Test Data Checklist

Version 1.0

Table of Contents

[1. Background 3](#_Toc280186864)

[2. Purpose 3](#_Toc280186865)

[3. Testing Types: Infrastructure 3](#_Toc280186866)

[4. Testing Types: Data Sets 4](#_Toc280186867)

# Background

Testing under ICD-10 will be extensive and include multiple system processes, business processes and analytic models. In order to minimize the risk of transition, test data will be required to provide the input to a number of test cases. The SMA should prepare this data to address varying risks based on a number of scenarios.

# Purpose

The SMA should consider the items addressed in this document when developing test data to support the ICD-10 migration process.

# Testing Types: Infrastructure

Test data development and management will have specific infrastructure requirements to facilitate the testing process. Table 1 identifies best practices for creating and evaluating testing infrastructure.

Table 1: Infrastructure Key Requirements

| Completed | Key Requirements |
| --- | --- |
|  | Create a test data repository with sufficient disc space to support all multiple sets of test data. A test repository should provide a database of test results which tracks useful information like what tests are failing, or which failures have the same backtrace. |
|  | Create a directory of test data with sufficient metadata to support:  A unique identifier for the test data set  A definition of the date range and volume of the data set  A description of the nature and purpose of the test data set  A method to track and audit data to define when the set was created and/or modified  A link to the appropriate data dictionary and format specification for the data set |
|  | Refer to the governance structure to assure appropriate controls over access and use of data sets |
|  | Training on access, use and limitations of data sets |
|  | Creation of appropriate documentation to define the content and limits of each data set |
|  | Utilize a de-identification process to assure that PHI is appropriately protected. Remove the 18 specific identifiers listed in the Privacy Rule and determine there is no other information that may identify the individual. The identifiers are:  Names  Geographic subdivisions smaller than a state  All elements of dates (except year) related to an individual (including dates of admission, discharge, birth, death and, for individuals over 89 years old, the year of birth must not be used)  Telephone numbers  FAX numbers  Electronic mail addresses  Social Security numbers  Medical record numbers  Health plan beneficiary numbers  Account numbers  Certificate/license numbers  Any unique identifying number, characteristic or code |
|  | Utilize a communication process to assure that any changes or modifications to the test data set are communicated to all stakeholders |
|  | Provide support for download and modification of test data sets for specialized purposes |
|  | Define a set of basic counts and calculations for certain parameters within the test data set to support the development of expected results for test cases |
|  | A repository of all code sets, GEMS, Reimbursement or crosswalk maps and specifications with support for robust query of code and crosswalk related files for test case development. |

# Testing Types: Data Sets

Test data will need to include a wide variety of data types to address specific testing needs. For some testing functions the content of the data may not be important for testing for throughput or other performance parameters, but for other testing needs, specific data content will be needed to test specific business scenarios and potential business risk areas (e.g. risk calculations). Table 2 identifies data types to consider in developing the repository of test data sets.

Table 2: Data Types and Value in Testing Process

| Completed | Data Type | Value in the Testing Process |
| --- | --- | --- |
|  | **Performance data** – A variety of sets of transactional or input data will be needed to test throughput and other performance metrics. | If there are rules that are applied in transaction processing that may result in processing lags, leveraging performance data will help identify specific data scenarios that stress test these rules. |
|  | **Analytic data** – Data sets to test various analytic models will need to include data content that specifically touches on the subject areas within the analytic or reporting model. | Tests the distributions of content to support testing of all filters, grouping and summary logic defined within the report. |
|  | **Rules/Edit data** – Rules or edits will have a specific set of criteria that would be triggered based on the content of claims or other transactions. | Test data will be required to test the rules/edits which trigger claims or other transactions. |
|  | **Predictive modeling data** – Data for predictive modeling of a variety of financial and risk models. | Test data will require specific data sets that are designed to include scenarios that test financial/risk calculations embedded within predictive or risk models. |
|  | **Policy related data** – Data sets that are designed to specifically test data driven logic as defined in policies. | The codes that are used in the policy definition will need to be contained in test data sets to assure that the intent of policies is realized in processing. |
|  | **EDI transactional data** – Test transactions for inbound and outbound transactions will be needed to test the data interchange processes. | This data will need to support testing of transactional validity at multiple levels. End to end transaction processing of specific business scenarios will also be needed as a test of both EDI transaction management as well as the appropriate application of end to end business logic. |
|  | **ICD-9 and ICD-10 based data** – Data set will need to include both ICD-9 and ICD-10 data to test. | Compare environments where dual mode processing is anticipated. |