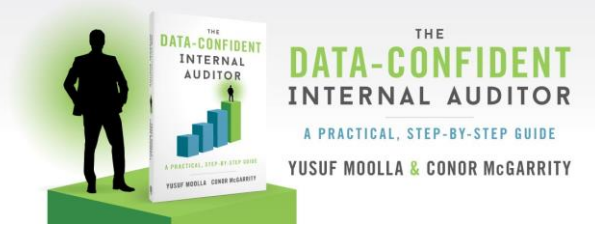


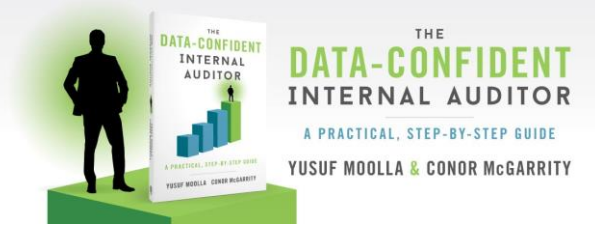
Chapter 6 | Bonus Resource Data Preparation



Example of Completed Data Quality Checklist

Type	#	Description	Data Set A	Data Set B	Data Set C
DQ1: Individual data sets	1	No duplicate records	All ok.	All ok.	2 duplicate rows. Remove one of them.
	2	No/few NULL values or empty cells	<5% empty or null. Year is null for some columns – fix these.	<5% empty or null. No key columns.	<5% empty or null. No key columns.
	3	No outliers	All ok.	All ok.	One negative value found, positive expected. Investigate.
	4	Values make sense (e.g. Yes/No) and within specified range (e.g. none before 2001)	All ok.	All ok.	All ok.
	5	Correct format for each variable	Fix years.	Fix years	All ok.
	6	No special characters	All ok.	All ok.	“©” found in column “XYZ”. Remove.
	7	No duplicate unique identifiers	All ok.	All ok.	All ok.
DQ2: Across data sets	8	Unique identifiers match across data sets – no orphans	All ok.	All ok.	All ok.
	9	Consistent data formats	All ok.	Fix date format in column “posted”.	Fix date format in column “posted”.
	10	Comparison values fall within the same ranges	Fix state names.	Fix state names.	All ok.

Chapter 6 | Bonus Resource Data Preparation



Data Quality Report

The structure and layout of your report will depend on your organizational templates – the table below contains some information to help populate that report.

Data Quality Report				
Audit	<i>Name of audit, reference number if applicable</i>			
Dataset A	<i>Description of the dataset</i>	Source	<i>e.g., data request number, source (system/warehouse name), extraction date</i>	
Dataset B	<i>Description of the dataset</i>	Source	<i>e.g., data request number, source (system/warehouse name), extraction date</i>	
Dataset C	<i>Description of the dataset</i>	Source	<i>e.g., data request number, source (system/warehouse name), extraction date</i>	
#	Description	Dataset, table, fields	Impact(s)	Remedial action(s)
Issue #	What we found (e.g. null values)	<i>Name(s) of dataset, table, field(s) in which this was found</i>	What the impact of the issue is (note: this is not the audit impact)	<i>What we have agreed with management on how the issue will be addressed, if it will be fixed</i>
1	The extract contained a duplicate row	Dataset C, across fields	<ol style="list-style-type: none"> Describe the impact of the duplicate record on recording/processing Describe the impact of the duplicate record when the data is extracted and analysed 	<ol style="list-style-type: none"> No action on the underlying data, this was found to be an extract issue The code for the data extract will be modified, this error will be checked against other extract code
2	The extract contained special characters	Dataset C, Field “XYZ”	This is a valid character in the underlying dataset; it was removed from the data for audit analysis.	None. Not applicable.
3	State: Differences in naming convention	Dataset A, Field “state” Dataset B, Field “state”	<ol style="list-style-type: none"> Describe the impact on recording/processing Describe the impact on analysis 	<ol style="list-style-type: none"> Actions to be taken to fix the underlying data (and the capture process) No action, this will be fixed at the front end.