

## Data Conversion and Migration

New York State is undertaking a separate but concurrent project to perform data migration for MAPPER data elements into an Oracle database. They will be performing initial data cleanup for data that will be provided to the Bidder for pre-population in the new TMS based on the Bidder's proposed approach. This document lists the types and quantity of data in the current architecture that the Department of Civil Service intends to convert from legacy systems to the new TMS. This information is being provided to assist potential Bidders in developing their plan/approach in terms of data migration strategy.

The Bidder must describe their data conversion and/or migration strategy proposal addressing in detail each of the following aspects of legacy data conversion/migration to the proposed solution based on, but not limited to, the below data descriptions;

### CONVERSION DESCRIPTION

The following data related to specific business processes will likely be converted to TMS:

**Standard Subtest List (SSL) (or Subject Area Table)** – The SSL file contains a list of standard subtest titles and related data and information for Subject Areas that may be included in a test. Data fields in this file include Standard Subtest List Code (a 25 character alpha/numeric field), distribution memorandum title, access designation (General/Restricted/Subcontract), number of items, review type, question type code (multiple-choice/simulation/in-basket/essay/notice-to-candidate/memory-resource booklet), the full subtest title description (148 character alpha/numeric field), and a variable length subtest description. All this data will be converted. There are approximately 3,200 rubrics and we predict they will grow to about 5,000.

**Test Items** – The Test Item file contains the text of test questions and related data and information. Data fields in this file include the key (7 characters), date changed (6 digit numeric date field), classification code (8 positions), macro owner (7 characters), subtest owner (7 character numeric), century code (2 positions), one correct answer identifier (1 character), item stimulus text, and all of the answer choices. A comment record follows the item choices. The last used version of each item will be converted. An approximate count of items to be converted is 150,000 with a predicted growth to 180,000.

**Macros** – A “macro” is a group of items that are linked together and share a common stimulus. The Macro file contains both the stimulus text and links to the related items as well as other related data. Data fields include the key, date changed, classification code, type designation (1 character), owner (7 characters), century code, and the keys of any associated items. All macros will be converted and the current estimation is 7,500 which will grow to 10,000.

**Directions** – A “Direction” is a brief passage of text that provides a basic instruction to candidates, coded to reflect its use with test items in a specific category. Data fields in the Directions file contain text and a classification code (8 characters). All directions will be converted and the current estimation is 1,500 which will grow to 2,000.

**Permanent Subtest** – A Subtest is a group of related items that is used to measure candidates' knowledge, skill, or ability in a specific Subject Area. A Permanent Subtest is a group of Test Items (as well as related macros and directions, if needed) in a specific Subject Area that has been assembled for repeated use as a set. As the Permanent Subtest is reused, subtest statistics are captured and stored to create a use history for reference. Data fields in this file include the key (7 digits), date changed, classification code/subtest number, approval status (i.e. prior),

general/restricted use status; owner (7 characters), century code, and the keys of the associated items. There are approximately 3,000 items to be converted with a growth potential to 4,000.

**Item Use History** – Item use histories are statistics related to the item’s performance. The record contains the key, classification code, series date, exam/series code (6 positions), number of candidates (five digits), percent difficulty (two digit), item reliability index (3 digits), twelve four-digit numeric fields representing high/low distribution of candidate choices across key and distracters and the item correct choice designation (1 character alpha) followed by the century identifier. Item histories up to and including ten years old will be converted. An approximate count is 300,000 with growth to 400,000.

**Permanent Subtest Use History** – Subtest use histories contain statistics related to the subtest’s performance. This contains the key, classification code, series date, exam/series code, number of candidates, mean (4 digit numeric field), standard deviation (4 digit numeric field), KR-20 (3 digit numeric field), number of questions (3 digit numeric field), a use field (1 digit alpha field), and century code, and the distribution of candidates across the number of items. Subtest histories up to and including ten years old will be converted. An approximate count is 15,000 with a growth prediction of 18,000.

**Images** – The Department has a library of images which are used in test booklets that are associated with Macros. We currently have approximately 7,000 and we approximate a growth to 9,500.