The Half-Life of Information at GIAC Enterprises

Group Discussion Written Project (GDWP)

Authors: Charlie Scott, cscott@infosec.utexas.edu
Stephen Strom, steve@stevestrom.com
Advisor: Stephen Northcutt

Presented: November 14, 2010

1. Executive Summary

Data is the lifeblood of GIAC Enterprises, and as such must be both protected and managed. The growth of GIAC Enterprises has made it increasingly apparent that some data loses usefulness with the passage of time. This loss of usefulness relative to time has been referred to as the "Half-life of information" (Burton & Kebler, 1960).

Managing data from the perspective of a half-life requires changes in the structure of our data management system. Up to this point in time, GIAC Enterprises has relied upon a relational database system, and this system has served the company well. But managing data according to a half-life model necessitates relating data from the perspective of time change. Such a time-related or temporal management model will necessitate design change to provide new capabilities. It is this requirement which has led the IT Strategic Planning Committee to embark on an initiative to transition our current database to a temporal XML data model.

This paper will begin by examining the concept of information half-life and determining some of the factors leading to temporal decay of data value. A data classification system will be presented, along with development of a metric to determine data half-life. The paper will proceed to present an initial XML container definition (which is SOAP 1.2 compliant) to support this temporal data model, as well as consider how temporal data management may impact the corporate strategic planning cycle.

It is noted through this document that the temporal XML transition initiative will have additional corporate benefits, notably in the ability to tune our existing DLP product and to look to future implementation of an XML gateway.

2. What is Information Half-Life?

The concept of information half-life, as coined by Burton & Kebler (Burton & Kebler, 1960), has its origins in the physical sciences. Originally used to describe unstable atoms, half-life was defined as the time it takes for a substance undergoing decay to decrease by half (Half-life).

When applied to information, half-life is defined as the time it takes for information to lose half of its value. For the purposes of GIAC Enterprises, we will define how to determine data half-life later in this paper.

Several key principles must be understood as GIAC Enterprises proceeds to develop a temporal data management model:

2.1.1. Data is worth less as it ages

As a general rule, data becomes less valuable to GIAC Enterprises as that data ages. For much of our corporate data, that aging will follow a general decay curve similar to that of the half-life of an isotope.

Other data loses value swiftly at a particular point in time. A case in point would be the value of cookie fortunes that are linked to a particular animal sign in the Chinese calendar. Animal signs are in a 12-year cycle used for dating the years. When a particular calendar year is over, any fortunes based on that animal sign become worthless until the 12-year cycle returns that animal sign to the calendar. This paper will refer to such data that is bounded by a particular time as "time delimited data."

2.1.2. Data value decays at different rates

An evaluation of GIAC Enterprises' data shows that while virtually all data decays in value, not all data decays in value at the same rate. As a case in point, the value of a particular customer list will not decay at the same rate as the value of procedures related to printing cooking fortunes. Understanding the factors that cause data value decay will help determine the rates at which such decay occurs.

2.1.3. Different factors impact the value decay of various categories

It also became apparent during the data evaluation that different categories of data are influenced by different factors leading to data value decay. The value of cookie fortunes again illustrates this principle. Fortunes that are of a general nature are more long lasting in value than calendar based fortunes.

Yet even these general fortunes suffer from factors that lead to data value decay. For example, the act of placing a general fortune into circulation begins the process of data value decay since it is only a matter of time until a regular customer receives a duplicate of that general fortune. Research has shown that customers receiving duplicate fortunes are not satisfied customers. Since customer satisfaction is a top priority at GIAC Enterprises, it is possible that temporal data management can increase customer satisfaction ratings.

Competition can also cause data value decay. It is a documented fact that certain competitors copy the cookie fortunes produced by GIAC Enterprises. As these counterfeit fortunes enter circulation, their very presence increases the probability that a customer will encounter a repeat fortune, thereby leading to decay in the value of that fortune data and resulting in customer dissatisfaction with GIAC Enterprises.

2.1.4. Information half-life is a key metric

Because of the fact that data does decay in value, it is important to try to quantify the length of that information half-life. Knowing the half-life of information is important from the perspective of managing the enterprise. Making key business process decisions using data that is value decayed will not help GIAC Enterprises perform at top efficiency. Today's competitive environment demands agile decisions based on current data. Knowing the half-life of such data will help GIAC Enterprises remain at the top of our industry.

2.1.5. Half-life varies for different data categories

Determination of information half-life will vary depending on the type of data

being evaluated. Half-life will both vary in how it is determined, as well as across data types. Because of this variation, GIAC Enterprises will need to develop flexible metrics that are effective in a number of situations and under varying conditions.

3. Factors that influence information half-life

Four factors have been identified as key components by which to evaluate information half-life. These factors were identified by a panel of GIAC Enterprises information consumers as the most important qualities they use to determine the value and relevance of data.

The panel also addressed how these factors could be determined and quantified. This notation will be included in each factor summary. It should be noted that not all factors will impact each category of data.

3.1. Data usefulness

The information consumer panel recognized that much data declines in usefulness as time progresses. Whether it is the freshness of cookie fortunes or the currency of a prospect list, the value of such data declines over time.

Just how fast it decays in value is often a subjective evaluation. Because it is a subjective analysis, it is the opinion of our panel that the best judges of the usefulness of a particular type of data are the data users themselves. These users are experts on both the use and usefulness of a particular data category, so we will convene these panels and quantify their expert judgment as an input to the information half-life calculations.

Some of the data controlled by GIAC Enterprises is subject to regular change. For example, the prospect lead list changes regularly with time, as does the current customer list. The Corporate Research arm of the corporation has indicated they can provide a subjective statistical analysis of the rate of that change for various data categories. This value is generated annually. For instance, if the current distributor list has a turnover of 30% each year, then the 30% annual would be the rate of change. Any given iteration of the distributor list would have 30% incorrect data in a year. This would give a rate of change value of 0.30. If the distributor list turned over twice for whatever reason, the rate of change would be 200% and the value would be 2.00.

Once that rate of data change is known, our data expert panel can use this information as an input, in addition to their opinion on usefulness, to come up with a usefulness metric.

3.2. Timeliness

Certain data categories are tied to specific periods of time and are more valuable during those periods of time. For example, cookie fortunes tied to the Chinese calendar year are primarily useful only during the calendar year for which they were produced. Such time delimited data gradually increases in value as the time delimited period approaches, peaks during the period itself, and then "bottoms out" rapidly when the

period is over.

Other data categories are linked to time, but not specifically bounded by time. The value of this data decays fairly evenly as time passes, and not in the same time bounded fashion as time delimited data. Both types of relevance loss will be calculated using the expert judgment of the corporate data consumers.

3.3. Loss to Competitive Advantage

The cookie fortune business is extremely competitive, and competition seems to grow every quarter. The most important factor in strengthening our market competitive advantage is the currency and value of our data. While various categories influence our competitive advantage in different ways, data currency is still the key to our market leading position.

Corporate Research has documented the rate at which our competitive advantage is eroded by various factors. These factors include, but are not limited to, the following:

- Changing customer preferences
- Corporate espionage
- Trade secret theft
- Insider data leakage

Because these erosion factors have been carefully studied and statistically quantified, it is possible using expert judgment to assign a value related to the Rate of Loss of Competitive Advantage to each category of data. This value then becomes an additional input to the information half-life calculations.

4. Data Classification

Analysis of the data held by GIAC Enterprises shows that a number of traits are possessed by that data. For the purposes of moving to a temporal XML data model, data has been delineated by three characteristics. While a fuller discussion of the data classification model can be found in Appendix A, it is worth presenting a summary of that model here.

Category is a marker that defines the general type of data that is involved. Some examples of Category would be Accounting data, HR data, and Time-Bounded Fortune data.

Each category of data is assigned a **Classification** rating. This classification system ranks data according to how that data should be handled within GIAC Enterprises. The basic classifications of Proprietary, Sensitive, and Restricted should meet business needs for the foreseeable future.

Timeliness is the final characteristic used to define the enterprise data. This is determined through the expert judgment of the corporate data consumers, and used to mark the XML schema for the associated data.

Table 1: Data Category, Classification, and Timeliness

Category	Classification	Timeliness
Time-Bounded Fortune	Proprietary	Very Timely
General Fortune	Proprietary	Timely
Production Procedures	Proprietary	Timely
Accounting	Sensitive	Timely
Human Resources	Sensitive	Timely
Customer List	Restricted	Timely
Customer Leads	Restricted	Somewhat Timely
Business Contacts	Restricted	Not Timely

5. Half-Life Metrics

From the preceding analysis of the factors that influence the half-life of GIAC Enterprises' data, it is now possible to develop a half-life metric for each category of data held in the corporate data warehouse. A detailed discussion of metric development can be found in Appendix B, with a summary of the important concepts behind this metric presented below.

Half-life calculation begins with the assumption that while all data has value, not all data is equally valuable. This value is known as the Qualitative Data Value (QDV), and will vary over time. If we begin by calculating the initial QDV (known as QDV_0), then recalculate the QDV each day over a period of time, we will be able to determine the point at which $QDV_n = QDV_0 / 2$. The number of days it takes to reach this point is known as the half-life. Such a methodology will enable GIAC Enterprises to empirically determine the half-life metric for each category of data.

It should be mentioned that the complete metric takes into account the time delineated nature of certain data. When such time delineated data exits the bounds in which it is valid, QDV will automatically go to zero, thereby noting the data as no longer valuable.

6. XML Schema

Appendix C contains the initial XML container schema for this temporal data model. While the XML given only contains elements related to evaluating the half-life of information here at GIAC Enterprises, converting to XML also brings with it a number of other advantages.

The schema definition as presented is compliant with SOAP 1.2. In the event that the company decides to implement a SOA structure, the data schema will already be in place to support such a deployment. This means that resources would not have to be re-allocated to bring the temporal XML data model into compliance with SOAP 1.2.

The use of an XML data model will also allow GIAC Enterprises to make better use of our existing Symantec DLP solution. XML will allow better tuning of the Symantec tool to reduce both false positives and false negatives, thereby maintaining better control over company proprietary information. Such tuning should also help reduce the load on the IT staff, since they won't have to spend so much time responding to false DLP alerts.

A final benefit of transitioning to an XML data model lies in enabling the future use of an XML Gateway. Figure 1 below shows a typical XML Gateway deployment. Transitioning our data now to an XML data model means that when the decision is made to deploy an XML Gateway, this data conversion will be already accomplished and will not have to be reallocated.

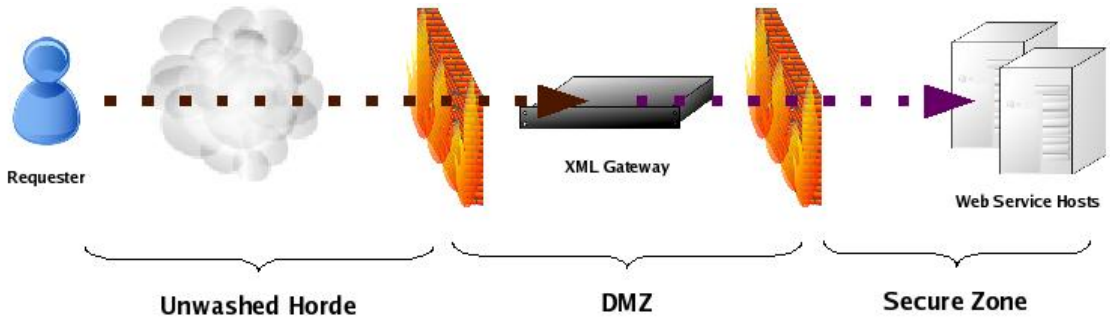

Figure 1: Typical XML Gateway Deployment View (Are XML gateways really the answer?)

7. Strategic Planning Cycle Impact

Strategic Planning has always been an important process at GIAC Enterprises. While various methodologies exist for determining when to begin a strategic planning cycle, one possibility is to link the planning cycle to the half-life of information.

With that in mind, a possible rubric based on information half-life is presented in

the following table. Such a rubric will obviously need tuning by senior administration. But a good initial starting point might be to initiate a strategic planning cycle when the total score for all categories of data is equal or greater than 24. This value in effect says that the current corporate data is declining in value, and a new planning initiative might be needed to refresh the corporate vision.

Table 2: Strategic Planning Rubric

Category	1	3	5
Time-Bounded Fourtune	Category has not yet reached its half-life.	Category has reached its half-life.	Category is beyond its half-life.
General Fortune	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Production Procedures	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Accounting	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Human Resources	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Customer List	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Customer Leads	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.
Business Contacts	Category has not yet reached its half	Category has reached its half-life.	Category is beyond its half-life.

8. Conclusion

Conversion of data from the existing relational database to a temporal XML data model presents many exciting possibilities for GIAC Enterprises. Using that temporal model to evaluate the half-life of information in the corporation likewise opens up many possibilities ranging from customer satisfaction to corporate planning.

While the details of such a project have not been fully determined, an initial project plan can be found in Appendix D.

Undoubtedly there are risks associated with such a major transition. But we hope that the elements of this paper will help the company move forward with this plan, as it promises to greatly enhance the ability of GIAC Enterprises to remain competitive in a fast-moving environment.

9. Bibliography

Alleman, G. (n.d.). *Risk matrix.* Retrieved November 14, 2010, from Herding cats: http://herdingcats.typepad.com/my_weblog/2010/07/risk-matrix.html

Are XML gateways really the answer? (n.d.). Retrieved November 14, 2010, from http://atownley.org/2006/06/are-xml-gateways-really-the-answer/

Burton, R. E., & Kebler, R. W. (1960, January). The "half-life" of some scientific and technical literatures. *Journal of the American Society for Information Science and Technology, 11*(1), 18-22.

Changing values in the published literature with time. (n.d.). Retrieved November 14, 2010, from The Free Library: http://www.thefreelibrary.com/Changing+values+in+the+published+literature+with+time.-a013213267

Container element. (n.d.). Retrieved November 14, 2010, from XML design patterns: http://www.xmlpatterns.com/ContainerElementMain.shtml

Data classification. (n.d.). Retrieved November 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Data_classification_%28data_management%29

Data classification toolkit. (n.d.). Retrieved November 14, 2010, from https://wiki.internet2.edu/confluence/display/itsg2/Data+Classification+Toolkit

Farmer, R., & Glass, B. (n.d.). *Time decay in reputation systems.* Retrieved November 14, 2010, from Building web reputation systems: the blog: http://buildingreputation.com/writings/2009/09/time_decay_in_reputation_syste.html

Frisk, J. (2009). *Project management and effective communication for security professionals and managers.* Baltimore: SANS.

Gao, D., & Snodgrass, R. (2003). Temporal slicing in the evaluation of XML queries. *Proceedings of the 29th VLDB Conference*, (pp. 1-12). Berlin.

Half-life. (n.d.). Retrieved November 13, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Half-life

Hertzel, D. (2003). Bibliometrics history. In M. Drake (Ed.), *Encyclopedia of library and information science* (Vol. 1, p. 318).

Information obsolesence and information half-life. (n.d.). Retrieved November 14, 2010, from http://www-ec.njit.edu/~robertso/infosci/obsolescence.html

Library journal use and citation half-life in medical science. (n.d.). Retrieved November 14, 2010, from Microsoft Academic Search: http://academic.research.microsoft.com/Paper/826487.aspx

Life cycle management manual. (n.d.). Retrieved November 14, 2010, from United States Patent and Trademark Office: http://www.uspto.gov/web/offices/cio/lcm/lcm.htm

May, J. (n.d.). *Does information have a half-life?* Retrieved November 14, 2010, from https://blog.coremedia.com/cm/post/1751367/Does_information_have_a_halflife.html

Mendelzon, A., Rizzolo, F., & Vaisman, A. (2004). Indexing temporal XML documents. *Proceedings of the 30th VLDB Conference*, (pp. 1-12). Toronto.

Newsfeeds, the half-life of information, and why we need Foursquare. (n.d.). Retrieved November 14, 2010, from http://channelyush.com/newsfeeds-half-life-of-information-usefulness-of-information-over-time-foursquare/

Obsolescence. (n.d.). Retrieved November 14, 2010, from http://www.iva.dk/bh/Core%20Concepts%20in%20LIS/articles%20a-z/obsolesence.htm

Overcoming obstacles to data classification. (n.d.). Retrieved November 14, 2010, from Computer Economics: http://www.computereconomics.com/article.cfm?id=1117

PEST analysis. (n.d.). Retrieved November 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/PEST

Porter's five forces. (n.d.). Retrieved November 14, 2010, from QuickMBA: http://www.quickmba.com/strategy/porter.shtml

Risk matrix. (n.d.). Retrieved November 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Risk_Matrix

SOAP version 1.2. (n.d.). Retrieved November 14, 2010, from W3C: http://www.w3.org/TR/soap12-part0/

SWOT analysis. (n.d.). Retrieved November 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/SWOT_analysis

Temporal database. (n.d.). Retrieved November 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Temporal_database

Temporal XML? SQL strikes back! (n.d.). Retrieved November 14, 2010, from UCLA Computer Science: http://www.cs.ucla.edu/~zaniolo/papers/time05.pdf

Wang, F., Zhou, X., & Zaniolo, C. (n.d.). *Using XML to build efficient transaction-time temporal database systems on relational databases.* Retrieved November 14, 2010, from http://www.csie.ntu.edu.tw/~tempdb2007/xml_build.pdf

XML schema tutorial. (n.d.). Retrieved November 14, 2010, from w3schools.com: http://www.w3schools.com/Schema/default.asp

10. Appendix A: Data classification

Data Classification, Categorization, and Timeliness Heat Map

Category	Classification	Timeliness
Time-Bounded Fortune	Proprietary	Very Timely
General Fortune	Proprietary	Timely
Production Procedures	Proprietary	Timely
Accounting	Sensitive	Timely
Human Resources	Sensitive	Timely
Customer List	Restricted	Timely
Customer Leads	Restricted	Somewhat Timely
Business Contacts	Restricted	Not Timely

DEFINITIONS:

Proprietary: Data loss would cause direct, exceptional damage to the company's business operations or revenue.

Sensitive: Data loss would cause direct damage to the company's reputation and lead to legal action due to pertinent laws or regulations.

Restricted: Data loss may cause indirect damage to the company's business operation or revenue, or may be recoverable.

11. Appendix B: Half-life metric development

The half-life metric is designed to take into account variations in company data while being easy to administer. Development relies heavily on the corporate expertise of both data consumers as well as the corporate research department. The model itself was built on the risk metric techniques used in project management frameworks (Frisk, 2009). For example, in project management risk metrics expert judgment is used to assign risk based on the opinions of experts in various arenas. In this model for calculating the value of information, the expert judgment of the data owners and consumers are used to determine the usefulness, competitive advantage, and timeliness of the data.

In order to better illustrate the inputs and the applicability of the metric, two data categorization examples are used throughout the discussion of the development process. The first is a Time-Delimited Fortune that is specific to the Year of the Rabbit in the Chinese calendar: *You will live in interesting times during the Year of the Rabbit.* The second is the fortune cookie production procedures, which falls under the Production Procedures data category.

The first input that is calculated is the ***Data Usefulness*** value. This is an initial starting value showing the relative usefulness of the data in question. By using a ranking register as shown below, the expert panel is able to assign a relative initial value to the various categories of corporate data.

Data Usefulness

Usefulness	Value
Extremely Useful	4
Very Useful	3
Somewhat Useful	2
Not Useful	1

GIAC's business experts in fortune cookie marketing and production have determined that the usefulness of the Year of the Rabbit fortune is "Extremely Useful," giving it a value of 4. The fortune cookie production procedures used, however, is only determined to be "Somewhat Useful," giving it a value of 2. The fortunes are unique and created in-house, whereas the production is done with well-known, traditional methods. The annual rate of change for both is 0%, or 0.0, so that is not a factor in determining usefulness.

Loss to Competitive Advantage is likewise determined by the expert corporate data consumers. This is also a relative ranking system designed to weigh the impact of losing the competitive advantage for a particular category of data. Using a ranking register as shown below, an initial Loss to Competitive Advantage value can be assigned to each category of data

Loss to Competitive Advantage

Loss	Value

High Loss	4
Medium Loss	3
Low Loss	2
No Loss	1

The expert corporate data consumers have determined that the loss to competitive advantage for the Year of the Rabbit fortune would be of "High Loss" (value of 4), since it is a key piece of intellectual property. The fortune cookie production procedures would represent "No Loss" (value of 1) because it is essentially the same traditional method used by many of our competitors.

Timeliness is the final value to be supplied by our expert data consumers. This will be an initial value showing how timely the data is determined to be at the time of initial half-life calculation. Using a ranking register as shown below, an initial Timeliness value can be assigned to each category of data.

Timeliness (Expert Judgment)

Timeliness	Value
Very Timely	4
Timely	3
Somewhat Timely	2
Not Timely	1

The Year of the Rabbit begins in February 2011. It is currently November 2010, within a few months of the fortune's delimited window, so the data is judged to be "Timely" (value of 3). Because the fortune cookie production method is a traditional one that GIAC has used for a long time, the timeliness of the data is considered to be "Not Timely" (value of 1).

From all of the inputs described, it is possible to calculate a ***Qualitative Data Value*** for each category of data. This metric is determined according to the following formula:

Qualitative Data Value = ((**Data Usefulness**) X (**Loss to Competitive Advantage**) X (**Timeliness**))

If the initial Qualitative Data Value is denoted as QDV_0, then the information half-life is the number of days it takes for $QDV_n = QDV_0 / 2$. Where *n* is the number of days.

Calculating the Qualitative Data Value each day for each category of data will allow GIAC Enterprises to empirically determine the number of days it takes, by category, for QDV_n to equal $QDV_0 / 2$. This day value is the ***half-life*** for each category of data.

The initial qualitative data value for the Year of the Rabbit fortune would be as follows:

$QDV_0 = (4 \text{ X } 4 \text{ X } 3) = 48$

The initial qualitative data value for the fortune cookie production procedures is:

$QDV_0 = (2 \text{ X } 1 \text{ X } 1) = 2$

Therefore, the half-life formula of the Year of the Rabbit fortune is:

$QDV_n = 48 / 2 = 24$

The half-life formula of the fortune cookie production procedures is:

$QDV_n = 2 / 2 = 1$

When $QDV_n = 24$ for the Year of the Rabbit fortune, it will have reached its half-life and GIAC will know the number of days (*n*) that it takes for the Time Delimited Fortune data category to reach its half-life. Likewise, when $QDV_n = 1$ for the fortune cookie production procedures, then the data will have reached its half-life and GIAC will know the number of days (*n*) that it takes for the Production Procedures data category to reach its half-life.

12. Appendix C: XML Schema

Schema example as a container for company data:

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

      <xsd:annotation>
            <xsd:documentation xml:lang="en">
            Data Categorization Schema for GIAC Enterprises
      </xsd:documentation>
      </xsd:annotation>

<xsd:element name="CompanyData" type="CompanyDataType"/>

<xsd:complexType name="CompanyDataType">
      <xsd:sequence>
          <xsd:element name="Data" type="xsd:string"/>
            <xsd:element name="DataCategory" type="xsd:string"/>
            <xsd:element name="DataClassification" type="xsd:string"/>
            <xsd:element name="ValueWindow" type="ValueWindowType"/>
            <xsd:element name=”Timeliness” type=”xsd:integer”/>
      </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="ValueWindowType">
      <xsd:sequence>
            <xsd:element name="BeginDate" type="xsd:date"/>
            <xsd:element name="EndDate" type="xsd:date"/>
      </xsd:sequence>
</xsd:complexType>

</xsd:schema>
```

Example of using the schema:

```
<CompanyData>
      <Data>
            You will live in interesting times during the Year of the
            Rabbit.
      </Data>
      <DataCategory>Time-Bounded Fortune</DataCategory>
      <DataClassification>Proprietary</DataClassification>
      <ValueWindow>
            <BeginDate>2011-2-3</BeginDate>
            <EndDate>2012-1-22</EndDate>
      </ValueWindow>
      <Timeliness>4</Timeliness>
</CompanyData>
```

13. Appendix D: Project Plan

Transitioning to a temporal XML data model requires a lot of planning and forethought. While not complete, the following outline presents some initial project planning around relevant tasks, milestones, resources and schedule. Note that this is not a well-developed plan, but simply serves as an initial starting point to further apply project management principles.

1. Tasks

- 1.1. Clarify data categories
- 1.2. Define factors that influence half-life
  - 1.2.1. Establish expert panel to determine data usefulness
  - 1.2.2. Establish expert panel to determine timeliness
  - 1.2.3. Compile statistical data on rate of change analysis for data categories
  - 1.2.4. Define aging boundaries for time delimited data
  - 1.2.5. Gather statistical data on rate of loss of competitive advantage
- 1.3. Categorize data
- 1.4. Define half-life metrics
  - 1.4.1. Initial half-life metric definition
  - 1.4.2. Trial analysis of half-life of data categories
  - 1.4.3. Refine half-life calculation metrics
- 1.5. Finalize XML Schema
- 1.6. Transition existing data to temporal XML model
  - 1.6.1. Validate & clean existing data
  - 1.6.2. Develop transition plan
  - 1.6.3. Test transition plan
  - 1.6.4. Transition data to XML model
  - 1.6.5. Validate data after move
- 1.7. Update & tune DLP for new data model

2. Milestones

- 2.1. Data Analysis & Preparation
  - 2.1.1. Data category definition
  - 2.1.2. Half-life factor determination
  - 2.1.3. Metric definition
  - 2.1.4. Data categorization
  - 2.1.5. Data half-life analysis
- 2.2. Data Transition
  - 2.2.1. Schema built
  - 2.2.2. Transition testing
  - 2.2.3. Data moved to new model

2.3. DLP system tuned

3. Resource Requirements

3.1. Data usefulness expert panel
3.2. Timeliness expert panel
3.3. Corporate research
3.4. Database development team
3.5. XML development team
3.6. DLP tuning team
3.7. Corporate documentation writers

4. Schedule

4.1. Data Analysis & Preparation (Jan 1 to March 30)
4.2. Data Transition (Apr 1 to May 30)
4.3. DLP Tuning (Jun 1 to Jun 30)