

Chapter 6

Legacy Data Cleanup

Overview

Importing legacy data will likely require data cleanup, either before the data is imported, or within the Toolkit once the data has been imported. Failure to cleanup legacy data will lead to subsequent problems within the Toolkit, so it is recommended that data cleanup be a top priority after import is completed.

Potential data cleanup issues

The following issues may present themselves in data that is imported into the Toolkit:

1. **Import of invalid records.** In most cases, the Toolkit will import records that do not contain required fields. These fields must be completed before any additional edits to the record can be saved. These invalid records can cause problems with operations such as merging of subject or name terms or merging of items in a lookup list.
2. **Import of repetitive terms in lookup lists.** When importing into fields controlled by lookup lists, the Toolkit will import data that does not match elements in the lookup list. For example, you may import "aat" into the Subject source field, where the value in the lookup list is "Art & Architecture Thesaurus (aat)." At times the terms may appear to be the same, but trailing whitespace causes the two strings to be, in fact, different. The Merge Item feature in the lookup lists can assist you with cleaning up this kind of inconsistency in your data.
3. **Import of punctuation repeated by built-in operations in the Toolkit.** In many instances, the Toolkit automatically provides punctuation for subject and name terms. If your subject and name data is imported with punctuation, there may be repetition in the Toolkit's displays and outputs.