CAPCO

DATA MANAGEMENT FOUNDATIONS FOR GETTING AHEAD AND STAYING AHEAD

AN INSURANCE INDUSTRY VIEW

ABSTRACT

A perfect storm of technological, economic and societal forces is driving the insurance industry towards accelerating investment in the digital and data science agenda. However, this is not without operational, legal and ethical risks. This paper examines the drivers for data management in the industry and how a lean and agile approach to data management reduces the burden of data curation and preparation, while simultaneously managing those risks and maintaining stakeholder participation and buy-in. We find that organizations that build lean and agile data management foundations can perform data science and digital innovation at greater scale, effectiveness and with quicker returns on investment. This provides not only immediate competitive advantage through enhanced business insights and customer engagement, but also the ability to sustain these advantages in the long run through uplifts in the data culture of the organization.

AUTHOR'S NOTE: Although there are subtle differences in meaning, the terms 'data science' and 'analytics' are used interchangeably throughout this paper as referring to the set of activities that exploit the semantic and statistical relationships in data to generate business insights.

1. INTRODUCTION

The business of insurance has been built on the ability to analyze data to assess risks, manage exposures, and set premiums. Indeed, while analytics has been a buzzword in the wider economy over the past decade, insurers have been methodically analyzing data for decades, if not centuries, so that governments, businesses and individuals can have the confidence to make investments and accumulate wealth by mitigating and pooling risks.

In recent times however, three things have emerged that threaten the status quo: (1) increasing quantities of structured and unstructured data generated through the increasing digitalization of the economy and of wider society, (2) the maturing of techniques and technologies capable of scaling the application of machine learning and artificial intelligence approaches to actuarial analysis, and (3) accelerated globalization that simultaneously lowers international barriers to entry (e.g. emerging players from Asia), while rendering supply chains more susceptible to global systemic shocks from the increased transmissibility of risk across borders, as we've seen with the COVID-19 pandemic.

Against a backdrop of claims inflation, increasing catastrophe loads, and market entrants from the 'insurtech' sector unencumbered by legacy infrastructure and thinking, insurers that fail to adapt centuries-old business models to enable the rapid, insightful and cost-effective consumption and analysis of data, will find themselves quickly falling behind the competition.

Whether it is to keep up or stay ahead of the competition, a general cornerstone strategy commonly adopted in industry is to rapidly advance the digital agenda. Coupled with the onwards march of global digitalization, this is generating a tidal wave of data that insurers are now increasingly looking to harness to power and sustain competitive advantage.

The ability to manage, provision and use this data in ways that go beyond traditional methods of data analysis in the industry, is fast becoming a key differentiator itself. Insurers are now seeking not only to improve operational efficiencies by breaking down and consolidating legacy data silos, but also for ways to make all their data work for more than it was originally intended.

To this end, advanced data science techniques and associated scaling technologies are now being increasingly applied to improve the quantity and quality of business insights that can drive better product structuring, price differentiation, claims management, customer engagement and risk profiling.

However, data science and digital solutions can only be as effective as the foundations on which they are built. Maximizing the returns on investment in data science and the digital agenda requires both a scalable and cost-effective means of harnessing data and a business culture that readily accepts and allows data science and digital innovation to influence its business strategy and disrupt the status quo.

Fostering such a culture provides the basis by which competitive advantage can be sustained in underwriting, exposure management, pricing, and general operational efficiency. It is our belief that a standardized and agile data management capability is crucial in achieving this objective.

2. A CHALLENGE OF SCALE AND COMPLEXITY

Data science is the collection of methods and tools that allows for a deeper understanding of what has happened in order to drive stronger predictions as to what might happen and make stronger prescriptions as to what actions should be taken. Businesses can leverage data science in a variety of ways to discover new opportunities, identify actionable business strategies and make data-driven decisions that would otherwise not be immediately obvious.

For insurers, some of the benefits include:

- Enhancing portfolio management through better visibility of business performance and exposures.
- Increasing agility of response to changing market conditions in relation to assessing and pricing risks.
- Improving claims management with respect to speed of response, fraud detection, and claims segmentation.
- Deepening the consumer experience through enhanced product tailoring, incentivization, cross-sales, and engagement.
- Identifying new business opportunities through analysis of consumer preferences and behaviors.
- Ensuring reporting requirements are met in a manner acceptable to external regulators e.g. reporting of CAT exposures.

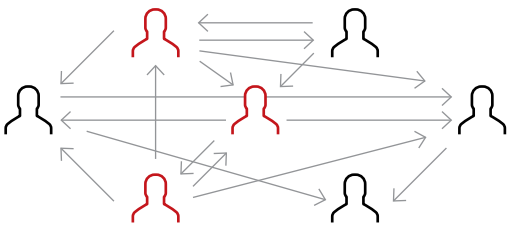
However, the intended insights and business value of the analytics can only ever be as strong and reliable as the data that underpins it. Improperly defined, inadequately sourced, poor quality data ('garbage in') leads to analytics outcomes that are incorrect, inaccurate, misleading, and otherwise meaningless ('garbage out'). Using such output at face-value causes non-optimal strategies and mis-guided business decisions at best, and financial and reputational damage at worst.

A production-ready analytical model or digital platform requires data that is properly described, sourced, cleansed, transformed and enriched. It is in overcoming these data curation and preparation challenges that insurers seeking to uplift their data science capabilities often struggle.

The problem is at least in part attributable to the way in which analytics has evolved in the industry, which is largely as an extension of price and exposure modelling. This has resulted in operating models where:

- The focus is on improving modelling capabilities while the development of a commensurate ability to curate and prepare data at scale is largely unrecognized or side-lined.
- The needs of traditional use cases dominate – outcomes remain largely beneficial to extracting maximum value out of existing models (e.g. finer location-based pricing), with some of the more disruptive use-cases remaining potentially unrecognized or unexplored.
- Data sourcing occurs in an unstructured peer-to-peer fashion through relationships and organizational knowledge.
- Data cleansing is duplicative, repetitive and time consuming, with data quality problems fixed at point of consumption rather than systemically at the root cause – no clear distinction of roles and responsibilities between data producers and consumers.

Figure 1: Unstructured data curation



The outcome is that analytics is often un-coordinated with silos of overlapping capabilities that exacerbate existing functional silos. This has inhibited the formation of unified platform strategies that would otherwise enable insurers to take advantage of the right technology that would support a truly data-driven organization. Instead, an over dependence on 'grey IT' (spreadsheets, end-user databases, etc.) has developed which significantly hampers the ability of an organization to manage data production, consumption and usage effectively.

Indeed, the tangled web of spreadsheet usage has meant that data curation and preparation have become critical yet extremely time and resource intensive activities at many insurers – by some estimates 80 percent of project time is typically spent identifying, sourcing, cleansing, transforming and enriching data, in preparation for modelling activities[1]. Even for an insurer actively seeking to become more data-driven, it is easy to see that this significantly raises the bar for economic viability which explains why data science remains difficult for insurers to scale across their organizations. Data science success stories still tend to exist in relative isolation when what is needed is a way in which beneficial results can be delivered consistently and sustainably.

2.1 LABS AND LAKES

Some insurers have implemented data science labs or 'centers of excellence' as a way of providing a funded ability to test the water in advanced data science such as artificial intelligence (AI) and machine learning (ML). Labs are ideal vehicles for bringing a data science concept from ideation through to delivery and production, perhaps even systematically resolving underlying data issues while doing so.

However, they are not intrinsically built to provide scalable and accessible data curation and preparation services for more business-as-usual functions in the wider organization, such as pricing and actuarial modelling. From the perspective of data management, unless appropriate steps are taken, a lab can in many ways be resemble yet another analytics silo.

Other insurers are looking to break down legacy data silos by consolidating data onto cloud-based data lakes that provide a platform for simplifying and streamlining the consumption of data. While this can have significant strategic benefits for scalable data provisioning, it is important to keep in mind that firm-wide lake migrations are costly multi-year initiatives and in practice, a lake must necessarily co-exist alongside existing legacy data sources for many years. To this end, two key factors bear consideration:

- The lengthy time-to-value of a data lake investment makes outcomes somewhat binary and therefore risky. The value of implementing a data lake lies in its ability to act as a single source of data and this value only starts to become realized when the lake becomes fully plugged into data processes that produce business outcomes (e.g. a modelling function). To this end, the cost of investment in standing up and continually expanding the scope of data ingestion, must be balanced against the time taken for the return on investment to be realized.

- Data co-location (whether physical or logical) does not imply business-friendly accessibility. Unless ingested data is managed and governed in a manner ready for business consumption, lakes run the risk of turning into swamps where there is little visibility as to what data is held, where it came from, its quality and the accountabilities / responsibilities involved. Building a data lake, in and of itself, does not mean data science and digital innovation will automatically scale – the organization must still implement associated cultural changes to ensure the lake is capable of supporting data usage at scale.

1. Harvard Business Review, hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists

Add in the business opportunities that can potentially be mined from the exabytes[2] of alternative data being harvested every day, and the burgeoning need to govern the legal and ethical 'flip-side' of analytics, it becomes clear that data science labs and cloud-based architectures are not enough on their own to sustainably deliver competitive advantage through data exploitation. An insurer's ability to do so depends on comparative advantages that can only be gained by building a data culture able to manage and govern data at scale across business lines, functions and the IT estate.

2.2 ALTERNATIVE DATA AND THE INTERNET OF THINGS

Alternative data, including unstructured data in the form of audio, video or free-text files, is an additional, rich source of information on consumers and internal operations. Coupled with the oceans of information generated by devices, sensors and controllers (the Internet of Things or 'IoT') and the rollout of 5G, society is in the middle of an explosion of data. Integrating all this data provides a significant opportunity for insurers to offer new services that are complementary to insurance products and help to service customers better and grow an insurer's top line.

The possibilities are endless, some of which include:

- Obtaining better claim ratios by leveraging the ability to derive structured information from pictures of the insured items, or telephone conversations held with insurance agents. If an insurer can combine this data with information held in policies and other structured sources, risks can be more accurately assessed and therefore underwritten, and fraudulent claims can be prevented before they occur through predictive analytics.

- Leveraging the ability to capture, store and use unstructured data to provide alternative engagement channels that enhance the customer experience. For example, it may make it easier for customers to get a quote by simply taking a picture of an insured asset, rather than having to submit structured descriptions of it.

- Enabling deeper insights into risks to provide finer risk-based price differentiation to customers. For example, using geo-location data to identify county-level tornado risks in the US.

- Leveraging IoT to offer more tailored products based on consumer habits and behaviors, and that can potentially include real-time premium adjustments. For example:

    - Monitoring driving behaviors and habits via in-car sensors to tailor car insurance products and real-time prices.

    - Monitoring batteries in fire alarms to modify fire coverage.

    - Using activity monitoring data from smartwatches to offer tailored health insurance products.

In general, integrating all the data captured for a specific customer (call centers, product portfolio, transactions, IoT), can provide a holistic view on preferences, issues, behaviors and commercial opportunities. Achieving these aims requires alternative data to be consistently curated, prepared and governed, including 'tagging' the information embedded in the data in a structured manner. Data management is key to ensuring that the metadata describing unstructured data sets, can be processed, aggregated and analyzed in a structured and scalable way.

2. Techjury.net/blog/how-much-data-is-created-every-day/#gref

2.3 MANAGING DATA LAWS AND ETHICS

The way in which data is used exposes insurers to both legal and ethical concerns around trust and fairness pertaining to the human impact, potential for misuse, and economic rights to the data. Increasing demand for data science and the sheer quantity of data potentially available for mining only exacerbates these risks. Therefore, it is fundamentally important to consider how the need to ensure legal and ethical data access, storage, sharing and usage, impacts how data should be managed and governed.

Figure 2: Laws and ethics

Some of the key legal risks include:

- Laws based on the physical location of data:
    - Data Localization / Cross-border sharing: Controlling the storage and transfer of data with respect to applicable laws governing the physical location that data may be stored in or transferred to / from.
    - Data Residency: This is where businesses choose, or data subjects request, to host data in a specific jurisdiction, thereby subjecting data to the laws of that country.
    - Data Sovereignty: Jurisdictions that claim data sovereignty apply all local laws to any data physically hosted in their country, this includes all data protection and data privacy laws.
- Data Privacy: Controlling the retention and usage of data with respect to the legal rights of data subjects.
- Data Security: Controlling who has and who should / should not have access to the data.

These considerations require operationally effective controls to ensure intended collection, storage, usage and cross-border transfers of data are properly risk assessed. However, from the sheer quantity of overlapping laws and regulations involved (due to the mix of location-based laws) it is possible for data to be subject to privacy and data security laws of more than one jurisdiction simultaneously which means significant challenges exist for managing and mitigating such risks, especially for those insurers with a global footprint.

Adding to this complexity are the ethical concerns that are yet to be enshrined within the legal or regulatory framework of a jurisdiction. This requires insurers to have a mechanism for performing value-based judgements on whether data usage breaches ethical standards.

A prominent area of ethical concern is in the application and influence of ML (Machine Learning) and AI (Artificial Intelligence) in business decision making. ML/AI outcomes depend implicitly on the scope of data used to train the system and there have been numerous stories in the wider economy of incomplete or unrepresentative training data causing unintended but controversial AI outcomes that exhibit ethically sensitive biases (e.g. towards race). It is therefore becoming increasingly critical that ML/AI implementations observe fundamental ethical principles of transparency and 'explainability' so that outcomes can be ethically audited. This requires (1) controls that review ethical risks and mitigations related to intended ML/AI outcomes prior to solution deployment and (2) mechanisms and techniques that enable the lid to be lifted on how ML/AI engines process

data to generate the outcomes (transparency) and the logic underpinning those outcomes (explainability).

A fundamental challenge of data ethics is therefore not only in applying judgement to what is in plain sight but to think through the potential second order affects and unintended ethical consequences:

- Tailoring products and prices to individual preferences, behaviors and price sensitivities is a double-edged sword. On the one hand it enables insurers to access more of the market by capturing customers who would otherwise not have been with a one-size-fits-all offering. On the other hand, it also increases the risk that well-intentioned outcomes can be perceived to be biased or based on unethical principles. As an example, strategies that price household premiums based on location may have unintended ethical consequences such as reinforcing racial inequalities or being perceived to be racially motivated. These can have severe impact on an insurer's reputation as well as incur costs through forced cancellation or modification of products and services.

- Price differentiation also has the potential to completely price certain consumer segments out of an insurance market (e.g. health insurance). This is an economic externality that can lead to significant swathes of the population left without critical insurance cover, which would then require some form of government intervention to address and as such, should be a consideration for private insurers with respect to future regulation as well as reputational damage.

Separately, it is also worth noting that the legal and ethical landscape pertaining to data usage is extremely fluid and fast changing – what is legal but possibly unethical one day, may not be legal the next, varies from country to country and between different types of data usage, and is often heavily principle-based and thus subject to interpretation.

What is clear however is that there is a significant international trend towards more law and regulation in this area:

- An increasing number of countries including China, India, Canada, Australia, Germany and Switzerland, have passed laws based on the physical location, or that constrain the cross-border transfer, of certain types of data. Indeed, the EU has firmly signposted[3] its intention for future data sovereignty laws across the whole union.

- Extending guidance[4] published the last year, the Council of Europe and some key observer states including USA, Canada and Japan, convened in July 2020 with a view towards establishing a legally binding AI usage treaty.

Given the impending growth in laws governing ethical data usage, it is therefore crucial for insurers to have appropriate governance and controls in place that can apply both legal and ethical judgements to ensure data usage is fair, lawful, and produces beneficent outcomes for the data subject. Furthermore, the complex and often conflicting and confusing picture of legislation also means that an integrated framework is required, as opposed to siloed solutions that are tailored for specific jurisdictions. That this should also be both auditable and scalable to the data science and digital ambitions of the organization, strongly underlines the need for a standardized approach that ensures such judgements are consistent and streamlined yet remain sensitive to local variations in legislation and societal norms.

3. European Digital Infrastructure and Data Sovereignty (European Institute of Innovation and Technology
4. Ethics Guidelines for Trustworthy AI (European Commission)

3. BUILDING SCALABLE FOUNDATIONS

What you get out of an analytics process is only as good as what you put in, and when business outcomes diverge unexpectedly from those predicted by the underlying analytics, the trust that decision makers place in data science becomes eroded.

This places an emphasis on the need to understand and control data science inputs at scale, which has never been more pertinent than with the recent and unprecedented coronavirus pandemic. This black swan event created an unexpected yet urgent need to rapidly and efficiently model and analyze portfolio exposures to business interruption claims, putting many insurers under significant operational strain in attempting to curate and make available clean, reliable and trusted data for the required analyses.

Here the concept of trust is key.

Simply put, unless data has been taken from known trusted sources, its meaning, provenance and quality should not be taken at face value and neither should the results of analytics based on such data. Otherwise there is increased risk of adverse analytics-based outcomes that would damage trust in the analytics capability itself, undermining efforts to build a data culture.

Therefore, for data science to be a successful as a strategic capability that provides a tangible differentiator for an insurer, it is not only necessary to grow data science skills and adopt sound scientific methodologies, but also develop a cost-effective mechanism of ensuring data can be trusted.

These can be summarized into two key data management requirements for controlling inputs to analytics processes:

1. understanding and obtaining the right data, and
2. fixing the data obtained.

3.1 UNDERSTANDING AND OBTAINING THE RIGHT DATA

For both model-led analytics such as a linear regression, and data-led analytics such as principal components analysis and k-means clustering, an understanding of data semantics and data provenance is critical for ensuring that the analytics has the right data:

- **Semantics:** The data that is needed (model-led paradigm) or being used (data-led paradigm) must be properly and unambiguously defined. For example, if the data scientist wants 'policyholders' to feed a pricing model, then an agreement must be made with the potential data provider as to what exactly is meant by this term: Is the requirement for policyholder names only, or are other policyholder attributes required? Does 'policyholder' include insured individuals, insured organizations or both? Does it include former policyholders? Is policyholder data required across all insurance products / business lines or to specific subsets?
- **Provenance:** Where the required data is sourced from must be determined. This not only increases trust that the data is from approved / authorized sources, but also that the data is representative of the population being analyzed. If the requirement is for policyholder data across household and commercial property business lines, then consuming data from sources that contain only household property policyholders would lead to improper outcomes, as would consuming policyholder data across property, automotive and health insurance products without the appropriate filters or analytical adjustments.

Data semantics and provenance become even more critical with the increasing application of artificial intelligence and machine learning methods in insurance. Not articulating and agreeing on data meaning, scope and sourcing of data leads to potentially libelous and reputationally damaging situations where an AI makes inferences based on the data it has been trained with, potentially leading to illegal or ethically inappropriate price discrimination based on gender, age, race, religion, or sexual orientation.

3.2 FIXING THE DATA OBTAINED

Once sourced, data may still contain errors or other quality issues. It is important that these are properly understood, so that they can either be resolved to the extent possible prior to consumption by an analytics process, be adjusted for in the analytical model itself, or be used to caveat the analytics outcomes.

Resolving and correcting for data quality issues is a data cleansing process that forms a critical part of the data preparation for an analytics process, as poor-quality data inputs manifests in outcomes that are biased, weakly explanatory, inaccurate, or otherwise simply incorrect.

Poor quality data is a primary limiting factor on the usefulness and reliability of analytics results. Having a cost-effective mechanism for understanding and resolving issues in sourced data is critical for improving the effectiveness of a strategic data science capability.

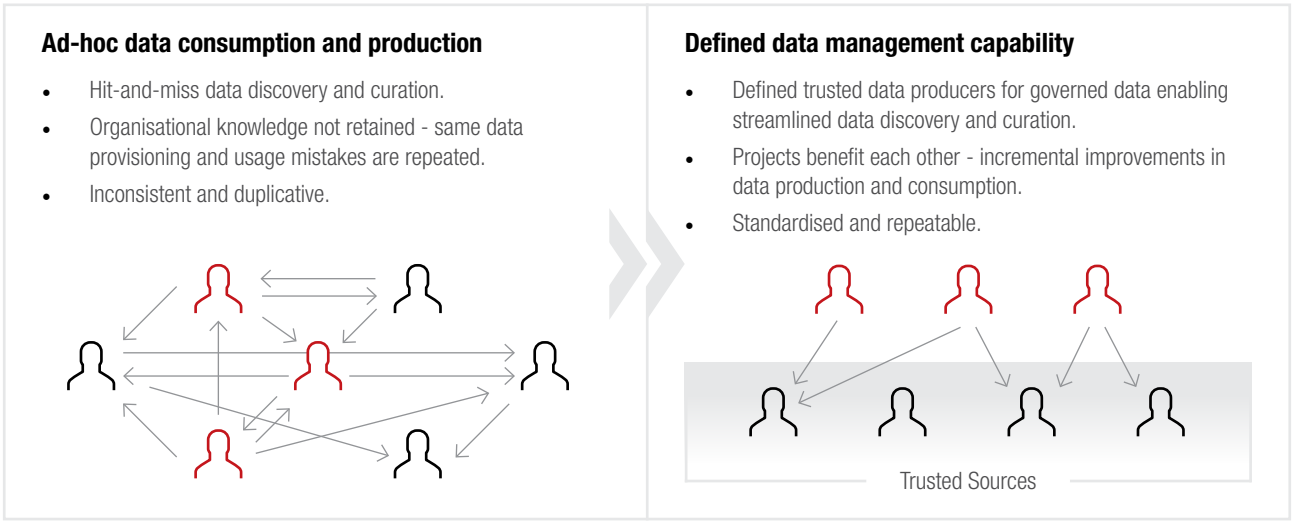
This therefore highlights the importance of ownership – identifying those that are accountable and responsible for not only defining the data but also in ensuring data is of acceptable standards of quality. Without clearly defined accountabilities for data quality, issues continually fail to be strategically remediated (at source) and are often just 'patched-up' at the point of consumption. Subsequent users of the same data are then forced to do the same cleansing operations, resulting in significant duplication of effort. On the other hand, an accountable owner that accepts and embraces the role takes steps to ensure data quality issues are fixed at source, and that controls that minimize data errors at point of entry are implemented and operationally effective.

3.3 SCALING THE SOLUTION

Without a vision for streamlining the servicing of these requirements, an organization's data science can easily devolve into a tangle of hit-and-miss, fact-finding engagements between analytics projects and potential providers, as each project seeks to find the right data from the right sources. Outcomes are inconsistent and effort is highly duplicative.

An integrated and standardized data management capability provides the hub of data services and expertise enabling all processes, analytics or not, to effectively outsource their data management needs. In such a setup, the integrated capability actively maintains a data catalogue, allowing service users to quickly understand what they need and where to get it, streamlining the discovery and fact-finding process.

Figure 3: Streamlined data provisioning



Such an approach allows for incremental gains as the knowledge (semantics and provenance) built from one project adds to the existing body of knowledge from others. From the data science perspective, the cost of data management is greatly reduced as data science projects benefit from the efforts of not only other data science projects, but the entire gamut of regulatory and transformational programs that occur in a modern insurance provider. For example, bad quality data is no longer remediated at the point of consumption by each data science project, but at the point of origination, therefore benefiting all future consumers.

Together with a consolidated data architecture, an effective data management capability provides the scale economies necessary for enhancing trust in data, and for making more data science projects cost effective.

Such a capability provides standardized and integrated mechanisms for describing what data is needed and how it is defined (semantics), for understanding where the data is sourced from (its provenance), for ensuring data can be trusted when it is used (data quality), and for the policies, process, accountabilities and responsibilities by which effective data management is defined, monitored, and enforced (governance).

4. TAILORING THE APPROACH

Lessons learned from the financial services industry

While the recent explosion of data management in the banking industry has been largely defensive in nature – driven mainly by regulation such as BCBS 239 (EU) and CCAR (US) as of writing, insurers have not had the same level of regulatory scrutiny with respect to the discipline of data management. Rather, data management initiatives at insurers have typically been delivered as small parts of wider IFRS 17 and Solvency II implementations, which have been largely been confined to finance, risk and actuarial functions with limited scope and integration with business lines. Such tactical implementations have resulted in silos of data management and expensive false starts, as these initiatives were never intended to be commercially relevant in a world where huge amounts of new data are being created every day.

Indeed, the increasing uptake of digital underwriting accelerated by the COVID-19 pandemic, coupled with market-led initiatives such as ACORD data standards, and common platforms for risk exchange and cover holder reporting (as examples), all reflect and drive the heightened need for integrated data management at insurers.

Although such market-led initiatives are of great benefit to the industry, they are however not primarily designed to provide lasting competitive advantage for participants, but to lower operational and transaction costs. As such, insurers will also need to complement adoption of market standards with an internal approach to data management that provides a cost-effective and scalable foundation for competitive advantage to be derived from data science.

This challenge is particularly acute at many traditional insurers where Microsoft Excel is still widely used, and data is often manually ingested into processes from disparate and ungoverned sources. Coupled with the lack of strong and mature data governance, the result is data that is used in silos and understood differently across the insurer's businesses and functions. It is perhaps not a surprise then that insurers spend disproportionate amounts of time curating and preparing data, e.g. CRM projects aiming to identify one version of their customer's names and addresses.

However, that insurers are generally behind banks in data management maturity should be viewed as a positive opportunity. The banking industry is littered with costly and monolithic data management implementations that have had varying levels of success and long-term benefit. Indeed, many data management programs in banking have resulted in output that have unfortunately been point-in-time and subsequently left on the shelf, inoperable, or simply wrongly scoped or specified for objectives other than regulatory compliance. Moreover, the scale of implementation has incurred, and continues to incur, significant costs.

Insurers are well placed to learn from and avoid these mistakes.

4.1 LEANNESS

Without clear traceability to business benefits, many components of a data management operating model are often seen as a 'tax' on the business. However, the lack of specific regulation in insurance forces the rationale for better data management to be framed in terms of the modelling / analytics / MI / reporting use cases it supports. This difference in rationale provides ideal leverage to align and prioritize data management activities to value drivers, and places emphasis on determining what is really important to the business.

In other words, not everything is a must have – it's better to do a few things well and build out from those strong foundations, rather than spread efforts too thinly at the cost of quality and sustainability of outcomes. The primary goal should not be to tick a regulatory box, but to ensure efforts have clearly articulated business use cases.

Data management implementations in insurance should be accordingly leaner than in banking. Practically this means three things:

- Insurers should prioritize data management capabilities and control the scope of implementation in a manner that focuses investment of time and limited resources into developing those that are most materially beneficial to supporting the successful delivery of business outcomes.
- A key part of delivering selectively is to ensure data management initiatives are directly tied to supporting the business outcomes of the rest of the change portfolio. This not only helps align with value, ensuring continuous stakeholder buy-in and participation, it also enables data management to be baked into the change portfolio 'by design', rather than as an afterthought.
- A value driven agenda means that it is often strategically sensible to combine the data management agenda with one that looks to consolidate the data architecture (e.g. a cloud migration) and/or one that looks to build out a data science capability. This allows business users to get a feel for an integrated data strategy and is again key to ensuring a 'by design' bottom-up rollout.

4.2 AGILITY

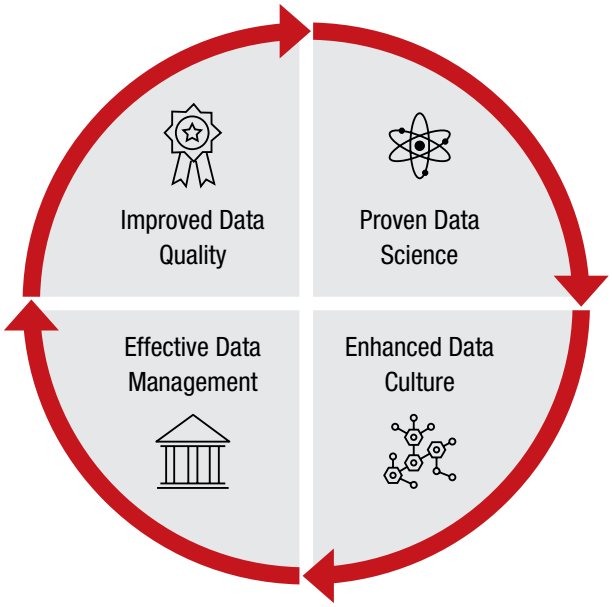
Stakeholder, buy-in, participation and motivation are critical. Due to the lack of outright regulatory drivers for such activity, it is important to deliver value early and frequency while the data strategy is being executed – a program that can take years.

This is more than just identifying quick wins. Rather than aiming for phased delivery of 'big-bang' functional drops, agile delivery principles should be leveraged to involve the stakeholders more closely and to iteratively prove out the data use cases.

Getting the ball rolling in this manner also has the substantial benefit of implicitly improving the data culture of the organization. Data science is effective when decision makers regularly make business decisions from the analytics insights, and the business outcomes are consistently in line with the expectations. Not only does an agile approach allow for continuous calibration of expectations and outcomes, it also actively and implicitly involves data owners and stewards, who are typically the same decision-makers that analytics projects provide insights to.

Strengthening the data culture facilitates the embedding of good data management behaviors, which drives up data quality, that in turn enables greater quantity and quality of data science, which further strengthens the data culture… adopting an agile approach thereby sets into motion a virtuous circle of self-reinforcing positive action.

Figure 4: Getting the ball rolling



As long as value drivers are being met, any technical inconsistencies of early deliveries with the strategic view can be resolved by reintegrating the solution further along the strategic pipeline.

Of course, there is a fine balance to be struck between the cost of reintegration and the benefit of a quick win. However, given that the most challenging aspect of a chief data officer (CDO) playbook is operational effectiveness and 'embedding', it is important not to underestimate the criticality of keeping stakeholders onboard in a data strategy.

4.3 CHANGE AND RISK MANAGEMENT

Adopting lean and agile ways of working are soft methods for keeping stakeholders involved and ensuring the build out of a data management capability supports the wider program portfolio. By contrast, a complementary and more top-down approach that pulls the wider program portfolio into line with the data management strategy involves integrating data governance with the organization's change / delivery frameworks:

- For agile methods this would be defined as mandatory data governance control stories that would need be adopted by all agile projects
- For waterfall methods, this would be implemented as data governance checkpoints in the requirements and design tollgates.

Such an approach provides a good mechanism for enforcing minimum data governance requirements as mandated by the data policy and enables projects to more formally think about incorporating the 'by design' considerations of data management.

Separately, key data management controls such as data quality management should be integrated into the organization's risk management framework. This ensures that the operational risks are formally mapped to these controls enabling a standardized mechanism for assessing their operational effectiveness and thus greater transparency as to the implications of control failure. This in turn provides another mechanism by which data management can be successfully embedded in an organization outside of a communications and training change plan.

4.4 DEMAND MANAGEMENT

Even with a lean and agile approach, it is likely that a data management function will not have adequate resources to meet all data management demands and build out the prioritized capabilities to the required scale. This is especially pertinent in the current economic climate where insurers necessarily need to focus on managing costs.

Being able to co-ordinate demand, manage expectations, and maintain tracking against a strategic and funded roadmap is therefore critical for ensuring a data strategy does not lose direction and become another false start, or otherwise become mired under its own weight.

To this end, it is important to ensure the data management office has an appropriately skilled program management capability, even in so-called 'business-as-usual', that is able to co-ordinate and maintain a short-term backlog of work without losing sight of longer term more strategic delivery themes.

This is especially important if adopting an agile mindset focused on short term deliveries. Without focus on the longer-term objectives, data management SMEs are likely to be increasingly asked to support data management activities in a haphazard and uncontrolled manner, defeating the whole rationale for an integrated approach – the SMEs may just as well be embedded with the functions / businesses / projects generating the demand.

4.5 LAWS AND ETHICS

Managing and mitigating legal and ethical risks related to data usage requires an integrated framework that provides for consistent results that can be appropriately calibrated to local variations in law and ethical norms.

A complicating factor is that data laws pertaining to the physical location of data and ability to share data across borders, will impact the teams that govern and manage data. Data that is required to be locally hosted, or has cross-border transfer restrictions, require local solutions, particularly for the management of data quality and data quality issues.

Ensuring streamlined governance of data usage in a manner that is sensitive to local legal and ethical risks, mechanically requires:

- Defining and governing policies and standards around fair usage of data and beneficence as a reflection of core values and the overarching code of ethics

- Ensuring appropriate data standards, data management skills, and supporting tools are in place for local teams

- Implementing standardized controls to ensure potential legal and ethical consequences of intended data usage, including data science uses, are reviewed, and appropriately mitigated prior to the usage. The multitude and confluence of laws, regulations and local ethics requires support for such controls to be performed consistently, objectively and at scale. It is therefore important to identify and categorize data in a manner that enables rules-based detection of legal risks and application of ethical decisions. Note the emphasis on locality with respect to ethics – it is important for ethical judgement to contain some local flavor, reflecting cultural differences

- Ensuring such controls are automated (to the extent possible) and building upon existing controls where they already exist. For example, ethical impact reviews could be built out from existing privacy impact assessment controls. This helps to reduce to reduce the burden of controls on the organization

- Enabling transparency and 'explainability' of processing by use of appropriate controls, tooling and training / upskilling on advanced design and implementation techniques that make this possible. This is particularly relevant for 'black-box' ML / AI -type applications where an outcome is the aggregate statistical result based on whatever data has been used to train the engine. In these circumstances, it is becoming increasingly important for appropriate data governance of AI usage, in order to retain the ability to 'lift the lid' on these applications so that the business rationale behind the outcomes can be fully understood and legally / ethically audited. Techniques that can be applied include local interpretable model-agnostic explanations (LIME), deep learning important features (Deep LIFT), layer-wise relevance propagation (LRP).

4.6 DATA LEADERSHIP AND CULTURE

Although data is at the front, back and center of a modern insurance business, there is often lack of accountability for data in a single senior leadership role. This dichotomy exacerbates the challenges of embedding good data management in industry and goes at least part way in explaining why insurers often struggle to embed data management on an organization-wide basis. A data leadership vacuum hampers the ability of an organization to internally 'evangelize' and champion the data agenda and stymies the development of a strong data culture as a result.

While the role of CDO is well established in banking, the role or its equivalent is yet to be fully embraced across the insurance industry with patchy uptake and many abortive attempts. Historically, many insurers have instead resorted to rolling accountability for data into other CxO roles, most commonly CIO, COO, or a chief data scientist / chief analytics officer (CAO) role where one exists.

While such an approach is understandable from the perspective of cost control and organizational simplicity, the disadvantage of doing so is that despite the best of intentions, this psychologically places data management as a secondary consideration after those of technology, operations or analytics. Furthermore, although data management may be logically perceived to fall within a technology, operations or analytics domain, the requisite skillsets are very markedly different. In such a setup therefore, it is crucial that data leaders personally gain the data management skills and experience that are necessary to maintain momentum in the data agenda and avoid the tendency to fall back on prior modus operandi and comfort zones.

If an organization is serious about moving ahead of the competition and grabbing pre-eminence and comparative advantages that come with establishing a strong data culture, data must be placed on an equal footing to other more traditional functions (such as finance, technology, operations). This necessarily means appointing a senior leadership role (e.g. a CDO) that can speak to the business imperatives, champion the data management agenda, and work in close tandem with the senior leadership team, particularly the CIO and CAO, to ensure the data strategy is integrated and reciprocally beneficial across data management, data engineering, digital innovation, and analytics.

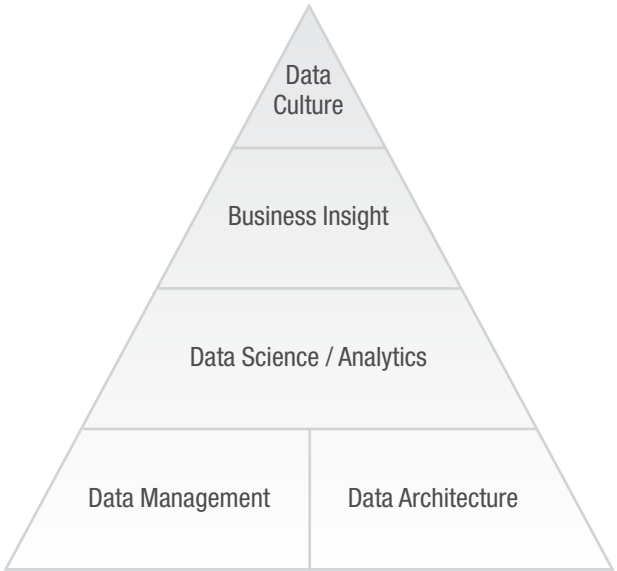
For more traditional organizations that are unaccustomed to such elevation of the data management agenda, this may at first seem a somewhat jarring change at both board and operational level. It is therefore crucial that the agenda is driven from the top, has the support and buy in of senior leadership, and is communicated and executed as a transformational journey that involves all levels of the organization.

5. CONCLUSION

Data is the lifeblood of a modern insurance business. In the hunt for rate adequacy, product innovation, and market differentiation, insurers are rapidly increasing investment in data science and digital platforms to leverage competitive advantage from the vast quantities of data that are potentially accessible.

Data is therefore at the heart of an insurer's ability to remain competitive in the long run. As such, it must be managed as an asset, not only to ensure that management of implementation, operational, legal and ethical risks do not overly 'tax' the organization, but also to build the foundation of a strong data culture.

Figure 5: Getting ahead and staying ahead – a pyramid of needs for building a comparative advantage



To get it right however requires a sustained investment not only in data science labs and consolidation of data architectures, but also in the organization, the people and the data processes involved. A data management capability works with a consolidated data architecture to provide a scalable, repeatable, transparent, and governed means for ensuring data is treated as a shared and trusted asset across the organization. Fundamentally, a data management capability enables the organization to understand and find the right data at scale and fix any issues with that data in a structured and 'reusable' manner, thereby enhancing scale economies and engendering a standardized level of trust in that data.

Data management initiatives can however incur significant change and run costs. It is therefore important to 'go lean' and get the ball rolling by focusing on delivering value early and frequently – this helps maintain stakeholder buy-in and participation which are particularly important considerations for insurers where there is a lack of regulation specific to data management.

An agile approach to building out data management capabilities that enables clear traceability of deliverables to value drivers, should therefore be adopted. Such an approach keeps stakeholders and engaged in the journey and the required cultural shifts. This not only enables outcomes and expectations to be closely managed ensuring delivery of useful insights that provide competitive benefit, but also helps to foster the data culture of the organization as an implicit part of change.

In the long run, it will be insurers that can most successfully build their data culture, that will be able to achieve the comparative advantages necessary to get ahead and stay ahead of the competition through data science and digital innovation.

CONTACT

Matt Hutchins,
Partner, Capco
matthew.hutchins@capco.com

Chris Probert,
Partner, Capco
chris.probert@capco.com

AUTHOR

Alvin Tan,
Principal Consultant, Capco
alvin.tan@capco.com

CONTRIBUTORS

Sanjiva Pereira, Managing Principal, Capco
Mauricio Sordo, Senior Consultant, Capco
Florie Kefalopoulou, Principal Consultant, Capco

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward.

Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and asset management and insurance. We also have an energy consulting practice in the US. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at www.capco.com, or follow us on Twitter, Facebook, YouTube, LinkedIn and Instagram.

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Gurgaon
Hong Kong
Kuala Lumpur
Mumbai
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

WWW.CAPCO.COM

CAPCO