Actuarial Data Management In A High-Volume Transactional Processing Environment

Joseph Strube and Bryant Russell, Ph.D., ACAS, MAAA

Abstract

The development and management of data resources that support property/casualty actuarial work are very challenging undertakings, especially in a high-volume transactional processing environment. In order to equip actuaries with the data resources necessary to excel in the performance of their functions, an Actuarial Data Management (ADM) support team is needed. It serves as a proactive, added-value conduit of business data and specialized technical support to an actuarial staff.

This paper examines the evolution of the actuarial data management function in the context of end user computing, and highlights the key roles and processes that comprise an effective data management operation in a modern property/casualty actuarial department. The paper also includes a case study that describes the development of the data management function in the Actuarial Department of Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan.

Keywords. Actuarial Applications and Methodologies; Data Management and Information; Actuarial Data Management; Actuarial Technician; End User Computing; Data Warehousing; Insurance Data Management.

1. INTRODUCTION

The goal of an Actuarial Data Management (ADM) unit is to equip an actuarial staff with the data resources necessary to excel in the performance of their functions. Having this goal, it is clear that data managers and their crews fill a service role. But they are equally manufacturers too --- of information products. They construct focused relational databases from ever expanding data warehouses, and multi-gigabyte, multi-dimensional data arrays from voluminous mainframe transactional processing systems. And on a less grand scale, they even electronically whittle desktop data files after much sifting, selecting, and aggregating.

The types of processes that render data products are really not new over the past 30 years, but the tools used to retrieve and manipulate the growing stores of raw data surely have evolved. Actuarial analysts have always needed to extract the data from wherever it could be excavated. People with longevity in this field will no doubt recall a manual preparation process using tabular worksheets on 15-column ledger paper. More fortunate actuarial technicians would copy numbers from “DP” (Data Processing) Department green bar reports, transcribing them into cascading loss triangle ledger pages. Less fortunate ones would need to derive the incremental differences before penciling in the new month’s or quarter's development of counts and amounts. Once the rote work of initial spreadsheet updating was completed, more derived data was generated. The spreadsheets from the previous process provided the input for the next series of manually computed and manually recorded numbers. This time the computations resulted in values approaching analytical data: frequencies, severities, and pure premiums. Eventually, the assembly line of data worksheets and computations concluded in a hand-off to the actuaries who would analyze the refreshed statistics and interpret their meaning in reference to their pricing recommendations, and the company's loss reserves and premium adequacy levels.

In the context of today's actuarial activities, this chain of events continues. But the details of the process are so radically different that anyone familiar with a modern actuarial department could vehemently debate the notion. The skill set required to prepare the data for actuarial analyses has escalated from a quasi-clerical level to an advanced blend of technical and business acumen. Pointing out the vast difference in those skill sets, however, doesn't detract from the valuable data quality and auditing contributions made by the technicians of the era preceding end user computing. But it can serve to highlight the multi-faceted role that a modern data technician must fulfill as the focal point for coordinating the data extraction, cleansing, transformation, aggregation, and deployment processes within a company's actuarial operations.

And that brings us to the crux and purpose of this paper. The incredible volume of raw data that is available to end users through the business processes and information technologies employed in large insurance companies today, requires an orchestration of human and technological resources to achieve the goal of equipping actuaries with the data resources necessary for them to produce excellent results. This paper will discuss the roles of the human resources, the data handling processes, and the technologies that are key to achieving an effective actuarial data management function in a high-volume transactional processing environment. For the purposes of this discussion, a high-volume transactional processing environment (HVTPE) is one where transactions added to the most granular actuarial data resource in the organization exceed one million per month.

2. BACKGROUND

2.1 The Evolution Of End User Computing

By the early part of the 1980s, end user computing had emerged as a technology field all unto its own. The college graduates of the '70s who majored in business, mathematics, and both the physical and social sciences were trained in their curricula to incorporate computers into their research, data analysis, and presentation tasks. Earlier on, this was limited to time-sharing on large remote mainframe computers. And as the decade advanced, mini-computers appeared in American business and industry. These smaller scale machines made computing more accessible to onsite personnel because the manuals to operate and program the computers were generally available to the business-based technicians that were capable of using the equipment. Further, the input/output media used by these computers (particularly 8" diskettes and data cassette tapes) were readily available to the non-Information Systems class of people in the company. The day had arrived when business users could input, retrieve, and record their own work, in essence restoring a degree of local control that they had taken for granted in the totally paper-based office. This issue of control would prove to be one that would resurface as a point of contention between business professionals pushing for greater computing capacity and the technology professionals that managed the hardware and software computing resources.

In larger insurance organizations where monthly policy, premium, and claims activity required a company's computer(s) to handle millions of transactions, Information Systems (IS) personnel were continually compelled to become familiarized with the next generation of mainframe CPU. Their task was to gauge when the next upgrade of hardware and software would be needed to deal with the burgeoning flow of data that the business side of the company said was necessary to achieve its sales, marketing, financial, and management goals, as well as meet regulatory requirements. A lack of vigilance to the company's computing growth requirements or a misjudgment of a technical solution's scalability could result in a crisis of "computer resource gridlock." This gridlock would be manifest with the off-hours (evening) batch processing jobs contending for more time than they were allotted, threatening the daytime online processes that supported the field organization's customer service activities.

Add to this ongoing struggle for equilibrium, a business end user community that was growing impatient with the continuing dependence upon the IS area for obtaining large data compilations and department specific computer applications. Actuarial end users, with their more analytically inclined orientations, were at the forefront of the push to put increasing amounts of computing power, data storage capacity, and next generation (end user) programming tools into the hands of people outside of the IS Department. Non-technical business professionals became end user activists, seeking to control their own departmental systems development and data management destinies. It was an unsettling epoch of events for the caretakers of a corporation's computing resources. Before that time the domain of systems programming was entered into only through formal education in computer science at the collegiate level. Now "unqualified" non-technicians insisted that some of the company's IS budget be spent on computing resources that they themselves would utilize.

2.2 End User Roles Emerge

As end user computing became commonplace in the insurance industry, users began to coalesce into groups differentiated by varied levels of technical interest and ability as well as functional role. In a research study conducted by John F. Rockart and Lauren S. Flannery of the Sloan School of Management at MIT, 250 people involved in end user computing from seven organizations (three Fortune 50 manufacturing companies, two major insurance companies, and two sizable Canadian Companies), yielded six distinct types of end users.[1] The profiles of these six user types (Non-Programming End Users, Command Level Users, End User Programmers, Functional Support Personnel, End User Computing Support Personnel, and DP Programmers) are displayed in Table 1. These categories of end users were defined in 1983 when the practice of end user computing was growing at a rate of approximately 50-90% per year in the organizations included in the study.[2] And they were predicated on the then current mode of end user computing --- mainframe-based computer software tools and data storage.

TABLE 1: Rockart & Flannery's Six Types Of End Users

Non-Programming End Users	Only access to computer-stored data is through software provided by others. They neither program nor use report generators. Access to computerized data is through a limited, menu-driven environment or a strictly followed set of procedures.
Command Level Users	Have a need to access data on their own terms. They perform simple inquiries often with a few simple calculations such as summation, and generate unique reports for their own purposes. They understand the available database(s) and are able to specify, access, and manipulate information most often utilizing report generators and/or a limited set of commands from languages such as FOCUS, RAMIS II, EXPRESS, SQL, or SAS. Their approach to the computer is similar to that of an engineer to a slide rule in days past. They are willing to learn just enough about the database and the software to assist the performance of their day-to-day jobs in functions such as personnel, accounting, or market research.
End User Programmers	Utilize both command and procedural languages directly for their own personal information needs. They develop their own applications, some of which are used by other end users. This latter use is an incidental by-product of what is essentially analytic programming performed on a "personal basis" by quantitatively oriented actuaries, planners, financial analysts, and engineers.
Functional Support Personnel	Sophisticated programmers supporting other end users within their particular functional areas. Individuals who, by virtue of their prowess in end user languages, have become informal centers of systems design and programming expertise within their functional areas. They exist today as "small pockets of programmers" in each functional organization … Provide the majority of code for the users in their functions. In spite of the large percentage of time that these individuals spend coding (several estimated over 80%), they do not view themselves as programmers or data

	processing professionals. Rather, … [their] primary task is providing tools and processes to get at and analyze data.
End User Computing Support Personnel	Most often located in a central support organization such as an "Information Center." Their exact roles differ from company to company. Most, however, are reasonably fluent in end user languages and, in addition to aiding end users, also develop either application or "support" software.
DP Programmers	Similar to traditional COBOL shop programmers except that they program in end user departments wishing to hire "contract programmers," to avoid high consultant/programmer fees, and to build a larger base of knowledge of end user language computing within the corporation.

2.3 End User Technology Evolves

From those early years until the present, the evolution of the technology supporting end users is nothing less than astonishing. As the mainframe tools and data storage devices matured, the introduction of stand-alone personal computers and simple stand-alone application software presented a new dynamic. These personal-level computing devices supplemented end users' arsenal of capabilities for processing, analyzing, and presenting data. PC's were connected to mainframes to facilitate the transference of data from one platform to the other. Stand-alone personal computers were connected to each other as well as other "sharable" resources (like printers and hard disk drives) to form local area networks (LANs). PC/LAN hardware and software escalated in throughput capacity and functionality. Simple "peer-to-peer" networks were replaced by "client/server" networks in which a single more powerful computer served up software and data files to less powerful remote units. The "graphical user interface" style of software, using a desktop metaphor on the PC screen and point-and-click icons, became the routine expectation when powering up one's personal computer. LANs were connected to other LANs in geographically remote office locations to form Wide Area Networks (WANs). The continuing cycle of hardware capacity growth followed by software functionality growth swept across the business world in waves that incremented in amplitude and increased frequency year after year. And as this phenomenon protracted, entirely new branches of data management technology erupted from the trunk of traditional Database Management Systems (DBMS) technology, e.g., Data Warehousing, Data Marts, Online Analytical Processing, Extraction-Transformation-Load Software, Metadata Repositories, Decision Support Systems, Data Profiling/Cleansing/Integration Software, and Data Mining.

Each of these new technologies advanced particular aspects of the relatively static processes of data collection, processing, storage, retrieval, and analysis. A general description for each of these Data Management terms is provided below.

- Data Warehousing refers to the development and maintenance of a collection of large databases created from (and separate from) an organization's primary business processing systems. These large databases address enterprisewide subject areas and provide flexibility for management and business analyst reporting requirements. Being separate system constructs, data warehouses do not impact the operation and performance of the business systems supporting the marketing, sales, product support, and customer service activities of a company.
- Data Marts are smaller, specialized databases, often created from data warehouses. They focus on particular department-level needs or a subset of subject areas.
- Online Analytical Processing (or OLAP) refers to a type of software that provides very rapid access to data stored in a database and enables users to view and analyze data as multi-dimensional arrays.
- Extraction-Transformation-Load Software refers to a highly specialized and powerful class of software that can perform data extractions from multiple source databases, translate and convert the data according to business rules specified by business and data analysts, then load the data to a target database structure for subsequent querying and analytical reporting purposes.
- A Metadata Repository is a special type of database containing information about another database, e.g., how the data in the other database was collected, transformed, and formatted, how frequently it is updated, and generally anything that can be useful to analysts that need to query data from that database.
- A Decision Support System (DSS) is a specialized database and an associated set of software tools that are dedicated to enabling management decision making processes.
- Data Profiling/Cleansing/Integration Software is a powerful class of software used to examine a number of characteristics of data in source databases, apply customized business rules to maintain or enhance the data's integrity and usability, and

consolidate multiple occurrences into a "best" version for loading into a single target database.

- ♦ Data Mining refers to a highly sophisticated class of database applications that detect hidden patterns and relationships in a collection of data in order to predict future behavior.

2.4 End User Roles Adapt

The extraordinary technological advancements over the past 20+ years have required a commensurate growth in the knowledgeability and skill sets of professionals participating in the discipline of actuarial data management. And depending upon the role or roles a person has needed to fill in the spectrum of end user types, the limits of his or her mental elasticity have been unequivocally tested. Given the incredible growth of transactional data to mammoth proportions (hundreds of gigabytes, even terabytes in size), and the diversification and elongation of data management processes (capture, inspect, cleanse, transform, qualify [create metadata], integrate, store, and distribute), there are substantially more steps and more time required to manufacture and deliver data ready for an actuary's consumption. It will be helpful at this point to update the characteristics of the end user types (described in Table 1) for current practices. It is a subset of these user types, operating as an efficient conduit of technology, which comprise the key roles necessary for truly effective actuarial data management. Refer to Table 2.

TABLE 2: Updated Characteristics of End User Types

Non-Programming End Users	Only access to data is through data products delivered by others. They do not program. Access to data is through menu-driven computer interfaces, standard desktop tools, or a strictly followed set of procedures.
Command Level Users	Have a need to access data on their own terms. They perform simple inquiries often with a few simple calculations such as summation, and generate unique reports for their own purposes. They understand the available database(s) and are able to specify, access, and manipulate information most often utilizing report generators and/or a limited set of commands from database query tools.

End User Programmers	Utilize command and procedural languages, as well as database query and creation tools directly for their own personal information needs. They develop their own applications and databases, some of which are used by other end users. This latter use is an incidental by-product of what is essentially analytic programming performed on a "personal basis" by quantitatively oriented actuaries, planners, financial analysts, and engineers.
Functional Support Personnel	Sophisticated programmers and data handlers supporting other end users within their particular functional areas. Individuals who, by virtue of their prowess in end user languages combined with an expert knowledge of the business data, have become informal centers of systems design and programming expertise within their functional areas. Provide the majority of data for the users in their functions. In spite of the large percentage of time that these individuals spend coding and managing data for others, they do not view themselves as programmers or data processing professionals. Rather, their primary task is utilizing tools and processes to get at, prepare, and distribute analytical data to their functional area.
End User Computing Support Personnel	Most often located in a central support organization such as a "Help Desk." Their exact roles differ from company to company. Most, however, are reasonably fluent in end user languages and assist end users with troubleshooting data retrieval and reporting issues.
DP Programmers	Similar to traditional COBOL (procedural) shop programmers except that they program in end user departments wishing to hire "contract programmers," to avoid high consultant/programmer fees, and to provide technical advice and support for data retrieval, transformation, and distribution. Usually specializes in one or more data handling tools, e.g., Database Systems, Data Warehousing, Extraction-Transformation-Loading software, OLAP, and Data Mining.

Generally, the emphasis of each role has shifted from being a consumer or producer of a computer application, to being a consumer or producer of a data product. This may seem like an unnecessary nuance to cite considering the usual byproduct of a computer application is a data compilation or hard copy report of some sort. But the advancements in client/server and desktop PC applications have made the task of accessing and working with data relatively elementary, even for novice business analysts. Consequently, data files of varied sizes and formats are now a norm as deliverables to actuarial analysts.

2.5 The Alignment Of End User Roles And Business Roles

In a large insurance carrier, one or more particular groups of people can be associated with each of the end user categories described in the updated table. This is seen as a natural development based upon a combination of business function, professional preference, and technical/analytical ability.

The Non-Programming End Users are typically those in management roles, executives and senior managers, who have others prepare data and presentations for their review. These end users may also include corporate educators and process modelers who use the data as presentation or communication devices.

Command Level Users include business analysts who have learned how to access the organization's data warehouse and, if available, selected data marts. Their use of data retrieval and reporting tools continues to be limited to what's essential to conduct their day-to-day jobs in areas such as human resources, accounting, market research, claims, and field operations.

End User Programmers continue to include people who are quantitatively oriented and utilize computer and data resources to build reliable and repeatable processes for satisfying their own information requirements in the context of their job responsibilities. These people are actuaries, actuarial analysts, planners, and financial analysts.

Functional Support Personnel include people who are individually a blend of businessperson and technician. Their role dictates they provide support for their department of residence and may be comprised of people who previously filled the roles of actuarial or financial analysts, but due to a technical expertise that developed over time, they transitioned into a specialized, quasi-technical position. In addition, this group can include entry-level personnel that are training in the data rudiments of the analytical and professional roles they are aspiring to fill. These people include actuarial technicians, financial technicians, data analysts, data technicians, as well as interns or college co-op students.

End User Computing Support Personnel include technical analysts that interface with the end user community to answer questions, act as points of contact for troubleshooting issues, provide or augment training, and manage user access to the organization's corporate level data resources (the data warehouse and cross-functional data marts). This group may be known as some type of Help Desk or User Hotline, or associated with the technology employed by end users, e.g., the Data Warehouse Support Group or the Decision Support System Team.

Lastly, the DP Programmers include internal or outside contractors that supplement and extend the technical competencies of the other end user categories on an as-needed basis. These people are often brought in to fill roles that require specializations outside the normal technical parameters of on-staff business end users.

3. KEY ROLES IN THE MANAGEMENT OF ACTUARIAL DATA

3.1 The Critical Trichotomy

All the end user categories described above are evident in either the administration of or the execution of actuarial data management processes at an insurance concern with a high-volume transactional processing environment (HVTPE). And more than one end user type can apply to different actuarial and actuarial support functions. There may be some crossover of categories between positions as well. For example, actuarial management typically cluster into the Non-DP Programmer and Command Level User groups, actuaries and actuarial analysts to the Command Level and End User Programmer groups, actuarial technicians to the End User Programmer and Functional Support Personnel groups, and lastly, help desk staff to the End User Support and DP Programmer groups. Technical contractors may function in a number of capacities that mimic the roles of functional support staff and help desk personnel besides filling specialized consultant roles.

To varying degrees, these functions each contribute to the development, maintenance, and management of effective and excellent actuarial data resources. Yet two of these functions, in conjunction with support from the company's Information Technology (IT) Department, are more crucial to the ongoing actuarial data environment than the others. In the modern insurance enterprise, it is the actuary, the actuarial technician, and the IT management function that form a critical foundation for an effective actuarial data management operation. Each of these three functions brings unique and indispensable elements into an alliance dependent upon mutual cooperation. This trichotomy will now be more fully described so as to compose an interaction model for accomplishing corporate actuarial objectives.

3.2 The Role Of The Actuary

3.2.1 HVTPE – The Good News And The Bad News

A high-volume transactional processing environment (HVTPE) presents the actuary with a classic good news/bad news story. The good news is that the actuary can rely on an extensive historical view of past transactions. The bad news is that the past transactional experience is extensive!

3.2.1.1 The Good News: A Lot of Good Information Available

Having a large volume of potentially credible data enhances the actuary's ability to analyze a company's historical experience.

- **Company-specific vs. industry experience.** The relationships between different classes, territories, etc. of the company's books of business can be examined by summarizing the data along various common dimensions. Company experience is readily comparable to industry-wide summary data.
- **GLMs.** The HVTPE is a Generalized Linear Model user's paradise. The information is available at a very granular level, and with proper summary, the GLM can be employed to compare multiple factors that bear on the performance of the overall book. Perhaps territory and deductibles are interactive. The GLM allows the actuary to consider both dimensions (and others) in one analysis.

- **Loss Distributions.** Loss distribution analysis is another area that relies on a highly granular data resource for detailed historical size of loss experience. The HVTPE is a particularly attractive resource for this type of statistical analysis. The sheer volume of historical data may permit the actuary to meet the parallel demands of homogeneity and credibility by restricting the view to aggregations of subsets with similar loss potential.

- **Reconciliation between functional areas.** Because the HVTPE provides a very granular level of detail, it becomes possible to directly compare actuarial aggregations (e.g., accident year, policy year) to other financial aggregations (e.g., calendar year, calendar quarter, underwriting year). This comparison can lead to a closer working relationship with other (non-actuarial) areas of the company. The various areas of responsibility no longer need to focus on which data is "right," but rather on what the various aggregations of data suggest about the book of business.

- **Cross-functional use of common data.** The actuary's requirements for data from a HVTPE will tend to be fairly detailed. In many cases, other non-actuarial areas of the company (e.g., claims, underwriting, and marketing personnel) find the resulting body of experience a valuable resource for their needs as well.

3.2.1.2 The Bad News: The Actuary May Spend Inordinate Time On Data Management

The HVTPE requires an "industrial strength" toolbox of hardware and software tools to store, aggregate, retrieve, and analyze the data. Consequently, "someone" needs to be concerned with the following items:

- **Hardware and software platform specifications**
- **Hardware and software procurement, maintenance, and updates**
- **Balancing**
- **Monitoring and maintenance of information quality and consistency**
- **Production vs. ad hoc environment**

Certainly, some actuaries are capable of providing the guidance for managing these elements of the HVTPE process, yet this is not the best use of the actuary's skill set. This data management role is better handled by a professional partner, the actuarial technician, working in tandem with the actuaries.

3.2.1.3 Involvement In The Development Of The Data Resources

It might be tempting to the actuary to place the responsibility for designing and implementing the actuarial data management processes of the HVTPE completely in others' hands. However, this is not realistic for many reasons:

- **The actuary cannot rely purely on canned reports.** Predetermined summaries of the various books of business, subsets, rating variables, etc., are helpful. However, every question raised by such reports will suggest further "deep dives" into the historical experience. If the data retrieval design is not clearly set out beforehand, this may result in an excess amount of time being spent on creating ways to get at the additional data. In the worst-case scenarios, the information being sought may not be accessible to the actuary.
- **Information that provides actuarial value may be seen as of secondary importance to other business areas.** For example, historical policy rating information may not be seen as critical to the sales area, and current policy rating information may not be of much use to the claims personnel administering claim payments. The pricing and reserving actuaries, on the other hand, would find value in both historical and current policy information. Actuarial involvement helps ensure that both types of information are made available for actuarial use.
- **The required level of detail for actuarial analysis is different.** In many cases, the actuary will require information that is more summarized than individual policy or claim level, yet is far more detailed than the financial reports required for management review of the business. Relying exclusively on operation reports (too detailed for actuarial analysis) or management reports (too summarized for actuarial analysis) prevents a full actuarial review of the programs.
- **Historical retention of useful data is important.** The actuary must help determine how long information should be made available for actuarial analyses. There is a balance between establishing the long-term patterns in claims emergence versus retaining data that no longer has any reasonable relationship to today's book. It is not just active policies and claims, nor just open tax years. The actuarial information requirements would most likely include these time frames and more. The actuary is best suited for determining this balance.
- **"User defined" information fields are needed.** Some actuarial aggregations involve both historical information and "user defined" information not available in other areas. For example, the definitions of "territory" and "symbol group" can change over time. Hence, the historical information may need to be re-rated or re-cast with the current or prospective view of the data. Such views are important for

actuarial analysis, yet of lesser importance to the other areas of the company. Without actuarial involvement in the definition of requirements, there is a risk of losing this capability.

- **Changes to the HVTPE have consequential impacts.** Actuarial informational requirements can be seen as a "next phase" item, rather than an upfront requirement. It is a natural tendency to ensure that all of the operational data needs are met before discussion of "back-end" reporting begins. However, many of the initial requirements set the precedence for long term informational deliverables. The order of delivery priority, i.e., delivering the operational needs first, is correct. However, gathering <u>all</u> of the informational needs, including the actuarial informational needs, should be done up front. Otherwise, there is a great risk of losing the opportunity to capture and incorporate extremely important data elements into actuarial analyses. Therefore, it is important for the actuary to be involved in this upfront process. This means a partnership is needed between the actuarial and non-actuarial areas that initiate projects that alter the HVTPE.

3.2.2 Qualifying Actuarial Data Requirements

The actuary relies on the ADM team to implement and maintain a successful system for capturing, storing, and retrieving HVTPE information. The actuary must clearly define and communicate to the ADM team specific data requirements.

3.2.2.1 Historical Data Requirements

The HVTPE information source can be extremely large. The actuary needs to start by identifying those components of the historical information that need to be captured and available. Typical items include those policy characteristics that shape the premium rate, as well as the claim characteristics that have a bearing on the claim size, frequency, emergence, etc.

3.2.2.2 Level Of Detail Requirements

Beyond "What data?" is the question of "How detailed?" This can be a very intensive, time-consuming effort. Does it make sense to go below "claim" level? What is a "claim"? If several transactions comprise a "claim," how important is it for the actuary to be able to combine the cumulative flow of transactions at the claim level? The same questions apply to multiple transactions that comprise events at the policy level. Together, the actuary and ADM team must consider the trade-offs between maintaining detail at a more detailed level versus the effort necessary to capture, store, maintain, retrieve, and aggregate the information.

3.2.2.3 Required Time Frames

Actuarial time frames of interest include transaction dates (e.g., report date, process date, accounting dates, etc.) as well as more intrinsic dates (e.g., loss incurred date, underwriting year, calendar year, etc.). Whenever possible, transaction dates should be stored explicitly. The reason should be clear, as information that "everyone knows" today becomes "no one remembers" as time passes. Storing the date information explicitly ensures that historical data retains its meaningfulness and its place in time as the data ages.

3.2.2.4 Aggregations And Summary Level Requirements

Some HVTPE information provides value at a very granular level, e.g., location of claim, while other information must be summarized to be of value. Examples include loss experience by class, premium by territory, etc. The ADM team can work to ensure that the data resource aggregations are achieved in an efficient manner.

3.2.2.5 User-Defined Fields That Change Infrequently

Certain user-defined information can be computed and stored along with the historical values themselves. Examples might include geographical regions built from ZIP codes, descriptions of deductible codes, and aggregations of business accounts into broad categories. The benefit is two-fold: a consistent definition and efficient summarization of key business information. The trade-off is the upfront time and effort required to build and store this information. This type of effort is best suited for information that does not change frequently. For example, building "state" from ZIP code is a fairly static, well-defined computation. The actuary can improve the usefulness of the data by requiring this information be pre-computed and available for retrieval.

3.2.2.6 User-Defined Fields That Change "On-The-Fly"

The actuary should require the ability and capacity for building summaries based on data fields that can be changed "on the fly." This is particularly necessary when the classes of business are periodically reviewed and re-classified. If the user-defined aggregations change, the assignment to each data element extracted from the HVTPE may change. For example, newly introduced vehicles are typically assigned to a vehicle class based on judgment. Once loss information has become available on a vehicle class, it is not unusual to find certain vehicles re-classified. The actuary may then analyze experience by currently assigned vehicle class, not by historical class.

The actuary's data is not purely historical in this instance. It is a combination of the historical experience extracted from the HVTPE (e.g., loss experience), combined with a user-defined aggregation (e.g., current vehicle class) that is derived from the historical data.

3.2.2.7 Support The Value Proposition Of The Data Dictionary

A data dictionary ensures a consistent view of what the information means across the organization. It provides a precise definition of what the field is called, what information is expected to be stored in the field, what the typical values for such a field might be, etc. The Data Dictionary can also document when a data field has only recently become available with useful information, or whether another data field has ceased to be populated with current information (e.g., if a program has been placed in run-off).

3.2.2.8 Historical Data Retention

The actuary's retention period for historical data will most likely be different from those required for other parts of the organization. Actuarial analysis may require use of policy year, calendar year, accident year, etc. For longer-tailed business, the aggregations may occur over many years of data. The actuary must balance between storing too little history, and storing more data than necessary, impairing the ability to efficiently retrieve useful information.

3.2.3 Critical, Actuarially Valuable, And Nonessential Data

The actuary must distinguish between information that is critical to the actuarial analysis, versus information that has potential actuarial use, versus nonessential information.

3.2.3.1 Critical Data Elements

Without the critical data elements, there is no reason to pursue construction of a distinct actuarial data solution derived from the HVTPE. Examples include loss information in sufficient detail, critical dates (e.g., loss incurred date), premium amounts, etc. An information system lacking these elements cannot provide sufficient detail for a full actuarial analysis of the corresponding business programs.

The actuarial data requirements for critical data elements must be clearly communicated to the responsible parties. It can be a fatal flaw in a HVTPE project to just assume that "everyone" is aware of the actuarial importance of this type of information.

3.2.3.2 Actuarially Valuable Elements

Other data elements can serve as valuable input to future actuarial analyses, yet cannot be considered critical. These data elements can be thought of as "actuarially valuable." Without these data elements, the actuarial landscape is bleaker and the analyses are thinner. Yet, the remaining actuarial information set will allow actuarial analysis to continue in some lesser capacity.

Because "actuarially valuable" lies one step removed from "critical," there will always be a question of whether or not it is justifiable to capture these additional fields. There is a trade-off between the cost of gathering, storing, and reporting on actuarially valuable fields, and the potential "what if" insight the data can provide. The actuary must be prepared to discuss what might be the potential value of each additional field captured.

The ultimate value of considering these data elements is in allowing the actuary to be proactive rather than reactive. The actuarially valuable data elements may not be examined in every analysis, but there is a time-to-market advantage in incorporating readily available information when needed. The additional data is helpful when supporting the introduction of revised rating systems. The actuarial valuable data elements also tend to show additional value when the actuary needs to investigate how or why a book of business deviates from its projected values.

3.2.3.3 Nonessential Data Elements

Finally, there are nonessential data elements. These are perhaps "nice to have," but not cost-effective from an actuarial point of view. The data elements are not crucial to actuarial analysis, or the potential value in an actuarial analysis is limited. For example, the color of a vehicle could be examined as to its interaction with claim frequency or severity, but it is hard to imagine how this might be incorporated into an actuarial pricing, rating, or reserving analysis.

It might still make sense to consider nonessential data elements, if the information is of value to other areas of the company. For example, claims adjuster ID, or sales agent number, may be valuable to another area. Inclusion in the HVTPE data extract would allow other areas to leverage the actuarial effort.

3.3 The Role of the Actuarial Technician

3.3.1 The Dual Roles

As a pivotal member of an Actuarial Data Management unit, the Actuarial Technician must always be focused on their top priority: ensuring client actuaries are provided with the data resources necessary to excel in the performance of their functions. This involves two types of support roles: one as a data facilitator, the second as a data supplier.

3.3.2 The Data Facilitator Role

As a data facilitator, the Actuarial Technician regularly monitors the corporate data resources that the actuaries depend upon, whether directly or indirectly. As issues arise related to the availability, accessibility, and integrity of the data, they are then in a position to advise the actuaries accordingly. Such advisories may be limited to simply notifying affected individuals about problems or pending circumstances. For example, these notifications would be made when data anomalies are observed in the data warehouse, data marts, or their source systems, or when the release of new data is accelerated or delayed. In other cases, the advisories may involve the relay of specific actions that must be taken to work around problems that have yet to be permanently resolved. Examples of these would include notifications to select certain instances of data elements over others due to problems that had surfaced, or perhaps providing details for filtering the data differently to avoid erroneous results. The point of these facilitating actions is to promote effective methods of obtaining the highest quality data possible, as well as enhance the productivity of the actuaries. Technicians can obviate the need for repeating tasks (or even whole processes) performed by the actuaries simply by being attentive to corporate data issues and circumstances surrounding the data systems and then conveying related information in a timely manner.

As a data facilitator, the Actuarial Technician also serves as an intermediary agent between the actuaries and the IT Department. Because the Technician's job is so dedicated to providing data to the actuarial staff, he or she is in a unique position for tracking the evolving needs at the local level. They are witnesses to the ongoing development of information requirements by virtue of being the only "first-tier supplier" of actuarial analytical data that is in close proximity to the actuarial consumer. This familiarity with data requirements is of particular value when new data requirements need to be conveyed to the IT area and formalized into a project. As discussed under The Role of the Actuary, data requirements must be determined by the actuaries; however, it is the task of the Actuarial Technician to facilitate the transfer of those requirements to those that must satisfy them when the supplier is the in-house IT Department or an outside technology vendor. The requirements transfer may involve editing documents, drafted by the actuarial staff, to expound on, clarify, or provide examples of the data being requested. Or it may involve providing interpretation of the request in terms that address corporate-mandated procedural requirements, e.g., forms and support documentation. If the data project deliverables are of a size or complexity requiring a phased in approach, then the Actuarial Technician should consult with the requestor(s) and discuss what options are offered by the supplier. And if any particular advantages or drawbacks among the options are evident to the technician, he or she should make the requestor(s) aware of the observations. Likewise, additional alternatives should be discussed if they would better serve the need and would be plausible for the supplier to accommodate. When the preferred options and priorities are decided, the Actuarial Technician should convey them to the supplier or facilitate a meeting of all parties. In the communications between the Actuarial area and the IT function, the Actuarial Technician is not impartial. He or she primarily represents the interests of the actuaries. With that said, when it becomes apparent that progress can only be made through a compromise of all parties' concerns, the Actuarial Technician should do their best to mediate a solution that achieves a balance between the contending positions without undue compromise to the actuarial position.

The last aspect of the Actuarial Technician as a data facilitator involves software and hardware tools. Again, because the technician's position is semi-business and semi-technical, they are in an advantageous position for researching and evaluating data manipulation tools that would achieve greater productivity not only for the ADM unit, but the actuaries as well. Database management tools were once considered to be exclusively within the IT staff's domain. But as the need for greater data manipulation capability evolved, these tools were adopted by the more technically-inclined end users, and eventually mainstream end users. This progression implies that the Actuarial Technician will remain at the forefront of the dissemination of software tools that will enhance the capabilities and productivity of the actuarial staff at large. Consequently, the actuarial technicians in the organization should keep other actuarial staff members apprised of tools that may provide such benefits.

3.3.3 The Data Supplier Role

As a data supplier, the Actuarial Technician is the fulcrum that allows an actuary to leverage his or her analytical abilities. Without the Technician to intermediate between the state of the data as it is stored in the corporate systems and its transformed state needed for statistical analyses, the time and skills of actuarial professionals would be heavily taxed. In an organization with high transactional volume, the proportion of time used to prepare data for analysis versus the time used to perform an analysis easily shifts from an 80% versus 20% proposition, to one of 20% versus 80%. And the shift cannot necessarily be discounted as the transition from older legacy systems to more accessible data warehouses or data marts occurs. For as the raw data is wrestled from older systems and kneaded into more refined and accessible chunks, the progressive requirements for analytical aggregations, as well as successive drill-down capabilities, emerge. In fact, the refinements involved with these types of data may expand the number of processes supported by the ADM unit, rather than simply replace them, because of the data's increased scope and the actuary's heightened need for its ongoing availability and accessibility.

As data suppliers, Actuarial Technicians perform both end user production activities and end user development activities. From the initial release of data that fulfills the input requirements of an actuary's periodic analysis, an implicit expectation materializes that the same data will be provided in an updated form in the future. The expectation may be communicated early on as part of the original request, or it may take the form of a "one-off project" that over time seems to recur in a variety of incarnations. In any case, the Actuarial Data Management group needs to maintain a production schedule for developing and distributing the data it is routinely expected to provide. That is job one. Unless the actuaries can count on the consistent and timely delivery of the input data for their recurring analyses, any new development work on the part of actuarial technicians is meaningless. The data as a whole will lose credibility not as a result of any inherent inferiority, but due to the unreliability of its providers. This consistency of timeliness is as important as the completeness and accuracy of the data itself in order to achieve superior data quality. So the support of ongoing production work must be the prime directive for actuarial technicians.

And the support includes addressing issues that would threaten the fulfillment of that directive.

Beyond the production activities, actuarial technicians need to allocate time towards new development efforts as well. These take the form of in-department projects, as well as corporate projects. The in-department projects represent enhancements to the existing data resources that the ADM team manages as well as new data development. Corporate projects can directly impact the data resources maintained by the ADM group either by altering the data that is fed to them or by altering the hardware and software infrastructure that supports them. In either case, the Technician needs to be involved and attentive to any negative effects by specifying, if not also performing, adequate user acceptance testing (UAT). Maintaining an ongoing presence during the course of the project by attending status meetings can often alert the Technician to hazards and issues that could result in detrimental consequences to their data systems that would not be apparent from a review of the project's business and technical requirements documentation.

3.4 The Role of the IT Management Function

3.4.1 The Information Technology Perspective

The extraordinary growth and advancement of the technology industry has compelled IT departments across the business landscape to expand, reorganize and reinvent themselves repeatedly in an attempt to meet the requirements of their business unit customers. The scope of the IT Management function now spans nearly every part of the modern insurance organization. The emergence of new technology, the drive for incremental improvement in business processes, and competitive pressures have propelled this expansion. The aspects of the role needed for an effective actuarial data management operation, however, are not quite so diverse. There are particular IT responsibilities that provide the key elements of support for achieving actuarial data management objectives.

3.4.2 Managing The Existing End User Infrastructure

First and foremost, the IT function must ensure the availability and functionality of the existing business computing infrastructure. This means more than simply troubleshooting problems after they've been reported to a help desk by users, but rather proactively managing the infrastructure. Is local area network (LAN) monitoring software in place to detect and alert IT management personnel about network traffic spikes and extended high loads that can be traced to substantial data transmissions? If a wide area network (WAN) is part of the data management infrastructure, does it have a traffic monitor with notification triggers? When network disk storage reaches a 75-80% utilization threshold, are warnings issued to archive and free up space so as to avoid abrupt interruptions? Also, if the opportunity arises for business personnel to be involved with infrastructure plans, it can be very beneficial for providing input. Sometimes decisions are made by IT management to reallocate resources that appear to be "on average" under-utilized. Summary level monitoring reports that are commonly used for such decision-making don't always present a valid picture. The resources in question may actually be utilized heavily for short periods at weekly, monthly, or quarterly intervals. A reduction of throughput capacity across a network or on a data server could seriously constrain the efforts to prepare and distribute updated data resources on schedule. The voice of a business user in the forum of an infrastructure planning meeting can make the IT area aware of that situation and avert a potential crisis.

3.4.3 Infrastructure Renovations And Innovations

Secondly, the IT function should facilitate the advancement of the infrastructure in such a way that the actuarial data management function as well as the actuaries can take advantage of already-installed technologies in new ways or adopt newer technologies that increase functionality and productivity. This can occur by providing access to current software remotely through dial-in and broadband channels. Because data management processes can require several hours, it would be helpful to have remote access so subsequent processes can be launched after normal business hours if automatic triggers are not available.

Also, the IT function can be especially helpful if they maintain a program for routinely upgrading versions of both server and desktop application software. In regards to new types of user software, it may be unrealistic to expect the IT area to keep current on products that would be especially beneficial to actuaries, unless a specific problem or a functional deficiency has been communicated. However, the actuarial area may become aware of new or enhanced products that promise to add substantial value to either the analytical or data management processes of the department. At these times, the IT area should be invited to jointly investigate the potential. The Actuarial area can assess the value of the products in terms of their business requirements, and the IT area can assess the cost of the products in terms of their installation and technical support requirements. Together, both areas can determine if the costs are justified by the anticipated values.

3.4.4 IT Project Management

Lastly, the IT function is needed to assist with the execution of projects that either 1) Offload work to an automated system developed and maintained by IT professionals, or 2) Exceed the ADM unit's domain of control.

In the first case, some of the data resources and programs created by the ADM unit for the actuaries may over time become basically static structures. That is, the architecture and computations contained in them do not require updating. It becomes enough to simply refresh the data for new increments of time. Yet the refreshment process may require several intermediate processes that require many days or even weeks to accomplish. Even though the programming is sound, offloading the routine work to a new production system built by the IT area would allow the ADM group to place greater focus on the more volatile and actuarially esoteric requirements of the department. In fact, ADM-produced prototypes of data marts and programs that satisfy routine requirements would both prove the concept of an IT project as well as serve to meet the actuaries' needs on an interim basis. Such prototypes can also serve as the basis for the requirements of a formal system development project.

In the second case, where a project exceeds the ADM unit's domain of control (or the group's scope), the IT area must be engaged to enlist and manage the necessary in-house and/or outside vendor resources. A common example of such a project is the addition of new data elements to the organization's data warehouse or interdepartmental data marts.

4. PROCESSES

4.1 Commitment To Succeed

Planning, designing, building, and maintaining actuarial data resources that house millions of policies and claims is not a simple undertaking. This deserves to be stated explicitly even though the majority of readers who have persevered reading to this point would never assume otherwise. However, there is no shortage of consultancies that sincerely profess they possess the knowledge, skills, tools, experience, and human resources to promptly craft a silk database from a sow-system's ear. And that is not to say that truly qualified consultants couldn't perform such a metaphorical miracle for a needy and adequately-budgeted actuarial manager. But the miracle simply can't happen without the dedicated, time-consuming participation of the actuaries and their support technicians who must articulate the data requirements in definitive terms, facilitate the collection and communication of critical technical information to the consultants, be willing to discuss data issues at length, and judge which, if any, and to what degree, compromises regarding the deliverables can be tolerated. As the proverb goes, you will only get out of it what you put into it. Whether the actuarial area's data resources are products of strictly in-house efforts or of consultants hired from outside the organization, key members of the actuarial staff need to be involved to whatever degree it takes to bring the initial databases online as well as sufficient commitment to oversee their maintenance ongoing.

4.2 Formal Vs. Informal Approach

With that said, data resource development efforts involve a number of processes that are generic. One or more of these processes can be approached within a formal system design methodology and using specially designed software tools, or they can be approached informally through a logical and incremental approach. Regardless of the magnitude of the project, when the development team needs to serve simultaneously as an operations team and a production support team, the logical, incremental approach may be the only practical means to accomplish a project's objectives. This is because production work must be regarded as paramount. As discussed under the Data Supplier Role of the Actuarial Technician, development work must be subordinate to production work or the credibility of current and future data deployments is diminished. Having sufficient resources to segregate ADM people between exclusively production and exclusively development work teams is without question a luxury afforded by few (if any) P&C insurance organizations. Consequently, the flexibility of following a simply logical and incremental approach to implementing data projects by an ADM team should be regarded as the norm, rather than the exception.

An example of approaching a data project using a simply logical and incremental approach would be rebuilding or refreshing an existing database or data mart using a new process or set of tools. When the opportunity arises to introduce new software tools or enhance a data development and deployment process, the ADM team can integrate the new tool(s) or reprogram the existing process(es) as time allows and run it in parallel with the current processes. This ensures that new methods and processes meet the existing standards as a minimum requirement, and provides continuity with past practices and data quality levels. A successful parallel test as well as continuity with past data deployments are very important for gaining the acceptance of new data products by actuarial data consumers.

In the case where IT Management leads a project and manages the development resources that are independent of the ADM team, the formal project methodology approach is expected to be the norm. In that instance, the technicians that design and build the project's deliverables can work as dedicated resources focused on the new development activities without engaging in the risky practice of placing production priorities in contention with development priorities.

4.3 The Processes Of Data Management

At a deeper level than the approach and management of an actuarial data project, the processes involved with fulfilling the requirements will normally imitate, if not actually parallel, those of data warehousing processes. Consequently, a survey of these processes are presented below to familiarize actuarial personnel with them.

4.3.1 Data Modeling, Metadata, And The Data Dictionary

The beginning process of actuarial data management is the identification, qualification, and modeling of the data required for actuarial analysis. Until the data requirements are sufficiently identified, defined, and structure added to them (as is done with data modeling), the deliverables conceived in the mind of an actuarial requestor may be very different than that of their support technician. The identification and definition of requirements are initially expressed in the data terms, descriptions, and valid values associated with the data elements that the actuaries need. Ultimately, the data requirements will be expressed in terms of data models, metadata, and a data dictionary.

Data modeling is the identification, analysis and organization of data elements into logical and physical database designs. Some data modeling software simply provide a means to build logical relationship diagrams among data entities for documentation purposes. Microsoft's *Visio* application is an example. Other data modeling software goes beyond mere designing of a database to the creating of a complete physical data model once the logical design is finalized. *Erwin* by Computer Associates is an example of that type of software.

Metadata is data about data in a system or data structure. Each data element has particular attributes that uniquely characterize it, such as its name, definition, description, data type, data length, format, valid values, domain ranges for values, source files and source data elements. When data is processed through retrieval, cleansing, conversion, and transformation stages, there is technical process metadata that is applicable to data elements as well, such as the date processed and business rules applied to derive the stored value. Supplemental business metadata is also valuable to capture for knowledge workers. Examples include descriptions of how the data values have changed over time or descriptions of how the data enters the system initially (via automated means or manual entry), and descriptions of how upstream business practices may have impacted the values observed in the field.

A data dictionary is a tool for displaying metadata to business and technical personnel. A data dictionary is important for expediting the transfer of knowledge regarding the meaning of data values stored in the data fields. Without a consistent point of reference for describing the meaning of codes as well as the sources and derivations of data elements, any analyst will be hampered in their efforts to build accurate queries and effectively analyze the queries' results. Likewise, data management technicians need to have a firm grasp of the technical metadata in order to build complex extraction and transformation processes.

4.3.2 Data Extraction, Data Profiling, And Data Quality

Data extraction is the process of selecting and copying discrete values from data fields resident in a system file or database. The system file or database from which the values are extracted is referred to as the data source. Interim processing can occur and the data is then stored in another system file or database referred to as the target. Any data retrieval and reporting tool can be used as an extraction tool provided it has the ability to store the retrieved data in a form that can be used as input to a subsequent retrieval tool. There are software programs, however, referred to as ETL (Extraction, Transformation, and Loading) tools that are specifically designed for this purpose. They generally provide the ability to access a number of different types of databases and data file formats.

Data profiling is the process for examining and analyzing characteristics of data to ascertain or improve the level of its quality. The importance of data profiling is heightened in those instances where allegedly identical (or predominantly similar) data is intended to be extracted from multiple sources, then integrated and stored into a single target structure. A basic data profiling approach would include data column analysis where specific properties are measured for individual data elements, e.g., minimum, maximum, and average field lengths, or minimum, maximum and mean for numeric values, precision and scale for numeric values, data type, data format, the number of distinct values for the field, and the number of occurrences of null/empty values. Depending upon the type of field and the count of its distinct values, a complete list of values compared to known valid values should be produced also.

Data quality is the process of rectifying data defects and improving the accuracy, integrity, and understandability of data. This is preferably accomplished by identifying and correcting inconsistent or erroneous data in a source operational system or data warehouse architecture. However, if it is not cost feasible (or perhaps organizationally feasible) to implement data quality enhancements at those upstream stages, then it falls to the Actuarial Data Management unit to effect a reactive data improvement process at the department level.

4.3.3 Data Integration And Data Transformation

Data integration is the process of merging data from different (and sometimes very disparate) sources. Source systems that contain what is thought to be identical data elements can, in fact, prove to contain different formats, data types, and representations of information. For example, four different systems may contain a data field identified as STATE. In one system, the value stored in a data record could be "Alabama", in another, "AL", in another "01", in another, the field may be null. To consolidate information from multiple systems in a manner that either retains or introduces integrity to the data, the source data must be understood. Likewise, to establish data integrity in a target data structure, target data elements must comply with the business and technical requirements driving the integration task.

Effective data integration involves the application of business rules (systematic procedures) for "cleaning" data and deriving valid and as-accurate-as-possible versions of it. This is necessary to ensure that actuaries who will eventually use the data as input to their analyses need not qualify their results with excessive error margins. The subprocesses of converting data to different data types and formats, applying data cleansing techniques, consolidating varied representations of the same values, e.g., "01" vs. "AL", and deriving new data elements from others, can be referred to as data transformations.

Data transformations that are used to populate operational data structures can be very different from those used for actuarial data structures. In the cases of operational systems and data warehouses, data elements are normally static with respect to historical occurrences. That is, when changes to the definitions or valid values of data elements in these systems occur, the changes are made from a current point in time and only for subsequent periods thereafter. Redefinition and valid value descriptions are not enacted retroactively.

In the case of actuarial data structures, it is not uncommon, especially for user-defined fields, to recast data definitions and data element values for all time periods. In fact, such redefinitions (and repopulating of the data in the structure) may occur regularly, each year, quarter, or even month. To non-actuarial data managers, this practice seems illogical and mistakenly viewed as a violation of proper data management rules and practices. To an Actuarial Data Manager or an Actuarial Technician, however, it represents a critical added value that he or she brings to the data resources they deploy to their actuarial clients. Where an actuarial staff is dependent upon non-actuarial data management resources, it will likely be necessary to explain the basis for such dynamically changing fields to obtain the necessary views of the data. Unfortunately, even with sufficient explanation, the normal flow of data projects through IT areas may not allow timely turnaround of such requests. Apart from the need for specialized subject matter expertise and attentiveness to data quality matters, this issue of addressing dynamic data requirements in a timely manner is a leading reason why an actuarial area needs its own data management unit.

4.3.4 Data Loading

The process of loading data into a structure for retrieval by actuarial personnel varies according to the type of database or retrieval tool that is intended to be used. For example, interim processing programs can save relatively small data tables as Microsoft Access database files, or even as comma-separated value (CSV) files for later importation by actuarial analysts. Oracle files can be loaded using native structured-query language (SQL) commands or using a more robust tool specifically designed for mass loading of data such as Oracle *SQL*Loader* (pronounced Sequel Loader). As the name implies, any ETL (Extraction-Transformation-Loading) software tool incorporates loading functionality into its design. But if such a tool were not already procured for the purposes of data extraction and/or transformation, it is likely be cost prohibitive to purchase it exclusively for loading data into a database structure.

5. CASE STUDY DISCUSSION

5.1 An Introduction To GMAC Insurance

In order to demonstrate a real-life scenario of the evolution of actuarial data management (ADM) functions within a high-volume transaction processing environment (HVTPE), the authors would like to describe a bit of history taken from their collective experiences at GMAC Insurance. And to place their experiences into the context of an evolving business organization, a brief background of GMAC Insurance is provided.

GMAC Insurance traces its roots to 1925 when the General Exchange Insurance Corporation (GEIC) was founded as a subsidiary of the General Motors Acceptance Corporation (GMAC). GEIC was established to fulfill the insurance needs of GM dealers and their customers. Initially, the focus was on physical damage protection for automobiles. A year after its incorporation, GEIC became the largest writer of automobile physical damage insurance in the United States and Canada. In 1939, GMAC established an agency company named the Motors Insurance Corporation (MIC), for which GM dealers became licensed insurance agents. And in 1960, GEIC and MIC merged, retaining only the latter's name.

During the 1970s, Motors Insurance Corporation pioneered mechanical repair protection. This optional coverage provided financial protection to customers of GM dealers for certain automobile repairs and services that were outside the scope of the traditional vehicle warranty. Examples of these repairs and services included warranty-type repairs occurring beyond the months and miles provisions of the warranty, rental car reimbursement for multi-day repairs, towing of the mechanically-disabled vehicle to a dealership repair facility, and a waiver of the applicable deductible for in-warranty repairs. The popularity of the mechanical repair protection programs catapulted the business line to a prominent status among MIC's writings by the end of the 1980s. The rapid growth of the mechanical repair protection business made it necessary to continually seek improvement of the data capture, reporting, and analysis processes managed by the actuaries and support people assigned to those tasks.

During that same era, the company expanded its products and services to transform into a truly multi-line property/casualty underwriter. Throughout the 1990s to the present, as opportunities have arisen to better support and add value to its parent organizations, General Motors Corporation and GMAC, the Motors Insurance Corporation has teamed up with other organizations through acquisitions and specialty insurance startups to form the GMAC Insurance Group, an A-rated (Excellent), Top 30 insurance group according to A.M. Best, with combined insurance writings approaching $3 billion per year.[3] GMAC Insurance remains the leader in the mechanical repair protection business, offering coverage on new and used, GM and non-GM vehicles sold throughout the United States and Canada.

5.2 The Rudiments Of Actuarial End User Computing

As one would expect, just as the mechanical repair protection business has substantially evolved over the past 25+ years at GMAC Insurance, so too have the end user computing practices and capabilities evolved that support its actuarial functions. In the beginning, the tracking of premiums and losses was an elementary task, although only very basic hard copy reports containing summary information were available for that purpose. At a system level, the corresponding contracts were recorded in a manner similar to automobile physical damage policies. This was a reasonable method for coding the information at the time, since losses were typically limited in size, and quickly paid once submitted. In fact, the initial pricing reports were produced in a format identical to the organization's automobile physical damage summary reports.

By the early 1980s, the pricing and reserving analysis work relied on very detailed reports produced by programs resident on the same mainframe platform as the mechanical business processing system. The reports required multiple boxes of computer paper to print out the detail needed as input for actuarial analyses. The required data typically resided many pages apart, and the sheer volume of paper to be stored required an excessive amount of physical storage space. Summary levels of the data had been predetermined and programmed into the reports. Consequently, if a subset of experience was needed, summarization proceeded manually, flipping through the pages and separately recording each aggregate. If a deeper level of detail was required, the only recourse was to request a programming change to develop another paper report. A revised report would normally require a minimum of a few weeks to a few months depending upon the complexity of the request and the workload (or backlog of requests) in the Data Processing Department.

The next innovation was the use of microfiche. The same volume of data took up much less space. Multiple timeframes were easily stored in a "shoebox" file, whereas the paper reports had required a storage room. However, the same data manipulation issues remained. Information was not easy to summarize, and if an additional level of detail was required, programming changes were needed.

A new method of obtaining new levels of summary from the microfiche reports was needed. An actuarial support person experienced in compiler languages (*Fortran* and *PL/1*) as well as fourth generation languages (*Focus* and *Easytrieve*) gained access to the master files of the reporting system. After creating extract files for each major segment of the mechanical business, summary level reports became readily available on an ad hoc basis. As needed, new levels of summary could be created and printed with one or two days of notice to the technician. This approach radically reduced the turnaround time to aggregate data by new criteria. If additional detail was required, however, a special request to the Data Processing Department was still necessary to effect changes to the production reporting system.

As the 1980s elapsed, the advent of personal computing began changing the analysts' landscape at the company. The ability of the end user to manipulate and summarize larger and larger subsets of information allowed pricing analysts to consider managing more detailed views of the business. Database programs on a desktop PC meant multiple summaries of a common set of data could be produced and compared to one another. Spreadsheet programs allowed the analysts to do more than just summarize larger sets of information. It was now possible to adjust the data for known influences, and thereby ferret out a deeper level of understanding of similar yet distinct segments of the business.

For the actuary, the PC provided a locally controlled, adaptable, "real-time" tool for analyzing loss triangles, exhibiting the policy year emergence of premiums and losses. Classification plans could be more frequently reviewed. "What if" analyses could be completed within the pricing and reserving functions, without the additional time burden of external programming efforts.

Unfortunately, with the increased use of personal computing, there came an increased demand for access to the raw material of data analysis, viz. data. It became quickly apparent that the process of re-keying data from mainframe-generated reports into the PC environment was costly, time-consuming, slow, and rife with the potential for input errors. Clearly, better and more efficient ways to gain access to the data were needed.

The next approach to improved data entry was to find a way to move from paper and microfiche reports to electronic versions of the reports. An early attempt entailed electronically capturing the online "print files" corresponding to the hard copy data reports. This information was then parsed and downloaded into a format accessible to the PC tools available at the time.

On the upside, the information no longer needed to be re-keyed, and the information was more easily balanced back to the original reports. Also, summarization proceeded more quickly via PC database programs. The PC environment allowed the actuary to adapt the analysis to reflect changes in the pricing and business environment. Grouping related segments of business for common analysis was much easier to do.

On the downside, the electronic data was captured at a highly summarized level. No greater level of detail could be extracted from these reports, without returning to the original sources of data, i.e., the mainframe online files within the mechanical business processing system. Without an apparent alternative, the actuaries and pricing analysts continued to rely on data programming professionals to accomplish refinements to the existing reports. This reliance led to large gaps in time between data requests and the subsequent retrieval and analysis of results. In addition, the process was slowed down by the need for the actuaries and analysts to explain fairly technical data requirements to the programmers who were not acquainted with actuarial analysis processes.

5.3 Getting Access To All The Raw Data

In the early '90s, it became evident that increasingly complex analyses required increasingly detailed information to support the analyses. Rather than continue the process of programming ad hoc subsets of the online files, it was proposed that a comprehensive set of mainframe data files be constructed that would make all of the contract level information available for actuarial analysis. The files were intended to meet the primary requirements of the pricing function and would capture the history of mechanical business as far back as could be retrieved. Consequently, the file set was dubbed the "Pricing History Data Files" or Pricing History Files (PHF) for short. These data files were to provide a very detailed, inception-to-date snapshot of all premium and loss records at a vehicle level as of each month end.

When contemplating the design of the PHF and the mechanics of building them, two approaches were considered. One was to construct files containing incremental transactions from the inception of the contract for each vehicle. Each month a new set of incremental records would be created and inserted into the files. The inception-to-date view of the business could then be derived by aggregating the records as of a given evaluation date. The second approach was to create an aggregate inception-to-date record for each vehicle from an existing "snapshot" master file, then combine additional incremental experience into the records as each month elapsed. The first option would allow the analysts to scrutinize the data down to the monthly operational level of detail. The second would not provide that granularity, but it would provide a level of detail sufficient for advancing the current state of pricing and reserving analyses. Ultimately, the pricing analysts chose the second option, due to the substantial additional cost (in terms of programming resources, data storage overhead, and ongoing processing time) expected to accompany the preferred first option.

The PHF System was created and controlled at the department level. The in-house system engineer contracted to technically design and build the data files worked side by side with the pricing analysts and actuaries on a daily basis. The resulting PHF data files were constructed in the remote mainframe environment, but extracts were summarized and downloaded for use in the local PC environment. Additional pricing details, such as descriptions of encoded values, were added to the PHF data. In the end, the PHF System enabled the analysts to "slice and dice" premium and loss experience across multiple time frames, blocks of business, vehicle types, etc.

With the programming and data resources now under departmental control, data extractions and summarizations were accomplished in a timelier manner. The analysts were able to communicate directly with a dedicated programmer, and so over time it became easier for both to collaborate on describing, accessing, and refining data requirements. As time progressed, several of the analysts became proficient in programming retrievals of their own data from the PHF System.

5.4 The Standardization Of The Data Extractions

Over time, it was observed that while many different analyses were created from the PHF System, there were common traits to a significant subset of data requests. For example, while a group of reports might have included loss amounts and loss counts arrayed by time intervals, the aggregations were constantly changing. For one analysis, aggregation might have been by vehicle type, while for another analysis the aggregation might have been by level of coverage. It was not uncommon to find an analyst describing their data request as "just like the last one except…" The upshot of this common thread was a realization that by creating standardized extracts from the PHF System, multiple data requests could be handled at once. (In today's world, these data extracts might be called "data marts," rather than databases.) These databases were created in a standardized format, at predetermined time intervals. The actuaries and pricing analysts would then summarize the databases to render useful information for their respective data analyses.

The business data needed for actuarial analysis of the mechanical repair programs was now accessible in a timely fashion. But to ensure the ongoing sustainability of the new data preparation process, a new role to complement the role of the pricing analysts and actuaries was needed. This role would oversee the data preparation processes from the routine regeneration of the PHF, to the subsequent extraction, summarization, downloading, and balancing of the data. And the new role would prove to be an impetus towards the next step in end user computing within the company's actuarial area.

5.5 The Emergence Of The ADM Function

The Actuarial Data Management function emerged in the mid-1990s. In its role as overseer of the actuarial pricing data preparation process, the ADM unit became responsible for balancing the PHF System output, fulfilling standardized database requests, organizing the resultant data sets, and creating complex applications in response to new analytical data requests by the actuarial analysts. The PHF, the standardized databases, and the processes and programs surrounding them were placed under the control of the ADM unit, comprised of a business manager with IT management experience, an actuarial technician, and two technical contractors.

What made this process work well was the division of labor. The ADM team was able to focus on the data development and delivery processes. The pricing analysts and actuaries were free to focus on using the information contained in the data. Both groups benefited from working together under a common departmental structure. Potential changes to data requirements could be discussed, tested, and refined based on direct interaction between the ADM team and the analysts.

The ADM team also provided the necessary bridge to enhance data management functions for actuarial use. For example, rather than rely solely on a suite of PC-based tools for working with small databases and spreadsheets, the ADM team investigated the use of OLAP (Online Analytical Processing) tools for storage and delivery of information. As a result of their efforts, an OLAP tool was procured and installed on a local PC server. This allowed for the routine deployment of "data cubes" to the actuaries as part of the routine PHF data refresh process.

The OLAP tool provided the actuarial analysts with fast access to multidimensional data. Much of the summarization was pre-computed, so that OLAP data retrieval was nearly instantaneous. Analysts gained the ability to query and retrieve multiple views of requested information, summarized along various dimensions or "cuts" of the data. Because the OLAP data was centralized, very little computing power was required at the local (client) PC. Rather, the OLAP server performed this function. Some have characterized the OLAP environment as an "industrial-strength pivot table." Suffice it to say that the OLAP tool implemented at GMAC Insurance came to be regarded as all that and more.

5.6 The Implementation Of The Data Warehouse

By the late 1990s, several other departments within the company began to look for ways to extract meaningful data to meet their informational needs. The views of the business data that were needed to serve their purposes, however, were different than those of the mechanical business pricing function. Claims personnel sought detail at the vehicle repairer level and at the individual claim level. Marketing required sales summaries by region and distribution channel. Finance needed detail to support their aggregations of company and line of business financial reports.

It also became clear that company management desired a "single source" for mechanical data reporting, which could replace a number of special purpose reporting systems with their individual support requirements and escalating maintenance costs. The benefits of such a strategy would include increased quality and consistency of information across departments, a reduction in the long term maintenance of the data, and centralized management and access to the data. The "single source" would need to be a "production" system, i.e., a resource recognized as part of company data processes, not a departmental data system. The company was on the threshold of designing and building a data warehouse.

As the data warehouse initiative was formally launched in 1997, the Actuarial Department stepped forward to provide key business support. The Vice President & Chief Actuary served as Project Champion, and subject matter experts from both the Mechanical Pricing & Reserving function and the Actuarial Data Management function participated. Project team members from other departments within the company participated in requirements capturing sessions and served as subject matter experts for their respective functions with the company. The IT area managed the project as well as the vendor used to design and build the warehouse. The Mechanical Data Warehouse rolled out in phases with the final phase deployed to production status in mid-2000. The completed project resulted in an extremely detailed and useful corporate data warehouse for the mechanical business line, with information stored at the transactional level. Since that time the Actuarial Data Management unit has served as the company's maintainer of the Data Warehouse Data Dictionary, and performs detailed balancing procedures as part of the data warehouse monthly refresh process. It also performs the company's user acceptance testing of data stored in the warehouse when changes are made to upstream (source) systems.

5.7 The Migration From The PHF System To The Data Warehouse

The deployment of the Mechanical Data Warehouse provided the foundation for re-sourcing the data used by the actuaries from the PHF System to the new and single corporate source for mechanical business data. The eventual migration of the PHF applications and the re-sourcing of the actuarial databases, which transformed them into true data marts, was a development project far surpassing any single ADM group effort to that point in time. The knowledge, experience, and advanced skills of the unit's senior actuarial technician and its technical contractor were critical for achieving a successful transition. Their efforts combined with that of the other support technicians made it possible to complete the project while maintaining a demanding ADM production schedule and providing data warehouse UAT support for numerous corporate initiatives that ran concurrently.

6. CONCLUSION

To conclude this discussion paper, the authors would like to reiterate the goal of an Actuarial Data Management unit, and provide some helpful perspective when attempting to establish or advance the function within an insurance organization.

6.1 The Goal Of Actuarial Data Management

As first stated at the beginning of this paper, the goal of an Actuarial Data Management unit is to equip an actuarial staff with the data resources necessary to excel in the performance of its functions. The ability of the ADM team to achieve that goal in a high-volume transactional processing environment depends upon many factors, some that are under the unit's control, and some that they can only exert influence upon.

Factors under the control of the ADM unit include attitudes and behaviors conducive to high quality work performance and high quality data development processes. These are best expressed by an orientation towards ongoing incremental improvement of the technical processes, persistence at resolving issues and problems related to data quality and timeliness, and perseverance during periods of heightened difficulties and setbacks. Another important factor under the unit's control is adequate testing of both interim and final deliverables created to satisfy actuarial data resource requirements.

Important factors upon which influence can be exerted are the requirements gathering process, interactions with actuarial clients and the IT area, education and training in the tools used to perform ADM functions (including participation in software user groups), roles in user testing of corporate projects that impact ADM functions and supported data resources, and the visibility of end user computing capabilities within the business and IT sectors of the company.

In some organizations, the actual tools used to perform ADM activities can be selected or at least influenced. In others, the tools are prescribed by standards established by the IT area or by corporate policy. Likewise, the degree to which technology can be applied to business problems by business people (who are not IT professionals) is sometimes governed by the IT area or corporate policy. In that regard, business-led actuarial data management activities can sometimes be perceived to press upon and even push the limits of what a business function should undertake. However, when a technical solution is within the capabilities of an ADM function and its developmental scope is departmental, not corporate, the IT area should recognize that their responsibilities need not extend further than administering the underlying infrastructure and providing advisory support as requested to resolve issues.

6.2 Actuarial Data Management Now And In The Future

Actuarial Data Management is a necessary function in a high-volume transactional processing environment. It serves as a proactive, added-value conduit of business data and specialized technical support to the actuarial staff. The greatest value will be realized by actuarial departments that foster within their ADM function a balance of subject matter expertise in actuarial data requirements and advanced technical competence with information technology. Such a balance not only makes it possible for the ADM unit to meet its prime directive of providing analysis-ready data to its actuarial clients as needed, but also to serve as a bridge between the actuarial staff and the IT area, facilitating much of their interactions and expediting mutual interdepartmental objectives.

The challenges of establishing and maintaining an effective actuarial data management function today and in the future will certainly persist and in some respects escalate. This is evidenced in the limited availability of highly qualified actuarial technicians, the growing need to ensure procedural integrity and consistency in compliance with federal regulations such as the Sarbanes-Oxley Act of 2002, the perpetual release of software and hardware innovations that support data management activities, the commensurate withdrawal of vendor support of past generations of software and hardware, and the continuing evolution of actuarial data research and analysis techniques. These current issues are compelling changes in the ADM function simply to maintain the status quo. When taken into account with the perennial drivers of change associated with an insurance organization (revenue/income enhancement, customer acquisition/satisfaction/retention, loss/expense containment, and marketplace competition), there can be no expectation of respite on the part of ADM professionals in support of their actuarial clients.

Endnotes

[1] John F. Rockart and Lauren S. Flannery, "The Management of End User Computing," Communications of the ACM, Vol. 26, Issue 10 (October 1983), pp. 776-784, ACM Press, New York, NY.

[2] Ibid.

[3] "U.S. Ranking by Assets – Groups" and "U.S. Ranking by Net Premiums Written – Groups", Best's Aggregates & Averages, Property/Casualty, 2004 Edition, A.M. Best Company, Oldwick, New Jersey, pp. 654, 658.

General References

Building the Data Warehouse, Second Edition, W. H. Inmon, John Wiley & Sons, New York, NY, 1996.

Building, Using, and Managing the Data Warehouse, Ramon Barquin & Herb Edelstein, Ed., The Data Warehousing Institute Series, Prentice Hall, Upper Saddle River, New Jersey, 1997.

Planning and Designing the Data Warehouse, Ramon Barquin & Herb Edelstein, Ed., The Data Warehousing Institute Series, Prentice Hall, Upper Saddle River, New Jersey, 1997.

General Motors: Priorities and Focus – Yesterday, Today and Tomorrow, First Edition, General Motors University, College of Leadership, General Motors Corporation, Detroit, Michigan, November 2000.

Abbreviations and Notations

ADM, Actuarial Data Management
CSV, Comma Separated Value
DP, Data Processing
ETL, Extraction-Transformation-Loading
GEIC, General Exchange Insurance Corporation
GMAC, General Motors Acceptance Corporation
HVTPE, High-Volume Transactional Processing Environment
IS, Information Systems
IT, Information Technology
LAN, Local Area Network
MIC, Motors Insurance Corporation
OLAP, Online Analytical Processing
PHF, Pricing History Files
SQL, Structured Query Language
UAT, User Acceptance Testing
WAN, Wide Area Network

Biographies of Authors

Joseph Strube is Sr. Manager of Actuarial Data Management at Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan. His team develops and maintains analytical databases for the Home Office actuarial staff and oversees the data warehouse for the organization's largest business line. A graduate of the University of Michigan, Mr. Strube's career in the property/casualty insurance field has spanned 30 years at multiple carriers and includes assignments in the Actuarial, Financial, and Information Technology areas. Mr. Strube draws upon his management and technician experiences in the IT area as well as his current management role over actuarial support functions.

Dr. Bryant Russell is a Sr. Team Leader in the Actuarial Department of Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan. He has had extensive experience in pricing and reserving for long-term automobile extended service contracts. He was one of the actuarial team members engaged in creating a corporate actuarial database for pricing and reserving of such service contracts. Dr. Russell later served as actuarial subject matter expert during the development of the organization's data warehouse. He achieved his A.C.A.S. designation in May 2000. Prior to entering the actuarial profession, Dr. Russell completed his Ph.D. in Mathematics at the University of Michigan.