CELENT

Benjamin Moreland

Data Mastery Spectrum Overview 2012

July 2012

A member of the Oliver Wyman Group

Executive Summary

Data Mastery, as its name implies, is the mastery of data available to an insurer that allows it to use data, both internally and externally, to gain business insight and improve decisioning in all aspects of the business. This report profile vendors that provide data mastery components: namely business intelligence (BI) presentation, analytics, data warehouse, ETL component solutions and BI content, such as insurance data models, dashboard and report templates, KPIs, etc. Data mastery also includes enterprise data mastery roadmaps to enable insurers to align their data efforts with their business goals and objectives and datasets. In addition, analytics has evolved to include big data and data virtualization.

This report describes the current data mastery space in insurance and profiles 27 data mastery vendors that provide at least one of the data mastery components to insurers. Each profile provides an overview of the data mastery components offered by the vendor. Some vendors offer multiple solutions, and each is listed within that vendor's profile.

Key findings from this research include:

- Insurers need a data mastery roadmap aligned to their business goals to avoid the costly mistakes of the 2000s.
- Data analytics is evolving, and insurers that can best lever- age both their internal and external data will be the leaders by 2020.
- Big data is a merging of the increasing data volume (terabytes to petabytes to exabytes) with variability (structured and unstructured data), velocity (real time response), and verity (validity) of the data. Data maturity will be key.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

Introduction

Data Mastery, as its name implies, is the mastery of data available to an insurer that allows it to use data, both internally and externally, to gain business insight and improve decisioning in all aspects of the business. This report profile vendors that provide data mastery components: namely business intelligence (BI) presentation, analytics, data warehouse, ETL component solutions and BI content, such as insurance data models, dashboard and report templates, KPIs, etc. Data mastery also includes enterprise data mastery roadmaps to enable insurers to align their data efforts with their business goals and objectives and data sets. In addition, analytics has evolved to include big data and data virtualization.

This report describes the current data mastery space in insurance and profiles 27 data mastery vendors that provide at least one of the data mastery components to insurers. Each profile provides an overview of the data mastery components offered by the vendor. Some vendors offer multiple solutions, and each is listed within that vendor's profile. Because data mastery initiatives involve many different elements, take many different forms, and can be focused on many different areas of an insurance enterprise, there are many different vendors that may be relevant to any individual project or strategy.

This report is designed to help insurers understand the general "vendor space" for data mastery solutions, and to provide an overview of the capabilities, areas of focus, and characteristics of 27 leading solutions. Details of each vendor's solution are presented in short profiles, and are compared to each other in a table designed to help insurers quickly assimilate information about this complex space and deter- mine which vendors are likely to be relevant to their specific initiatives.

This report does not include rankings, evaluations, or recommendations, since the space is so heterogeneous, and each insurer's needs differ by their strategic goals and current infrastructures. However, Celent believes that this report will make the process of vendor short- listing and evaluation much more efficient and effective for insurers.

Vendors' data mastery offerings are shown in Table 1.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

Table 1: Vendor Data Mastery Components Offered

Vendor (Solution)	Enterprise Data Mastery Roadmap	BI Presen-tation	Analytics	DW/DM	ETL	BI Content/ Data Sets	Big Data
4Sight Business Intelligence		Z	Z	Z	Z	Z	
Accenture	Z	Z	Z	Z	Z		Z
Aon eSolutions	Z	Z	Z	Z	Z		Z
Clearwater Analytics		Z	Z				
CoreLogic		Z	Z				
Cover-All	Z	Z	Z	Z	Z	Z	
CSC	Z	Z	Z	Z	Z	Z	Z
FICO	Z	Z	Z		Z		Z
IBI	Z	Z	Z	Z	Z	Z	Z
IBM		Z	Z	Z	Z		Z
Infobright		Z		Z	Z		Z
Informatica					Z		
Innovation Group	Z	Z	Z	Z	Z	Z	
InsFocus Systems		Z	Z	Z	Z	Z	
iPartners		Z	Z	Z	Z	Z	
LexisNexis		Z	Z	Z	Z	Z	Z
Munich RE Automation Solutions		Z	Z	Z	Z		Z
NiiT Technology				Z	Z	Z	
Oracle		Z	Z	Z	Z	Z	Z
Pegasystems	Z	Z	Z				
Polaris Software	Z	Z	Z	Z	Z	Z	Z
Robust Designs	Z	Z	Z	Z	Z	Z	
SAS	Z	Z	Z	Z	Z	Z	Z
SunGard		Z	Z	Z	Z		
UCT		Z	Z		Z		
Valen			Z				
Yodil	Z	Z	Z	Z	Z	Z	

Source: Celent

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

The vendors' profiles are broken into two sections in this report: those that offer all or most of four of the five systems components (ETL, data warehouse/data marts, analytics server, and BI presentation layer), and those that specialize either in a couple of the data mastery components or a niche business area. Most of the vendors that offer the four system components do so as an integrated suite.

Big data and data virtualization solutions are still in their infancy stages within insurance, but Celent expects that with the growth of telematics and social media data, as well as better use of structured internal and external third party data, big data solutions will be implemented by insurance leaders to give them a competitive advantage. Celent's view of big data is provided in the next section. This report does not describe the vendors' big data solutions, just if they currently have one.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

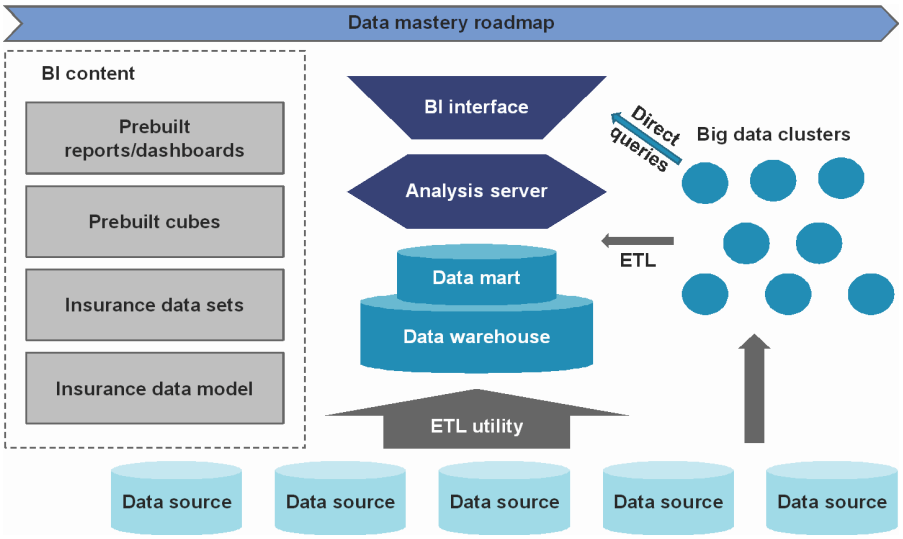
Data Mastery Systems Overview

Definition

In a June 2003 report, Celent coined the phrase "data mastery" to cover a broad range of technology solutions that are related to leveraging their data in various efforts, including data warehousing, data marts, CRM, unified presentation layers, reporting, data quality and integration, and business intelligence. The purpose of data mastery is to leverage business insight from various data sources to make better business decisions. Historically this insight was used in a reactive manner to improve business processes and build predictive models for underwriting and claims fraud, but the use of analytic results in business decisions has become real time.

There are eight major components to data mastery shown in Figure 1, of which some are system based, namely the data sources, ETL (extract-transform-load), data warehouse/data marts, big data clusters, analysis server, and BI interface. The components that are not system-based are the data mastery roadmap and BI content. Each the components is described below.

Figure 1: Data Mastery



Source: Celent

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

Data Mastery Roadmap

Based on the efforts of insurers, we can observe that the most important component to any data mastery effort is the definition of a data mastery roadmap. Many data mastery projects, especially master data management (MDM) programs, have been IT-led with the belief that, if an insurer could create a data warehouse with all its internal data consolidated and possibly supplemented with complementary industry data, it could be in a position to provide any dashboard, report, or predictive model for its business users. These efforts failed under their sheer size and complexity, but also due to a lack of understanding what business problem the insurers were solving.

Over the last decade, insurers spent millions—large insurers spent tens of millions—on data mastery programs with little to no realized business value. The missing ingredient in their efforts was a clear vision and direction of how these programs aligned with their business strategy, goals, and objectives. Insurers are flooded with data: smaller insurers have or are approaching terabytes of data internally, and larger insurers have or are approaching petabytes of data. With the addition of telematics, external industry data, and social media data, many carriers will soon have exabytes of data at their disposal, with zettabytes around the corner (within the next three to five years).

This data can be leveraged across all areas of an insurer's business, including:

- Financial
- Customer Service
- Sales & Marketing
- Operations
- Agents
- Field Staff
- Claims
- New Business/UW

This list is not meant to be exhaustive, but to show the breadth that the data can be used within insurance to improve each area. With all this data, and the ability to collect it, consolidate it, cleanse it, analyze it, process it, create information from it, distribute it, and query it—insurers need to have a strategy and direction on how to best do this. Insurers cannot take the time to manually, after the fact, chase down the data to provide the necessary information.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

In addition, each business area can also be divided into subareas, such as claims fraud and claims operations. Businesses have business goals and objectives, and their data mastery efforts need to be aligned with those business goals. An insurer's data mastery program must be built on a foundation that is scalable and flexible to meet those changing requirements. In a tough economy and competitive pressure, efficiency and effectiveness are critical, and a data mastery roadmap allows an insurer to know how to invest in tools and programs to meet those needs.

Insurers can, and many have, chased their tails trying to get value out of their data due to a lack of focus, direction, and alignment with business goals.

A data mastery roadmap begins with the business goals and objectives, such as market growth, IT expense reduction, reduced claims fraud, improved customer retention, etc. Business-driven KPIs can be created that will allow the business to know if it is moving in the right direction. Next, the insurer needs to select a mature insurance data model and create a data reference architecture. Finally, as it builds the foundation from which it can grow, it can begin its data mastery implementations based on business/project priority knowing that it can scale and expand its efforts iteratively.

BI Content

Any data mastery effort is going to be driven towards providing business insight in one of three areas: reports, dashboards, real time analytics, predictive analytics, and data mining. Historically, insurers have used data to drive reports on business metrics for finance and executives, and many insurers' underwriters have leveraged predictive analytics. Reports tend to be historical views of some aspect of the business, in many cases one to two months old. Recently, more advanced insurers in this area have used data mastery for real time dashboards and real time analytics that are fed back into their core systems to have a more immediate positive impact on business performance.

Underlying any mature data mastery system is BI content, such as insurance data models, report and dashboard templates, business KPIs (key performance indicators), and insurance data sets. In addition, prebuilt cubes for different data perspectives and insurance industry data sets are used to provide IT fast and agile delivery options. Data sets are a collection of related data in a specified business or IT area. Normally, each column represents a particular variable, and each row represents

an element or member of the data set. Data sets can be built by the insurer from its own data or acquired from third party data information providers.

Data Sources

One of the most difficult parts of a data mastery project is determining exactly which data sources will need to be accessed. Relevant data for any initiative may live in a wide variety of systems, and a wide variety of technological and syntactical formats. Data may be stored in flat files, outdated database formats, modern database formats, XML, or other formats. In addition, while insurers historically have primarily used just internal data, more are using external data that provides insurance benchmarking data, customer risk data, social media data, and many others. Insurers must address the quality and value of the external data and how it can be used for better decisioning and effectiveness results.

Some of the relevant data may be semistructured or unstructured, freeform text. Different systems may also use different syntaxes. As a very basic example, some systems might record a customer name in a single field, while others use separate fields for first and last names. Different systems may use different identification numbers for the same customer, policy, event, claim, etc. The first stage in data consolidation is to determine what data needs to be consolidated, where it is now, and what format it's in.

Since many insurers have gone through some modern data initiatives, it is likely that at least one data source (if not multiple data sources) will be relational databases. While a modern relational database will make gathering data from the data source easier, insurers will still often need to deal with multiple databases utilizing different data models.

This report does not include a review of vendors providing external industry data/information or databases or other data source solutions, though some of the vendors may indeed offer database technology in addition to the components listed in their profile.

ETL

Once the data sources have been identified and the data models and repositories selected, it is necessary to choose an appropriate method of data exchange. Data exchange involves two main steps:

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

- Data mapping and cleaning to map the source data to the target data model.
- Data transfer to physically copy the data from one system to the other.

Data mapping and cleaning is by far the most challenging part of any data mastery project. It typically consumes at least 80% of the implementation time and resources, and about 40% of the overall project from planning to training and maintenance. While the process of data mapping and cleaning is generally very time-consuming and labor-intensive, the use of ETL tools designed specifically for the insurance industry can provide significant efficiencies. Specialized data cleansing vendors can also ease this burden. In the insurance industry, often data mastery projects fail (or fail to achieve their full goals) because of issues at this stage.

Data transfer is a matter of copying the data from its sources to the repository. For asynchronous systems, this is typically accomplished via a batch transfer. For synchronous systems, this typically requires connectivity through real time methods such as the use of an EAI (enterprise application integration) platform, or an internal Web service.

Data Warehouse / Data Mart

Once the necessary source data has been determined, a decision needs to be made about where and how the consolidated data will be stored and accessed by the analytic component. Is it an operational data mastery project that requires constant updates from production systems, or is it an analytic project that needs snapshot or historical data that is updated less frequently? Is the best solution a data warehouse, a virtual data warehouse, operational data store, or a unified presentation layer?

In what format will the data be stored? What will the data model and schema be? Celent recommends that insurers consider purchasing a predefined insurance data model from a data mastery system or data master services provider with deep experience in data mastery for insurance. Even if these models are not comprehensive for the needs of the individual carrier, using them as a base can save a tremendous amount of time and allow a carrier to scale with each project and not be overwhelmed trying to build out the entire data model first.

Celent also recommends choosing a data model that is compatible with the ACORD XML standards for the US and UK geographies.

Analytic Tools

The goal of unifying diverse data in a single repository is to be able to subject that data to analysis using some form of analytic tool and extract useful information from it. Analytic tools encompass a broad spectrum, from simple presentation of related data (for example, a unified view of all products associated with a single customer) to highly complex statistical analyses of things like profitability, claims rates, or customer segmentation.

Choosing an analytic tool depends entirely on what information needs to be extracted from the unified data, and who needs to be able to use the data. Is the end user a low-level knowledge worker or a high-level financial analyst? Is synchronous or asynchronous data required? What depth of reporting is necessary? How much training will analysts require on the system? All of these issues should be considered prior to selecting or building an analytic tool.

Different vendors provide different components to allow insurers to get the most analysis from their data. The analytic tools can be further broken out into three categories. Some vendors provide only one, some provide two or three.

Analysis Server

If an insurer is just looking for a simple presentation of related data (such as a unified customer database), an analysis server may not be necessary. But if the insurer is looking to perform complex statistical analyses and run custom queries and reports, a relational database such as used by the data repository is not designed for the real time analysis across large amounts of data.

An analysis server allows the manipulation of data into structures better suited for this kind of business intelligence. The analysis server might use OLAP cubes or snowflake schemas, though the precise methods are not the subject of this report. The analysis server, combined with the other analytic tools, allows an insurer to report on and query data in ways not possible with the normal relational database.

Business Intelligence Presentation

What good is all the complex statistical analysis if the correct people can't see or understand it? Executives will want to see higher-level dashboards that summarize enterprisewide trends, while a financial analyst will need to view detailed reports run against many different variables. Others might need an interface complex enough to generate and run custom reports, while still being intuitive enough to handle

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

without IT help. Despite the naming given in this report, it's important to understand that the business intelligence presentation is also a form of insurance-specific content. For a view to be effective, it needs to be designed specifically for users asking insurance questions or viewing insurance reports.

Many vendors that are selling business intelligence solutions are really providing the user interface that allows the right people to view the right data in the right way. The insurer may still need to choose an analysis server from a different vendor—and possibly set up its own data cubes—in order to utilize the vendor's business intelligence utility.

Big Data Clusters

Big data is a relatively new development within data mastery. Not only has the size or volume of data greatly increased over the last 5+ years, from gigabytes to terabytes and now petabytes, but the data has changed in several other ways as well. Historically, most carriers have created data mastery solutions with structured, primarily transactional data from their systems. However, carriers are now dealing with semi-structured or unstructured data (variety), real time requirements (velocity), and variability within the security models, metadata schema, and varying sources. Thus big data can be defined by volume, velocity, variety, and variability.

Big data adds dimensions to dealing with data/information challenges. It also adds a new component to data mastery which can act as multiple components of past solutions. Big data adds clusters to the previous data mastery definition. These clusters are collections of semi-structured (XML tags) and unstructured (video) data. Through big data techniques, such as Hadoop, these clusters can be used to put enough structure to the data to act as an ETL for semi-structured data and place it in a more traditional data warehouse. The data can also be used to query directly for insights and information, so it can act as a data warehouse of semi-structured data. Finally, there are query techniques that allow the data to be analyzed directly, so it can act as the analysis server as well. For more information on the use of Hadoop techniques, see the Celent report *Big Insurance Data: Drawing Lessons from Amazon, Google, and Facebook*, September 2011.

Big data will present new challenges to insurers along each of the four dimensions: volume, velocity, variety, and variability. Most data master vendors have continued to scale their solutions to deal with the larger volume of data (i.e., large terabytes to petabytes of data); however, most of them lack solutions that address the other three dimensions. As vendors make their solutions more mature, insurers will find internal

challenges as well, including big data skills, such as Hadoop, MapReduce, NoSQL, and other techniques to query semi-structured or unstructured data. What makes these new technologies unique is that they are not based on the traditional RDBMS.

Insurers will also be challenged on how to mix their traditional structured data with the new, and in time vaster, semi-structured data. As described above, there are approaches to provide structure to semi-structured data and feed that into the more traditional data warehouse approaches. But the challenge for insurers will be how to mine the big data clusters to guide their "insight" investigation and research. Newer, in-memory analytics solutions will provide some help in this area, but not the full solution or answer insurers will be looking for.

Insurers will also have to deal with the quality and reliability of the external data. Not only will be it be in a different format, but in some cases the external data may be more reliable than their internal data (eg., telematics), but in other cases, such as social media, there will be a learning curve on how to best leverage this data. Insurers tend not to trust data originating outside of their firewalls. This will change as the leading insurers figure out how to best leverage the vast amounts of external data that will be available to them.

Data Virtualization

Data virtualization is not listed as a separate component within data mastery because it is focused on the abstraction of the source and format of the data, as opposed to centralizing all the data in one data warehouse in a standard and unified data model. Some carriers would prefer to not aggregate all their data into one central data warehouse, then leverage data marts for different areas to access the data. These carriers would prefer to leave the data at the source and manage retrieval of data from multiple sources and then manipulate it for the desired result.

As the amount of data (structured and semi-structured) grows, continuing to centralize and sync the data in a central data warehouse is getting harder and harder, and data virtualization provides an alternative solution if the vendor supports this approach.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

SAS Inc.: SAS Business Analytics

Company and Product Information

SAS Inc. (SAS) is a privately owned company with headquarters in Cary, NC. It was founded in 1976 and has over 12,400 employees. Its target market is Tier 1-3 global insurers for both life and P&C.

SAS provides all of the data mastery components in its SAS products, as shown in Table 19.

Table 19: Data Mastery Coverage Areas

DM Area	Solution
Analytics	SAS Business Analytics
BI Presentation Layer	SAS Enterprise BI Server
ETL	SAS Data Management
DW/DM	SAS Insurance Analytics Architecture
Data Sets	Z
Enterprise Data Roadmap	
Big Data	SAS High Performance Analytics
Data Virtualization	SAS IT Resource Management

Source: Vendor RFI
Z indicates vendor offers functionality; a blank cell indicates it does not.

SAS does not offer a data mastery roadmap for insurers independent of its product offering, however it does offer an integrated suite of analytical software on a single environment. Its products provide the ability to apply predictive modeling to both structured and unstructured data. The speed and scalability of SAS enables insurers to analyze huge quantities of data quicker and more easily.

SAS' BI presentation layer provides insurers with easy-to-use dynamic visualization capabilities integrated with world class analytics results. The business user is not dependent on IT to service its ad hoc requests and can explore the right data on its own, see trends, and communicate the results quickly where needed across the entire organization. Users can drill deep into content and explore a virtually unlimited number of columns and records to discover new insights that drive business value.

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

SAS offers a unified data management platform. By combining data integration, data quality, enterprise data access, and master data management into a unified development and delivery environment, organizations can maximize each phase of the data management process. Using a single framework with an intuitive interface empowers organizations with data that is consistent, accurate, and timely.

SAS' data warehouse includes an insurance-specific, logical, and physical data model and prebuilt ETL functionality. The company offers superior data management capabilities to consolidate data from multiple legacy systems into a single version of the truth and automatically validating data as part of the data integration process. SAS can easily integrate with any carrier core system to populate the Insurance data model. SAS data sets include SAS/ACCESS and SAS Enterprise Miner.

To SAS, Big Data refers to empowering organizations to transform and analyze both structured and unstructured data into business value using analytics.

SAS Visual BI or SAS JMP provides dynamic business visualization that enables business users in an easy-to-use interactive environment to explore ideas and information, investigation patterns, and discover previously hidden facts through visual queries.

The SAS integrated suite of products can be used in multiple functional areas within a carrier, including customer data, fraud, distribution/channel management, underwriting, policy service (limited), claims, pricing and enterprise risk management.

SAS data mastery solutions are one of the most widely used in production across insurers today. Their customers are presently using SAS implementation to support personal lines (auto and homeowners), commercial lines (auto, property, and all other standard commercial lines products), E&S and specialty, reinsurance, individual life and annuities, individual health, group health, and group benefits.

Five percent of SAS customers (which is fairly large) are using their solutions on an ASP and SaaS basis, respectively.

SAS products are supported on multiple operating systems including Windows (preferred option), z/OS, Solaris, UNIX AIX, HP-UX, and Linux for x86. The preferred databases are Oracle, Microsoft SQL Server, and Sybase. DB2, MySQL, SQL on UNIX. SAS/ACCESS® engines provide seamless and transparent read/write/update access to more than 60

Copyright 2012 © Celent, a division of Oliver Wyman, Inc.

data sources, including relational and non-relational databases, PC files, and data warehouse appliances, including Informix, ODBC, OLE-DB, SQL on UNIX, Teradata, MySQL, Netezza, and others.

The preferred application server is Oracle WebLogic, and JBoss and WebSphere are also supported.

The code base for features, functions, and screens for business users is 100% Java for SAS EBI Server web-based clients and 100% .NET for SAS EBI Server desktop applications. The code for developers and configuration is 100% Java.

CELENT

For more information please contact info@celent.com or:

Benjamin Moreland
bmoreland@celent.com

North America

USA
200 Clarendon Street, 12th Floor
Boston, Massachusetts 02116
Tel.: +1.617.262.3120
Fax: +1.617.262.3121

USA
1166 Avenue of the Americas
New York, NY 10036
Tel.: +1.212.541.8100
Fax: +1.212.541.8957

USA
Four Embarcadero Center, Suite 1100
San Francisco, California 94111
Tel.: +1.415.743.7900
Fax: +1.415.743.7950

Europe

France
28, avenue Victor Hugo
75783 Paris Cedex 16
Tel.: +33.1.73 .04.46.20
Fax: +33.1.45.02.30.01

United Kingdom
55 Baker Street
London W1U 8EW
Tel.: +44.20.7333.8333
Fax: +44.20.7333.8334

Asia

Japan
The Imperial Hotel Tower, 13th Floor
1-1-1 Uchisaiwai-cho
Chiyoda-ku, Tokyo 100-0011
Tel: +81.3.3500.3023
Fax: +81.3.3500.3059

China
Beijing Kerry Centre
South Tower, 15th Floor
1 Guanghua Road
Chaoyang, Beijing 100022
Tel: +86.10.8520.0350
Fax: +86.10.8520.0349

India
Level 14, Concorde Block
UB City, Vittal Mallya Road
Banglalore, India 560001
Tel: +91.80.40300538
Fax: +91.80.40300400