

Publication: Data Management Review

Column Title: Data Warehouse Delivery

Author: Douglas Hackney

Headline: Surviving the Curse of Legamarts

Issue: September 1997

[begin copy]

Almost everyone has existing, non-integrated, non-architected data marts humming away in the business. Some are even consciously pumping them out in an effort to stave off fast arising business problems. (A special place in hell awaits these souls...) For all of us, the every present question is, "what do we do with these things, and how do we get them integrated into a data warehouse/data mart environment?". Obviously, the best choice is to let them die a natural death, to be replaced by real, integrated, architected data marts. For many, however, simply letting the LegaMarts of the business wither away on the vine will not be an option. They will be the unlucky ones sitting in the room when someone farther up the food chain issues the long dreaded directive to "integrate yon LegaMarts." Thus will begin a long, painful journey. Between the time you are assigned this unenviable task and the last LegaMart is integrated you will spend many hours disparaging the talents and ancestry of any and all involved in the creation of these non-integrated monstrosities. LegaMart integration is a difficult, challenging and often ugly task. While I can offer you little comfort in this effort, I can offer you a road map to success. The process is:

1. Identification of unique LegaMart elements

The first step is to identify all unique elements in the LegaMarts. These are dimension members, metrics and calculated aggregations that are unique to that LegaMart and are not integrate-able with other LegaMarts. For example, a U.S. based headquarters staff might utilize a LegaMart with a columns labeled "Domestic Sales", "International Sales" and "Canadian Sales." These columns would not be entirely integrate-able with the Canadian LegaMart's columns labeled "Domestic Sales" and "International Sales." Although the "Canadian Sales" column is common in both LegaMarts (albeit with different semantic terms "Canadian Sales" and "Domestic Sales"), they both have a unique view of "International Sales." Each of the unique attributes you identify will need to be re-labeled in the post-integration environment. In this example, the Canadian LegaMart's columns would need to be re-labeled "Canadian Sales" and "Canadian International Sales", while the US headquarters LegaMart's columns would need to be re-labeled "US Domestic Sales" and "US International Sales."

2. Identification of common elements

Second, you must identify all common elements across all LegaMarts. Start with the dimensions and then proceed to facts, aggregations, histories and status tables. It is important to note that you cannot expect to find common semantics across these LegaMarts. You will need to examine the underlying sources, integrations and business rules in order to determine if "Gross Sales" in one LegaMart is actually the same thing as "Gross Sales" in another. Welcome to the pain of LegaMart integration. This process is very time and resource intensive, and will easily suck all available resources for a very long time. As the days turn to weeks, and the weeks to months, and your team is still slogging through line after line of COBOL to decipher undocumented business rules and arcane logic, it is a good time to reflect on the fundamental rightness of an enterprise data mart architecture.

After the seasons have changed, probably more than once, you will finally have a definitive list of all common elements in the enterprise's LegaMarts. At this time, you will most likely need to do a second pass through them, since they will have continued to evolve and mutate in the months since you started this effort. The good news is, it goes faster the second time, since all you're looking for is delta changes in business rules and transformations since the first pass. About three quarters of the way through the second pass is another excellent time to ponder the righteousness of building incremental data marts under the auspices of an enterprise data mart architecture.

Finally, you will have a current list of all common elements across the entire collection of LegaMarts in the organization.

3. Lay back of common dimension keys

Your long journey through the second step will have led to the discovery that dimensions using the same name in multiple LegaMarts have wildly different keys. Your next job is to establish common dimension keys for all LegaMart dimensions, and to lay the new key structure back out to the LegaMart dimensions. This is not always as simple as it sounds, because at this point you are going to start altering the schema of the LegaMart itself. Up to now, the owners and operators of the LegaMarts have only had to issue you a user ID and password so you could examine the data. Now we begin the process of you actually taking control of the destiny of the LegaMart. You can expect some raging turf wars in this stage of the process. The best approach is to minimize the amount of structural change as much as possible. For instance in the case of the layback of common keys, simply append an additional column to their tables with your new key rather than eliminating their existing structure.

At the conclusion of the process you should have all LegaMarts with common primary keys in all dimensions, this puts you into the position to perform the next big step of bringing all the dimensions into one place.

4. Integration of dimensions

Assuming that you have not suffered any mortal wounds at the hands of the protectors of the LegaMarts you are ready to integrate all the LegaMart dimensions into common dimensions for the enterprise. This involves eliminating redundant dimension members and beginning the semantic battle. You will find that although the sales LegaMart has a dimension called Product, it will be very different from the manufacturing LegaMart dimension also called Product. They will contain different hierarchies, different descriptive attributes and different metric attributes. It is your job to eliminate the duplicates and build a master product dimension acceptable to the constituencies of all LegaMarts.

In order to accomplish this you will need to duplicate the hierarchies of each group and include all pertinent dimension metrics from each LegaMart constituency. Expect to end up with many columns containing various flavors of the same basic attribute all labeled with their source LegaMart, i.e. "Sales Product Type" and "Manufacturing Product Type". Build the common dimension masters in a separate RDBMS designated to hold the "enterprise master version of reality", which will become your reference architecture for the integrated data mart environment and the master source for all dimensions to be replicated out to the subsequent independent data marts.

In many cases, you will be unable to gain consensus on a single source system of record for the business dimensions. For example, the Sales group may only consider the order entry system suitable as a source for the customer dimension, while the Service organization will only allow the service dispatch system to be the basis for their view of "customer." In this case, you will need to turn to one of the point solutions that are available that perform integration and de-duplication on multiple sources and output a single master file.

5. Establishment of common semantics

By far the greatest challenge in any enterprise wide initiative is gaining consensus on common semantics. Your goal is to gain as much conformity as you can. Don't expect to ever get 100% buy in for all the common semantics across the enterprise. You will always face a small group of users that "have always called total income Net Sales, and that's what we'll always call it, period!" Fight the good fight, and do what you can to get the business to agree that the sky is blue, and the official term will be blue, not indigo, robin egg, deep sea and royal. Best of luck.

If you do happen to get close, don't be afraid to work additional rounds of meetings to drive to consensus. If you can get down to only a few attributes with disparate semantics with only a few hold out user groups you can play the "peer pressure" card and force them to play along, however reluctantly.

Once you have done what you can in the what will be referred to far into the future as the "great semantic wars", apply the common semantics across all the master dimensions you have built. For all attributes that you were unable to gain consensus on, you must carry duplicate columns, each labeled individually.

6. Establishment of common business rules

Now that you have established what the common business dimensions are, and established what the common semantics will be, all that remains is to establish common business rules for the calculation of all metrics in the system. And you thought common semantics was tough! Now you get to try and convince the Sales LegaMart constituency that they should adopt the method of calculating "Net Profit" used by their arch enemies in the Marketing group for the Marketing LegaMart. You won't win all these battles either, so expect to be carrying multiple columns populated with various forms of the same basic metric. You can expect to end up with "Sales Net Sales" and "Marketing Net Sales" side by side on reports in the future. People become very wedded to their particular method of calculating "the truth", and are loathe to adopt anyone else's definitions. Again, your goal is to gain the greatest level of consensus as possible. Also, don't be afraid to make another pass at consensus when certain bull headed personalities are removed from the equation, even if it is months or years after your first efforts.

7. Integration of facts

Once you've got the business rules established, you are ready to integrate the various versions of the transaction detail facts found in the collection of LegaMarts in the organization. Again, the process is to create a master integrated environment built from the LegaMarts that is separate from all LegaMart environments. If you don't create this separate system you will most likely fall prey to internecine political and turf warfare among the competing LegaMart groups. After you are finished with the master integration, you can easily provide integrated versions of each group's LegaMart.

Facts are usually fairly painless to integrate, as they are reflections of the base OLTP transactions. Your biggest worry will be variances in scrubbing algorithms and among the various LegaMarts.

8. Re-create aggregations

Since you have fully integrated dimensions and facts available, it is a fairly straight forward process to re-create the aggregations found in the LegaMarts. This process is very rarely problematic. Your biggest challenge will be in consolidating multiple data mart users in one multi-dimensional aggregation, should you choose to do so.

9. Scrub and re-build histories

Unfortunately, this one is more challenging than spinning up the aggregations. You will need to re-cast the existing history tables using any new business rules established in the process and apply the common semantic standards of the integrated environment. For simple relationship histories, such as Product History that tracks Product, Cost and Price the effort is very low. For time series snapshots that capture the entire dimension, you will have a formidable effort ahead

of you if you choose to re-build the entire history. If you can get “buy in” from the user community, it is obviously much easier if you simply start fresh with the new integrated dimensions and begin the historical snapshots at the time of deployment.

Conclusion

If you have made it to this stage of the process with your wits still about you, it is time for a well deserved celebration. The successful integration of LegaMarts is one of the most challenging tasks you will ever face. It requires high levels of technical, leadership and political skills. To say nothing of an iron will, boundless determination, endless energy and very little need for sleep.

The LegaMart integration effort will yield a consistent, reliable enterprise data mart environment that shares common sources, dimensions, semantics, business rules and metrics. As the astute reader will no doubt notice, these are the exact characteristics of an enterprise data mart architecture. Avoid the pain and cost of LegaMart integration, build your data marts using an enterprise data mart architecture. You will never regret the investment.

Excerpted from “Understanding and Implementing Successful Data Marts”, a forthcoming book from Addison Wesley Longman Publishing. Further information is available at www.entergroupltd.com.

[end copy]