

Data Quality in the Insurance Market



watertrace
Management Consulting

Contents

INTRODUCTION	1
WHAT DATA PROBLEMS COULD AFFECT YOU?	2
Data Specifications.....	2
Data Integrity Fundamentals.....	2
Duplication	2
Accuracy	3
Consistency and Synchronisation.....	3
Timeliness and Availability.....	3
Ease of Use and Maintainability.....	3
Data Coverage.....	4
Presentation Quality	4
Perception, Relevance, and Trust.....	4
Data Decay.....	5
Appropriateness.....	5
Security	5
WHAT CAN BE DONE?.....	6
Define	6
Measure	6
Analyse.....	6
Improve.....	6
Control.....	6
Data Specifications.....	7
Data Integrity Fundamentals.....	7
Duplication	7
Accuracy	7
Consistency and Synchronisation.....	7
Timeliness and Availability.....	7

Ease of Use and Maintainability 7
Data coverage 7
Presentation Quality 7
Perception, Relevance and Trust 7
Data Decay 8
Appropriateness 8
Security 8
CONTACT INFORMATION 9

Introduction

The London Market produces, stores, and uses data in every single aspect of its operations. That data forms the basis of the information that is used in the decisions, which shape the long-term plans of each company, as well as the day-to-day decisions made by underwriters, brokers and other individuals. Imagine the effect upon those decisions if just a *small* proportion of that data is of a poor quality: underwriting decisions will be based upon incorrect information; reserving assumptions could prove to be inadequate or too cautious; future growth targets for insurers and brokers become unachievable or too easily reached; or strategies for the future have their foundations built upon an unreliable representation of the status quo.

Now imagine the consequences of a *significant* proportion of the data being of poor quality.

The business impacts of poor quality data are, or have the potential to be, severe; a view that is shared by the regulators, as they focus upon the quality of the data used throughout an organisation, and within the requirements of Solvency II.

Data quality should therefore, be an integral part of the operational considerations of all companies involved in insurance.

This document has been produced to assist insurers and brokers to understand the types of problems that can beset their data, how to find and measure them, and finally how to address them.

What data problems could affect you?

Believe it or not, there are thirteen different attributes of the quality of the data used by an organisation, and it is likely that most of them will affect most of the data stored by most organisations. All thirteen are detailed below with illustrations in the sidebar relating them to the data used in the London Market.

Data Specifications

Data needs to be defined along with the manner in which it is created, updated, moved, and used, some or all of which actions will be dependent upon the type of user. The specifications provide:

- The context for interpreting the results of the data quality assessments
- Instructions for manually entering data or designing data load programmes
- The rules for amending or updating information
- Instructions for developing applications

Do your data specifications exist? Do you believe that they are of a sufficiently high quality, and that the documentation is adequate? Are you using industry standard data specifications?

London Market Relevance

Example – Gross Premium

What does Gross Premium mean? How are deductions calculated and treated?

How should the data be recorded on the systems?

If the premium changes, how should it be updated and how should the system manage those changes, and present them to the users?

How should the data be retrieved and used by third party applications?

Data Integrity Fundamentals

Data has fundamental characteristics of existence, validity, structure, and content, each of whose quality is integral to all their subsequent uses.

Techniques are used to analyse the data's essential measures, such as completeness/fill rate, validity, lists of values and frequency distributions, patterns, ranges, maximum and minimum values, and referential integrity.

This stage is often referred to as data profiling and its results should be documented and analysed.

Duplication

Duplicate data can exist either as fields within a record or as duplicate records. They cause problems for the users; the custodians and are associated with many hidden costs.

Potentially duplicated records can be identified with a data-cleansing tool but user intervention is essential, as decisions need to be made regarding which of them are real duplicates, which record should survive, and which pieces of records should be carried into a new combined record. This is an iterative process that is highly labour-intensive and therefore expensive.

Duplicated policy records create confusion for the users, as premiums and claims could easily be misaligned.

Duplicate fields within records create problems for users and automatic load programmes, as it will be unclear, which field should be populated.

All duplicates create issues within the IT and data-warehouse teams, as they won't know whether to load them all, discard some and retain others, or merge them all into one unified record.

Accuracy

Data accuracy differs to fundamental integrity, as it a measure of the correctness of the data, which requires comparing the data to the authoritative source of reference; for example, Dun & Bradstreet or Standard & Poor.

Accuracy checking is usually a manual process that takes the form of a survey or an inspection.

Example – Property portfolio

In the property example above, does the street exist in that county, and that county in that state?

Example – Reinsurance purchase

Is the reinsurer correct and have the rating agency ratings been correctly assigned?

Consistency and Synchronisation

Consistency and synchronisation are important because the same data are often stored in many different places in the company. Any use of the data should be based on those data having the same meaning and value.

The assessment process entails identifying the databases where the data are stored, mapping each incident of the same datum, comparing to the original, analysing and reporting the results - including the potential impact, root causes and recommendations.

London Market Relevance

In how many different places do you have your data stored & are you sure that each incidence has the same value?

Examples:

(1) Policy records contain inception and expiry dates that are/could then be replicated throughout the system(s) in credit control, claim ledgers, reinsurance schedules or a data warehouse.

(2) A premium transaction is recorded in the financial transactions of a policy record, the outwards reinsurance schedules and the financial ledgers.

Timeliness and Availability

Timeliness and availability are measures of the degree to which data are current and available for use as specified, and in the timeframe in which they are expected. If the gap between the time that the real-world objects change, and when the data that represent them are updated in a database and made available for use is too long the value of that data diminishes. Assessing whether the data suffers from poor timeliness and availability, entails using a randomly selected sample of records, and trace them through the data lifecycle, to determine the time elapsed between the steps in the process for each of the records.

Timely policy entry, including the updating of low-level aggregate data for catastrophe modelling is crucial, and any delay in either entry or use of the data, exposes the insurer to an unknown element of risk.

Premium data entered is less valuable if not available in time to facilitate the credit control function.

Claim movements can take several days from the time of agreement, through to notification and processing. Delays can affect underwriting decisions or impact upon cash-flows by delaying potential reinsurance recoveries.

Ease of Use and Maintainability

Closely aligned to 'timeliness and availability' is the dimension of 'ease of use and maintainability', that measures the degree to which the data can be accessed and used, and the degree to which data can be updated, maintained, and managed. The dimension is most probably measured by interviewing a selective sample of data users – creators, custodians and customers - to see how long it takes them to complete their relevant processes.

Reports based upon data that is stored but highly complicated to retrieve, can take many man-days to extract, manipulate and format. The report is expensive to produce and may be too late when actually delivered.

Databases that have complicated structures can entail convoluted processes, and lengthy times to update fields and require great care in the storage and movement of the data.

Data Coverage

Coverage is concerned with how comprehensive the data available to the company are in accounting for the total population of interest.

The total size of the population or data universe must therefore be estimated, and the size of the database population measured so that the coverage can be calculated (by dividing the number of records by the estimated total population), analysed, and documented.

London Market Relevance

How relevant is the experience of your motor portfolio to the overall experience of the UK market? Are you making decisions based upon a relevant representation of the market, as a whole or is your own portfolio too small to be extrapolated.

Estimate the number of policies within a specific class, (say, 700,000) and count the policies you have issued in the current 12-month period (say, 20,000). Divide 20/700 and you will understand that your data coverage is about 2.8%.

Presentation Quality

The quality of the presentation of data refers to whether the format and appearance support the appropriate use of the data.

The assessment process requires the data to be identified within the context of the information life cycle, determining where it is collected and used and assessing the information itself and the quality of the format.

Two perspectives are involved and compared: the assessor's perspective, which

necessitates understanding the context within, which the data will be used; and the perspective of those using the information.

How do you receive the data relating the binding authorities you either broke or underwrite? Is the data presented in a manner that enables it to be readily understood, processed and represented?

Does your MI have a consistent style that facilitates easy use of reports from many different departments?

Perception, Relevance, and Trust

Perception, relevance, and trust are subjective assessments reflecting the confidence in the quality of the data and its importance, value, and relevance to the business needs.

The data creators, custodians and customers are questioned, using surveys that are prioritised for data with high importance to the business. Results of the survey will be pointers to possible data cleansing exercises, or better communication between stakeholders.

Probably best illustrated with a hypothetical example:

Policy rate movements are entered on to every single policy by underwriting staff. The data is managed by the IT team, and used by the underwriters and monitoring teams.

The data is used with confidence by the underwriters, but in one class of business, totally disbelieved by the monitoring team.

Survey outcome: the underwriters in that one class are completing the field using a different mathematical logic.

Data Decay

Data decay, also known as data erosion, is a measure of the rate of negative change to the data. It is a useful measure for high-priority data that are subject to change, as a result of events outside a system's control.

Examples of such changes would be:

- Changes not captured;
- system upgrades;
- New uses of the data not envisaged at the design stage;
- Loss of expertise in the data stakeholders; or
- Process automation.

Measuring this attribute is not as important, as understanding the likely causes within your own environment, and building the capability to keep the data up to date.

London Market Relevance

Using the same bullet points examples would be:

- Endorsements not noted or Renewal chains not maintained;
- System upgrades or wholesale changes of suppliers;
- Gross premiums recorded prior to the advent of GAAP reporting; and
- Data entry is outsourced, Dedicated IT staff move on or Promotion to management of SME's

Process automation causes decay in two ways: firstly removing the effects of human intervention, and secondly changing the performance metrics; both of which result in a loss of comparative value Data, used for regulatory reporting may need to be 'cleansed' before being used, as it has been inappropriately processed. The feedback loop is important to make, either the data creators aware of the inappropriateness of the actions taken, or system technicians aware of software 'glitches' to minimise further instances.

Appropriateness

Data is important only if it can be used to produce the desired business outcome, the contextual analysis for which comes from the business requirements.

The data must be tested to ascertain the extent to which they meet business requirements and, if discrepancies exist, a feedback loop to the creators built in to the test and/or the data requirements updated, and the data retested if necessary.

Irrelevant data is just as inappropriate and clogs up the database unnecessarily. If there is no business outcome being served by the data, it is inappropriate to hold it. For example, in a motor policy, what business outcome is served by recording the colour of the vehicle?

Security

If the security of your data is circumspect, the design and execution of all of the previous checks could be irrelevant.

The assessments of data security will examine the quality of the security protocols, the extent to which monitoring exists to detect unauthorised transactions, and analysis of penetration testing, that measures the efficacy of the environment safety features.

For example, do policies exist regarding the use of terminals (such as leaving them unattended) or passwords? Does the system automatically close down inactive sessions? Do authority levels exist?

Is analysis conducted on transactions & users?

Are regular penetration tests conducted?

What can be done?

So, now you know what sort of problems can affect data in general, are you ready to ask the questions of your own data?

Watertrace developed a methodology to assist in examining your data, and answering any questions you may have about the extent to which your data is fit for purpose. The methodology is built around the five Six Sigma phases: define, measure, analyse, improve and control, which together constitute a 'data improvement lifecycle', and is supplemented with a 'lessons learned' exercise, conducted at the end, that feeds back into the 'define' phase of the next iteration of the programme. These activities will help you to determine the quality of the data you already have within your organisation and how to improve it in situ. Also consider the implementation of ACORD industry data standards to reliably share data with your business partners and reduce the burden of re-keying and cleansing.

Define

This phase has three parts:

- Understanding the business strategies and goals, and the issues and opportunities that arise from them, enabling you to understand why the data is important
- Prioritising the business issues, to ensure that the biggest threats are addressed first
- Planning the project using standard methodologies

Measure

There are three distinct parts to the Measure phase, more detail of which is contained within the appendix:

- Analyse the information environment;
- Evaluate the quality of the data in each of the attributes discussed in the previous pages;

- Evaluate the business impacts of the data issues discovered; and

Analyse

There are four distinct parts to the analysis phase, more detail of which is contained within the appendix:

1. Analyse the information environment;
2. Evaluate the quality of the data in each of the attributes discussed in the previous pages;
3. Evaluate the business impacts of the data issues discovered; and
4. Investigate and identify the root cause(s).

Improve

Having completed the analysis and discovered some data issues, an improvement plan needs to be implemented that addresses, correcting the current errors and, using the root cause analysis, amending the systems and/or processes that cause the errors to prevent them arising in the future.

Control

The objective of this stage is to implement ongoing monitoring, to ensure that new solutions have appropriate data quality controls, and metrics in place to verify that the improvements have had the desired effect upon data quality. In addition, a programme of continuous, standardised and documented improvements will help avoid the organisation returning to old processes and behaviours.

The Impact on Data Quality Attributes from Implementing ACORD Data Standards

Data Specifications

ACORD provides the global insurance industry with documented data specifications. These specifications are developed and maintained through an open, transparent process, open to all companies in the insurance industry. The specifications are available through www.ACORD.org/downloads

Data Integrity Fundamentals

ACORD's data standards define what data items within a message are Mandatory, conditionally mandatory (i.e. must be supplied dependent upon a particular business condition) or optional. The standards define the codesets (e.g. Country codes, Currency codes, Line of business codes, Perils codes) and metrics to be used (e.g. Percentage of whole, Kilometres etc.), the interchange (or "handshake") points along a business process when particular messages should be used, and how they should be responded to.

Duplication

ACORD's data standards are developed to ensure that data is not double-counted or duplicated within a message. The messages are typically de-normalised.

Accuracy

ACORD standards do not guarantee accuracy, but the use of a well specified message does improve the chances of accurate data and makes any manual checking process more efficient.

Consistency and Synchronisation

ACORD's RLC messages are built from what is essentially a data model (the data dictionary). This model ensures consistency and synchronisation. Greenfield implementers are recommended to use the data dictionary as a reference model; brownfield implementers are recommended to map their legacy systems to the model to

ensure that data can be seamlessly re-used and the benefits of straight-through-processing realised.

Timeliness and Availability

ACORD standards provide a reliable specification for sharing data between organisations. Removing the uncertainty and re-keying of data inherently speeds the process up.

Ease of Use and Maintainability

ACORD standards provide a reliable specification for sharing data between organisations. Removing the uncertainty and re-keying of data inherently makes it easier to use.

Data coverage

ACORD data standards provide the amount of data that the industry deems to be appropriate for the business processes involved.

Presentation Quality

ACORD XML standards are designed to be "machine consumable" and "human readable". The data is in a "raw" state, but companies can employ their own presentation templates (XSLTs) to display the data as required. In short, though, ACORD standards provide the data required and the consumer is free to display it in the manner they find most appropriate.

Perception, Relevance and Trust

Quantitative data which does not have to be re-keyed by the sender can be trusted as being their view of the data. While standards remove inefficiencies, they cannot guarantee qualitative data as this is subjective. Nonetheless, data definitions reduce the chances of mis-understanding between business partners.

Data Decay

As processes change and data requirements grow, previously captured data may not provide all the data points required today, or in the future. Using standards does not overcome this, but does enable the easier transition to new requirements as well as managing the gap between the two states, as per data migration and business change processes.

Appropriateness

ACORD data standards provide the amount of data that the industry deems to be appropriate for the business processes involved.

Security

ACORD messaging standards enable the secure transfer of data messages between companies' systems. The security of data behind company firewalls is part of companies' own overall network security policy.

Contact Information



London Underwriting Centre
Suite 1/3, 3 Minster Court
Mincing Lane
London EC3R 7DD
United Kingdom
Phone +44.207.617.6400
Fax +44.207.617.6401

Two Blue Hill Plaza
3rd Floor
PO Box 1529
Pearl River, New York 10965-8529
United States
Phone +1.845.620.1700
Fax +1.845.620.3600

www.acord.org

watertrace
Management Consulting

118 / 119 Fenchurch Street
London
EC3M 5BA
United Kingdom

Phone +44 (0) 20 7221 3273

<http://www.watertrace.com>