Event-Driven Database Maintenance:

A Process for Increasing Value and Lowering Costs

Table of Contents

White Paper

by Matt Manning
President
Information Evolution Inc.

Last Updated: April 2008

Overview

Currently accepted best practices do not maintain databases cost-effectively. Update methodologies are typically based on legacy processes that haven't changed much apart from the addition of online self-updating mechanisms, which often yield marginal results and take a lot of time and effort to clean up. These processes are commonly scheduled as "stop and start" batch efforts tied to print publication schedules or, increasingly, to other annual or periodic events. The records to be updated are also often not prioritized in any sophisticated way, so that records that should be really be treated quite differently often receive the same amount of attention and are updated using the same type of research methodology.

Given that online end-users increasingly demand deep, accurate, and timely data, it is critical that database owners wake up to the reality of the changed marketplace and reengineer their processes to ensure that they deliver more value.

Given that online end-users increasingly demand deep, accurate, and timely data, it is critical that database owners wake up to the reality of the changed marketplace and reengineer their processes to ensure that they deliver more value. This doesn't mean increased expense—in fact it can often mean the opposite—but is does mean change.

This white paper explores the importance of database maintenance, current best practices deployed to meet that need, and an alternative process designed to increase end-user value and reduce operating costs.

The Need for Database Maintenance

Databases "age" on average at a rate of 20 percent per year. While most database managers are well aware of this rule of thumb, it is important to delve deeper into what causes data to change and which records are most likely to change. Some subsets of records have at least one field change every year. Other records only change every few years.

Events

The events that cause a record to require updating fall into two general categories, routine and major.

Routine events include:

- Personnel updates
  - a new job title for an individual listed in the record
  - a new hire
  - a new individual in an existing position

- Product/service changes

  - a new brand/product offering

  - expansion/change in products/services offered

  - divestment of brands/operating units

- Branch offices/distributors/sales reps

  - new branch offices

  - new partner firms

Major events include:

- Ownership changes

- Headquarters location changes

- Company name changes

- Executive leadership changes

- Contact information changes

Because large entries in databases have more personnel, more branches, more products, and generally *do* more, they are far more likely to require more frequent updates. Prominent records, like biographies of Nobel Prize winners or company data on venture-backed start-ups in high-growth industries, are also far more active than other records and demand more editorial attention.

Large or prominent records only make up less than 20 percent of a database on average, but they have a high value to end-users because of their large budgets and constant activity. They also age far more rapidly than other records because they have more frequent "routine" changes. For company records this means personnel changes, product/service changes, new or changed branches, divisions, plants, and distributors. For biographic records this means new accolades, publications, presentations, and added board memberships. In any case, all these routine changes are extremely important to end-users. Reaching a new employee before the competition has done so is critical to a marketer's success and presentations or reports based on old biographical data are a source of embarrassment.

"Regular" records may not change from year to year, but when they do, the event is more likely to be something major such as a closure or a move for a company, or a death for a biography. Regular company

Large or prominent records only make up less than 20 percent of a database on average, but they have a high value to end-users because of their large budgets and constant activity.

records are often long-standing operations that need a range of services, or they are growing operations that may be hiring and making many different types of purchases. Because of this, they have a high value to many customers. Regular biography records are often tenured professors, incumbent politicians, and senior executives who might not do much that is noteworthy in a given year but who make major decisions and wield substantial influence nonetheless. When run-of-the-mill records change, the change is usually major and customers need to know what has happened as soon as possible. Timely action prevents end-users from spending marketing dollars on dead records or cluttering up CRM databases with obsolete information.

Of course, a substantial percentage of regular records don't change at all over the course of a year. This can be upwards of 25 percent depending on the database, and it is critical that these records are left to last and given a smaller share of the database maintenance budget.

Current Best Practices in the Database Maintenance Process

Maintenance, meaning updating and enhancement, of databases of company, biographic, and product information usually are handled as follows:

- *Frequency:* annual; project length of 3-9 months

- *Record selection:* batches "tiered" with priority records (VIPS, oldest records) updated first and more frequently

- *Resources:* outsourced

- *Processes:* multi-stage

   - Email/web "self-updating"

   - Internet research

   - Telephone research
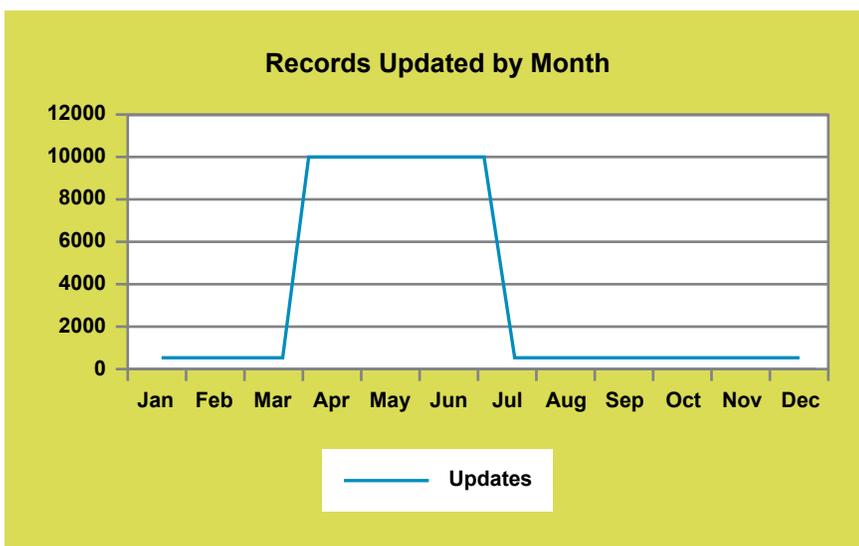
Frequency

An annual updating cycle is a legacy of the print production schedules intended to achieve "just-in-time" updating right before packaging and delivery of the content to the client. A typical update schedule looks like this for a database traditionally printed in September.

Records Updated by Month

12000	
10000	
8000	
6000	
4000	
2000	
0	Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

Updates

Record Selection

Prioritization of the records to be updated is logical because limited resources should be devoted to records with the most value to end-users. Most prioritization schemes use a combination of a record's popularity and age.

Popularity is usually measured by the "size" of the record (i.e., a large company with lots of people, products, and other data; a prominent person with a deep records of accomplishments), although savvy online publishers also look at usage data to see which records are the most accessed (i.e., private companies; people or companies in high-growth industries or geographic regions). Age is easily measured by the record's last date of updating.

Resources

Research work is increasingly expensive to conduct with resources in "Western" economies. Good talent is hard to find and expensive. There is often a high rate of staff turnover. Seasonality of publishing projects can make human resource management challenging; full-time employees can be difficult to justify for projects that aren't active year-round and it is not easy to locate, train, and retain high-quality part-time talent.

Processes

Email/Web Self-Updating: This has been in use for several years now, although "self-updating" is a misnomer because user-generated content still requires an editor to make sure content conforms to editorial standards before approval. Response rates to requests for listee-provided information have also dropped substantially as spam filters increasingly block deliveries.
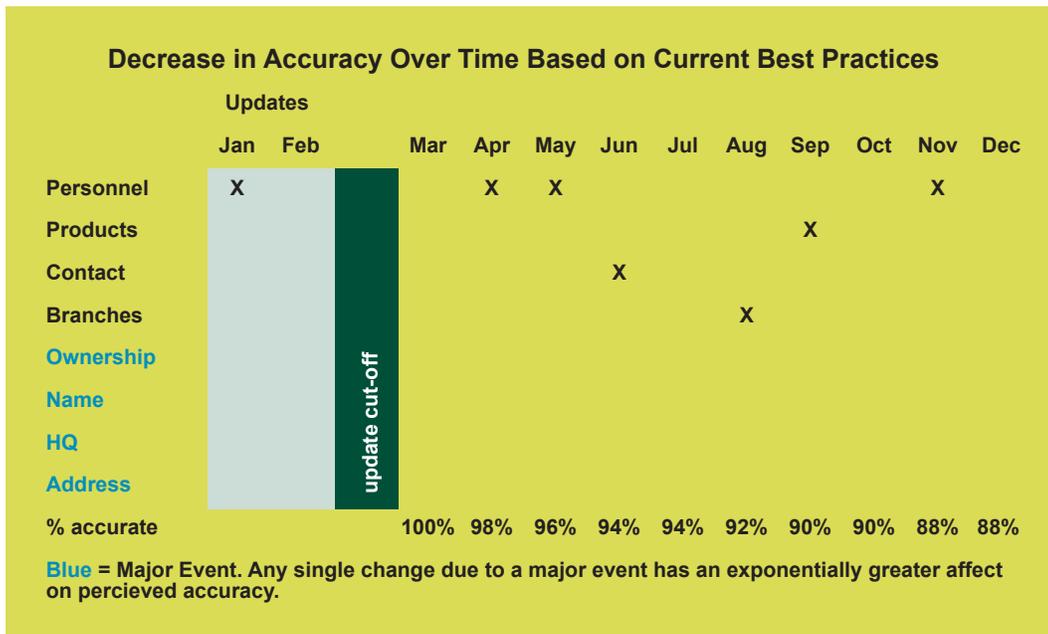
Internet Research: A labor-intensive approach that is more and more successful as the depth and reliability of primary source content on the web increases, although the risk of gathering obsolete data remains a major consideration. It is not effective for certain data points, for example mid-level management contacts, especially for IT positions or highly personal bio data, and often requires the use of premium sources, custom harvesting routines, and extensive quality control checks.

Telephone Research: The most expensive way to update a record, but the only alternative for specific types of records (those with telephone but no URL; older records; records requiring contact names). Because of cost and quality control issues, it is best used sparingly as a targeted tool.

An Alternative Database Maintenance Process

Improving the current methodologies used to update databases involves the reallocation of resources and fundamental reengineering of current processes. Competition from free services and the increase in the price of premium information services requires the managers of those premium services to provide a higher level of quality (timeliness, depth, accuracy) or they risk losing their client base.

Frequency

Decrease in Accuracy Over Time Based on Current Best Practices

Updates	Jan	Feb	update cut-off	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Personnel	X				X	X						X	
Products										X			
Contact							X						
Branches									X				
Ownership													
Name													
HQ													
Address													
% accurate				100%	98%	96%	94%	94%	92%	90%	90%	88%	88%

Blue = Major Event. Any single change due to a major event has an exponentially greater affect on percieved accuracy.

Annual updating schedules are inherently flawed given the seismic shift from print to online and must be migrated to year-round schedules. This change also solves problems related to short-term update processes: managing on-again, off-again research teams; poor quality at the start of processes; and the long lag time between events and when the change is reflected in the database.

Costs of not reflecting major event updates in database products include:

- Higher customer service costs

  - Fielding incoming complaints

  - Issuing credits and refunds

- Lower renewal rates

- Higher cancellation rates

- Bad publicity and word-of-mouth

These are significant factors even taken in isolation, but combined they are too potentially damaging to ignore.

Record Selection

The current types of record prioritization are very useful, but still leave customers with out-of-date high-profile records and dead records.

Prioritization needs to be based on making updates as soon as possible after a change occurs. The best way to get close to "real-time" updating is to set up a series of "monitors" that check news sources and specific databases like the Social Security Death Index, which reports deaths, or new business filings from the State of Delaware every day. Incoming data then needs to be vetted, edited, and pushed live.

Costs of not reflecting major event updates in your database product are:

- Higher customer service costs

  - Fielding in coming complaints

  - Issuing credits and refunds

  - Lower renewal rates

  - Higher cancellation rates

  - Bad publicity and word-of-mouth

Resources

Year-round updating means that researchers can be hired based on their domain-specific experience and they can be allowed to develop even deeper knowledge of their particular data universe over time.

The use of offshore resources is imperative but requires careful training, supervision, systems, and quality control. Batch processes don't allow enough time for resources to become familiar with the data and this is especially true with offshore teams that frequently bounce from project to project. Year-round updating means that researchers can be hired based on their domain-specific experience and they can be allowed to develop even deeper knowledge of their particular data universe over time.

Processes

Email/Web Research: This tool can be a dangerous distraction and can foster unrealistic expectations in terms of cost, response rates, and accuracy. The most effective use of this mechanism involves emailing records one at a time, including the record to be updated in the body of the email (not as an attachment or accessible via a hyperlink), and giving the recipient a simple "reply to this email" option.

Internet Research: An increasingly valuable source of information, especially when combined with automated web spiders that gather information prior to the human research effort.

Telephone Research: This tool is as effective as your product's brand. If people want to be listed in your database because they respect your brand name and see a tangible value to cooperating with the research process,

telephone research can be very effective. It's best to focus telephone researchers on confirming data, especially contact names, not adding new data. Using the telephone to confirm dead records is also very effective.

Conclusion

In order to deliver valuable information to their end-users, database managers need to set up event-driven alert mechanisms, develop editorial management systems to allow offshore research teams to work year-round to interact with the incoming data, add quality control mechanisms with detailed "audit trails" of all changes made, and client-side quality control mechanisms. All of this can be built in-house or commissioned. It can be integrated with or coexist alongside existing editorial management systems.

However it's done, it is critical that content owners take these steps to turn their attention to the truly challenging issues of replacing dead records, creating new records, and adding depth and value to their databases.

About the Author

Matt Manning is the president of Information Evolution, Inc., a firm that designs and implements efficient research and editorial processes for content companies.

About Information Evolution, Inc.

IEI provides human resource and technology services to companies, primarily in publishing or related industries, that manage large databases in real time. For more information, call (512) 650-1111 or visit www.informationevolution.com.