Synchronizing Data Management Technologies to Integrate Actuarial Processes

Omar D. Kouatly,
Mark W. Littman, FCAS, and
Aleksey S. Popelyukhin

Synchronizing Data Management Technologies to Integrate Actuarial Processes

Omar D. Kouatly, Mark W. Littmann, Aleksey S. Popelyukhin

ABSTRACT

Actuaries are under increasing demands to provide more analysis of relevant information required for business decisions in a shorter time frame. Inefficiencies inherent in departmental boundaries and responsibilities, such as between reserving and pricing, stress each group's ability to deliver on these ever increasing demands. Further, the increased demands often expose dated data handling techniques and apparent inconsistencies in the results of separate, but related actuarial functions. As a consequence, there tends to be less time available for analysis, with more energies expended on managing the flow of the underlying data and results and on reconciling and explaining apparent inconsistencies.

Actuaries should take heart, however. Advances in technology offer new solutions for not only data management, but also the actuarial analysis phase of the overall process. The two main points of this paper are:

1) The actuarial process is in the midst of a broader flow of data and information. Multiple technologies are needed in order to integrate the actuarial process to (upstream) original data sources and to (downstream) management reporting systems and in order to meet the requirements of actuarial analysis

2) The solution to 1) offers actuaries a realistic opportunity to "read from the same page" by sharing data and results, with a consistent yet flexible toolkit for actuarial methods and reporting. Such possibilities offer actuaries across the organization the opportunity to re-assess what they do and how they do it.

Omar D. Kouatly, ARM

Omar is a Manager of software development in the Insurance Software Division of Price Waterhouse LLP. He has achieved the Associate in Risk Management designation. He earned his Bachelors degree from Rutgers University. Omar made contributions to Chapter 1 and Chapter 2.

Mark W. Littmann, FCAS

Mark is a Senior Manager and Actuary in the Actuarial Consulting Services group within the Insurance Industry Practice of Price Waterhouse LLP. He is a Fellow of the Casualty Actuarial Society and Member of the American Academy of Actuaries. He earned his Bachelors degree from Valparaiso University. Mark was the primary contributor to Chapter 2.

Aleksey S. Popelyukhin, Ph.D.

Alex is a Lead Software Developer in the Insurance Software Division of Price Waterhouse LLP. He earned his Ph.D. in Mathematics and Mathematical Physics from the Moscow State University (Russia). Alex was the primary contributor to Chapter 1.

The authors are listed in alphabetical order; they respectfully request that any reference to the paper cite all authors.

ACKNOWLEDGEMENTS

The authors would like to thank Krystine Cramer, for her assistance in editing this paper; Boris Privman, FCAS, for sharing his views on the current state of actuarial process automation; and Fran Lattanzio, FCAS, for his vision and inspiration to integrate actuarial functions through shared technology.

Synchronizing Data Management Technologies to Integrate Actuarial Processes

INTRODUCTION

This paper is organized into two chapters. The first chapter contains the text of a technical paper, describing a data management view point on the actuarial process. The second chapter contains an illustrative dialogue between the presenter and the audience at the CAS Ratemaking Seminar. The dialogue is intended to clarify some of the points raised in the paper, as well as describe a sample of practical applications of the blue-print for an integrated actuarial data management and analysis process.

We offer the following actuarial and managerial issues for the reader to consider as he/she reads the technical paper. These issues will be discussed in the second chapter.

- The data sets used for pricing analysis are typically segmented by state and coverage on a total limits and basic limits basis, while the data sets for reserving functions may be summarized for all states on a gross and net of reinsurance basis. The different data sets are naturally related to each other, representing different "slices" of the original data set.

- Actuaries are frequently called upon to provide input to the business planning process, including but not limited to projected loss ratios and runoff of reserve balances. We view these as natural by-products of the pricing and reserving functions.

- Insurance organizations may deploy their actuaries throughout the business units or in a centralized unit. In either case, the chief actuary needs to ensure consistent actuarial quality and minimize "transfer costs" among actuarial units. We view this situation as warranting common yet flexible toolkits for analysis, and a mechanism for sharing data and results among the units.

- Financial management is looking for consistency in the information generated throughout the organization. Actuarial rate indications may suggest a deficiency of 15%, while the current financial statements show that the company is generating substantial net income. Actuaries are being called upon to reconcile the two measures. We see this situation as an apparent disconnection between the data and information generated by the pricing unit, the reserving unit, and/or the finance department. This highlights the need for everyone to read from the same page.

- Recognizing that actuarial analyses involve both science and art, a consulting actuary's client is asking the actuary to quantify the impact of each. For instance, a prior study generated an estimate of X. The current study provides a revised estimate Y for the periods covered in the prior study. The actuary is faced with the question of how much of the change was due to actual experience and how much was due to changes in actuarial judgments. We view these as factors giving rise to multiple iterations of the analysis -- one with last period's judgments superimposed on the current experience and another with current judgments. The challenge relates to the calculation engine's ability to run multiple iterations and store the results from each.

CHAPTER 1 – THE TECHNICAL PAPER

It is one month before the annual CAS Ratemaking Seminar, and the light blue book that contains the papers to be presented at the seminar has just been published and distributed. The text of one paper submitted in response to the call on Data Management topics follows:

Information, Please:
The Actuarial Process from a Data Management Point of View

Preface

As the information revolution advances, more information is being processed at even greater speeds and with even greater efficiency than was dreamed of only a few decades ago. Bold new approaches to information handling are being developed every day that enable business users to arrange seemingly unrelated oddments of information into usable and cohesive databases. These databases can then be optimized and utilized across many business departments.

Advances in database technology will prove to be invaluable for the automation and enhancement of actuarial processes as a whole. These advances offer solutions to help actuaries meet the increased demands placed on them. We observe, however, that because of the varied demands actuaries pose during consecutive stages of the actuarial process, no **single** database technology may suffice. Rather, an amalgamation of several of the latest data management techniques may provide the solution. We believe that the **integration** of a few major technologies, most notably Data Warehousing (DW) with OLAP (On Line Analytical Processing) and Object-Oriented Databases (OODB) with OML (Object Modification Language), may provide not only a satisfactory automation solution, but also a viable platform for future advances in technology.

Following is a discussion of these recent developments, and how they apply to the actuarial processes in the property/casualty industry. The elements to this discussion include:

- An analysis of the steps involved in the actuarial process from a data management point of view and the requirements imposed by this process on the ideal database solution
- An examination of existing database technologies in order to find the one(s) that fit better.
- A description of the "ideal" actuarial system, which can utilize technologies existing today to satisfy all the requirements discussed in the first section

A Data Manager's Problem:

Contradictory Requirements of the Actuarial Process

The actuarial process, like many analytical processes, consists of three stages: input, calculate, and report. This process presents a perplexing problem: three main stages, equally important to the ultimate goal, yet making often disparate demands upon a database.

Stage One: Input (Gathering Data and Building Objects)

Actuarial data such as losses paid, case reserves, allocated loss expenses, premiums, claim counts, etc., usually come from different sources of a transactional nature. This stage would benefit most from technologies that enable actuaries to:

- Reach legacy systems,
- Extract, clean and relate data, and

- Possibly store extracted data in some organized fashion.

Thus actuaries need a technology that would enforce data cleanup, the matching of codes from different sources and provide storage optimized for future aggregations. Many insurance companies have this gathering process in place: they accumulate data into well-organized tables which do not require rebuilding from scratch every time, but rather are append to regularly (quarterly, monthly, etc.). Naturally, companies do not wish to give up this process. In this paradigm all existing libraries of programs written in COBOL, PL/1, APL, and SAS programs (which read transactional data from tapes, disks, or cartridges to create datasets on mainframes or files on personal computer networks or hard drives) can be adjusted to serve as conversion tools for this type of storage technology.

From a data management point of view, most actuarial objects are results of aggregation either by time (for instance monthly data to annual or quarterly aggregations) or other dimensions (across policies to product lines, across claims to statutory lines, or across states to countrywide, etc.). On top of that aggregation, actuarial loss development triangles are essentially cross-tabulations; indeed,

- A paid loss development triangle is an aggregated cross-tabulation of payments summarized by a Loss Period (Accident, Report, or Policy-based) dimension by a Valuation Date (accounting date) dimension.

  ⇒ A cross tabulation of data by Loss Period by Valuation Date produces a right justified triangle (as in Schedule P of the statutory annual statement blank); to illustrate,

Loss Period	Valuation Date	Data
1993	1993	1225000
	1994	20000
	1995	37500
	1996	50000
1994	1994	90000
	1995	110000
	1996	800000
1995	1995	825000
	1996	1800000
1996	1996	3300000

Cross-Tabulation →

	Valuation Date			
Loss Period	1993	1994	1995	1996
1993	1225000	20000	37500	50000
1994		90000	110000	800000
1995			825000	1800000
1996				3300000

⇒ For a left justified triangle, the data are cross tabulated with Loss Period by Age (equal to the difference between the Valuation Date and the beginning of the Loss Period); to illustrate:

Loss Period	Age	Data
1993	12	1225000
	24	20000
	36	37500
	48	50000
1994	12	90000
	24	110000
	36	800000
1995	12	825000
	24	1800000
1996	12	3300000

Cross-Tabulation →

	Age			
Loss Period	12	24	36	48
1993	1225000	20000	37500	50000
1994	90000	110000	800000	
1995	825000	1800000		
1996	3300000			

Therefore, this task requires adequate tools optimized for aggregations and cross-tabulations. However, collections of data that has been gathered, cleaned, and pre-aggregated into triangles or other objects are not yet ready to support reporting and decision-making. The data needs to be processed by actuaries' sophisticated methods or algorithms.

Stage Two: Calculate (Actuarial Analysis)

No two companies process actuarial data the same way. Yet we can still observe certain commonalties in the implementations of the generic process.

- First, it is usually high volume, due to the data objects being organized into many segments, to improve homogeneity of each set of data analyzed. These segmentations typically reflect a company's profile of coverages, geographic, and customer-based market definitions.

- Second, the process involves a limited number of algorithms, in the sense that the many data objects will be processed through just a few algorithms. (Some algorithms are designed to perform diagnostic testing on the data to measure trends or provide insight as to which estimation method(s) may be appropriate or not. Other algorithms process the data in order to determine numerical outcomes.)

- Third, these algorithms have a tendency to be changed from time to time. We are not concerned whether the changes are enhancements or radical departures from prior techniques; it is the **ability** to change that is imperative for the actuarial process. In other words, actuaries require an open processing system.

- And fourth, the process generally generates a finite, standard set of outcomes (estimated ultimates, IBNR, Loss Ratios, etc.), which usually have to be preserved (saved) after calculations are completed.

While many companies maintain differences, these four commonalities appear to be consistent across the industry.

The actuaries' interaction with the data and methods also sets the actuarial process apart from other, linear processes. The actuaries utilize their training and experience to interpret data and make judgments regarding certain factors. They may do any or all of the following:

- Feed their algorithms with the data, parameters and sometimes manual selections

- Calculate outcomes

- Repetitively adjust their selection, thus updating outcomes

- Store final outcomes (in either spreadsheets or printouts) either for use in further algorithms (ultimate losses for ALAE or rate adequacy analysis) or for future summarizing and reporting.

In light of the discussion above, this process demands a high volume, computationally complete, and open system, which provides the persistent storage of results and which allows for iterations and interactions.

Actuarial calculations are usually applied to matrices or vectors, which can be viewed as data objects (as opposed treating them component by component). For example, a loss development triangle is generally treated as a whole **object**; the individual elements that comprise it, on their own, do not hold much interest to an actuary. However, as part of the broader collection of elements that comprise the object, they are of great interest. The calculations that are applied to actuarial objects are generally performed on many objects, and most likely, multiple times (it would be quite extraordinary for the actuary to be satisfied by the results of the first selection).

From a data management point of view, the act of performing standard loss development analysis is a **transaction** (for our purposes, an operation involving several items). As such, it is not different from crediting the interest to a savings account in a bank. For instance:

New (or modified) Interest =

Average Balance Calculation Method (Account Balance, APR Rate)

Compare:

New (or modified) Ultimate =

Loss Development Method (Paid Triangle, Selected Development Factors, Tail Factor)

Note, that a Paid Triangle (and for that matter the column of daily Account Balances) is treated as a whole object. In both cases some modification to an object (as a function of a few other objects and, probably, some parameters) occurs.

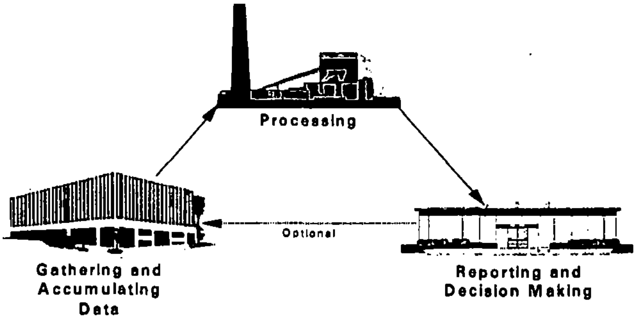
Thus, the database solution for Stage Two should be optimized for transactions on objects.

Stage Three: Report (Summarizing and Communicating Results)

Once the data has been gathered and calculated into results, it must, of necessity, be communicated to others in the actuarial department and to management across the company. The reporting aspect of Stage Three means summarizing and disseminating results. This allows actuarial management to review not only the final estimates, but also diagnostic summaries of the critical parameters of the process. Reporting to other departments will likely contain high level summaries of results that are relevant for corporate and line management for decision making.

The results of actuarial analysis, usually, are spread among hundreds of spreadsheet files, printouts or APL produced flat files with no mechanism to bring them together. Therefore, Stage Two needs to accumulate results in a centralized repository.

Furthermore, because the results of actuarial analysis are objects, they must be broken down into their component parts in order for them to be used for summarizations through aggregations and cross-tabulations. Thus, while breaking down these objects into their components, Stage Three should build an entity optimized for slicing, dicing, summarizing and reporting.



Technologies of Interest

We believe that few if any companies have achieved automation and integration of all three stages. The reason for that is apparent when you observe that the actuarial process as a whole has quite contradictory requirements.

On the one hand, during Stage One, the process requires a large storage facility optimized for aggregations and cross-tabulations, in order to generate triangles, which are essentially a cross-tab. Like Stage One, Stage Three also involves aggregating results, for instance, across lines and locations. These two stages would benefit greatly from the data warehousing technology.

On the other hand, the demands of Stage Two of the process are quite different. This database must efficiently store and retrieve objects (such as loss development triangles and other actuarial matrices). Its processing engine must be flexible enough to accommodate different methods and adjustments and should be optimized for sequential query execution (for processing multiple lines of business or multiple contracts) rather than aggregation or cross-tabulation. Finally, it should have an interruptible calculation process for interactive actuarial selections and judgments.

Out of all existing technologies there are two that appear to have properties that satisfy the actuarial demands outlined above: data warehousing with OLAP and object-oriented databases with OML.

Data Warehouse and OLAP

As defined by the "father of Data Warehousing" Bill Inmon (see Inmon, W.H., Using the Data Warehouse, QED, 1994), a "Data Warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision making process," where

- **Subject-oriented** means that the Data Warehouse is data-driven and organized around high level entities of the enterprise.
- **Integrated** means there is consistency in naming conventions, measurement of variables, encoding structures, physical attributes of data and so forth.
- **Time-Variant** means the database accumulates historical data over time, and
- **Non-volatile** means no updates, only initial and periodic batch loading and access to the data. Every update triggers a massive rebuilding of pre-summarized sets. This restriction on the frequency of updates helps to optimize DW for aggregations.

A Data Warehouse is usually comprised from Operational Data Storage (ODS) which resembles a traditional data table structure, complemented with a Multi-Dimensional Database, on which OLAP tools operate. OLAP (On Line Analytical Processing) is a technology which allows the user to perform:

- Mathematical operations on aggregated and cross tabulated data elements (for example, IBNR = Ultimate - Reported),
- Roll-up and drill-down type of queries, and
- Pivoting, i.e., the easy exchange of horizontal and vertical dimensions in two-dimensional slices.

Lotus Improv, Essbase, Holos and Excel's Pivot Tables are famous members of the OLAP category of tools.

A database technology similar to a Data Warehouse is a Data Mart. Roughly speaking, a Data Mart is a departmental Data Warehouse, as opposed to an enterprise-wide one.

Object-Oriented Databases and OML

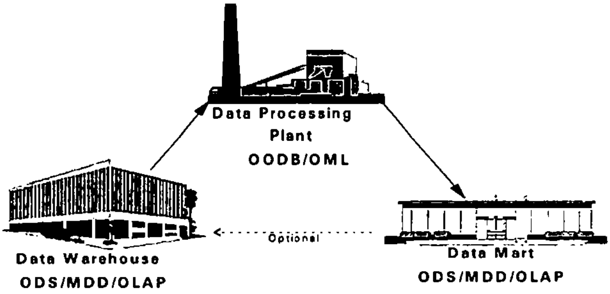
An object-oriented database provides persistent storage of uniquely identified objects, complemented by a computationally complete Object Modification Language (OML), where:

- **Persistent** storage means that objects remain after the creation and/or modification processes have terminated.
- **Unique identifier** reflects the object's place in the class hierarchy rather then being just the next available value as in traditional transactional databases.
- The **computationally complete** property of the language is emphasized because the standard Structured Query Language (SQL) is computationally limited. Also, SQL operates on a set-by-set basis rather than a record-by-record approach, which is convenient for object processing purposes.

An OODB, with the help of an OML, provides a means for the creation and manipulation of objects and a mechanism for the storage and retrieval of those objects. Typically, an OODB is optimized for fast retrieval and updates and provides all the fundamental benefits of a transactional database to object-oriented applications: a **transactional database** for objects.

IBM DB2 Common Server, Oracle 8's cartridges, and Informix Universal Server's Data Blades are the latest commercially available additions to an OODB family.

A complete definition of OODB can be found in "The Object-Oriented Database Manifesto," by M. Atkinsos, F. Banchellon, D. DeWitt, K. Dittrich, D. Maier, S. Zdonik in the Proceedings of the First International DOOD Conference, Kyoto, Japan, 1989.



"You are what you are optimized for."

There is a clear distinction between Data Warehouses and transactional databases: namely the type of optimization they utilize. The contrasts between the data stored in a transactional database and a Data Warehouse are:

Transactional data	Data Warehouse
Frequently changing	Static
Requires record-level access	Data is pre-aggregated into sets
Repetitive standard transactions and access patterns	Ad hoc queries with some periodic reporting
Event-driven: process generates data	Data-driven: data governs process
Updated in real time	Updated periodically with mass loads

An object-oriented database has all the properties of a transactional database outlined above, plus additional properties highlighted in the following summary of the differences between object-oriented databases (appropriate for Stage Two) and Data Warehouses (for Stages One and Three):

Object-Oriented database	**Data Warehouse**
Stores objects	Stores elements
Object identifier (non value based)	N/A
Computationally complete language	Not required

Data Warehouses are optimized for aggregations and cross-tabulations that are perfectly suited for Stage One of the actuarial process in which triangles and other objects are accumulated. A company may mass-produce triangles from one large Excel pivot table (an example of a Multi-Dimensional Database (MDD)) by slicing it differently for various profiles of coverage and geography. The pivoting and aggregation properties of a Data Warehouse with OLAP are also invaluable for the reporting and decision making features of Stage Three.

Judging the properties of an object oriented database as discussed, we can see that this technology meets the demands of Stage Two for a high volume, computationally complete system providing persistent storage for calculated results. Issues of openness and interactivity will be discussed below.

The Ideal Actuarial System

Thanks to the recent breakthroughs in the areas of Data Warehousing, On-Line Analytical Processing, and Object-Oriented Databases, an ideal integrated actuarial system can be built today. However, not one of the technologies described above may satisfy actuaries completely; Data Warehousing is not optimized for transactions (triangle retrieval with saving back ultimates from the Loss Development Method is essentially a transaction), while in the OODB paradigm, it is not easy to summarize or cross-tabulate objects.

The nature of the ideal solution parallels the nature of the three stages in that there would be three main elements to the ideal data management system: the Core, an Input Converter, and an Output Generator.

The Core

In the core of an "ideal" actuarial system would lie an object-oriented storage/retrieval database, which would store every actuarial object (triangle, matrix, row, column or scalar) as a single record in the database as opposed to the traditional approach, which is to store elements of the triangle as separate records. The advantages of such a database are numerous:

- **Speed**: Retrieval or update of one record in the database is always faster than the same operation on the multiple records. Speed is a significant factor even from an actuarial point of view. For companies that perform quarterly analyses of entire portfolios including multiple tests, either for diagnostics or estimation, speed is crucial.

- **Integrity**: Failure to retrieve or update a record with one element of the triangle (in the traditional database) may render the whole triangle unusable

- **Consistency**: Once stored, the triangle is not a subject for change because of adjustments to triangle creation algorithms or changes in line definitions. That makes it invaluable for auditing and similar purposes.

- **Diversity**: Such a database can store objects of different shapes and sizes. An annual development triangle, for example, would occupy one record; a quarterly one would also occupy one record, as would a vector of ultimates. This approach is radically different from traditional "by-element" storage solutions.

- **Data Retention**: All objects, data and results are available as a starting point for the "next time" analysis or auditing.

- **Effective Storage Space Utilization**: Utilizing sparse matrix technology, triangles can be stored very effectively. (And unlike traditional "by-element" storage solutions, where descriptive information is repeated multiple times (once per element), in OODB descriptors are stored only once per object.)

- **Reuse**: Objects stored in such a database may be reused in time (next reserve test) or in related actuarial applications (ultimates from reserving in pricing, loss ratios from pricing in reserving, ultimate losses for ALAE estimate, etc.). This reuse (sharing) of information ensures that actuaries and analysts do not expend time and effort "reinventing the wheel."

- **Application Optimization**: Most actuarial methods treat a triangle (or vector of ultimates or loss development factors) as whole indivisible objects.

The selection of a storage/retrieval system can not be considered separately from the applications. To meet the requirements of Stage Two of the actuarial process, the system has to be optimized for object manipulations and high-volume sequential query processing. In order to describe actuarial modifications of the triangles or objects of other shapes, the system needs an Object Modification Language. We assert that the language must have the following attributes:

- **Sophistication**, because some actuarial methods require complex calculations;

- **Flexibility**, because actuaries seem to perpetually tweak their methods;

- **Interactivity**, because actuaries need to get instant feedback for different assumptions, and

- **Familiarity**, because there are already too many languages to learn: APL, SAS, Visual Basic, PL/1, etc.

Actuarial algorithms expressed in this language should be size-invariant and the system should provide a mechanism for accepting objects of different sizes.

Though the demands may seem exacting, there is such a language: a spreadsheet. Spreadsheet ranges may be designated for communications with the object oriented database (accepting objects stored there and storing new, or updating existing, objects back there from spreadsheet ranges). The ideal system would provide a means to designate such ranges for exchange with the main storage facility and a re-sizing utility for adapting algorithms expressed in the spreadsheet to differently sized objects. Spreadsheet functions that are expandable through add-ins would provide a syntax for an OML. In addition, a spreadsheet's ability to interactively accept user input serves the actuarial selection and judgment process perfectly. The spreadsheet environment's printing, formatting and charting facilities only add value to an already near-perfect match.

Note, that a spreadsheet in this scenario is treated only as a language (algorithms depository) and NOT as a file-based storage solution (as it is treated now in many insurance companies) which we suggest is non-effective. In the ideal actuarial system, effective storage for data as well as results is provided by the object oriented database.

Input Converter

For the core processing system to be a part of an integral solution, there should be a tool for generating pre-aggregated actuarial objects (mainly triangles and vectors) from traditional "by element" storage systems currently existing in every company.

That is where a Data Warehouse comes into play. A Data Warehouse, with its cleansed data and descriptive dimensions and members, provides the perfect platform for generating triangles and pre-aggregating other actuarial objects. Take into account that data warehousing technology is optimized for cross-tabulation and aggregation and that makes it even better suited for the task.

One way to take advantage of a Data Warehouse's Multi-Dimensional Database is to use it as a triangle generator: slices of a properly organized Multi-Dimensional Database will be exactly the triangles for different lines or other segments of the book of business. Therefore, the availability of an automated routine for triangle generation is the foundation for the ideal actuarial system. The good news is that such a tool can easily be built.

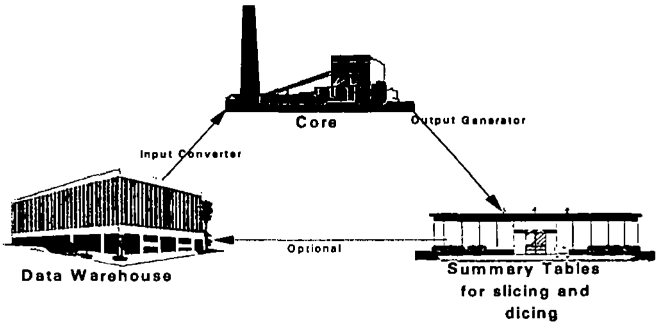
In an OODB, data objects are organized by a simple structure, most likely using the same descriptors that were used in the original Data Warehouse. The structure, if carefully designed, should meet the needs of most companies yet be expandable if necessary.

The only question that remains is how to retrieve the objects that were created during the actuarial process and stored in the database (for example, payout patterns or estimated IBNR)? Stated another way, how is reporting and summarizing accomplished?

Output Generator

For summarizing and reporting (Step Three), the advantages of data warehousing and OLAP technologies will again be needed. The Output Generator's task will be to break down the actuarial objects representing results (ultimates, IBNR, loss ratios) back into individual elements and create a traditional Data Mart to support decision making. Such an output generator would complete the linkage of the OBD storage system back to the data warehousing solution.

The resulting Data Mart could be used by itself or provide one more source for feeding the original Data Warehouse. As an example, the actuarial estimates of ultimate losses by line by state by accident year could be stored back to the Data Warehouse for subsequent use by field office management in reviewing profitability trends. Or, the estimated IBNR by annual statement line by state could be sent along to the financial reporting department for generating Page 14 of the statutory annual statement. In addition, the original Data Warehouse dimensions would not be changed by either the Input Converter, core OODB or Output Generator; therefore, such a task (feeding back results from the Data Mart to the Data Warehouse) would not be a problem.



Future Enhancements

There are a number of innovations that are emerging that may prove useful to the casualty actuary. One of the most interesting for the purposes we are discussing here is Data Mining (sometimes called Knowledge Discovery). The ideal data management solution we have outlined above would provide a transition platform into these upcoming advancements.

Data Mining

Data Mining is a new technology on the horizon with great actuarial potential. This area of database technology applications deals with the search for regularities not known prior to the search. It examines well-organized databases (like Data Warehouses) for clusters of similar data (and other patterns). Used properly, data mining may prove to be invaluable for determining homogeneous data subsets, which may provide clues to actuaries regarding the creation of new sub-lines or combining a few existing lines into one.

Flexible Structure

Future technology developments (Data Mining and other) may generate one more requirement for an "ideal" actuarial system: a flexible dimension-member structure, that has the ability to introduce (or delete) new dimensions or members (for example, new sub-lines) into the system.

Conclusion

The synchronization of several technologies that we have described in this paper is possible now. Once implemented, the "ideal" system empowers actuaries to increase productivity tremendously and (most importantly) boost creativity. The impact of such a system would be comparable to the impact of the introduction of computers, PCs and spreadsheets themselves.

CHAPTER 2 – QUESTIONS AND ANSWERS AT THE SEMINAR

Actuaries and other interest parties are attending the presentation on the above paper at the Ratemaking Seminar. The presenter's prepared remarks have concluded, and the floor is now open for questions and answers.

Integration of Pricing and Reserving

Audience: I work in the pricing unit at a primary insurance company. You mentioned how results of pricing work could potentially be recycled by the actuaries in the reserving unit. Our analysis is typically done by state, while the reserving actuaries tend to look at countrywide data. And besides, we look at total limits and basic limits data, while they mostly look at direct and net data. How do you think these are re-usable?

Presenter: What you have described are analyses that look at different slices of the grand Data Warehouse. If we accept the premise that the data are consistent and reconcile with each other, then your estimates of basic and total limits ultimate losses that you generate in your rate adequacy analysis should have some relationship to the estimates generated by the reserving actuaries. Now, for pricing, you don't stop at historical ultimates, but continue on and make adjustments for trend, underwriting actions, benefit levels, and so forth. But come back a step or two in your process. You've prepared estimates of basic and total limits historical ultimate losses. By state. Probably by coverage, too. At a minimum, wouldn't it be a good idea to compare your estimates on a total limits basis to the estimates generated by the reserving actuaries on a direct basis? If the sum of the states doesn't equal (or come close to) the countrywide estimate, there may be further investigation that's warranted.

Audience: Our company's vice president of underwriting has been persistent in asking us how we account for the shift in state mix in our countrywide reserves studies. We try to look for trends in the loss development factors, and give more weight to the recent periods, but it's really a judgment call. It might be worth trying to look at the data on a state by state basis, then add it up, and compare it with the estimate based on the review of countrywide data. But you know, we just don't have the time or resources to do all that work.

Presenter: That's a common lament of actuaries. They're being asked to do more work with either the same staff or even less staff. That leads to actuaries and analysts spending most of their time down-loading data, running macros, printing exhibits, and adding up the results for grand totals. The analysis phase always seems to be the one that gets squeezed.

So how do you respond? We suggest that you need to look for different ways of doing things. Automating existing tasks is only part of the solution. One area of efficiency is gained by not "reinventing the wheel" and by maximizing the flow of information from actuarial group to actuarial group. And by having efficient means by which to communicate your results to other departments and systems.

A major component of the solution is a common repository for actuarial data and results. A properly designed and implemented repository will remove barriers to information -- the key information that you need to do your jobs efficiently and effectively. However, the difference between a group of repositories and a single repository is the consistent basis by which the data is stored in the single repository. Consistency is one of the critical factor that enables actuarial efficiency and effectiveness and provides the ability to move to the next level.

Sharing of resources, information, and analysis leads to enhanced productivity and to a higher quality work product. Productivity is enhanced because needless re-doing of work is eliminated and quality is improved by increasing agreement and reconciliation within the actuarial group. So, coming back to the most recent comment from the audience regarding the extra work of doing reserves by state, I would suggest that you meet with your pricing colleagues to see how what they're doing could be leveraged in the reserving process.

Who has control?

Audience: Who has control over this system? The actuaries or the I/T people?

Presenter: That would depend on the corporate culture and organization. The solution leaves all the options open. If we consider the three main stages of the actuarial process, the calculate stage would seem to clearly fall to the actuaries. But the first stage could be within the actuarial group or I/T group, and the third stage within the actuarial group or a management group.

What do actuaries need to know to use these tools?

Audience: You've mentioned quite a few buzzwords and acronyms today. What does an actuary need to know to use them?

Presenter: We have mentioned a number of the buzzwords and their acronyms so that you are familiar with how they can help address the different demands of the actuarial process. It's important for the actuary user to know what that the tools exist and what they can do with them, rather than fully understanding how the tools do what they do. That's an area of expertise for the I/T professionals.

Recycling Reserving Results for the Next Period's Analysis

Audience: You've mentioned sharing between pricing and reserving within an insurance company. I'm a consulting actuary and performing regular reserving studies for my clients. I'd like to recycle the results of my last review and incorporate them as input items in my current review. For example, I'd like simple comparisons of prior estimates with current estimates. And, somewhat more difficult, I'd like to evaluate the components of the change -- actual experience versus my selections. How could I do that?

Presenter: Great question. To meet that request, your object database would need to include a time dimension. That is, valuation would be one of the fields, and the entries for the field would be, for example, September 1996, December 1996, and so forth. When you're ready to do the March 1997 review, you could design a range in the spreadsheet that looks to the database and retrieves the ultimates or selected factors from the December 1996 review. The spreadsheet would also be accessing the raw data for the March 1997 review, and ultimately sending back the "new" ultimates from the March 1997 review, whether they be the same or different that the ones from December 1996.

This offers a number of interesting possibilities. You could use last period's development factor selections and apply them to the current data; this would quantify the effect of any changes in the estimates due to changes in the data. Then, you can use current development factor selections to develop a current estimate, and evaluate the impact of change in factor selections.

By-products from pricing and reserving for business planning

Audience: In the introduction to your paper, you mentioned how actuaries provide certain items as input to the business planning process. Examples include projected loss ratios and reserve runoffs. These go beyond the traditional scope of pricing and reserving functions. Can you elaborate?

Presenter: Projected loss ratios within a five-year business plan, for example, frequently start with the current loss ratio, and rolled forward for expected impacts of rate changes and loss trends. The algorithm for deriving on-level factors for rate adequacy can be used again in this planning context; the rate change histories could include actual and expected rate changes, with an argument of the function describing the date for the projection. If the function generates only on-level factors, they are easily adjusted to earned impacts of rate changes.

As for cash flow projections of reserve balances, this is a natural by-product of a reserves study. The investment department is an interested recipient of the liability flows, for their investment management decisions. In fact, a schedule of expected IBNR and payments by calendar period creates one-half of a scorecard on the actuaries' estimates. As the actual data emerge, the second-half can be filled in and compared to the expected values. We've found that such projections, especially for IBNR, make these intangible actuarial estimates more "real."

Connecting rate adequacy indications to published financial results

Audience: In your introduction, you also mentioned reconciling rate adequacy indications and calendar period net income. Can you elaborate upon that?

Presenter: The example given was that an actuarial rate indications may suggest a countrywide deficiency of 15%, while the current financial statements show that the company is generating substantial net income. We see this situation as an apparent disconnection between the data and information generated by the pricing unit, the reserving unit, and/or the finance department. It may be that the reserving actuaries' estimates are more optimistic than the pricing actuaries. Or it may be that there was a very recent rate decrease which the pricing actuary is now saying was too much of a decrease. Or, it could be that the basic limits rates are deficient by 15%, which is what management heard, but the full story is that the total limits indication suggests that rates are reasonable, implying that the increased limits pricing is redundant. Or it could be that the reserving actuaries and pricing actuaries agree, but the finance department is recording reserves at a level less than the actuaries' indications.

There is no simple answer to the example as described. This highlights the need for everyone to read from the same page. We see the actuaries as being responsible for ensuring consistency among the various signals and answers that actuaries give each other and to company management. A common data repository or clearing house is a key component of the solution.

A spreadsheet? That's too simple.

Audience: You described something called an object modification language as being at the core of your system, and you said that a spreadsheet fit the bill. By that you mean a package like Lotus or Excel, right? That sounds too simple. When I work in spreadsheets, I enter in formulas cell by cell, which would be comparable to calculations on elements of objects, not the entire object at once. Can you clarify further how you think a spreadsheet serves this role?

Presenter: Certain spreadsheet packages can work with two-dimensional arrays as a single argument in typical cell-formulas. So if you have two ranges that are named, each of which is, say, a 10 by 10 triangle, you could create a third object, the quotient, for instance, by typing in range A divided by range B. Sort of like what the APL language could do. And spreadsheets can link to conventional programming languages to do some of the more complex data manipulation and calculations. For example, Excel has an ability to talk to and accept results from C++, Basic, Pascal, or even Fortran functions.

Why not just a DOS-based file storage system?

Audience: When I save a spreadsheet, I save the formulas and the data together, in one place. In your system, the data and results are stored separately from the spreadsheet files. How is this better?

Presenter: There are two immediate advantages to separating the algorithm from the data. First, you save on disk space for storage, because you avoid saving the algorithm over and over again. Second, and more importantly, modifications and corrections to the algorithm can be done in only one place, rather than over and over again.

Organizational structure of an object-based storage/retrieval database

Audience: You mentioned that the objects in an object oriented database are uniquely identified, by an object's place in the class hierarchy rather than being just the next available value as in traditional transactional databases. Can you tell me more about that? Just what does that mean for me?

Presenter: The objects are organized within a simple hierarchy, most likely using such descriptors that were used in the original Data Warehouse. For example, the fields might be line of business and state. Within the "line" field, the different entries may be personal auto liability, personal auto physical damage, homeowners liability, and so forth. Other fields would be used to describe "what" the object was; by that I mean, whether it was losses or expenses, either paid or reported, or amounts or counts.

Not just data, but also other things

Audience: Some of the objects, though, aren't really data. For our pricing studies, we need to include tables of rate change histories by state, and tables of expense. Plus, the profit loading comes from a simultaneous evaluation of cash flows in a total rate of return context. Where would these kinds of things fit into your solution?

Presenter: The database of objects does not only have to contain real data generated by business monetary transactions and the results of your analysis. The same database can contain a number of "miscellaneous" objects. Using the same conventions for describing the data objects, you can describe these other items. The table of rate changes could be viewed as a matrix of "n" rows and four columns, where the entries for each row might be the month and date of the change, the amount of the change, and a code to indicate whether the change applied prospectively or to all-inforce business.

The actuarial algorithm would need to anticipate this object coming from the database and be ready to perform the calculations needed to generate on-level factors. That would probably be done through a user-defined function, as it would probably get very messy using spreadsheet formulas only.

Similarly, you could also store factors for expenses, either in one table for all states or state by state. You also mentioned the need for a linkage to a rate of return model, for evaluating the profit loading. If I am correct in assuming that the rate adequacy model and the rate of return model are both implemented in a spreadsheet, there's nothing to stop the interaction between those models. In fact, if one of the elements of the rate adequacy model is loss development analyses of paid and reported losses for deriving ultimates, then the payment pattern could be directly tied into the rate of return model.

I still like APL

Audience: For hard-core actuarial work, I still prefer APL. Our company's I/T people tried to make me abandon all my APL mainframe programs in favor of working in spreadsheets, but I told them we couldn't survive without them. They just don't understand how many calculations I have to do. So, now I'm running with a PC version of APL.

Presenter: One of the actuaries who has worked with our system has actually described it as having "the flexibility and power of APL but in a spreadsheet." I think you'd be pleasantly surprised at what spreadsheets can do today. And if they can't do something that's particularly complex, you can link them to user-defined functions that can do the complex job for you.

Customized analysis for individual reinsurance contracts -- how does it fit the model?

Audience: I work at a reinsurance company, in the underwriting area. We do our analysis on a deal by deal basis; every deal is different. It's mainly long tail excess casualty business. We update our ultimate loss estimates quarterly, and then provide the reserves by contract to our accounting department. Each of the analyses is unique, and we have spreadsheets for each deal. We've been growing recently, so now the quarterly process of updating and compiling the estimates is beginning to overwhelm us. I don't think your system could help us. What's your reaction?

Presenter: You mentioned that each deal is unique, and that you have a customized spreadsheet for each. I would suspect, however, that a significant proportion of the basic analytical algorithms that are used is common, like evaluating historical claims development data or projecting future cash flows. The unique terms and conditions of the contract are the likely sources of the unique programming in the spreadsheets.

The objects of interest to the accountants are probably things like estimated IBNR, whether discounted or not. The objects that you aggregate may have different date attributes, but they share a common substance and form. It's IBNR as of a particular date. It can be uniquely described. And so, we could save those IBNR vectors or scalars, wherever they are in the maze of spreadsheet tabs, back to the object database. And then use aggregation tools for summarizing and reporting to accounting. If the deals are done in a number of currencies, the summarization could aggregate by currency, for feeding to an exchange rate conversion algorithm either in accounting or actuarial. The tools could produce not only the paper summary but also a data table that could feed into the general ledger system.

Actuarial management issues

Audience: I am the corporate actuary at our company, and our actuaries are spread out throughout the operating units, for both pricing and reserving work. The business units have managed to hold on to their own legacy systems, none of which generate actuarial data in a consistent format. Further, we have students and actuaries rotating to other departments every now and then. From my perspective, I need to ensure that the quality of actuarial work throughout the company is consistent. Plus, I need to have a mechanism to consolidate the actuaries' data and estimates from all over the company, for my own reporting to the board as well as for preparing the statement of actuarial opinion. I think what you've described can help me, but it's not clear exactly how.

Presenter: You've mentioned a number of troublesome areas. Let me briefly comment on each. First, you have multiple sources of data that generate data in different formats, and I would infer that the actuaries probably use different tools for their analyses. Am I on the right track?

Audience: Yes. Some actuaries still have their favorite APL programs that they wrote five years ago, some have dubbed themselves masters of Lotus macros, while others proclaim the wonders of Excel. It's a nightmare.

Presenter: I won't disagree with you on that. As for the multiple data sources, it would seem your information systems are ready for a major overhaul, with data warehousing technology being the most likely replacement. Once that is in place, the ability would exist for the summarizations and cross-tabulations of historical data that are needed for creating the actuarial objects of interest. These can be relayed and stored in one or more object-oriented databases on the LAN's throughout the company, so each actuarial group could have a customized database for their business unit.

Another issue you face is the multiple tools that the actuaries use. We know that actuaries are very proud of the applications and model that they build. And rightly so. The trick is to find a toolkit that is powerful enough to do the calculations and yet flexible enough for the actuaries to customize the application and model to the project at hand. I suggest to you that a spreadsheet package, with a library of customized actuarial functions, can satisfy those requirements. This consistent toolkit would also address the "learning curve" issues that arise from periodic movement of students and actuaries throughout the company.

You also mentioned a need to consolidate the actuarial data and estimates from across the company. Summary files, created by the output generator, using data warehousing-type tools, could be produced by each area, and then consolidated at the corporate level.

Is this fiction or reality?

Audience: This all sounds well and good, and you've talked a nice game here, but "where's the beef?" Is this all vapor-ware? What can you show us?

Presenter: Let me turn my laptop PC on . . . . .