

“Going Global” with Data Warehousing

By Stephen Dine, President Datasource Consulting, LLC
& Matthew E. Pohl, President Datavision Consulting, LLC

Abstract

On top of the common challenges of creating a data warehouse, the need to source data from multiple countries into a “global” data warehouse adds new complexities. Using lessons learned from a global data integration project at Sunrise Medical, Inc., this article addresses some of the unique challenges of “going global” with data warehousing. It covers both technical and non-technical stumbling blocks you will likely face, and provides possible solutions.

Introduction

As any data warehouse professional can attest, there are always many challenges to building an enterprise data warehouse. These challenges range from ensuring the right hardware capacity is planned, trying to select the best ETL approach, creating a robust data model, dealing with bad data, and fighting internal political battles.

To take data warehousing “global,” there is another layer of complexity added to a data warehousing effort. Having recently finished the implementation of a global data warehouse at Sunrise Medical, Inc., we thought other data warehousing professionals might benefit from knowing some of the uniquely global issues we faced and how they were overcome.

Project Background

Sunrise Medical, Inc. is a \$650 million manufacturer of durable medical equipment. Its products range from wheelchairs (e.g., motorized, sport, and pediatric) to furniture for nursing homes. Like many companies in the medical equipment industry, Sunrise has grown, in large part, through acquisition. As such, many Sunrise divisions have operated independently for many years and conduct business on a number of different transactional systems.

Sunrise is also a multinational company, having manufacturing operations in seven countries and distribution divisions in six more. The company is based in the U.S., so financial information is consolidated every fiscal period into a common currency (U.S. Dollars). Unfortunately, the information accumulated for financial reporting was not of sufficient detail or frequent enough for the Global Product Management Group. The

Global Product Management Group is responsible for monitoring sales across the globe by the different product lines. To accomplish this, sales information needed to be available down to the item number level of detail, and be more timely.

In the fall of 2000, the Global Data Warehouse project (GDW) was initiated. The first three months were used to formalize the business requirements for the project. Joint application development (JAD) teams were used as the primary vehicle for collecting information in order to limit the number of interviews that needed to be scheduled. Development began in March 2001, and lasted for the next seven months. In order to get global data, it was determined that 12 systems would need to be sourced. The Sunrise Global Data Warehouse went into production in October 2001.

Global Project Management

Project Communication

While not unique to data warehousing, communicating with team members from different countries requires special attention. Since English is often a second or third language for these team members, communication must be very detailed, specific, and comprehensive. We found early on in the project that our informal, abridged style of communication was causing more confusion and lost time than making the extra effort to be exact and succinct in what we wanted. Since a team in the U.S. was leading the project, most of the project tasks were being originated and managed from North America. Due to the significant time differences, team members initially relied on emails and voice mails as the primary methods of communication. Most messages were sent with requirements laden in slang and acronyms with the assumption that the resources in other countries would understand what was being communicated. In the best-case scenario a day or so was wasted giving further clarification, but in some cases weeks were lost due to misinterpretations and bad assumptions. In the end, what we found to work best was to follow up e-mails with phone calls and regularly scheduled audio conferences.

A major project asset is having a team member that can speak all or most of the languages the team will encounter. The Sunrise GDW Project had such a resource, and it helped in a variety of ways. One simple benefit was being able to send this resource sample data from the different systems to translate and decipher. Another way it benefited the team was that this multi-lingual resource could communicate to non-English speaking resources, allowing us to access personnel we would not have been able to otherwise.

Work Schedules

One challenge that might be obvious at first on global projects is the differences in work

schedules. This includes both the differences in time zones and the differences in vacation/holiday schedules. Not only can it affect the project schedule, but also the team dynamics as well. Between the Western U.S. and Western Europe, there are only a few hours of overlap between workdays. When issues arise in the afternoon in the U.S., they have to wait until the next day for resolution. We learned that we needed to set expectations about using the hours of overlap wisely (e.g., regularly scheduled calls). We also needed to be sensitive to when we scheduled meetings. Many times we would schedule meetings at a time when we needed our resources in Europe to be focused on meeting a deadline. While it was convenient for the U.S. part of the team to have a meeting as the first order of business when we arrived to work, this was in the afternoon for European team members, a time of day when they were being the most productive. We also used this time to frequently call the European members with ad hoc questions. Our most successful approach to this issue was to have regularly scheduled conference calls with just the team members affected by items on the agenda.

While the U.S. only averages less than 14 official days of vacation and holiday time away from the office, it is not uncommon for many European countries to average more than 30. As much as we argued, we were unable to convince management that the U.S. portion of the team should benefit from these global benefits, since it was a global project. Instead, the team had to adjust for the fact that most Europeans take long breaks during the summer, and that the Sunrise facilities in Italy (including IT personnel) basically take the month of August off. This cultural difference must be kept in mind when planning a global project, as it may directly impact your ability to meet your deadlines. Therefore, be sure to understand planned time off for your project resources, and do not act surprised when they respond “all of August.”

Technical Challenges

WAN Connectivity

In the U.S., we seem to assume that other countries operate with the same communications infrastructure with regard to technology. One area that is easily overlooked is WAN capacity and support.

Often, smaller non-U.S. sites are connected to corporate via ISDN, small frame pipe, or potentially even local dial-up. These types of connections make scheduling daily extracts much less reliable or, at the very least, more complex. Be sure to get a high level topology of the systems you are considering early in the project to see if some systems will require upgrades before they can be readily sourced into the data warehouse. Like most data warehouse projects, a detailed data volume analysis will need to be conducted to ensure that your load window is sufficient given the bandwidths available.

Usually the primary reason for low bandwidth is cost. It can be prohibitively expensive to

lease a mere 128Kbps frame connection in many countries. Fortunately it is often less expensive to implement a high speed connection to a local ISP. With tools now available you can transmit your data across the Internet to your local server and decrypt the packets upon arrival. You can also take advantage of data compression to help move more data across smaller pipes, but it will add another layer to manage in your ETL process.

Assuming you have the influence to get your organization to bump up the bandwidth to your other divisions, you will also need to consider how these lines will be managed. You don't want your service level agreement (SLA) with the organization to be more aggressive than the SLA that your carriers have with your telecom group. You won't realize how many different companies are responsible for the delivery of your data packets until you are unable to contact one of your external servers. While you probably won't have much influence in getting the network back up any faster, you can plan for alternate methods of data transfer. During our implementation, one of the major carriers in Canada was sold and the other two carriers refused to carry any of their traffic until a new agreement was reached. Since we relied on a VPN connection via the Internet to carry our traffic, we had to wait until the companies worked out a new agreement before receiving data again.

Code Pages

Before this project, the term “code page” was not part of our vocabulary. Code pages control what characters are displayed for each ASCII value. Different code pages are used in countries that need additional, or extended, character sets. Examples of these characters include the German umlaut and the Spanish tilde.

Our first run-in with code page issues came when we first sourced data from Spain. They were using a European character set that was not supported by our ETL tool. This led to letters in customer names and product descriptions getting replaced by special characters (e.g., “smiley faces”) when loaded into the database. The challenge was that the European code page was required for some of the other jobs running in the same batch as our extract routines. These jobs involved the printing of invoicing reports that were required to run in the same batch as our jobs.

Eventually, we were able to standardize on the IBM-850-DOS code page. The extract programs were modified to specify this code page during the extract, overriding the European code page used by the rest of the schedule. We had to move a number of jobs to new batches in the source systems to make this work, which was more of a political challenge than a technical one.

End-User Tool Requirements

Many of the issues faced during the end-user tool selection process were similar to a normal data warehouse project. For example, users wanted a Web-based, easy-to-use application with integrated reporting and OLAP capabilities. Sound familiar? However, there were a number of requirements that were unique to meeting the requirements of supporting multiple countries. These included multi-currency support, multi-lingual interface, and global technical support and training.

Aside from the challenge of automating the exchange rate process, the GDW team had to figure out a way that users would be able to analyze sales information based on a variety of exchange rates. Users wanted to view sales information against monthly average rates, weekly average rates, and static rates across time. In addition, they wanted the ability to see these different rates across a number of currencies. Fortunately, the OLAP tool we selected was able to handle this quite well with little work on the database side. By connecting exchange rates to the Time Dimension in the multidimensional cube, users can access a single cube and see sales in all twelve of our applicable currencies. We were even able to use rates other than daily rates (e.g., average for the fiscal period) to convert global sales into a common currency. This allowed a single set of reporting/analytical functionality to be used across the organization, instead of having to create different reports for users wanting to see various exchange rates.

While the language of the data in the GDW stayed source-specific (except for common codes, where English was used), users wanted the front-end tool to be presented in their native language. The GDW team also felt that it was important to meet this requirement in order to standardize on a common application. While many vendors boasted about being able to support multiple languages, most required significant development and support on the GDW team’s end to make it work “seamlessly.” Since we were looking to implement a Web-based solution, sending out different versions of an application was not viable. One large vendor required that we install multiple versions of their product on an application server and use IP redirect to try and funnel users to the correct language. Fortunately, the Web portal selected provides seamless multi-language support for users. You can either specify which language the user sees or it will default to whatever language the operating system is presented. Therefore, our Spain users see buttons and menu items in Spanish. In the end, we were only able to find one vendor that could meet these specific requirements.

Local training and support by the tool vendor is also a common issue in data warehousing, but making this requirement global made it more difficult for tool vendors to meet this requirement. We found that most did not have a support presence in all of the countries where we would have users. The differences in time zones and holiday schedules (mentioned above) also impacted the level of support availability. There was

also the need to support multiple languages, given that at some of Sunrise’s smaller facilities the IT staff only spoke their native language. It is becoming more common for software companies to channel their tech support via a knowledge based Web site. When the knowledge base is available only in English, the effectiveness decreases for a global user base.

Data Issues

Currency Exchange (FX)

One of the obvious challenges of going global with a data warehouse is working with multiple currencies. Since the Sunrise GDW was extracting data on a daily basis from the source systems, the GDW was required to also obtain daily exchange rates.

There was a considerable effort made to identify a source for currency exchange information. Sunrise found a service from OANDA.COM that allowed a text file to be downloaded each day using a PERL-based FTP process. Obtaining the data daily allowed the GDW to also calculate a variety of average exchange rates that users desired for reporting.

In addition to the U.S. Dollar data stored in the GDW, we also stored the sales information in local currencies. This was necessary in order to more precisely reconcile back to the original source data. Had the GDW only stored the converted amounts, it would have made it very challenging to match GDW data to source system information.

Language of the Data

The decision was made early on in the project that English would be the primary language of Sunrise’s GDW. This, however, did not mean that we would be translating all source data into English. Data like customer names and product descriptions would stay consistent with how they were stored in the source system. English was used to standardize various codes (like units of measure) and for table and field names in the database. Wherever possible, ISO Codes were used to replace local codes. For example, each system had its own way of representing currency with a code, so these codes were mapped to ISO Currency codes. We also had to translate certain “local” codes to English in order to group related codes together. This was challenging since many times foreign words and codes did not translate perfectly into English. Our approach was to use different codes when the meanings were not nearly identical, instead of hiding the potential nuances in a single code.

In our situation, Sunrise was able to agree upon making English the primary language of the data warehouse. In many companies, this decision is not so easy. While there is no silver bullet for resolving this issue, it is important to recognize this potential pitfall up front. It is very costly to translate source data (both at the beginning and ongoing) and meta data into multiple languages. If this is an issue for your upcoming project, make sure that you have allowed for it in your project plan.

Reporting Hierarchies

To make sense of global data, one needs a common way to organize the data across a global company. Before the GDW Project, each transactional system had its own way of organizing customers, products, and geography.

It was decided that in order for the global sales information to be useful, new hierarchies would need to be established for these three. In order for this to occur, three major tasks need to be accomplished:

1. Businesses needed to agree on the structure —this was obviously not a technical challenge, but a political one. We spent many hours working through various approaches to the customer and product hierarchies, attempting to meet the global needs. What we found is that it was not possible on the first attempt to meet every local reporting need in the global hierarchies, especially when it came to classifying customers. Therefore, the global customer and product hierarchies were established to meet truly global reporting needs, and local classifications would continue to be supported in order to meet local reporting needs. The geographic hierarchy was easier, since it was based on ISO standards for Continent, Sub-Continent, and Country codes.

While this task falls mainly on the business to accomplish, it is important to have someone with data modeling skills involved to ensure the classifications are mutually exclusive, that each member of a level is really at the same level logically, and that the hierarchy can be modeled in the database.

2. Classification effort —Since the global hierarchies were totally new, all Sunrise customers and products had to be reviewed and classified. This effort required intense support and coordination with business personnel that had the functional knowledge to perform the classification. Be sure your business users have signed up for this task, or it will not get done and reflect poorly on the data warehouse project.

3. Each of the source systems needed to be modified —impacting a source system is something that data warehouse projects try to avoid. But in order for the new classifications to be managed across Sunrise’s organization, it was necessary for source system fields to be identified for storing the necessary classification codes. In

some cases, existing fields that were not being used for the transactional system were selected. For other systems, new fields were added specifically for this functionality.

Conclusion

With the number of companies that have international footprints growing each year, global data warehousing will likely be an area of growth for our profession. Sunrise’s GDW project identified a number of issues that are unique to global projects. Our hope is that the solutions we have identified will help you be more successful when it is time for your data warehouse to “go global.”

Authors’ Bios

Steve Dine founded Datasource Consulting to help customers implement successful, maintainable data integration and business intelligence solutions. He built and led a successful global business intelligence program for a major global medical manufacturing company and led the business intelligence practice for a Colorado based consulting company. With more than 10 years of business and IT experience, Steve combines both hands-on technical experience with a strong business acumen. He helps companies design and implement high level, strategic data integration and BI architectures that align to the unique requirements of each individual organization. He has also helped companies by providing vendor neutral assessments for new and existing BI programs. Steve earned his bachelors degree from the University of Vermont and an MBA from the University of Colorado at Boulder.

Matthew Pohl is the founder of DataVision Consulting, Inc. DataVision’s specialty is helping companies realize tangible benefits from their data warehouse investment through practically applied data analysis and OLAP functionality. DataVision also serves in more traditional roles such as data warehousing project management, architecture and design, reporting/OLAP tool selection, and reporting/OLAP development. Matthew is the former Director of Business Intelligence at Corporate Express. During his 4-year tenure with Corporate Express, Matthew and his team were able to dramatically increase the number of users of CE’s data warehouse to more than 500 users. Matthew also spent six years with Arthur Andersen’s Economic and Financial Consulting practice in Denver, Colorado. While at Andersen, he was involved in numerous complex data analysis projects that covered a variety of industries and functional areas. Matthew has an undergraduate degree in Decision Sciences and an MS in Information Systems.