

1 A&R HISTORICAL DATA MIGRATION: YEARS 2017-2019

1.1 MOTIVATION

The migration of legacy attendance & revenue records into the new A&R System will make historical data available for A&R user consumption and for potential future endeavours like variance tracking.

1.2 PROBLEM

To be able to rely on the integrity of the historical data when it is migrated, we must ensure that historical fields are correctly mapped to the appropriate field in the new A&R System. The implementation team can do the legwork of this map, but the data owners should review and sign-off the map before it is implemented.

1.3 PROCEDURE

There are a handful of considerations to be made when creating the map:

1. The new system has introduced the concept of an 'activity'. Simply put, 'activity' is the term used to describe a type of data record. There are 7 activities in the new system:
 - a. Frontcountry Camping
 - b. Frontcountry Cabins
 - c. Backcountry Camping
 - d. Backcountry Cabins
 - e. Group Camping
 - f. Day Use
 - g. Boating
2. Any new record migrated into the new system must conform to the record data structure of the new system. This hierarchy is structured as follows:
 - a. Park
 - i. Subarea
 1. Activity
 - a. Month/Year (actual record)

In detail, records are organized by month/year at the most granular level. Every record belongs to an activity. Every subarea has at least one activity. Every park has at least one subarea.

3. There are 83 columns and over 167,000 rows in the historical dataset spreadsheet. The historical data has no concept of 'activity' - each row of the spreadsheet corresponds one month of data for a particular subarea, so a single row may contain multiple activity types.

Each column in a row represents a datapoint for that park/subarea/month/year.

To migrate historical data into the new system, we must infer which activities are contained in each row. We can do this by reviewing which columns in the row are populated with datapoints,

so long as we know which activity each column belongs to. Given any row, if any column belonging to an activity is populated, we can assume that row has that activity.

Some columns do not appear to belong to any activity. To account for these cases, a new activity, 'Legacy Data', will be created to house the datapoints.

4. From the activities extracted from each row, we can create new records in the system that follow the hierarchy in (1), and therefore the new system will remain organized in a way that
 - a. doesn't alter the way the current system handles data, and
 - b. makes historical data available for system users and future refinement.

We need to be certain this approach is valid and approved by the business before carrying out the migration.

Statement	Approved
The migration procedure outlined above is approved.	[YES]

1.4 MAP

1.4.1 Map Fields

The proposed map dictates how the columns from the historical dataset spreadsheet will be interpreted and assigned when they are migrated into the new system. When the map is implemented, each column of the historical dataset spreadsheet is 'mapped' into the new system with the ruleset imposed by the map:

1. **Legacy Name:** The name of the column in the historical dataset spreadsheet
2. **New Name:** The name of the field in the new system with the assumed equivalent definition of 'Legacy Name'.
3. **Activity:** The inferred activity associated with the column
 - a. Each column must belong one and only one activity.
 - b. Note: 'Meta' refers to columns that contain metadata. It is not an activity and can be ignored.
4. **Type:** The datatype of the column's datapoints. The Type can be either 'string' or 'number'.
 - a. Strings are for text: names, dates, descriptions, notes etc. Numbers are for quantities and monetary values.
 - b. It is important to get the column type correct, as the new system will interpret the value of the column differently depending on Type.
 - i. A string type incorrectly cast as 'number' will fail to migrate.
 - ii. A number type incorrectly cast as 'string' will successfully migrate but will do so with an incorrect value.
5. **Calculated:** True or false: Some columns appear to match a field in the new system, but the new system does not actually store that field; rather, it is calculated at run-time from other fields.
 - a. Legacy records created from historical data will also calculate these fields at run-time.

- b. These historical data columns will still be migrated, but will not be represented the same way as the new system's calculated field.
 - i. An example: in the new system, any 'net revenue' value is calculated from the 'gross revenue' value at run-time and is not saved in the new database.
 - ii. The historical value of 'net revenue' will still be migrated, but as a new field (eg, 'Legacy – Net Revenue'), and will be represented separately on the site as a read-only value.
 - iii. The legacy record will now have two fields that should represent the same idea but may not have the same value depending on how the value was calculated historically ('Net Revenue' vs. 'Legacy – Net Revenue').

For this first pass, we are only focusing on migrating historical data from 2017-2019, or ~22,000 rows from the historical dataset spreadsheet.

1.4.2 Map Verification

For each column ('Legacy Name') in the historical dataset spreadsheet, review its corresponding row in the map and review the following statements.

For each row in the map:

Statement	YES/NO
1. Is the definition of 'New Name' in the new system the same as the definition of 'Legacy Name' in the historical data? a. The A&R admin portal is a good way to view how the new system defines fields.	Yes, looks like it, some unit differences in the numbers for calculated fields
2. Does the column belong to the 'Activity' and only to the 'Activity'?	Yes
3. Is the 'Type' of the column correct? Will every value in that column always be of that type?	Yes- Park identification (ie Region, section, park, sub-area), month and variance notes will all be strings. Everything else is a number.
4. Where applicable, is it appropriate to represent this column with a calculation? a. The A&R admin portal should show the formula for how the new system calculates any values.	Yes

If the answer to any of the above questions is 'no', then the map may need altering. When no further altering is necessary, the map can be approved.

Statement	Approved
The migration map outlined above is approved.	[YES]

1.5 OTHER CONSIDERATIONS:

The following considerations should be reviewed and approved by the business.

1.5.1 Null Interpretation

The number '0' and the concept of 'empty' (null) fields in the historical dataset spreadsheet should be defined by the business. Review the following statements.

For a given row in the historical dataset spreadsheet:

Statement	YES/NO
1. If one column value is 0 and another column value is empty (no value), do both columns represent the same idea?	Generally, yes- a zero with <i>no variance note</i> is the same as null, because we regard it as meaning as data not available
2. Where a column value is '0', does this '0' represent a recorded value (<i>ie</i> was this '0' recorded consciously by a human)?	Depends on if there are variance notes- in rows without a variance note, a zero is the same as a null value for analysis
3. If the answer to 2) is yes, do we want this '0' column value migrated?	We want the zero migrated if there is a variance note in the row.
4. Inversely, where a column value is null, does the absence of a value imply it was never recorded?	Yes.
5. If the answer to 4) is yes, can we ignore migrating this column value?	Yes.
6. If a row does not contain any recorded values, can we ignore migrating this row altogether?	Yes.

Examples of where the above questions might apply:

- The column for 'Regular Frontcountry Camping Parties' is 0. Does this mean an operator has consciously recorded the subarea had 0 regular frontcountry camping parties that month? Or does it mean that 'Regular Frontcountry Camping Parties' value for that month was never recorded? Do we care about the difference between these two interpretations? Do we care about showing this 0 in the new system?

Answer: For the migration of historical values, zero is the same as null because we cannot confirm that Park Operators were following a set practice for actual zero attendance. That is, a park that had an actual attendance of zero may have been entered as a zero by the P.O., or simply left blank/ had no entry. There is no way to tell unless there is a variance note explaining the zero (for example the

variance note may say “park closure due to wildfire,” in which case we know the attendance was zero and not just a missing month of data. –[REDACTED].

- A row from the historical dataset spreadsheet only contains a mix of 0s and null values for every non-metadata column. How should we interpret this row?

Answer: In the future we may establish a set practice with Park Operators regarding null values verses 0s. However, we can’t reliably interpret them even in a row like this- for example, did the Park Operator start entering zeros and then decided to leave the rest blank? The only interpretation factor between null/ zeros is if there is a variance note. –[REDACTED].

Because we must infer activity type from populated columns, we must know whether 0s and empty values count as ‘populated’.

1.5.2 New Parks and Subareas

Some rows in the historical dataset spreadsheet belong to parks and subareas that are not currently in the new A&R system. These legacy parks and subareas will be added to the system.

Note: if a row is missing any of the following items, it cannot be made to follow the hierarchy of the new system and will not be migrated:

1. Park (ORCS + Park Name)
2. Subarea
3. Year
4. Month

Statement	Acknowledged
It is acknowledged that any row in the historical dataset spreadsheet that is missing mandatory fields will not be migrated.	[YES]

1.5.3 Park/Subarea Management Areas (metadata)

Parks and subareas in the new A&R system includes a ‘Management Area’ metadata field. The historical dataset spreadsheet is missing this column, so in instances where a new legacy park/legacy subarea is to be created, this field will be empty.

Statement	Acknowledged
It is acknowledged that any legacy park or legacy subarea created as part of the migration will be missing a ‘Management Area’ value.	[YES]

1.5.4 Legacy Data Flags

Any new database item added to the new system will carry a ‘legacy’ flag for quick identification as a historical item. This distinction will help us display historical items clearly, and in a worst-case scenario, back out the entire migration to preserve the integrity of the current data in the new A&R system.

1.5.5 Failure Documentation

Any row that fails to migrate will be documented so we have a record of which migrations were successful and which were not.

1.6 ADDITIONAL COMMENTS

Please add unit to name of calculated fields that involve a unit conversion (ie parties or vehicles to people). The migration plan looks good, assuming if an error is found in the mapping or other migration element that we will be able to correct course.

1.7 APPROVAL

Statement	Approved
Regarding migrating the historical dataset between 2017-2019, the migration procedure, migration map, and additional migration considerations outlined in this document are acceptable and approved. This document can be relied on as a ruleset by developers who will orchestrate the migration.	[YES]