

Treetimes

A Publication of Treehouse Software, Inc.

This Issue

Legacy Data Migration: DIY Might Leave You DOA1

Editor's Notes2

Joint TSI/DataMirror Event2

CA's Legacy Renewal Solution.....6



TSI provides the ADABAS component to CA's Legacy Renewal Solution. (see page 6)

Download A Free Trial of DPS X-Link, the Instant XML Gateway for ADABAS.

Just go to:

treehouse.com/dpsxlnk.shtml



Treehouse Software, Inc.

Legacy Data Migration:

DIY Might Leave You DOA

by Wayne Lashley (This article first appeared in z/Journal's z/Flash e-Newsletter)

In any application migration/renewal project, data migration is usually a relatively minor component in terms of overall project effort. Yet failure of the data migration process can and does cause failure of the entire project. So why do project managers fail to take essential steps to mitigate this risk?

Often, data migration is perceived as “the customer’s problem”, or a “problem to be considered later”, the expectation being that someone will write a suite of extraction programs that will dump the legacy data into flat files to be loaded into the new relational database management system (RDBMS). However, anytime a particular data representation is expressed in a foreign representation, there is great opportunity for mistranslation, so such a simplistic point of view opens the door to potential catastrophe.

A comprehensive legacy data migration strategy must promote a number of important tactics and features:

1. **Discovery and analysis of legacy source data structures**, including comparison of logical vs. physical definitions and comparison between processing environments (e.g., development vs. production).
2. **Insight into actual data content**, including domain analysis, unused elements, “seasonal” or “state change” elements, record typing, index usage and validation of application assumptions.
3. **Analysis of proprietary legacy database datatypes and features** that must be fit appropriately to the new RDBMS schema.
4. **Data modeling and design of the RDBMS tables**, often with the goal of not only reflecting the structure and content of the legacy source, but also its behavior.
5. **Specification and implementation of appropriate mappings and transformations** from source to target.
6. **Capture of all source and target metadata** into a repository.
7. **Native interfaces to both the source database and the target database** to maximize throughput and ensure correct interpretation and formatting.
8. **An evolutionary methodology for refining the mappings and process**, allowing for repeatable and comparative testing.
9. **Minimized extraction impact** on production systems.
10. **Support for phased project implementation**, including Change Data Capture and possibly bidirectional-update capabilities.

(continued on page 3)

Editor's Notes

by Joseph Brady

Free, On-line Demos of TSI Products

Would you like to see TSI products in action before you request a trial? To set up a live, on-line demonstration of any TSI product, simply fill out the short form on the Treehouse Software Web site at www.treehouse.com/webexform.html. All you need is an Internet connection and a current Web browser to see how TSI products work right on your PC screen.

Become a Beta Test Site for New TSI Product Releases

We are looking for current North American customers to become Beta test sites for new releases of TSI products. If you are interested in becoming a Beta test site, please fill out the short form at www.treehouse.com/betasites.html, and a TSI representative will contact you.

Find out the Latest on TSI Products

To find out about current versions of TSI products, compatibility (operating systems, languages, etc.), and support information for all of our products, view the TSI Product Status Matrix on-line at www.treehouse.com/prodstatus.shtml.

Tree·times

▶ **Editing, Writing, and Design**
Joseph Brady

▶ **Contributors**
Wayne Lashley, Orion Wolff

▶ **Production and Distribution**
Terri Hammerschmitt

Back issues available upon request. Documentation for all products is available in hard copy or on CD-ROM.

Hard Copy Circulation: 8,000

Joint TSI/DataMirror Event

by Orion Wolff

DataMirror®
SUCCESS COMES IN REAL-TIME

Recently, TSI hosted our first ever co-marketing event with DataMirror Corporation in Sacramento, CA. We are very pleased to say that our first event was a major success!

The goal of our event was to showcase data integration and Change Data Capture (CDC) solutions for ADABAS and beyond. That is, we wanted to demonstrate how applications across multiple heterogeneous environments (including ADABAS) could be easily integrated using joint solutions from TSI and DataMirror.

We had a very nice cross-section of attendees from industry, state, and local government organizations.

Wade Wiant, Channel Manager with DataMirror, spoke about the strategic partnership and expressed DataMirror's excitement to partner with the leading ISV in the Software AG community. This was very well received by our audience and helped generate interest for DataMirror's presentation.

Dave Randall, Director Channel and Alliance Engineering with DataMirror, (*seen speaking to attendees in the photo below*) gave an excellent presentation featuring the real-time data integration products and their integration points with Treehouse. Since the theme was integration, Dave focused on DataMirror's **Transformation Server** and **Transformation Server/Event Server**. The highlight of the presentation was the customer success stories, of which there were several described, including FedEx Ground and Tiffany & Co. The DataMirror representatives did an excellent job of conveying the business benefits of real-time data integration and CDC.



Chris Rudolph, TSI Technical Representative, also presented various options for ADABAS real-time data integration, including **RDBMS-to-ADABAS**, **DPS X-Link**, and **DPSync**. This part of the presentation demonstrated our ability to provide a 360° view of an organization's data. The audience asked lots of good questions about licensing, product support, and maintenance, then joined us for a buffet lunch and project-specific conversations.

In case you missed this event, you can download the joint presentation form the Treehouse Web site at: www.treehouse.com.

Legacy Data Migration: *DIY Might Leave You DOA* (continued from page 1)

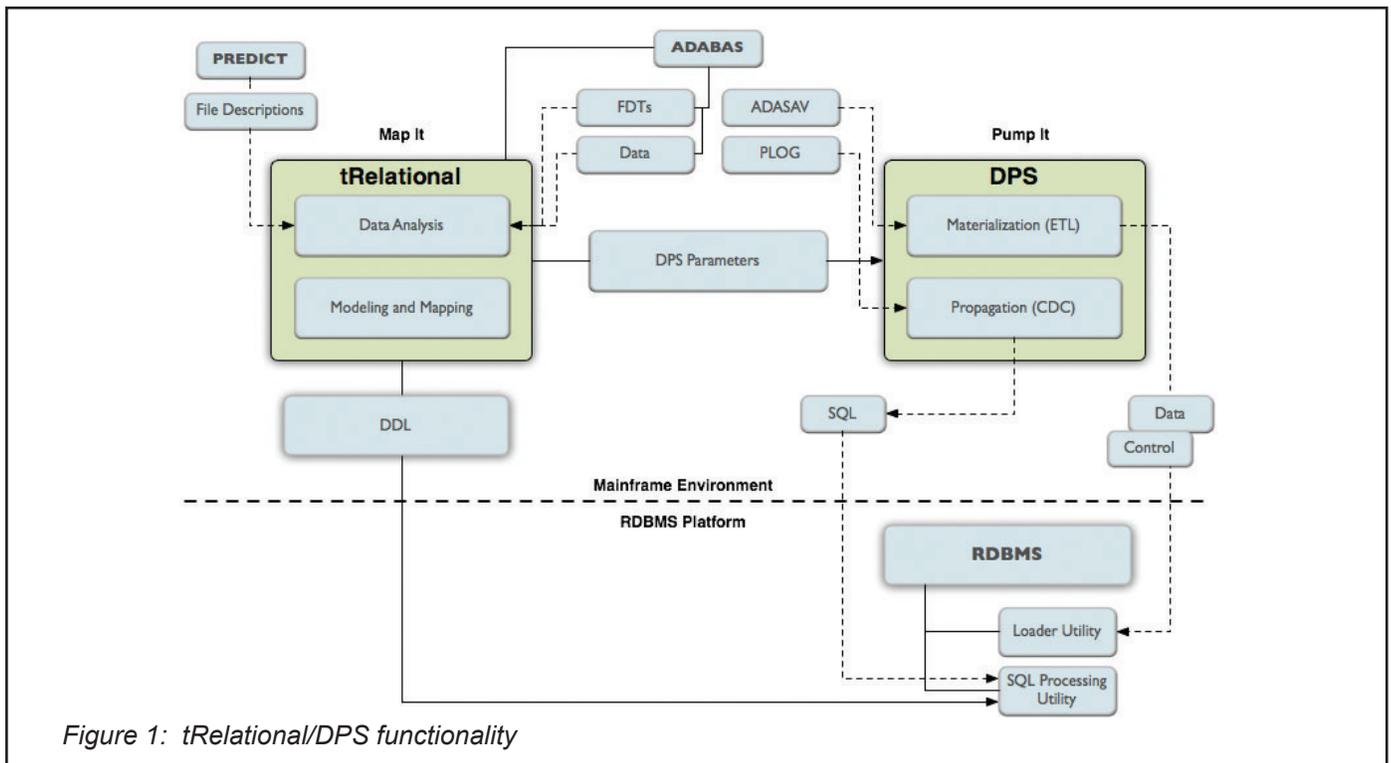
It would be a daunting challenge for a migration project team to design, develop, test and deploy tools to support these tactics and features. Furthermore, to reinvent the wheel for each new project is simply a waste of resources. Fortunately, this is unnecessary, since software vendors offer mature, proven, feature-rich and cost-effective products that contribute substantially to project success.

As the leading provider of data migration, replication and integration products and services for Software AG's ADABAS DBMS, TSI has over 11 years of production experience in the field. TSI's **tRelational/DPS**, first introduced in 1995, has been used to support many data transfer and migration projects. Each customer's project has led to enhancements and refinements in the product set, to the point that today **tRelational/DPS** handles virtually any conceivable ADABAS data migration requirement.

As an illustration of the potential pitfalls of legacy data migration, consider some issues that TSI has had to confront with respect to ADABAS:

- ADABAS has no concept of “transaction isolation”, in that a program may read a record that another program has updated, in its updated state, even though the update has not been committed. This means that programmatically reading a live ADABAS database—one that is available to update users—will almost inevitably lead to erroneous extraction of data. Record modifications (updates, inserts and deletes) that are extracted, and subsequently backed out, will be represented incorrectly—or not at all—in the target. Because of this, at TSI we say “the only safe data source is a static data source”—not the live database.
- Many legacy ADABAS applications make use of “record typing”, i.e., multiple logical tables stored in a single ADABAS file. Often, each must be extracted to a separate table in the target RDBMS. The classic example is that of the “code-lookup file”. Most shops have a single file containing state codes, employee codes, product-type codes, etc. Records belonging to a given “code table” may be distinguished by the presence of a value in a particular index (descriptor or superdescriptor in ADABAS parlance), or by a range of specific values. Thus, the extraction process must be able to dynamically assign data content from a given record to different target tables depending on the data content itself.
- As the “adaptable database”, ADABAS provides some surprising “features” that customers have employed, either deliberately or inadvertently. For example, ADABAS permits a program to assign a value to a field where the value length exceeds the field's defined length. If one were to rely strictly on field definition, without examining actual data content, one would probably err in designing the target schema, and value extraction would be inaccurate.

tRelational/DPS offers invaluable insight and capabilities in dealing with these and other issues (Figure 1).



(continued on page 4)

Legacy Data Migration: *DIY Might Leave You DOA* (continued from page 3)

For **discovery and analysis of legacy source data structures**, **tRelational**, the “Map It” analysis and design tool, provides facilities to compare logical ADABAS structures, as documented in Software AG’s PREDICT data dictionary, to the physical structures as described in ADABAS Field Definition Tables (FDTs). Note that PREDICT is a “passive” data dictionary—there is neither requirement nor enforcement that the logical and physical representations agree. However, in order to ensure the integrity of data migration, the target schema and transformations must be based on an accurate understanding of the source structures.

tRelational also provides **insight into actual data content and use of proprietary legacy database datatypes and features**, through several different types of analyses that focus on proprietary aspects of ADABAS that must be considered when migrating to an RDBMS. For example, ADABAS files often contain repeating structures (periodic groups and multiple fields, or PEs and MUs) that are not easily represented in a normalized RDBMS schema. **tRelational** analyzes the actual number of used occurrences of these repeating structures in order to guide the RDBMS design—normalized, denormalized or some combination of the two. These analyses almost invariably reveal instances that violate basic assumptions about the applications that maintain the data. Learning of these early on in the data migration process avoids rework later.

Productivity in the **data modeling and design of the RDBMS tables** is greatly enhanced with **tRelational**. In seconds, the product automatically generates a complete, native, fully-functional yet completely-customizable RDBMS (IBM DB2, Oracle, Microsoft SQL Server, Sybase, etc.) schema based on a specified ADABAS file (Figure 2). It even generates the RDBMS Data Definition Language (DDL) to instantiate the schema. Furthermore, **tRelational** generates **specification and implementation of appropriate mappings and transformations** for converting ADABAS datatypes and structures to corresponding RDBMS datatypes and structures. These are expressed as “**DPS Parameters**” and drive the “Pump It” process.

It is important to note that the schema, mappings and transformations that result from auto-generation can be tailored to any specific requirements after the fact. It is even possible to import an existing RDBMS schema and retrofit it, via drag-and-drop, to the source ADABAS elements.

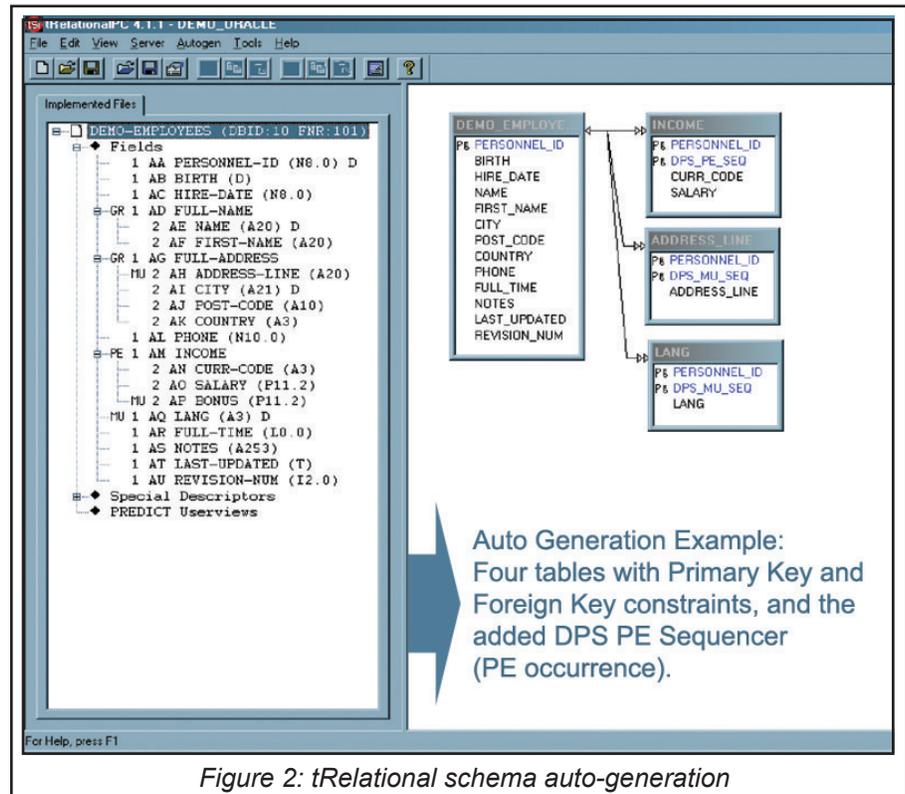


Figure 2: tRelational schema auto-generation

In the new release of **tRelational**, TSI developers are providing additional support to migration projects through enhancements designed to help mimic ADABAS proprietary features. For example, ADABAS allows for definition of a superdescriptor, a “composite index” defined on combinations of entire fields, or *specific portions of these fields*. In the case of an RDBMS such as DB2, there is no provision for an index that is based on only portions of column values, so **tRelational** can automatically generate an additional, indexed column containing the appropriate combination of values, along with triggers to maintain the column value when changes are made to the underlying columns. For an automated application migration, where code written using Software AG’s NATURAL 4GL is migrated to Java, for example, the new Java application can emulate NATURAL’s use of a superdescriptor by using the derived index.

tRelational captures **all source and target metadata** into its own repository, which can be exploited in several different ways. For example, a particular source ADABAS field may be cross-referenced to the RDBMS columns into which it is populated or transformed. At any time, the specifications for source-to-target data migration are at the project team’s fingertips, and where applicable can be exported for use by the application-migration toolset.

(continued on page 5)

Legacy Data Migration: *DIY Might Leave You DOA* (continued from page 4)

DPS (Data Propagation System) offers highly-efficient batch ETL (Extract-Transform-Load) processing, known as *Materialization* in TSI terminology. Materialization is optimized by use of **native interfaces to both the source database and the target database**: a static native ADABAS data source (a utility ADASAV backup of the pertinent database files), and native formatting of output that is RDBMS-loader-ready, including automatic generation of the loader utility control instructions, with a high-throughput transformation engine in the middle (Figure 3).

With these analysis and productivity aids, and an optimized ETL process, an **evolutionary methodology for refining the mappings and process** is enabled. TSI's advice to data migration customers is: "migrate early and often". The sooner that data can be populated into the prospective RDBMS, the sooner that data quality and assumption issues can be discovered, confronted and resolved, and capacity and performance planning commenced. Moreover, very early on in the process a test bed can be generated for modeling and prototyping the new application.

Using a static data source allows ETL processing to be repeated and compared between various mapping scenarios and data samples. Thus, the RDBMS schema and mappings can be iteratively refined over time as requirements are developed and discoveries made.

Customers also benefit by minimized extraction impact on the operational ADABAS database and applications. ETL can be resource-intensive and costly, particularly in scenarios where database calls are charged back to the user and batch windows are tight -- but this is not the case with DPS, which places no workload whatsoever on ADABAS itself.

Due to pure logistics, most migration projects require **support for phased project implementation**. Only the smallest projects can risk a "big bang" approach. **DPS** users benefit from *Propagation*, TSI's term for Change Data Capture (CDC), in that it enables implementation of the new RDBMS for certain tables, to be used in a read-only manner (generally speaking) in the new system, while maintenance of the source ADABAS data continues in the legacy system for a period of time.

Any changes made to ADABAS (reflected in the ADABAS protection log, or PLOG) are periodically extracted and transformed to appropriate SQL statements, according to the established mapping rules—the same ones as are used for ETL—to update the target RDBMS tables (Figure 4). This can even be done in real time using TSI's **DPSync** product. For very complex project implementations, TSI's **RtA** real-time RDBMS-to-ADABAS solution, in combination with **DPSync**, delivers real-time *bidirectional* CDC.

It is impractical to discuss all the features and capabilities of **tRelational/DPS** within an overview discussion, so suffice it to say

that we have only touched on those most critical ones here. Given the maturity, wealth of functionality and relative low cost of tools like **tRelational/DPS**, as compared to the effort, complexity and risk entailed in "Do It Yourself", there is no reason why a legacy renewal project should run aground on data migration. •

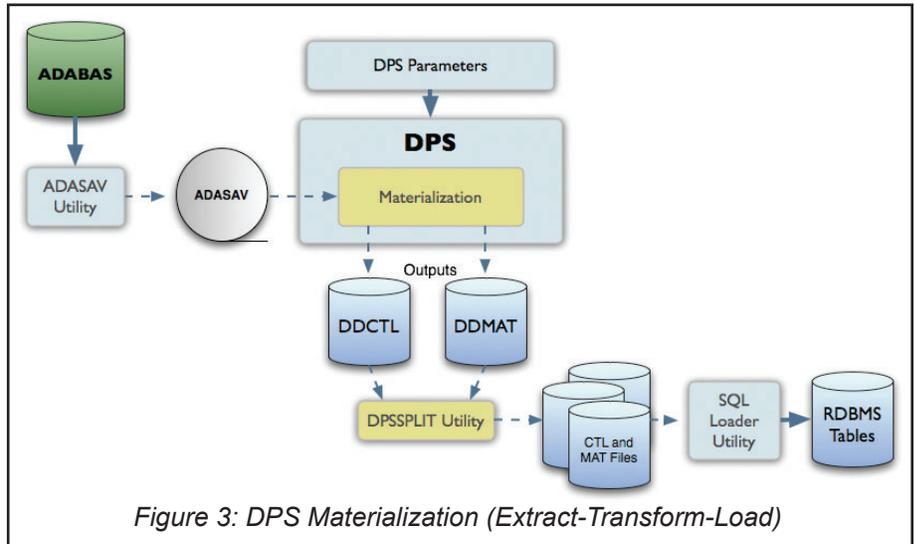


Figure 3: DPS Materialization (Extract-Transform-Load)

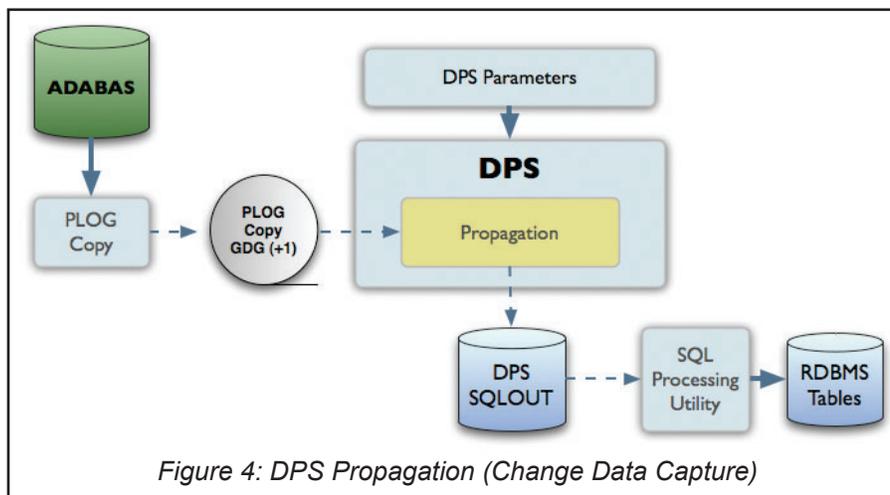


Figure 4: DPS Propagation (Change Data Capture)



CA's Legacy Renewal Solution: An Automated Approach to Legacy Modernization

Excerpts from the CA Whitepaper created by Mike Helft and Cindy Peake

Today's businesses demand new and improved software to enable business growth and the ability to rapidly develop and deploy new products. At the same time, IT organizations face continued pressure to speed application delivery while reducing costs. As if that's not enough, technology change is happening at a rate faster than ever before, and organizations are continually challenged to assess these new technologies and then implement those technologies that will benefit their business. This creates the ongoing challenge of balancing investments in the existing software infrastructures with investments in future technology opportunities. Legacy modernization is one solution to this challenge. Legacy modernization is about retaining and extending the value of legacy applications, reusing the intellectual property buried in these systems and transforming them to more modern architectures. There are numerous drivers that compel an organization to explore legacy modernization, as well as numerous approaches to accomplish this task. Each approach has various pros/cons and results. This paper explores these approaches and introduces a cost effective, low risk alternative that combines knowledge based services and automation tools to move your mission critical legacy applications into the 21st century.

Why Modernize?

There are numerous drivers that compel an organization to explore Legacy Modernization strategies:

- **Application Integration** (integrate applications from either diverse parts of the organization or across organization boundaries)
- **Platform Migration** (strategies to move away from mainframe based solutions to distributed environments)
- **Regulatory Compliance** (requirements associated with legislated standards such as Sarbanes Oxley and HIPAA in the U.S. and BASEL II in Europe)
- **Cost Reduction** (hardware infrastructure as well as costly software license fees)
- **Complexity Reduction** (many legacy systems have been modified so many times, by so many different developers that the code has become very complex and inflexible)
- **Speed to Market/Flexibility to Adapt** (software to enable business growth and the ability to rapidly develop and deploy new products)
- **Technology Change** (assess new technologies and then adapt their software applications to leverage those technologies that will benefit their business)
- **Skills and Business Knowledge** (migrate legacy systems, which were hand written in mainframe languages such as COBOL and Natural, into an environment that

can be understood and maintained by non-mainframe skilled system analysts and developers)

CA's Legacy Renewal Approach

Most organizations fully understand that their legacy systems contain not only useful, but vital business information and business rules. But the task of extracting this value from these legacy systems has remained difficult and time consuming, and therefore many organizations have simply avoided this "daunting" task. The key to overcoming these barriers is to automate as many of the tasks associated with legacy modernization as possible. Substantial automation of the process makes legacy transformation an economically attractive and low risk proposition compared to other modernization options such as rewriting or replacing the legacy application. Automation, combined with a repeatable methodology will simplify the tasks as well as speed the process of modernizing legacy solutions.

The Legacy Renewal Solution from CA uses automated extraction and a model driven approach to successfully transform legacy COBOL or Natural code to new solutions that implement modern technologies and architectures such as J2EE, .NET, Web services and SOA.

Automated Business Rule Extraction and Documentation

A key step in successfully modernizing applications is to analyze the existing legacy code. CA uses Evolveware's S2T™ Technology to automate the extraction of business rules, data and workflows from legacy COBOL or NATURAL/ADABAS code. The extraction process uses an iterative approach to maximize the automation — on average, as much as 70–85% automated conversion of legacy data and logic. Thorough reports are created to assist in manually converting the remaining legacy code.

There is often an assumption that all code should be converted, however this is typically an incorrect assumption, because some of the legacy code may be irrelevant to the modernized solution. During the extraction process, reports are generated that enable the organization to better understand unknown or poorly documented processes, as well as simplify complex logic constructs that have likely been developed over several years and by several developers.

Legacy Data and Logic Interpretation. One of the biggest challenges in a migration exercise is to translate potentially meaningless code into well understood business terms. A key to success of a legacy migration project is to rename and reorganize the legacy data components that have been identified, very early in the process. If the data interpretation

(continued on page 7)



CA's Legacy Renewal Solution: An Automated Approach to Legacy Modernization

(continued from page 6)

is done upfront, then the translation of the actual logic can focus on a better understood data structure, which can significantly speed the process. Similarly, logic interpretation involves translating meaningless COBOL statements into terms that better describe the business processes and can be understood by all members of the team. This translation can also be accomplished in the model driven development environment, however, experience has shown that intercepting the code early on during the conversion process is more efficient and speeds the overall conversion process.

Legacy Conversion Output. Results of the legacy conversion process include two key deliverables:

- **A Business Model.** This contains all the elements that have been extracted from the legacy source code and translated into objects that are understood by the modeling environment. It is fully documented so it is possible to trace each object in the model to the originating piece of code.
- **Documentation.** Equally important is the documentation, which summarizes what has been converted and just as vital, what has not been converted. This document has proved very useful during the forward engineering phase of the project.

Deliver Modernized Solution

Once the business rules and data have been successfully extracted from the legacy system, they are automatically converted into a business model. From that point you can use a model driven approach to enhance the business rules to reflect current business practices, and then forward engineer those processes into new application architectures and modern technologies. CA's premier model driven development environment, AllFusion®Gen, provides code generators that enable organizations to deploy their new applications to numerous modern platforms and architectures such as J2EE, .Net, UNIX, Linux, Web Services and SOA. And for organizations who wish to remain on the mainframe, the robust code generators in AllFusion Gen enable organizations to modernize their applications without getting off the mainframe. This flexibility helps reduce the risks of modernization by providing platform independence, where systems are 100% generated for multiple environments from a single model, reducing risks by providing an evolutionary path towards modernization.

After the data and process models are created, the new application logic and databases are generated. Another significant challenge associated with application modernization is data migration. The Legacy Renewal Solution offers flexible data modernization options which can leverage existing database environments, or transition to new ones, or a combination of both. If data migration is required, **CA uses data transfer technology from partners such as Treehouse Software, Inc.** to facilitate conversion of data from one database environment to another.

Copyright © 2006 CA. All rights reserved.

TSI is the "Go to" ADABAS Partner for CA's Legacy Renewal Solution

Like other TSI strategic technology partners, CA recognizes the value of having TSI's best-of-breed data transfer products as part of any ADABAS data conversion project.

Contracts for ADABAS data conversion can be obtained easier, aggressive pricing can be used, conversion tasks will take less time and/or fewer people, and profits can increase by possessing mature ADABAS data transfer tools that competitors possibly do not have.

What does 11 years get you?

Since its establishment in 1982, TSI has become a leader in systems software development and marketing, with approximately 700 customers in the U.S. and many foreign countries. Additionally, TSI is the leading provider of data migration, replication, and integration products and services for ADABAS, with **over 11 years of production experience in the field.** TSI's **tRelational/DPS**, first introduced in 1995, has been used to support over 70 ADABAS-to-RDBMS data transfer and migration projects, worldwide.

TSI's product set extracts data out of legacy ADABAS files and migrates it into RDBMS-based data stores, data warehouses, or packaged applications. Some sites want to convert the data once, while others intend to keep the data in RDBMS tables synchronized with the data in ADABAS files. The **tRelational/DPS** product set is the proven solution for migrating and propagating legacy ADABAS data into RDBMS-based Business Intelligence/Data Warehouse applications.

tRelational auto-generates complete RDBMS schema from existing ADABAS files and, alternatively, allows easy mapping of ADABAS fields to pre-existing data warehouse or ERP schemata. After **tRelational** does the mapping, **DPS** can then materialize (initially load) and propagate (subsequently keep synchronized) the ADABAS data into an RDBMS. All this is done without requiring direct access to ADABAS, and therefore, production systems are not impacted.

Treehouse Software Products

ADABAS-to-RDBMS Data Transfer:

DPS - ADABAS-to-RDBMS data materialization (ETL), replication, and propagation (CDC) software

DPS X-Link - Instant XML-based read/write access to ADABAS

DPSync - Real-time ADABAS-to-RDBMS data propagation (CDC) software product set

tRelational - ADABAS modeling, mapping, and data analysis tool; DPS parameter generator

tRelationalPC - Windows-based graphical interface to make the tasks of modeling and mapping even simpler

Treehouse Remote Access (TRA) - Middleware that allows tRelationalPC to communicate with tRelational on the mainframe.

NatQuery - GUI-based tool that intelligently generates NATURAL code to handle all of the complexities of data extraction from ADABAS

NatCDC - Add-on to NatQuery designed to create immediately-usable data out of the ADABAS PLOG

UNIX:

SEDIT - XEDIT and ISPF/PDF compatible editor for UNIX and Windows

S/REXX - REXX-compatible language for UNIX and Windows

S/REXX Debugger - Optional graphical debugger for S/REXX programs

Software AG Related:

ADAMAGIC - Tool for converting mainframe ADABAS files into ADABAS for UNIX/Linux/Windows, flat file, or comma-delimited formats

ADAREORG - File reorganization tool for ADABAS

ADASTRIP - Data extraction utility for ADABAS

AUDITRE - Generalized ADABAS auditing facility

CHART for NATURAL - NATURAL application analysis and documentation tool

N20 - NATURAL application change management system

N20/3GL - 3GL support within N20 for PANVALET, LIBRARIAN, ENDEVOR, and PDSs

PROFILER for NATURAL - NATURAL quality assurance and testing tool

SECURITRE - ADABAS and NATURAL security interface to RACF, ACF2, and TOP SECRET

TRIM - ADABAS and NATURAL performance monitor

TSI is a DataMirror DataWorld partner, and is authorized to market the full line of DataMirror products, including:

Integration Suite™ - Combines all of DataMirror's real-time integration software and global services.

Transformation Server™ - Real-time multi-platform data integration and transformation with zero-programming required.

iCluster™ - Manage clustered iSeries environments.

LiveAudit™ - Capture all data that is inserted, updated, or deleted to create real-time audit trails.

PointBase™ - Java relational database, mobility and synchronization products.

iReflect™ - Mirror ORACLE database transactions in real-time from the primary system to the recovery system.

Transformation Server/ES - Detects events as they occur in mission-critical production applications and creates useful business information to feed into the message queues of leading EAI, BPM, and SOA environments.

DB/XML Transform™ - Powerful Java-based, XML-driven engine for bi-directional data transformation between XML, EDI, database and text formats.

Constellar® Hub - Powerful ORACLE-based EAI tool.

TSI is a BEA Select Partner, and is authorized to bundle solutions involving DPS X-Link with BEA products including:

BEA WebLogic Server 9.0 - The world's leading J2EE application server.

BEA WebLogic Integration - Converge two otherwise disparate activities-application integration and application development.

BEA AquaLogic™ - The broadest line of Service Infrastructure products for successful SOA deployment. It consists of:

BEA AquaLogic Service Bus™

BEA AquaLogic Data Services Platform™ (formerly BEA Liquid Data)

BEA AquaLogic Enterprise Security™

BEA AquaLogic Service Registry™

Phone: (412) 741•1677 Fax: (412) 741•7245 E-mail: tsi@treehouse.com Web: <http://www.treehouse.com>

TREEHOUSE SOFTWARE, INC.

409 Broad Street, Suite 140
Sewickley, PA 15143 USA

POSTAGE
INFORMATION
HERE