed. Distribution comments for these cases will facilitate this review.

APPENDIX B

QC NOT PERFORMED BY THE FUNCTION

Certain error checking will still be done outside of the function. For example:

In EST records where there are **multiple rows** in the distribution data grid, secondary errors will not be detected. As long as one row combines with rank and hybrid data to form a valid combination, other invalid data will not be identified by the function. Invalid combinations of distribution data alone (i.e., within one row of data) can be identified by running queries on the TAXON_SUBNATL_DIST table alone. Internally invalid distribution data combinations can be found in the D_DIST_INVALID_MATRIX where rank = "ANY".

In addition, this means that if the rank is **SNA** and at least one row of distribution data suggests that a rank is not applicable (e.g., if the element is exotic in part of the subnation) the DDF result will be **VALID DATA COMBO** *even if* there is another distribution data row that indicates that a rank is applicable (e.g., if the element is native, regular, and confidently reported in another part of the subnation). Additional checks can be done on records with more than one distribution data row (especially where rank is SNA) to detect these records.

Animal ESTs with all values of multiple rows in the distribution data grid the same except Population (e.g., if there are native, regular, confidently reported, currently present breeding and nonbreeding populations, there should be just one row in the grid with Population = Year-round) will not be flagged as invalid unless all of the rows are invalid combinations. Cases of this created during data conversion to Biotics are being corrected at data exchange.

Plant ESTs with a value in the **Population** field (the field d_population_id is in the database table for plants even though it is not visible on the data entry window) are not flagged as valid or invalid by the function. Valid plant combinations have d_population_id = null so if this field has a value in it, the function result will be "NOTHING VALID OR INVALID FOUND". (Note that the function evaluates a null in the d_population_id field as "0" (zero); this is why the plant entries in the valid matrix have "0" for d_population_id.)

The following query can be used to clean up plant records with a value in the population field. Before running the update, it is a good idea to run just the select portion of the query to see how many records will be affected. If there are more than a few, you may want to dig deeper to find out why. The update statement should be run by an administrator in SQL+, logged on as the biotics_user. To be cautious, the update statement can be run without the commit line, after which the user can issue a rollback if things look off or the commit to make the changes permanent.

```
/*
QUERY NAME: Remove population value from plant element subnational
tracking records

DESC: Query to update PLANT element subnational distribution records
containing any value in d_population_id
DELETES ANY VALUE IN D_POPULATION_ID

('plant' includes vascular plants, nonvascular plants, fungi, and algae)
```

```

AUTHOR: G. Davis
DATE: 6/25/08
*/

```

```

UPDATE taxon_subnatl_dist
SET rec_last_mod_date = SysDate,
    rec_last_mod_user = 'script',
    d_population_id = null
WHERE taxon_subnatl_dist_id =

    (SELECT tsd.taxon_subnatl_dist_id
     FROM element_subnational est,
     element_national ent,
     element_global egt,
     scientific_name gname,
     taxon_subnatl_dist tsd,
     d_name_category name_cat
     WHERE tsd.d_population_id is not null
     and tsd.element_subnational_id = est.element_subnational_id
     and est.element_national_id = ent.element_national_id
     and ent.element_global_id = egt.element_global_id
     and egt.inactive_ind = 'N'
     and egt.gname_id = gname.scientific_name_id
     and gname.d_name_category_id = name_cat.d_name_category_id
     and name_cat.d_class_framework_id = 1 /* plant */)
;
commit;

```

Hybrids with Srank not SNA

If global data or the subnational name suggests an element is a hybrid and therefore a rank *may* not be applicable, as long as the *subnational* 'Hybrid Without Conservation Value' box has NOT been checked, a rank other than SNA will not cause the function to call the record invalid (unless there is some other reason that the rank should be SNA). Member programs may want to store subnational ranks for these elements, however, it is possible that a rank has been applied to such a taxon in error or that the taxon has simply not been flagged as a hybrid without conservation value in the subnational tracking record and the rank has not been updated to SNA (many of these elements are currently ranked SNR).

The following query can help find and review records where the subnational tracking record field 'Hybrid without Conservation Value' is not checked but where the subnational name or global data suggest that the element is a hybrid.

```

/*
QUERY NAME: Find hybrids without EST 'hybrid without conservation value' and SRANK
not SNA

DESC: Query to select element_subnational records where global data
or subnational name indicates the element is a hybrid
but subnational 'hybrid without conservation value' is N and Srank is not SNA

Data which indicate the element is a hybrid:
    EGT Hybrid is 'Yes' (tg.d_hybrid_id = 1)

```

EGT Hybrid Without Conservation Value is checked (tg.g_hybrid_ind = 'Y')
Global scientific name contains '% x %'
Subnational scientific name contains '% x %'

AUTHOR: G. Davis

DATE: 6/25/08

**/*

```
select est.element_subnational_id,
nation.iso_nation_cd nation,
subn.subnation_code subn,
d_name_category.name_category_desc name_category,
sname.scientific_name sname,
ts.hybrid_ind s_hyb_wo_cv,
est.s_rank srnk,
gname.scientific_name gname,
d_hybrid.hybrid_desc g_hybrid,
tg.g_hybrid_ind g_hyb_wo_cv,
egt.g_rank grnk,
validate_est_dist(est.element_subnational_id) dist_data_check

from element_subnational est,
taxon_subnational ts,
element_national ent,
element_global egt,
taxon_global tg,
scientific_name gname,
scientific_name sname,
d_name_category,
subnation subn,
nation,
d_hybrid

where est.sname_id = sname.scientific_name_id
and est.element_subnational_id = ts.element_subnational_id
and est.element_national_id = ent.element_national_id
and ent.element_global_id = egt.element_global_id
and egt.element_global_id = tg.element_global_id
and tg.d_hybrid_id = d_hybrid.d_hybrid_id (+)
and egt.gname_id = gname.scientific_name_id
and egt.inactive_ind = 'N'
and sname.d_name_category_id = d_name_category.d_name_category_id
and est.subnation_id = subn.subnation_id
and ent.nation_id = nation.nation_id

and d_name_category.d_class_framework_id in (1, 3) /* plant, animal */
and (tg.g_hybrid_ind = 'Y' or tg.d_hybrid_id = 1 or sname.scientific_name like '%
x %' or gname.scientific_name like '% x %')
and (est.s_rank not like 'SNA' and est.s_rank not like 'PNA')
and ts.hybrid_ind = 'N'

order by name_category,
sname.scientific_name,
nation,
subn
```