

Entry Concerns with the MLS Database

By Paul R. Bynum-Market Analyst

The Northwest Arkansas Multiple Listing Database is, of course, critical for all the reports generated on residential properties. Like most databases compiled by human hands, it is not perfect.

In a former article I touched on the problems associated with late entries after the current reports have been generated. (See the article: [Changes in Reported Data](#) on the [Mount Data Website](#)). This article covers other problem entries that affect reports and what I personally do about these entries.

I have been downloading MLS data and analyzing it since 1992. In correction with this, I have developed a "problem correction" visual basic program. All downloaded data is run through this problem correction program before reports are generated.

All problem entries in the database are the result of human input. Many, if not most of these problems could be eliminated by a few simple changes in the software program that accepts the input data.

Example, if a property is listed for \$175,000 and then a sold price is entered for \$1,750,000 (not unusual), it is obvious that an extra zero was added by the person entering the final price. The existing software program allows such entries to take place. Yet, a pop-up box could be displayed when the sold price was entered saying: ***"Sale price is larger than list price. Are you sure you want to continue with this entry?"*** Doing this would help eliminate such problems.

Such entries if not caught, will skew any number of reports. When reporting the entire region, such entries could be part of a huge number of data entries and thus tend to be insignificant. ***However, for individual price ranges within a particular city, such an entry could be very significant. Some programs that generate reports will run incorrectly without valid entries.***

Ideally, a database should be left alone-untouched and pristine. For practical reasons this is not always possible and in my work, what ***can*** be corrected with reasonable expectations of accuracy, ***is*** corrected.

The table below indicates the monthly reports generated. The table following contains a description of the problems, occurrences per 25,000 records, the source, the reports affected and what is done to eliminate or mitigate such problems.

I invite any interest parties to comment on this article and share your own methods of dealing with such entries. Email me: [Comments](#).

Until next time and by the numbers
Paul Bynum
October 24th, 2007

Reports

<i>Report</i>	<i>Abbreviation</i>	<i>Fields Used</i>
Months of Inventory	MOI	County, City, Status, Type, List Date, Pending Date, OMD*, List Price
Days-on-Market Current	DOML	County, City, Status, Type, List Date, List Price
Days-on-Market to Pending	DOMP	County, City, Status, Type, List Date, Pending Date, List Price
Days-on-Market to Closing	DOMC	County, City, Status, Type, List Date, OMD, List Price
Price per Square Foot	P/SF	County, City, Status, Type, OMD, List Price, Square-Footage
% Sold to Listing Ratio	%Sold	County, City, Status, Type, List Date, List Price
% Existing to New Construction Units	E/TAI NC/TAI	County, City, Status, Type, List Date, List Price, Year Built
True Available Inventory	TAI	County, City, Type, List Date, Pending Date, OMD, List Price
Volume	V	County, City, Status, Type, OMD, List Price, Sale Price
Units Sold	US	County, City, Status, Type, OMD, List Price, Sale Price
Median Sale Price	MSP	County, City, Status, Type, OMD, Sale Price
Sale Price to List Price Ratio	SP/LP	County, City, Status, Type, OMD, List Price, Sale Price
Homes per Buyer	H/B	County, City, Type, List Date, Pending Date, OMD, Sale Price
Market Activity Index	MAI	County, City, Status, Type, List Date, Pending Date, OMD
Sellers Market Index	SMI	County, City, Status, Type, List Date, Pending Date, OMD, List Price, Sale Price
Zip Code Studies	ZCS	County, City, Status, Type, List Date, Pending Date, OMD, List Price, Sold Price
Special Reports	SR	County, City, Status, Type, List Date, Pending Date, OMD, Original List Price, List Price, Sale Price, Square Footage, Bedrooms, Baths, Lot Size, Acreage, MLS#, Listing and Selling Office, Listing Agent, Sales Agent, Address, Subdivision, Year Built, Zip Code

*OMD (Off Market Date), includes Sales, Expires, and Withdrawn Date
 Status=Current, Pending, Sold, Existing, or Withdrawn.
 Type=New Construction, Historical or Existing

Table of Database Problems

(#affected is per 25,000 records all current, pending and OMD Oct 2006-Sept 2007)

<i>Problem Description</i>	<i># Affected</i>	<i>Fields</i>	<i>Reports Affected</i>	<i>Process Rationale and Method</i>
Date entered as shorthand: 87 for 1987, 04 for 2004, etc. Or field is left blank.	1,158	Year built field.	E/TAI, NC/TAI	The needed first two figures are normally added. Thus 98 becomes 1998. If the year field is blank, there are no prior listings, and the "Age" field shows new construction, the current year is substituted for the blank field.
Blanks in the "Age" column	3,390. (New Construction increased from 6,322 To 9788. October 2007)	Age field	E/TAI, NC/TAI	Blank "Age" fields are considered existing homes. Since a large number of new construction homes are entered with blank "Age" fields, they will be processed as existing rather than new construction homes. All blank "Age" fields that are new listings, built within the last two years and never before been listed on the MLS, I change to "New Construction."
Inconsistencies with names in all caps or proper format	884	County and City fields.	Some of the Basic programs use filters that are "case sensitive"	All names in both fields are transformed into "Proper" format. This means for example: FAYETTEVILLE becomes "Fayetteville"
Inconsistencies with field formatting.	Numerous	All fields affected.	Some of the Basic programs use filters that are "case sensitive".	Date formats, currency fields, text fields and number fields are corrected for consistency in format.
Zip code does not match city of property.	156	City and Zip Code fields	ZCS	City is compared to zip codes. If there is an inconsistency, the zip code is set to the city's known zip code. Only towns with single Zip Codes are changed.
Square footage in excess or left blank	49	Square Foot field.	P/SF, SR	Impossible to accurately deduce. Field is set to "UNK" for blanks. SF in excess of 10,000 is looked at and if suspicious, not used in program calculations.
List price field is blank	15	List price field	Virtually all records	If there is an original list price this is copied into the list price field. Otherwise the record is not used.
Sale price field is blank and status field is set to "Sold"	8	Sale price and status fields	V, Us, MSP, SP/LP, HB, SMI, ZC, SR	List price is copied into sold price field and record used for counts.

Table of Database Errors

(#affected is per 25,000 records all current, pending and OMD Oct 2006-Sept 2007)

<i>Problem Description</i>	<i># Affected</i>	<i>Fields</i>	<i>Reports Affected</i>	<i>Process Rationale and Method</i>
Time Share Entry	13	Type filed	None, as I do not track	Time shares are eliminated from the database.
Sales price is Significantly greater than list price	8	List and sales price fields	V, US, MSP, SP/LP, HB, SMI, ZCS, SR	Counted if $SP > LP * 1.3$. If adjustments are obvious, they are made, otherwise, they are not used.
List price significantly greater than sales price	58	List and sales price fields	V, US, MSP, SP/LP, HB, SMI, ZCS, SR	Counted if $LP > SP * 1.3$. If adjustments are obvious, they are made, otherwise, they are not used.
Listing date is greater than off-market date	1	List and OMD fields	MOI, DOML, DOMP, DOMC, %S, E/TAI, NC/TAI, TAI, HB, MAI, SMI, ZCS, SR	Since these records are discovered monthly, the OMD is set for current month and day.
Listing date is greater than pending date	1	List and pending date fields	MOI, DOML, DOMP, DOMC, %S, E/TAI, NC/TAI, TAI, HB, MAI, SMI, ZCS, SR	Since these records are discovered monthly, the pending date is set for current month and day.
Pending date is greater than off-market date	871	Pending and OMD fields	MOI, DOMP, DOMC, P/SF, TAI, V, US, MSP, SP/LP, HB, MAI, SMI, ZCS, SR	Most often, I reverse the dates, or set the OMD to the current date. Otherwise, they are unusable.
Off-market date is greater than the current date	27	OMD field	MOI, DOMP, DOMC, P/SF, TAI, V, US, MSP, SP/LP, HB, MAI, SMI, ZCS, SR	Often mistake is made when homes are withdrawn before scheduled withdrawn date and date is not changed. Since these records are discovered monthly, the OMD is set for current month and day.
Listing and/or pending date is greater than the current date	Rare	Listing or pending date field	MOI, DOML, DOMP, DOMC, %S, E/TAI, NC/TAI, TAI, HB, MAI, SMI, ZCS, SR	Not common. Since these records are discovered monthly, the future dates are set for current month and day.

Table of Database Errors

(#affected is per 25,000 records all current, pending and OMD Oct 2006-Sept 2007)

<i>Problem Description</i>	<i># Affected</i>	<i>Fields</i>	<i>Reports Affected</i>	<i>Process Rationale and Method</i>
Sale record with no pending date	Rare	OMD, status fields	MOI, TAI, HB, MAI, SMI, ZCS, SR	Occasionally occur. Reports are affected but little can be done.
Duplicate MLS numbers	Rare	All fields	All records	Yes! Upon occasion these occur. Have no idea how they got or stay that way. I delete the oldest record. If exact, I delete the duplicate.
No MLS numbers	Rare	None	None	Yes! Upon occasion these occur. Have no idea how they got or stay that way. I delete them as spurious entries.
Listing date equals pending date	318	Listing and pending date fields	MOI, DOML, DOMP, DOMC, %S, E/TAI, NC/TAI, TAI, HB, MAI, SMI, ZCS, SR	Although not impossible for list date and pending date to be identical, it is highly unlikely. If applicable, I set pending date to current date. If not, they are unusable. Reports are affected but little can be done.
Pending date equals off-market date	192	Pending and OMD fields	MOI, DOMP, DOMC, P/SF, TAI, V, US, MSP, SP/LP, HB, MAI, SMI, ZCS, SR	Virtually impossible for a home to become pending and close the same day. Also impossible to correct. Shows thoughtlessness by person imputing the data. If applicable, I set sale date to current date. If not, they are unusable. Reports are affected but little can be done.
Double Entries	253	All fields	All Reports	Entries are in the database having different MLS numbers but having the same address, list date or off market date, listing agent and company. Sometimes they both have the same status (such as sold), and sometimes they have different ones (such as pending and current). Some can result from agents changing offices and re-listing entries, and the previous record is left on the database. Some are just double entries with different MLS numbers. I delete either the oldest record, or one of the double entries. Condos are not deleted as individual unit addresses are often not entered into the database, making it difficult to discover duplicates.