C.A.MOTZKO & ASSOCIATES
Sensible Solutions for Test Equipment Management

**Reprint from the
January – February 1999
Issue of Cal Lab Magazine**

USING DATA NORMALIZATION TO GET THE MOST FROM YOUR EQUIPMENT DATABASES

Charles A. Motzko, P.E.
C. A. Motzko & Associates

Whether at work or at play, we constantly store information or data. At work, you probably file information about calibrations and instruments. And at home, many of us store our personal inventories in some database.

If you don't enter and store your data **consistently**, you will experience problems or delays when you try to retrieve it. For instance, if you can't find information about various instrument calibrations and custodial responsibility, you cannot execute an effective calibration recall system. Moreover, if you can't find the serial number or other information on your personal possessions, your insurance claim may not be honored.

The same concepts apply to your test equipment databases. To make your databases as efficient and easy to use as possible, you'll want to take a hard look at how they store information. In this paper, we'll discuss the concept of data normalization, and we'll show some techniques and approaches on how to employ data normalization to get the most from your current databases.

We need to point out that the issue of database design is beyond the scope of this paper, but is a critical aspect in the overall operation of any database. This paper does not address the realm of entity relationships, redundant data, unnecessarily recurring data or how the fundamental relationships are established.

DATA GONE BAD

As we've pointed out, illogically or inconsistent data can cause a number of problems. This is especially true for relational databases. Poorly entered data may provide erroneous information, be difficult to use, or even fail to relate properly. Most of these problems are the result of anomalies in the way data is entered. Where an anomaly is defined as any occurrence that weakens the integrity of your data due to irregular or inconsistent data entry.

For instance, there are nuisance anomalies. A manufacturer name is entered in a number of different ways. You'll then have to find and update each record individually, as manufacturer's name query can not group like assets (models). Of course, you could use a FIND command and then manually replace each of the small found sub-sets, but the best solution is to arrange your database so that you have only one data field to find and modify.

Another instance shown in **Figure A**, the example database titled: *Calibration Recall,* contains several assets that are the same manufacturer and model and some basic product and recall information. As you can see, the example database also contains certain information for each item that is different even though each asset is exactly the same.

CALIBRATION RECALL DATABASE (initial)							
Asset No.	Manufacturer	Model	Options	Description	Serial No.	Date Calibrated	Date Due
12345	HPC	HP6205B	w/opts	Pwr Sup	NSN	12/1/98	12/1/99
56789	HEWPAC	6205B	009/011	Dc Power Supply	2412A	11/15/98	11/15/99
98765	HARLAB	6205B-09-11		Dual Supply	67890	7/9/97	7/9/98
43210	HARLAB	6205B/009/011		DC PS	1218A85210	NCR	NCR
13467	HP	6205B	009-011	Dual Dc Power Supply	2236A12345	11/29/98	11/29/99

Figure A – Source Database

These types of entries may look familiar to many of you as you possibly have had to contend with this type of (mis)information for years in your property, asset or calibration records. Legacy databases promote this type of activity with their fixed field length requirements, denied delimiter use, lack of data entry verification and restricted word tables.

Where data entry verification can be as simple as setting a field to text or numeric entry only or as sophisticated as mixed field entry based on a format template. Restricted word tables are specifically used to describe the same types of instruments specified data fields in a **consistent** manner.

In **Figure A**, even though the assets are identical, none of the assets have the same manufacturer, model and option data entry. Each time an additional asset is entered, one is probably repeating some or all of the information regardless of the correctness. These types of files could become a mess rather quickly in a large database with hundreds or tens-of-thousands of assets from many different manufacturers.

This is further exacerbated by the fact that the company that is stamped on the name or serial plate may not be the manufacturer that is producing and selling the same product today. Case in point is that Harrison Labs, Sanborn Instruments, Dymec, Mosley and Boonton Radio Corporation, just to name a few, were acquired by Hewlett Packard, which continued to produced many of the same model numbers after acquisition.

Figure A is an example where both Harrison Labs and Hewlett Packard manufactured the 6205B DC Dual Power Supply.

NORMALIZING YOUR DATA

In the database realm, we use a process called *normalization* to produce the most efficient and functional databases. *Normalization* is the storage of data where it uniquely belongs. Doing so creates a secure, robust and reliable structure for your application.

Let's consider the Calibration Recall Database again. Each record contains three fields of unique data: *Serial Number, Date Calibrated* and *Calibration Due*. No record should *collectively* repeat this data. By collectively, we mean that no other record should contain the same information al all three fields. We didn't include the *Asset* field in this breakdown because it's an automatically entered serial field and will therefore be unique to each record.

This doesn't mean that the data in each field must always be unique to each record. In fact, as you can see, the *Manufacture* and *Model* fields often repeats data, and the *Option* and *Description* fields may also repeat data. Of course, the same could be said of the *Manufacturer* and *Model* fields. However, there's one big difference: the three fields (*Serial Number, Date Calibrated* and *Calibration Due*) are mutually exclusive (unique) to each asset, while the *Manufacturer, Model, Option* and *Description* fields are mutually dependent (repeatable) are not.

DETERMINING DEPENDENCY

This discovery brings us to an early step in normalizing our data: determining relationships, otherwise known as *dependency*.

To be dependent, similar records must affect one another. In other words, they must relate to one another. In the Calibration Recall Database, each asset's *Calibration* and *Calibration Due Date* values depend on the *Asset* field. The products are DC Power Supplies and we can determine the *Calibration Due Date* by looking up by asset for individual items or by manufacturer - model number for functional or class groupings. Furthermore parametric searches can be done on a structured noun descriptor to further refine the found set of assets.

This type of data flexibility and reporting can be used to promote and leverage standards assignment, support equipment requirements, sparing and provisioning, workload planning, scheduling and deployment, usage and utilization … Etc.

However in the example Calibration Recall Database, there's no dependency between the Model fields and the Manufacturer field. For instance, the *Manufacturer* designated "HARLAB" supplied two assets. But you can't determine similar product's *Calibration Dates*, or *Calibration Due Dates* using the *Manufacturer* or *Model* fields. This information isn't unique to either record—it doesn't affect or depend on any of the product fields.

Therefore, we can conclude that the supplier and product information need to be normalized for best use of the database.

NORMALIZING YOUR DATA

Now that we've determined that the example data really needs to be normalized we need to set up a manufacturer database and a product catalog database? A sample manufacturer's lookup table is provided in **Figure B**.

Manufacturer Database	
Abbreviation	**Acronym**
HPC	HP
HEWPAC	HP
HARLAB	HP
HP	HP

The manufacturer database is a simple lookup table that takes the abbreviation in use then looks up and returns the normalized manufacturer's acronym or name.

This table must also take into account the transition from the serial plate manufacturer to today's current supplier or OEM. In practice, this table will become quite large and complex with a variety of abbreviations pointing to a single normalized manufacturer's acronym or name.

The product catalog database is a much more complex matter in terms of development and implementation.

The first normalization decision is to determining how one need's to display installed options. To a large extent this will be determined by the constraints built into the legacy database being normalized.

The question of extended model number (with delimited options) or a Bill-Of-Material (BOM) approach will be determined on an individual basis. To illustrate the differences and the underlying database support examine the following manufacturer–model–option number data fields are shown in **Figure C**:

Data Fields Required

Mfg	Model	Opts	Comments
HP	6205B/009/011		This is an example of an extended model number with the root model followed by the delimited option numbers. The key constraint is the number of characters in the fixed length field. Simple to set up, but very maintenance intensive.
HP	6205B	009 011	This is an example of a BOM approach where the root model and options are separate data fields (linked by asset/serial numbers). The key constraint is the ability to hold the separate option fields with the link to the asset's root model number. Complex to set up with very low ongoing maintenance requirements.

Figure C

One now needs to employ a set of data cleanup routines that will:

- Convert the existing manufacturer's abbreviation to the standard acronym
- Strip out the root model number (remove extraneous comments and options, where necessary)
- Deal with the installed options and maintain the referential integrity back to the unique asset or serial number.
- Concatenate the manufacturer acronym and root model number to produce a single manufactures - model number occurrence that is unique for that combination.
- The concatenated manufacturer – model number will now be used as a index key for the data lookup and replace routines

In the simplest form, this has prepared and started the normalization process for the source database. The Calibration Recall Database (initial), shown in Figure A, is now compared to a Product Catalog File to normalize the targeted data fields.

The development of a standard product catalog for use in the normalization routine is available from a variety of sources. For example, one can use cleaned sub-sets of their own equipment database, standard or

approved equipment listings, manually normalized equipment databases, manufacturer's listings or subscribe to one of the commercially available equipment cataloguing services.
The product catalog is usually the source of the approved information that will be used for normalization; it is considered an essential element in the process. The reason for this is that the product catalog acts as the repository for ALL the information known about a particular manufacturer – model – option(s) set independent of the asset – serial number. This is the standard that benchmarks the usefulness of your information system.

Figure D shows a screen shot of one page of an existing commercial product catalog that is available on a subscription basis. The purpose is to show some of the breadth and depth of information that is available to employ for normalization of your own data.

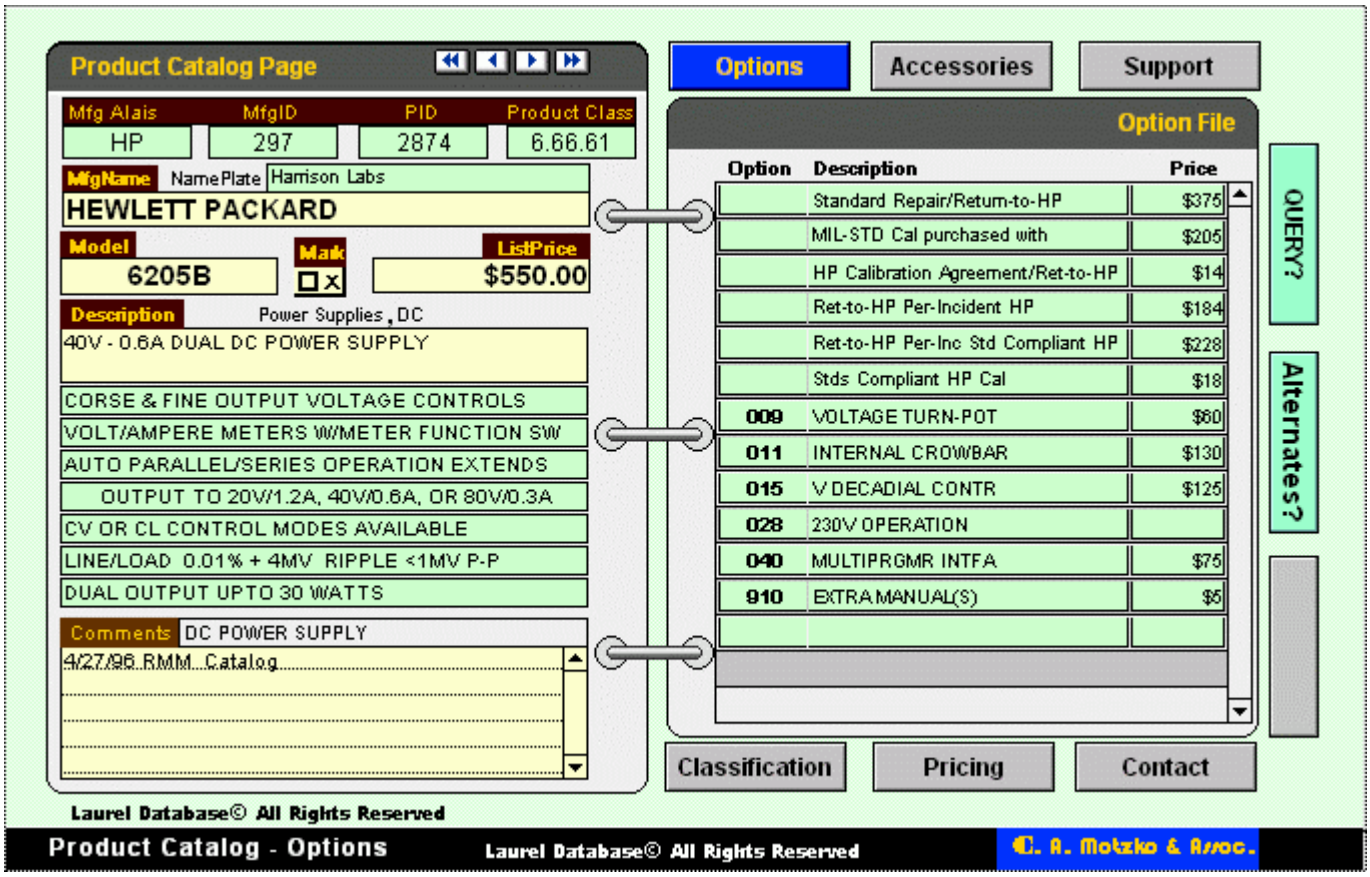

Figure D

RELATING THE FILES FOR NORMALIZATION

At this point, you might consider returning to the Calibration Recall Database and deleting the four fields that have been determined to require normalization — **don't do this**.

You must relate the source databases to the product catalog by using one common field. If you don't, there's no way to know which manufacturer supplied which model number. In this case, the most logical field to use for relating the product catalog to your database is the concatenated manufacturer/model number that is built for the normalization process.

So, instead of deleting all five fields, return to the Calibration Recall Database and run the cleanup routines, which will look up the normalized data from the product catalog and populate the proper fields in your database. You will then end up with two files: (1) the source or initial file and (2) the target or normalized file.

In addition, all of the standard pre and post validation processes are carried out on the source and target file such as:

- Total item count
- Duplication check
- Asset or tag uniqueness test
- Value summations
- Date summations
- Manufacturer – model summations
- Classification summations
- Custodial summations
- Location summations
- … Etc.

with a reconciliation process to correct any found discrepancies.

One of the key benefits from data normalization is the inclusion of "structured noun descriptors". This is extremely useful as it provides a standard description for the classification routine and promotes full parametric searches on the descriptor field.

A typical structured noun descriptor, for a normalized data line, has a basic structure would look much like the following. However, one must always keep in mind the legacy system's data field structures and length.

Group Code **HP** 6205B **009/011** 40V - 0.6A Dual Dc Power Supply
| | | | |_Structured Noun Descriptor (within system constraints)
| | | |_**Option(s)**
| | |_Root Model Number
| |_**Manufacturer (full name or acronym based on system constraints)**
|_ Unique "Group ID" or index used by the application for classification and selection to select & display possible alternatives or candidates for substitution The Group ID is an integral part of the product catalog and is usually invisible to the casual user.

The actual process of normalization is tailored to the requirements of both the source and target database. A simple process chart that illustrates this is shown in **Diagram 1**:

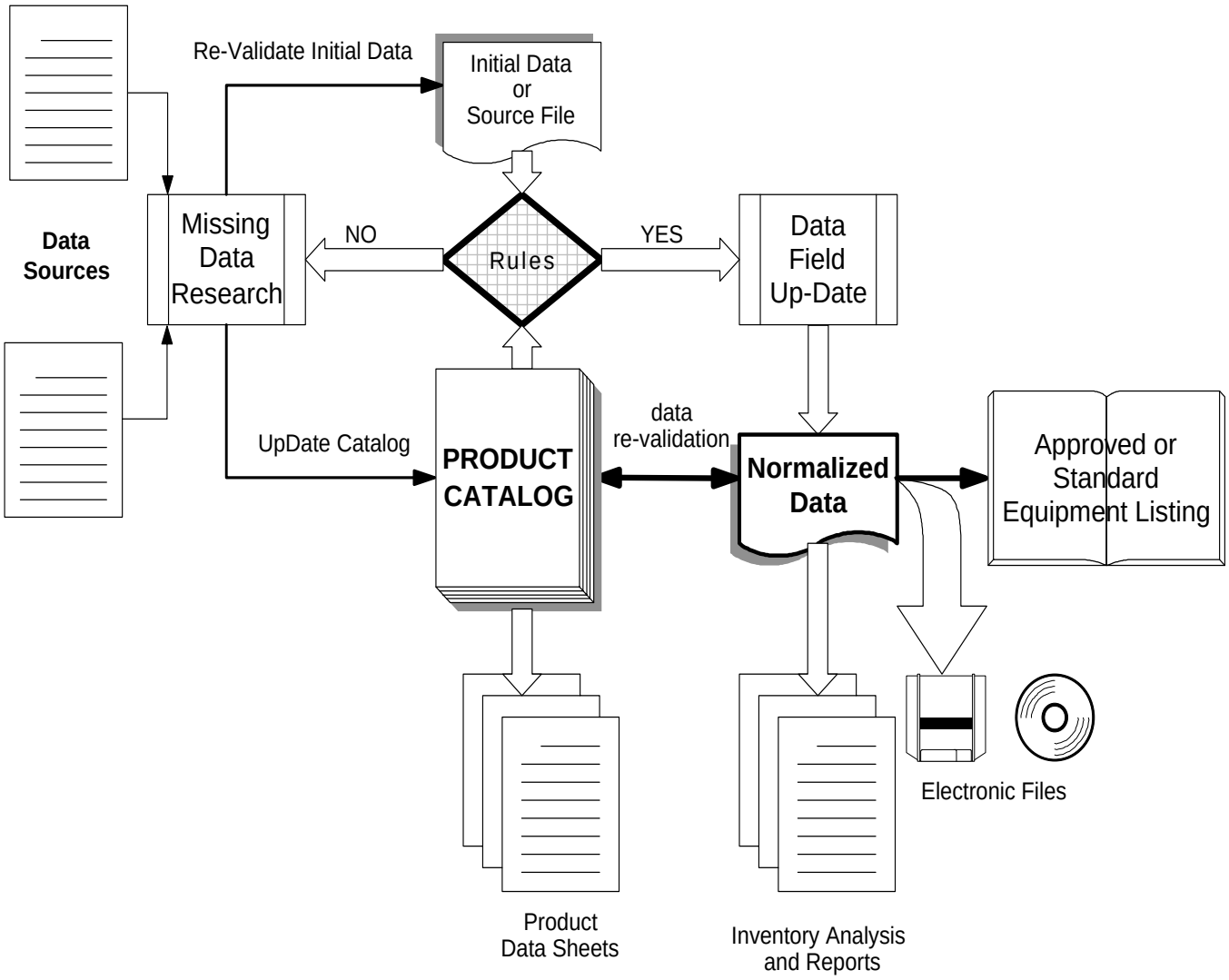

Diagram 1

BRINGING IT ALL BACK TOGETHER

We have now reviewed the normalized Calibration Recall Database, validated the information and examine the changes. This is modified data set is shown in **Figure E.**

CALIBRATION RECALL DATABASE (normalized)							
Asset No.	**Manufacturer**	**Model**	**Options**	**Description**	**Serial No.**	**Date Calibrated**	**Date Due**
12345	HP	6205B	009-011	40V - 0.6A Dual Dc Power Supply	NSN	12/1/1998	12/1/1999
56789	HP	6205B	009-011	40V - 0.6A Dual Dc Power Supply	2412A	11/15/1998	11/15/1999
98765	HP	6205B	009-011	40V - 0.6A Dual Dc Power Supply	67890	7/9/1997	7/9/1998
43210	HP	6205B	009-011	40V - 0.6A Dual Dc Power Supply	1218A85210	NCR	NCR
13467	HP	6205B	009-011	40V - 0.6A Dual Dc Power Supply	2236A12345	11/29/1998	11/29/1999

Figure E - Target Database

Here the validation was as simply as the item count is the same, manufacturer – model review matches, no duplicates occur and dates by unique asset number correspond. It is evident that more complex databases will require a greater range of validations to take place before acceptance.

Now that you've finished, you'll find that the layout looks exactly like the original, however the change is that you now have data that is meaningful and supports your management information and reporting system. The bottom line is that your test and measurement inventory in **consistently** described, in understandable terms, and prepared for inclusion into whatever classification scheme you have chosen.

Any subsequent changes in the four common data fields are done once! Not by individual asset or serial number. Queries, with global change, or lookups to a product catalog (and the product catalog's attendant update tables) can be used to modify the common data fields with a minimum of effort and user training.

It must be noted that the author does not know of an easy fix to deal with missing or incomplete serial numbers other that a physical inventory and this can become VERY labor and time intensive!

One possible approach, which has met with some success, is to validate and record serial numbers on every asset movement or calibration. Over time one will touch the majority of the active inventory. For inactive inventories or items in storage, again the only known fix is a full physical inventory.

CONCLUSION

As you work more with database design and normalization, the task will become intuitive. However, data normalization can take a little getting used to. In this article, we've discussed the need for data normalization and some of the expected benefits from the process.

With careful planning and a concise implementation plan, database normalization projects should only have to be done once (Hurray!).

Working within a legacy system's data field structures and length, surprisingly good results can be obtained.

In many cases, the Test and Measurement Equipment Inventory and Calibration Recall System is "Mission Critical" to the organization. This is where you need to seek the full involvement, advice and assistance from of your in-house data experts, computer department or contract with an outside consultant so as not to compromise the mission.

Some of the side benefits of the normalization process that are not readily apparent are:

- All dates can be adjusted to a Year 2000 compliant format, *if your database is ready for this step*!
- Manufacture – Model classification into Family, Group, Characteristic and Parameter can be accomplished within the normalization routines. This can be a simple three schema architecture or as complex as a seven level architecture. The question of depth is one of resolution and resources to assign to the classification effort. This step supports the full equipment classification process required for inventory analysis and equipment management.
- The structured noun descriptor fully supports parametric searches in the *Description* field.
- With grouping and structured noun descriptors, one is ready to develop a substitution or alternative candidate listing of the existing inventory (for improving asset utilization) and is a major step towards model standardization or approved equipment listings. This benefit alone will provide the cost justification for any test and measurement equipment data normalization project.
- Ability to group and accomplish "Interval Analysis" on like or similar assets
- "Uncertainty Budgets" can be established by equipment classifications or selected manufacturer – model groupings
- Inventory reconciliation and valuation efforts are greatly reduced with vastly improved accuracy.

Charlie Motzko has been directly involved with test and measurement equipment for better than three decades. He is a State of California Registered Professional Engineer, in Quality, and is senior principle with C. A. Motzko & Associates which specializes in "Sensible Solutions for Test Equipment Management"©

All questions are welcome, contact us at:

C. A. Motzko & Associates, 546 Keelson Circle, Redwood City, CA 94065-1212
Tel (650) 595-8878, Fax 650.631.3808, solutions@motzko.com