Data & Information Quality in a Rapidly Changing World

CAMAR Spring 2009 Meeting - June 4, 2009

Agenda

- **Introduction**

- **Data & Information Quality**

- **The Shifting Focus of Insurance Information and the Impact on Data & Information Quality**

- **Data Quality Evolution**

- **Addendum: CAS Data Management Educational Materials Working Party (WP5)**

Presenter

Pete Marotta AIDM, FIDM

Enterprise Data Administrator

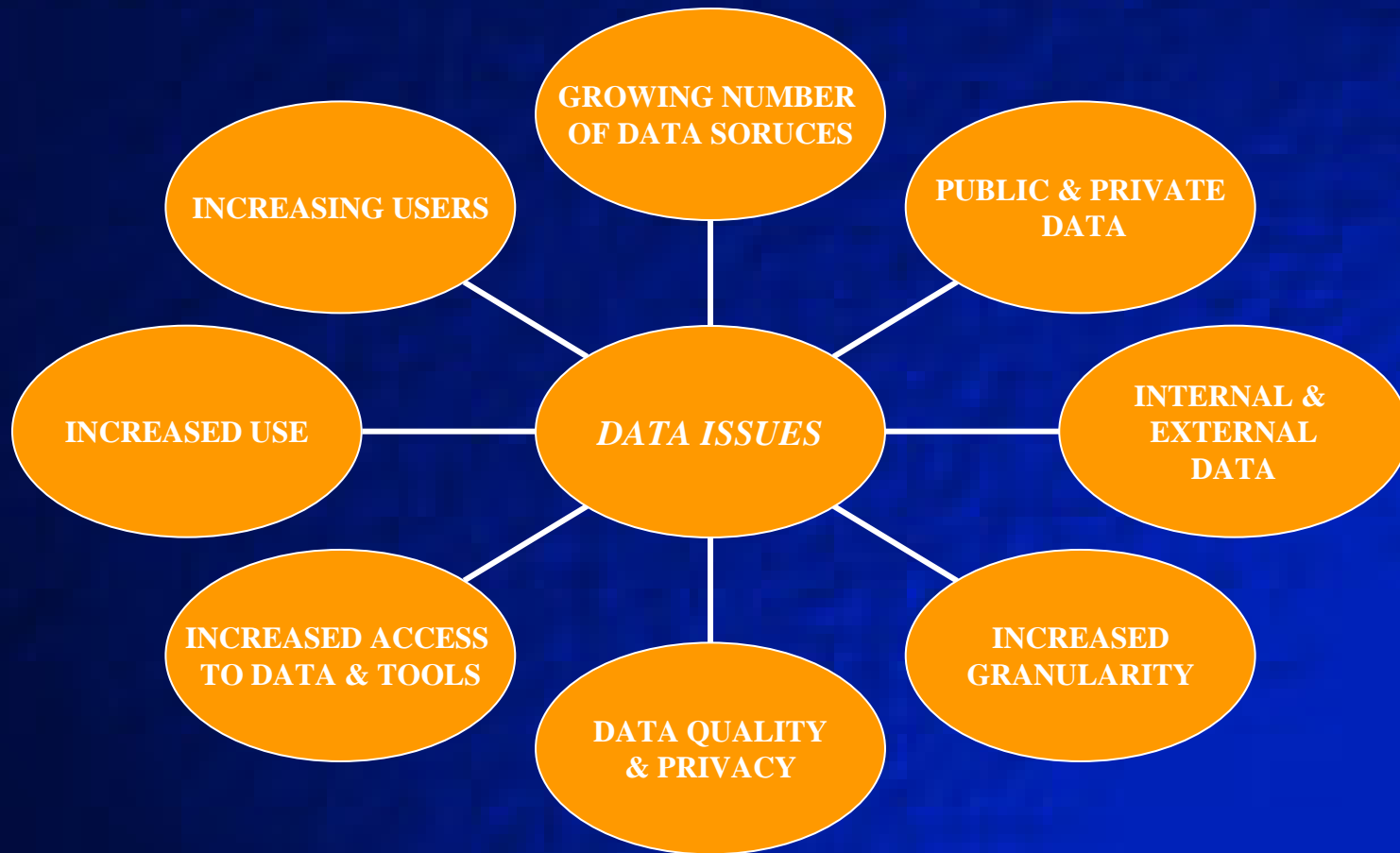**Insurance Services Office**

pmarotta@iso.com

201-469-2205

Introduction: Data Issues

"(Insurance professionals) … need solutions that will help them gain insight into risk, cause of loss, and resulting claims. They need to model ways to predict such eventualities … Success in these pursuits will increase profitability and much depends on the quality of the data. So, … why isn't everyone who is dedicated to growing profits equally concerned about data?"

Sharon Schwartzman, Editor-in-Chief, Techdecisions, "Culture Shock?" an editorial in the August 2007 edition of Techdecisions

Introduction: Data Issues

DATA ISSUES

- GROWING NUMBER OF DATA SORUCES
- INCREASING USERS
- PUBLIC & PRIVATE DATA
- INCREASED USE
- INTERNAL & EXTERNAL DATA
- INCREASED ACCESS TO DATA & TOOLS
- DATA QUALITY & PRIVACY
- INCREASED GRANULARITY

Introduction: Data Issues

- **GIRO working party on data quality survey:**
  - On average any consultant, company or reinsurance actuary:
    - spends **26%** of their time on data quality issues
    - **32%** of projects affected by data quality problems

General Insurance Research Organization (GIRO, part of the British Institute of Actuaries - Data Quality Working Party)

Introduction: ERM Road Map

"Post-mortem analysis of troubled insurance companies often reveals bad data as a root cause … these companies lacked an integrated data highway leading to a central ERM destination"

Make a Situational Assessment – how the data is generated, how it flows internally and externally, how transformed by users, how stored and retrieved, data quality issues

"Insurers Eye Road Map for ERM Highway" , Richard Hershman , National Underwriter, October 29, 2007

Data and Information Quality

"The only people who need not worry about data quality are those who neither create nor use data. No one participating in any modern economy can make that claim."

Data Quality: The Field Guide, Thomas C. Redman, Ph.D. Digital Press, 2001

Data & Information Quality: Data Quality – Key Characteristics

- Accuracy
- Validity
- Timeliness and Other Timing Criteria
- Completeness or Entirety
- Precision
- Reasonability
- Absence of Redundancy
- Accessibility, Availability and Cohesiveness
- Privacy
- Transparency

Data & Information Quality: Data Quality vs Information Quality

Information quality takes into account
- Data quality,
- Processing quality, and
- Reporting quality

> **DATA + ANALYSIS = RESULTS**
>
> *T. Dasu and T. Johnson, Exploratory Data Mining and Data Cleaning, Wiley, 2003*

Data & Information Quality: Information Quality Components

- **Information Definition Quality: definitions, specifications and business rules**
- **Information Content Quality: "raw materials" of information – key characteristics are completeness, validity, accuracy, precision**
- **Information Presentation Quality: the "finished product" – key DQ characteristics are accessibility, timeliness, presentation intuitiveness and freedom from bias**

"IQ Characteristics: Information Definition Quality" by Larry P. English, April 2006 Vol. 2, Issue 2, The Information and Data Quality Newsletter

Data & Information Quality: What Is Data Quality?

"**Quality data is data fit for its intended use**"

Data Management Best Practices, ISO Properties, Inc., 2003

"**In the end, the customer determines quality.**"

Improving Data Warehouse and Business Information Quality, Larry P. English, Wiley Computer Publishing, 1999

So, another way of defining Data Quality is meeting customer's expectations.

"**… suitable for the intended purpose of an analysis and relevant to the system or process being analyzed**"

Actuarial Standard Of Practice #23

Data & Information Quality: Understanding Customer Expectations

- **Identify customer needs**
- **Recognize complicating factors, for example**
  - in many cases the data underlying a service or product goes unrecognized by the customer
- **Recognize that the same data may have differing customer expectations relating to different customer perspectives, for example**
  - An actuarial perspective
  - A financial perspective

Data & Information Quality: Expectations - An Actuarial Perspective

Actuarial Standard Of Practice #23: Data Quality

- **Purpose is to give guidance in:**
  - Selecting data
  - Reviewing data for appropriateness, reasonableness, and comprehensiveness
  - Making appropriate disclosures
- **Does not recommend that actuaries audit data**
- **Due consideration to the following:**
  - Appropriateness for intended purpose …
  - Reasonableness and comprehensiveness …
  - Any known, material limitations …
  - The cost and feasibility of obtaining alternative data …
  - The benefit to be gained from an alternative data set …
  - Sampling methods …
- **Definition of Data**

# Data & Information Quality: Expectations - A Financial Perspective

- **Accountability, Quality and Transparency Regulations**
  - Sarbanes Oxley:
    - Processes and controls
    - Data control and reconciliation
    - Systems testing
    - Testing and assessment
    - Data Quality and Data Transparency are key
    - Documentation
    - Strategic Planning
    - Compliance
  - Solvency II
  - Reinsurance Transparency

The Shifting Focus of Insurance Information and the Impact on Data & Information Quality

- Regulation
- Data Analysis
- Technology
- Globalizations

Regulation: Changing Environment
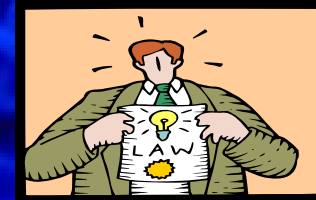
- **From Annual Statement to Market Conduct Annual Statements to NAIC Databases**
  - Financial Data Repository (FDR)
  - National Insurance Producer Registry (NIPR)
  - Fingerprint Repository
  - On-Line Fraud Reporting System (OFRS)
  - Uninsured Motorist Identification Database
  - NAIC Insurer Climate Risk Disclosure Survey
- **From financial data used to monitor solvency to financial, statistical data and analytics used to monitor enterprise risk**
- **From US driven privacy regulations to internationally driven privacy regulations**
- **Solvency and financial reporting: NAIC (National Association of Insurance Commissioners) to NAIC and IAIS (International Association of Insurance Supervisors) and IASB (International Accounting Standards Board)**

Regulation: "Regulate Me, Please"

New York Times, April 15, 2009, "Regulate Me, Please", artwork by Lewis Scott
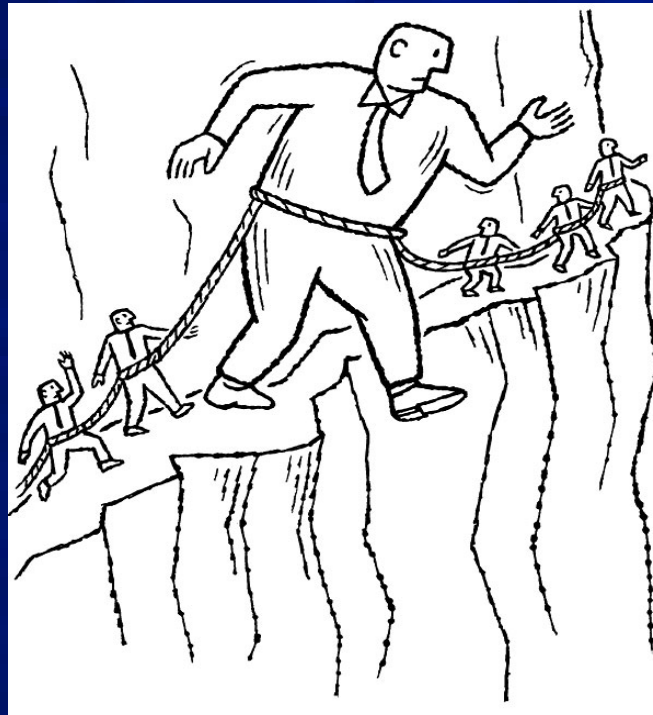
NAIC Insurer Climate Risk Disclosure Survey

http://www.naic.org/documents/committees_ex_climate_090224_survey.pdf

- Goal is to provide regulators, shareholders and the public with substantive information about the risks posed by climate change to insurers and the actions insurers are taking in response to their understanding of climate change risks.
- Disclosure is important because of the potential impact of climate change on insurer solvency and insurance availability and affordability across all major categories of insurance: property casualty, life and health.
- Survey contains a set of questions to help regulators assess insurers' risk assessment and management efforts and follow up with questions as necessary, subject to applicable examination and confidentiality provisions.
- Will also provide additional information for consumers to incorporate into their purchasing decisions.
- Not required to report quantitative information.
- The objective is for initial disclosure by some insurers for Financial Reporting Year 2009 with the first report due May 2010.
- Mandatory reporting if premium > $500M in 2009 and >$300M thereafter; optional for all others.

NAIC Insurer Climate Risk Disclosure Survey: Questions 1 to 4

1. Does the company have a plan to assess, reduce or mitigate its emissions in its operations or organizations? If yes, please summarize.

2. Does the company have a climate change policy with respect to risk management and investment management? If yes, please summarize. If no, how do you account for climate change in your risk management?

3. Describe your company's process for identifying climate change-related risks and assessing the degree that they could affect your business, including financial implications.

4. Summarize the current or anticipated risks that climate change poses to your company. Explain the ways that these risks could affect your business. Include identification of the geographical areas affected by these risks.

NAIC Insurer Climate Risk Disclosure Survey: Questions 5 to 8

5. Has the company considered the impact of climate change on its investment portfolio? Has it altered its investment strategy in response to these considerations? If so, please summarize steps you have taken.

6. Summarize steps the company has taken to encourage policyholders to reduce the losses caused by climate change-influenced events.

7. Discuss steps, if any, the company has taken to engage key constituencies on the topic of climate change.

8. Describe actions your company is taking to manage the risks climate change poses to your business including, in general terms, the use of computer modeling.

Regulation: Impact on Data and Data Management

- **Promoting the interoperability of data and databases at the local, state and national levels**
- **The need for clear, standardized, comparable information**
- **The need for well defined data reporting requirements and data dictionaries, promoting consistency across lines of business and time**
- **Awareness of regulations beyond the insurance space and beyond the US borders**
- **Recognition of the regulatory issues associated with re-purposing data – financial, statistical, operational, demographic, etc.**

Data Analysis: Changing Environment

- **From traditional underwriting and pricing - using traditional data sources (risk data, industry statistics) to predictive modeling and analytics - using non-traditional data sources (demographics, GIS, 3rd party data, non-insurance data, non-verifiable data sources, etc.)**

- **From risk-specific risk management to enterprise risk management**

- **From a stable risk control and claims environment to a dynamic environment of new hazards - mold, terrorism, computer viruses, cyber terrorism, etc.**

- **From traditional actuarial pricing methodologies to use of models – notably catastrophe models**

- **Use of non-insurance specific data used for pricing and underwriting - credit scores, insured occupation, household data, etc.**

Data Analysis: Impact on Data and Data Management

- From a data quality focus on validity, timeliness and accuracy to a data quality focus on completeness, transparency and accuracy
- New, different and more granular data
- From data available on a periodic basis to data available real-time
- From statistical plans and edit packages to data dictionaries, schema and implementation guides
- From structured data to structured and unstructured data
- From static geographic data points – ZIP Code, Territory, etc. – to dynamic Geographic Information Systems (GIS) and real-time data continuums
- From internal data sources to internal, industry and third-party data sources
- Increased use of analytical tools

Data Analysis: Impact on Data and Data Management

- Reducing the cost and time associated with data collection, storage, and dispersal, making data available more quickly
- Promoting the interoperability of data and databases, allowing for better data integration thereby giving the users more options for how data can be used
- Managing data content and definition across the organization which promotes consistency across business units and across time – internally and externally
- Ensuring the quality of the enterprise data, enterprise communication among the various data sources
- Recognizing the issues associated with re-purposing data – contractual, regulatory, technological, data, data quality, etc.
- Mapping of data across disparate sources – documenting data gaps and significant differences

Technology: Changing Environment

- **From centralized highly controlled technologies to ASPs, the, Internet, XML, LANs, PCs, etc.**
- **From technology as a business enabler to technology as a business driver**
- **From mainframes to LANS and high powered PCs**
- **From data collection to ETL (Extract Transform, Load)**
- **Data and access using new technologies, for example -**
  - Handhelds
  - VoIP
  - Smart Phones
- **Telematics, GPSs, Black Boxes, RFIDs, weather data, etc.**

Technology: Impact on Data and Data Management

- **Managing data over many moving/continuous data points v. data points fixed in time**
- **Measuring the quality of these new types of data**
- **From data available on a periodic basis to data available real-time**
- **How to use and store new types of data. The need for data "trigger points"**
- **Protecting data from inappropriate use**
- **New methods for protecting the privacy and confidentiality of data**

Technology: Impact on Data and Data Management

- **Balancing the need for more granular data with the cost and time associated with data collection, storage, and dispersal**

- **Managing both structured data and unstructured data**

- **From static geographic data points – ZIP Code, Territory, etc. – to dynamic Geographic Information Systems (GIS) and real-time data continuums**

- **Recognizing the technological implications of repurposing data**

Globalization: Changing Environment
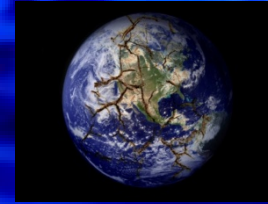
- **Outsourcing IT, data management, business functions and the need to educate foreign staff about US issues**

- **Expanding business beyond US borders and the need to educate US staff about foreign issues**

- **Cultural differences**

Globalization: Impact on Data and Data Management

- **Expanding the data quality focus to recognize cultural differences**
- **The need for procedural manuals, edit packages, data dictionaries, schema and implementation guides to recognize differences in terminologies and definitions**
- **The need for cross-border transparency**
- **Increased emphasis on compliance with international rating, reporting laws and solvency regulations**

Globalization: Impact on Data and Data Management

- Recognizing the cost and time associated with international data collection, storage, and dispersal
- Recognizing differences in technologies across borders
- Promoting the interoperability of data and databases, allowing for better data integration thereby giving the users more options for how data can be used
- Increased emphasis on industry data and data exchange standards
- Managing data content and definition across borders

Impact on D & IQ

Changing Expectations

- **Data transparency: documentation, controls, completeness, accuracy, etc.**
- **Privacy, confidentiality, compliance, solvency**
- **Issues associated with re-purposing/re-use of data**
- **Regulations beyond the insurance space and beyond the US borders**
- **Interoperability of data and databases**
- **Integration of data from multiple sources**
- **Increased use of third party data**
- **More granular and different data, including real-time data continuums**
- **Data prep for use in analytical tools and models**
- **Use and storage of new types of data**
- **Protecting data from inappropriate use**
- **New and improved access to data, data exchange and analytical tools**
- **Technology convergence – across industry, across country**
- **Outsourcing – in-country and abroad**
- **Expanding business beyond national borders**
- **Recognizing cultural differences**
- **Terminology & language differences**
- **Timing: real-time, time zones, etc.**

IMPACT OF D & IQ	REG	ANAL	TECH	GLOBAL
METADATA	X	X	X	X
DATA MODELS	X		X	X
DATA & PROCESS FLOWS	X		X	X
MASTER DM		X		
MAPPING DOCUMENTATION	X	X	X	X
DATA STANDARDS	X	X	X	X
DATA TRANSFORMATION. & GENERATION	X	X	X	X
DETAILED SPECS	X		X	X
DATA SOURCES: INTERNAL & EXTERNAL	X	X		X
REG. MONITORING	X			X
AUDITS & CONTROLS	X		X	
DATA & TEXT MINING		X		
UNSTRUCTURED DATA	X	X		
VERSIONING	X		X	
SCIENTIFIC MEASURES	X	X		X
BI & KNOWL. MGMT.	X	X		X

Data Quality Evolution
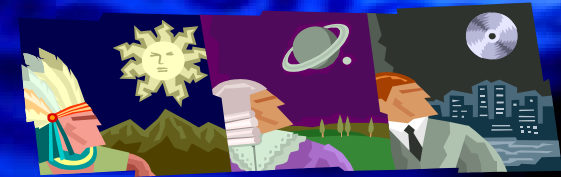
- **Information Quality Stewardship**
- **Strategic Data Planning and Data Governance**
- **Controls – internal and from 3rd party data sources**
- **Measures - internal and from 3rd party data sources – especially timeliness, completeness and redundancy**
- **Master Data Management**
- **Chief Data Officer v. Data Quality Officer v. Chief Information Officer**

Data Quality Evolution

- **Identify data sources**
- **Metadata – internal and from 3rd party data sources**
- **Mapping**
- **Data confidentiality and security -**
  - Encryption – data in transit and data at rest – security v. cost.
  - Fobs, etc.
- **Data re-use v. DQ levels**
- **Changing use of data v. DQ levels**
- **Data standards – business, industry, cross-industry, cross-border, technology**

Data Quality Evolution: Master Data Management

- **Master data management (MDM) is the set of processes to create and maintain a single view of reference data that is shared across systems. It is used to classify and define transactional data through the use of a centralized integration manager.**
  - It leverages policies and procedures for access, update and overall management of this central resource and its coordination with other participating systems across the enterprise.
  - Areas such as customer data integration (CDI), which involves management of customer reference data and product information management (PIM), which includes management of product and supplier reference data, are domain-specific subsets of MDM.

Data Quality Evolution: Chief Data Officer

- **"The Body Has a Heart and Soul, Roles and Responsibilities of the Chief Data Officer" by Thomas C. Redman, Information and Data Quality Newsletter January 2007, IAIDQ:**
  - Applies corporate data strategy and data polices defined by corporate data stewards/data council
  - Leads the data quality program
  - Leads the application of corporate data strategy and data polices to data suppliers
  - Owns and houses the metadata process
- **Need for a Chief DQ Officer?**

Data & Information Quality in a Rapidly Changing World

Questions and Commentary

Addendum: CAS Data Management Educational Materials Working Party (WP5) - Actuarial IQ

- **Introduction to Data Quality and Data Management being written by the CAS Data Management Educational Materials Working Party**
- **Directed at actuarial analysts as much as actuarial data managers:**
  - What every actuary should know about data quality and data management

Addendum: CAS Data Management Educational Materials Working Party (WP5) - Why Actuaries?

- **Both:**
  - Information consumers and information providers
- **Both:**
  - Have knowledge of data and high stakes in quality
- **Both:**
  - Skillful and influential

Addendum: CAS Data Management Educational Materials Working Party (WP5) - Actuarial IQ Key Ideas

- **Information Quality concerns not only bad data:**
  - Information which is poorly processed or poorly presented is also of low quality
- **Cleansing data helps:**
  - But is just a band aid
- **There are plenty of things actuarial analysts can do to improve their IQ**
- **There are even more things actuarial data managers can do**
- **Actuaries are uniquely positioned to become information quality advocates**

Addendum: CAS Data Management Educational Materials Working Party (WP5): Working Party Publications

- **Book reviews of data management and data quality texts in the Actuarial Review starting with the August 2006 edition**

- **These reviews are combined and compared in "Survey of Data Management and Data Quality Texts," CAS Forum, Winter 2007, www.casact.org**

- **Upcoming paper: "Actuarial IQ (Information Quality)" to be published in the Winter 2008 edition of the CAS Forum**