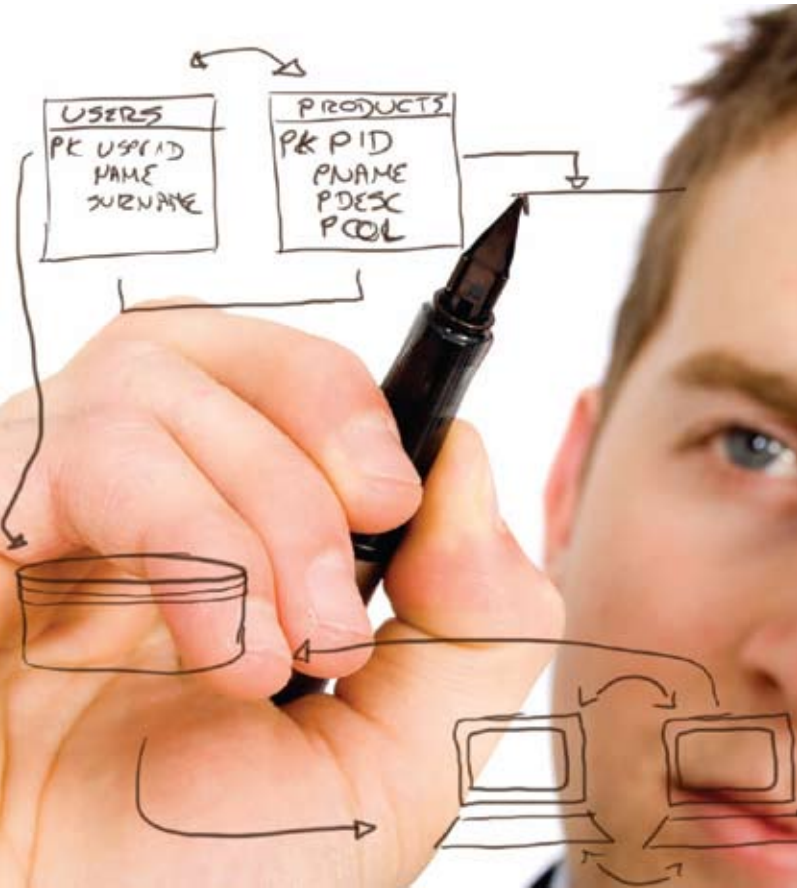
The Road to Conversion

The critical value of clean legacy system data



By Thibault Dambrine

Anyone who has been through a software package conversion knows that, even in the best of cases, these types of assignments can be stressful. There are a number of reasons for this.

- The new package implementation is essentially a bet by management that a new, bigger, better, typically more complex, more sophisticated packaged application will work better than the one currently being used. The new package will typically cost more than the previous one, the conversion will entail a one-time project budget, there is a lot of money at stake, and failure is not usually considered an option.
- It is always good to remember that software is really about the business first and the people who run it. Converting to a new software package also means a change in process, in the way the business will go about their day-to-day activities. Change is stressful, for the users as much as for the IT staff.
- From a technical, purely "software conversion" point of view, there are typically a number of unknowns at the beginning of the project. Items such as the state of the legacy system data, how the mapping will be done, and the conversion effort estimation are just a few. All these will have to be determined at the beginning of the project, typically with educated guesses, more than hard data.

In this article, I will expose techniques and principles designed to minimize the stress inherent to most conversion projects. The aim of these pre-conversion steps is to ensure the conversion stage would start with the best possible chances of success. While aiming to start a conversion project on strong foundations sounds like common sense, the tasks necessary to achieve this state of readiness are not necessarily obvious or trivial.

Picture yourself in the following situation: You have a new job, in a different city, on the opposite coast. Your move will be billed by the pound and you will pay for half of it. Would you want to take everything you have in your current house? You may ask yourself: "Is everything I have accumulated still useful?" and most importantly: "how much will it cost to move items that are not necessary useful and that will take space?" Typically, this would be a good opportunity for tossing, selling (garage sale), or even giving away as many obsolete or unused items, as possible before the actual move.

What makes easy sense in the example above does not always jump to the top-of-mind awareness as the first step in a systems conversion project. The perception, often, is that if the data is "good enough to work with, it is probably clean enough". The reality is that most likely, a significant percentage of the data no longer has any significant value. Worse, it will cost more to convert and will probably become a liability. To avoid this, the following steps can be taken:

1) Obsolete data can be identified in advance—and thus can be completely ignored—and not subsequently converted.

2) Valid data destined to be converted can be cleaned. Toward this end one must:

- Ensure that the data is complete. Examine each column/row that will be exported to the new system. Ensure that this data does not have blanks or NULL's.
- Ensure that the data is accurate. Make sure that it is valid and current.
- Ensure that the data is consistently positioned within free-form fields. For example, if going from a system where the address fields are free-form to one where each part of the address has a discrete field or column – e.g. apt #, street #, street name, city, county, state or province, zip or postal

code and country, you may want to ensure that the data is consistently placed in the same position within the free-form fields. This way, you will be able to pick up information at predictable positions and save time.

- Ensure that there are as few exceptions as possible in the data if a rule should apply. For example, if in your legacy tables, plant X item descriptions should start with "X", let there not be any "Y" item numbers made available for this plant. Simple exceptions such as these are enough to cause more work at conversion time.
- Ensure that the data is free of duplicates. For example, look for duplicate descriptions, addresses, manufacturer's part numbers, names or telephone numbers. There are many ways to look for duplicates and every one that is removed will be a future saving in the new system. One less piece of lint to worry about.
- Ensure that the data matches reality, by involving the business is involved early on. They work with the data every day. They will be the best people to be able to say "what makes sense". If you tell them I have cleaned up the Bills of Materials and only 300 out of 3000 will be converted, ask the business: does this make sense? If your business analyst thinks your figure does not pass the smell test, you may want to have a second look at your selection criteria.

First Things First – Create a "Data Cleanup" Project before the conversion can start. To achieve the expected benefits and savings anticipated at conversion time, a pre-conversion data cleanup project cannot be casual. Best practice would be to write a formal Statement of Work (SOW) identifying the goals and rules that will guide the project activities. Management will want to make a conscious effort to review, understand and approve this SOW. They will also allocate a budget to finance this effort and if possible, engage a project manager to run and monitor this task. Here are some things to look for in a Statement of Work...

Divide and Conquer: Identify your "Data Objects"

To understand what data to clean and what changes to make, one must know something about both the origin and the destination of the data. The assumption here is that both old and new software systems will have similar functions; e.g. an older ERP system to a newer ERP system. Since the purpose of the new system is similar to the old one, but likely more powerful. It may have similar data, but not necessarily arranged or organized exactly the same way. Expect also that while the general concepts may remain, the specific terminology in the new system may be different from the old one. Mapping the old file names, field names, and sometimes even module names early on will ease the understanding.

A simple and effective way to match the data between the old and the new systems in view of the cleanup and subsequent conversion is to use the concept of "data objects".

Data Objects are "logical groupings of related data sets".

An Example of these could be "Item Master", which can actually be a set of tables, rather than a single table, covering generic items used across all plants, items used at specific plants only, items representing labour as opposed to physical items. An other example could be "Customer Care", a set of tables covering information on customers such as name and addresses, previous purchases, discount structures, memberships etc. Identifying data objects has some immediate benefits:

- It allows a division of labour. The cleanup exercise can be divided along data object boundaries. Each data cleanup sub-team can tackle a separate data object or data set. Responsibilities are clearly delimited.
- Data objects in the legacy system can be matched to data objects in the new system. For example the vendor master data object on ERP Package A should be comparable for the one in ERP Package B, even if their respective physical implementations (data structures) are different.
- Understanding the mapping of old to new data system at the data structure level will enable the data cleaners to do more than just cleanup duplicates or obsolete data. With this understanding, the legacy data cleanup can be used as an opportunity to optimize the legacy data to facilitate the conversion.

The Big Rules and Scope

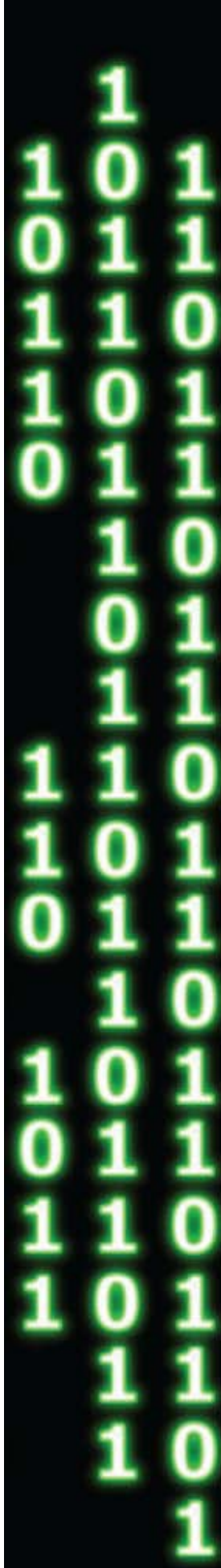
Now that the data objects are identified, the second item to attack is the "Big Rules". "Big Rules" can be likened to general guidelines that will help both the business and the stakeholders to make decisions as to how the data will be either deemed obsolete (no need to clean!) or how the cleanup needs to be done. Here are some examples:

- The conversion scope will span data created within the last 24 months.
- Any purchase order that can be closed should be closed prior to the conversion. No closed POs will be converted.
- Any vendor with no activity within the last 12 months will not be converted.
- Any cleanup activity should be done by the business, in the (legacy system) production systems wherever possible.

Cleanup Methods: the "How" and the "Where"

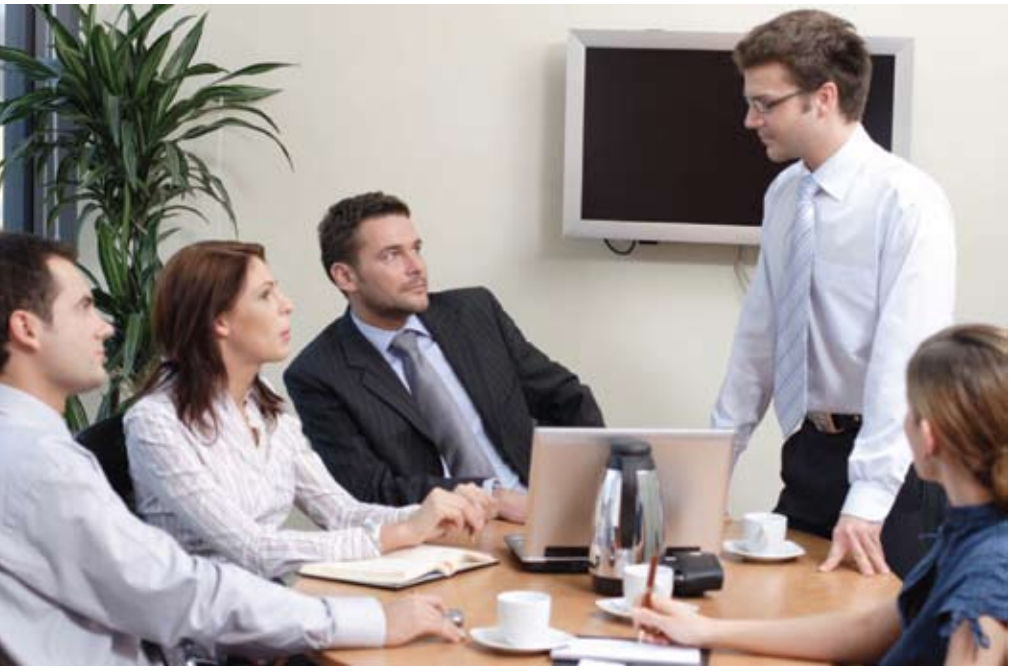
The data cleanup, when done in the legacy system, can be done manually, with some measure of automation (read programs) or a mixture of both, depending on the data set:

- **How: Manually.** Clean up should be done



manually if the data volume is very small and all of the data to be cleaned up can be accessed by the business using their access screens, or if the data is highly unstructured and has to be eye-balled to "make sense" before cleansing

- **How: Using software.** Clean up with programs if the data volume is larger than what can be economically done by hand and simple rules can be applied.



The business must be involved early in the Big Rules decisions

- **Where: In Production.** If the data is cleaned up in production, chances are it will stay clean and will not have to be cleaned again. Each new data extraction from production should be cleaner than the previous one—and ideally, data should not have to be cleaned more than once.
- **Where: Using a staging area.** In this situation, the data is extracted from the legacy system into a staging database, cleansed and reformatted to be ready for the conversion outside of the legacy system. If there are multiple practice conversions before the real deal, every practice will require a re-run of the cleanup routine in staging.

Putting it all together, you will likely use a mixture of all of these methods, but keep the following guidelines in mind: As much as possible, ensure manual or labour-intensive cleanups are done in production. All repeatable, program-run type of cleanups can be on either production or staging, but being able to clean in production will (as a rule) save time. Yes, I have written this once before—it is worth repeating. The business will not necessarily like having extra work, but it remains worthwhile. Extract from the legacy system, clean the data in a staging database, transform and export to the new system – OR – clean the data in the legacy system, then extract, transform and export to the new system. It could also be a combination of both solutions, depending on the data objects being converted.

A good rule of thumb is to decide early which method would suit your conversion situation best for each individual data object. To decide, a good test is to ask is, "Does this data have dependencies?" When changing or removing any existing data within a complex system, one has to be careful not to "orphan" any data—a purchase order without vendor for example. Typically, this type of situation is easier to monitor in the legacy system than when doing the cleanup outside in a staging database, which will not likely have all the same triggers and referential integrity structures as the legacy system.

A note on resource planning: Physical resources: planning a staging area to do the cleanup is no different than any other system. It will require disk space, processing power and people to do the job. The business people already have jobs and you will hear this more than once if you ask them to start cleaning and standardizing data. The only mitigating factor here is to plan early. Data cleanup does not come naturally and the ramp-up is typically long.

Define your Cleanup Approach

The following is a list of the different components to consider in the approach to the data cleanup:

1. Determine data extraction criteria

Minimize— as much as possible—the scope of data to be converted. This will affect every subsequent cleanup activity, as in "make them less painful". Figure out criteria that will help minimize the effort. Do not cleanup data from divisions that are not operational. Do not cleanup data that is past a certain age. Any identifiable criteria you may find and can validate with the business to reduce your scope will help.

Standardize the data—for example, the same description should be used to describe the same material even if it was in two different locations.

Ensure consistency. For free-form fields to be converted to discrete columns, one of the biggest benefits of a pre-conversion data cleanup is to make sure the data consistently follows strict rules. For example, if the legacy system has four "free-form" lines for the address and the target system has discrete fields for each part of the address, the cleanup opportunity is to ensure that even in the free-form legacy system, each part of the address is stored in a consistent position. This will ease the conversion process by reducing the effort to take the data from the old system to the new. Your SQL conversion code can then pick up each data element in a specific spot without having to "fish" it out with a LOCATE operation. For dependencies that may not be enforced with database level constraints, ensure (for example) that there would be

no children without a parent (e.g. order details without order headers). One of the side benefits of good consistency is that it enables the spotting of duplicates, which are undesirable.

Determine if the cleanup scope needs to span the data coming via interfaces. In other words, should data provided by other, (external or internal) systems via interfaces be part of the cleanup scope? Note that it may or may not be possible, especially if that data comes from outside the company.

Identify and remove duplicate records. None of them should be left before the actual conversion starts. Note that spotting duplicates can only be done well after all the previous steps are completed.

2. Close out completed transactions

Ensure that housekeeping procedures (if they exist) have been adhered to and are up-to-date. Examples are:

- Verify current min/max levels for stocked items.
- Close out POs linked to Work Orders with status complete.
- Set all completed work orders to status complete. This is an opportunity to easily identify data that should not be ported over in the conversion.

3. Reconcile financial / inventory balances:

For any item with a financial balance, ensure the balances reconcile between different ledgers. For example:

- Non-reconciled open items (GL, AR)
- Bank reconciliation
- Reconciliation between accounting systems and legacy systems
- For any item with an inventory balance, ensure that physical inventory has been taken within an acceptable time period. Financials are a particularly sensitive area.

A conversion will be deemed questionable or even failed if the financial picture is affected significantly for reasons that are difficult to explain. This part has to work. The company depends on it.

4. Complete missing information:

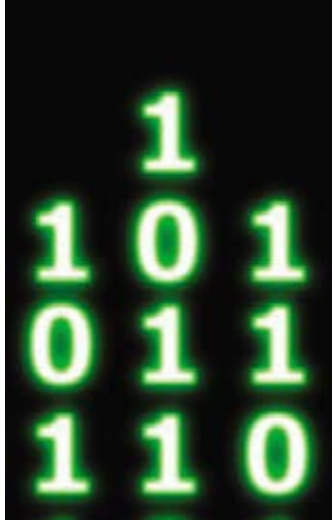
Identify any data that have been created with incomplete, incorrect, or outdated settings or content. The business will not stop running while the IT department is cleaning and preparing the data for the conversion. For that reason, it is also important to ensure that as little as possible "new bad data" is created.

To ensure this is the case, you may wish to update actual procedures to set up data in the source system, and create verification procedures to monitor that the new data entered conforms to the updated procedures, i.e. to ensure that the data remain consistent, complete, and correct.

5. Set measurable effort and result goals:

In his original observation (early 1900's), the Italian economist Vilfredo Pareto noticed that 80% of his country's land was owned by 20% of the population. The Pareto principle, as it is now named, also known as the 80/20 rule, will apply in data cleanup assignments.

- While working through data cleansing and migration projects, there is a cost benefit to achieve a balance between effort and accomplishment. True to the 80-20 rule, a high percentage, say, 80% of records to be migrated will likely require little or no modifications. The remaining 20 percent on the other hand may require a combination of committee decision making, meetings, exception processing, extra system time and ultimately—solutions requiring creative data manipulation.
- When analyzing the 20% of the data which will require more effort, do not forget to understand what areas are critical to cleanup. These should be done first. The benefit of maximizing the return on effort in this area in particular is critical. Typically, there is a finite conversion budget and it is important to use it wisely.

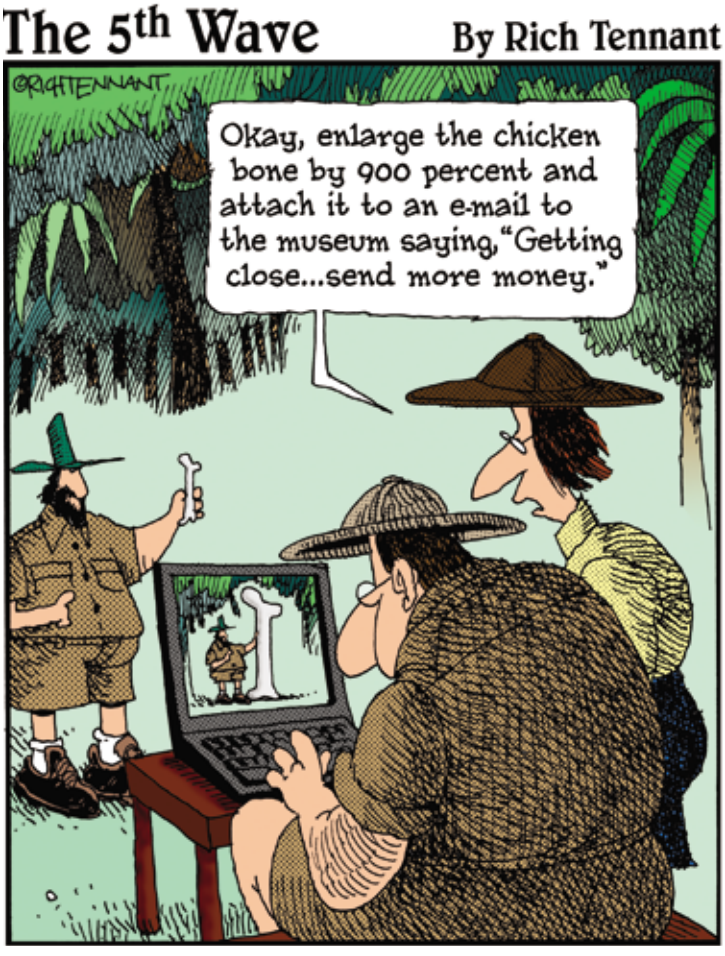


Engage the Business

The implementation of a new data processing packaged system is above and before all a BUSINESS DECISION —not just an IT decision. The business must be involved early in the Big Rules decisions, in recognizing what cleanup methods will be used, in deciding what is the scope of the data cleanup will be, and in deciding who will be involved, from the Business point of view as well as IT. Implementing a RACI chart (Responsible, Accountable, Consulted, Informed) early on will help define who will need to do what during the project. For each cleanup activity, figure out who needs to be involved, what method of cleanup will be used, and who will signoff on the cleaned-up data— in effect, who will be the "data owner".

Initiate a Data Cleaning Project Plan— and Stick to it!

Cleaning data prior to a conversion is typically not a simple task. There are a number of tasks and a number of stakeholders and priorities in the mix. It should be planned as a proper project. It is desirable to assign a project manager to put the tasks in order and monitor their progress to complete the task on time. Start Early. You will never do enough to be completely ready. Unless your business is bankrupt, you will constantly have the challenge of dealing with new data coming in while cleaning up what you can see is worth cleaning. Starting early will help.



In Conclusion

Idle data is not good to keep in the best of circumstances. It literally uses disk space, backup time and resources, requires attention, all the while adding little or no value. In the case of a conversion project, it actually adds negative value by costing more to convert and often forcing more exception processing to ensure the data goes over.

Software conversions are typically complicated projects. Stephen Covey, in his first two habits (out of 7) of highly successful people, said it best: **Habit Number 1:** "Be Proactive." This is what a pre-conversion data cleanup is all about! **Habit Number 2:** "Begin with the End in Mind." Never lose sight of the final goal, which is a smooth, successful conversion. With that in mind, BEFORE the conversion: 1. Make every effort to reduce the data to convert (identify old data, scope it out). 2. Make every effort to ensure that the data to convert is clean, consistent, and complete. The net effect of a pre-conversion data cleanup should be a much smoother, less stressful conversion effort—one that would not give any surprises and that can use a minimum amount of rules. TUG

Thibault Dambrine *works for Shell Canada Limited as a senior systems analyst. He holds the ITIL Foundations as well as the Release and Control Practitioner's Certificates. His past articles can be found at www.tylogix.com.*