

Use an Integrity Report to Find Errors in Your Data

Hand your list to the data processor and move on to the next task. *Or*, hand off your list and ask for a data file integrity report, then move on to the next task. What is “data file integrity” and why is it important?

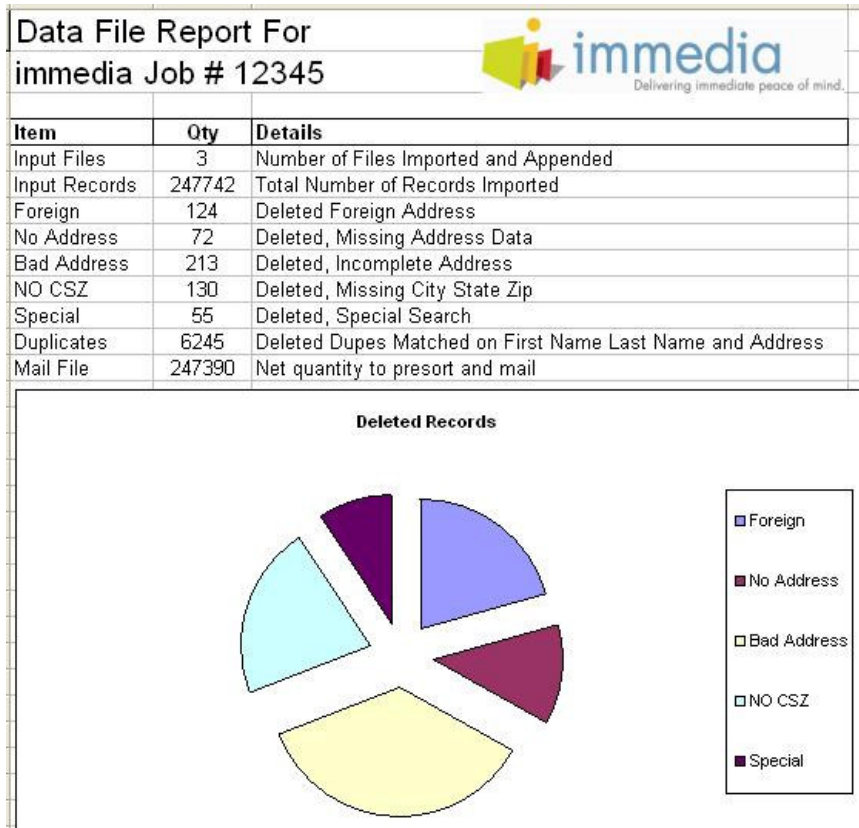
“Integrity” is defined by Wikipedia as:

The condition that exists when data is unchanged from its source and has not been accidentally or maliciously modified, altered, or destroyed.

The goal of data quality is to develop guidelines that help businesses improve data accuracy and validity.

The integrity report is referred to more often as a *quality report* see figure 1 below. Whether feedback received is in great detail or a summary, the report shows where the problems are in your data file. Mailing list, account numbers, sales history – whatever the data file contains – an integrity report can bring to light degradation trends, or simply errors to fix. Feedback on your data also provides guidelines for your data collection source. This leads to building a strong training base for employees that handle your company data.

We are familiar with human errors with data entry, but how about errors that occur when data is transferred from one computer system to another? For example, you enter information using a non-typical font, or coming from a *legacy system*. Weird things happen to the data when a system attempting to interpret the data finally accesses the data. A data file may begin on a legacy system, then output to an old DOS



platform, then imported to a Mac, sent through the web to a PC running Windows on a network with a home grown server. During that translation you start with usable data and end up with hard-return characters, windings and symbols and more.

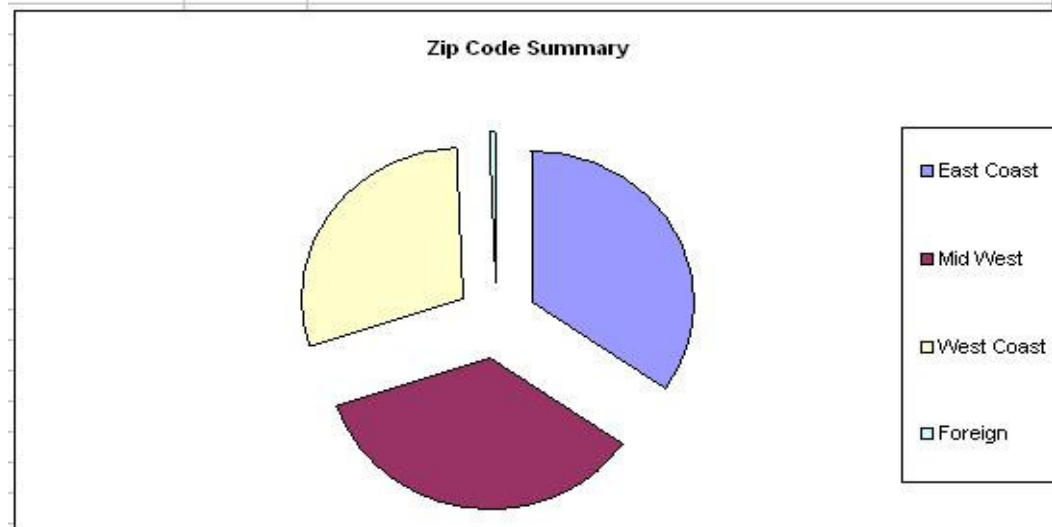
Software bugs and viruses scramble data as you pass it back and forth through email, or experience even a hardware malfunction. This happens without your knowing, because the errors are buried deep within your data and can only be found if your data file is “combed” through.

An integrity report can be general or specific, as mentioned earlier. You may ask for statistics on errors to see how many records were missing last names, phone numbers, secondary contacts and PO Boxes. The larger your data file, the more impact data quality has on postage savings for presorted mail.

In another view, you may want to know (see figures 2 and 3 below):

- How many and which records do not have a 14-digit account number
- Redundant data in multiple fields of the same record
- Facility names in the address field
- Percentages by geographical area

Data File Report For		
immedia Job # 12345		
Item	Qty	Details
East Coast	86710	Eastern State Zip Codes
Mid West	86710	Midwestern State Zip Codes
West Coast	74323	West Coast Zip Codes
Foreign	1200	Non US Domestic Zip Codes

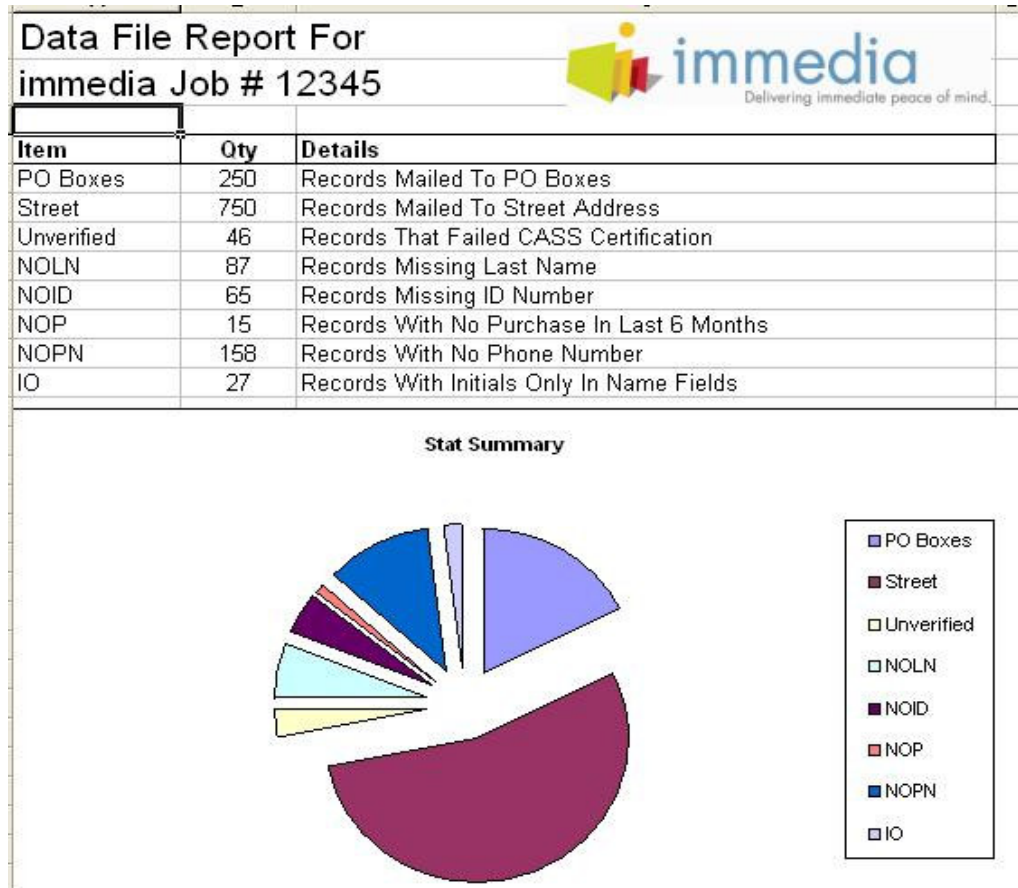



How can you use an integrity report to benefit your organization? If data integrity and quality is a value, especially in an era where information is an asset, it is one more loose end tied up.

When was the last time you analyzed your data for quality or statistical information?

Ask for a data quality report in detail, summary, or charts and graphs. It will add a few dollars to your project, but will prove to be worth it. If your quality report comes back clean, consider it an A+. If it isn't clean, use it as a map to find and

improve your data quality. immedia has the capability and expertise to provide this valuable feedback.



 Wael Lazar – Data Specialist