



Data Protection
When Legal Meets
Data Analytics

Kenny Tung & Glenn McCarthy

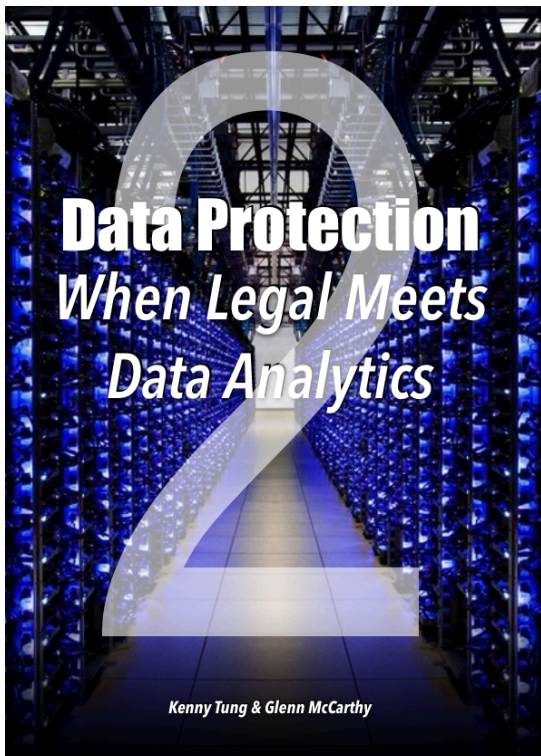


© *Kenneth Tung & Glenn McCarthy | 2022*
Cover photo: Google Data Center
Published by LegalBusinessWorld

Summary	4
IV. Bringing It Together	5
Benchmarking Is Only a Step to Finding Your Game	5
The New IT Is IS	8
First Steps in Data Protection	18
Notes	33
About the Authors	34

Complete Series

Read the complete series *Data Protection - When Legal Meets Data Analytics* by Kenny Tung & Glenn McCarthy. Go to the Legal Business World [eBook library](#) and read the book online or download your copy. (As of March 2022 available: *Part 1 & Part 2*).



Summary

This book follows a discussion of the data protection challenge from the perspectives of Legal and business strategy in Part 1 and an investigation of the post data-proliferation trade-offs experienced by businesses in Part 2. After arriving at the notion that each company must find its own game that fits its reason of existence in an ecosystem, this Part explores a data analytics pipeline that is integrated with business strategy and operations with an illustration of the roles of stakeholders in the organization and ecosystem. Despite how underdeveloped data governance technology is relative to numerous other technologies deployed, the last section provides recommendations for the data protection function to start with, not the least having an HR roadmap. Also included is an insert on lessons learned from a cautionary tale. Finally, to help review the key points in the three parts of the book, readers can find a list of questions at the end to stimulate their thoughts on next steps.

This (e)book series is based on a workshop taking an interdisciplinary perspectives of data protection delivered at a Legal Function Transformation Round Table subgroup on Oct.22, 2021.

IV. Bringing It Together

Benchmarking Is Only a Step to Finding Your Game

The above are some salient considerations for corporations in striking a) the right balance between their in-house data infrastructure and third-party systems as well as b) the delicate economic and strategic balance between a digital go-to-market DTC e-commerce vs. leveraging marketplaces and platforms for data. Some platforms like AWS also offer third-party data systems over the cloud.

To find the balance right for it, a corporation will need to be holistic to make strategic decisions on not only what are core activities but what constitutes core data, and overlay the strategic stance with operational and financial decisions to allocate the capital to build or subscribe to the necessary data technology across access, analytics and insight to arrive at strategic and operational outcomes. Based on the financial profile of the business case, the ratio of operating expenses to capital depreciation and amortization, investor's timeframes, objectives, and return thresholds will be key variables for these critical decisions.

In these trade-offs, corporations will decide the appropriate level of vertical integration for data analytics just as they do for manufacturing or logistics. Development of third-party data companies, "Data-aaS"

and "Insight-aaS," to serve purposes beyond traditional system integrators is already changing the economics of the data analytics value chain. Just as a transportation fleet does not drill for oil or own a refinery, a pipeline and a network of gas stations, not all businesses need to integrate scraping and crawl engines or cleaning, structuring, and tagging raw data into their operations. For many businesses, the cost and compliance risks of dealing with data further upstream will not be worth the returns.

The degree of integrating throughout the data supply chain depends in part on the company's core competences, the economies of scale and state of play in the data analytics market and how these two spheres may intersect. In some industries, disruptions and innovation convergence may create the opportunity for a player to take a swing to integrate. The auto industry, well known for a long and complex supply chain, is witnessing a leader in full self-driving (FSD) collecting massive on-the-road scenarios and designing its own supercomputer silicon to complete training data sets to round off the long tail scenarios.

Success here will be rewarded with an algorithm that wins rapidly across new geographies and markets, feeding a network effect and transforming industries as Apple has hastened disruptions in numerous legacy physical products like phones, music players, cameras, calculators, etc. The "first-mover" advantage in FSD will extend beyond the auto market to enable entrance into the "robo-taxi" business,

and an FSD champion will lead in an industry S-curve larger than that in electric vehicles. The value propositions in situation awareness and navigation will also be part of the principal competence if Tesla is intent on transforming robotics and adjacent fields. Such a vision and strategy should be worth all the marbles to play for with an integrated data analytics value chain.

Yet, not every company needs to or can be Tesla, provided that it does not throw the proverbial baby of interacting with and engaging customers out with the bath water of non-core data-to-insight operations. Trade-offs in more complex or interconnected contexts may appear as wicked problems, impossible to solve. Nevertheless, whether an organization positions itself for an infinite game or limps along from one short term metric to another, the finer data analytics can still enable it to chart a course guided by minimax solutions or yielding the minimal negative outcomes. In the meantime, staying in touch with the JTBD at the business and product level engages a different calculus complementary to data analytics. At any rate, the fledging data analytics infrastructure and supply chain still need to serve the purpose of the organization's existence.

What help decision making in the C-suite are insights to ask the right questions and predict answers toward strategic outcomes. Data analytics and all accompanying costs and consumption of management and resources give direction to deliver that. In the search for outcomes, not all data play equal roles. Therefore, much of the non-core

acquisition, processing, and analysis of data may be accomplished by market intelligence firms. As the knowledge-to-insight meme graphic in Part 1 demonstrates, the further along the data analytics process (Data→Information→Knowledge→Insight→Outcome Oriented Impact), the closer the use case and data approach the mother ship of the data analytics fleet, and the more proprietary and controlled it will be.

The New IT Is IS

Sorry, it is not Information System, but Integral to a Sustainable strategy and ecosystem. The forces of Industrial Revolution laid the ground in manufacturing efficiency (e.g., Henry Ford) and management by metrics (post-W/W II private and public sectors). These forces also empowered the finance department to evolve from a profession that last invented double-entry accounting into a strategic function in an organization as industries climb the quality and efficiency curves and navigate supply chains with metrics. From teams of accountants armed with slide rulers and gridline papers, then calculators and spreadsheets, to masters of enterprise resource management systems, Finance enhances its role in decisions with better KPI guidance in business processes and at the corporate level.

The same transformation is taking place for Information Technology. Its reason for existence is now far beyond maintaining colleagues' devices, email and access to and service level of corporate systems.

The convergence of technology like computing power, the IoT, ML and other advances in AI, joined at the hip with decades of codes and other tools, is enabling clients and colleagues to wield the proverbial telescope and microscope to see further and better with connectivity, granularity and approaching real time. Complemented by discipline of insights such as the JTBD, business can now hone the art of prediction with more efficiency. Similarly, the rise of BI from a supporting cost center (some with roots from corporate librarians) to the star in many board meetings tells an adjacent story at the tip of the spear.

Much in this progress is attributable to data analytics transforming the economics of decisions, resulting in part from the precipitous drop in the cost, and hence the rise of efficiency, of generating options and predictions. Yet, equally critical is management and deployment of data, just as other important resources, in ways integral to the strategies sustainable to successful ecosystems. Keeping eyes on the prize helps leaders to appreciate that the drill is just one way to make a hole, or to avoid seeing everything as a nail with hammers in hand.

A Cautionary Tale in Deploying Data Analytics

Around the time of completion of this article, IBM announced the sale of its Watson Health data and analytics business for a reported amount around US\$1 billion and put an end to much speculation and hand wringing in the media for over a year. While IBM will continue Watson supercomputer's AI solutions in other sectors such as

robotics process automation and Legal, this retreat from healthcare goes beyond the more than US\$4 billion of data assets like the largest insurance claims data that IBM acquired after rolling out Watson.

Watson Health represents a classic case of putting technology as the cart before the horse, in this case, of user value propositions.[1] All too often, even in a case of a sophisticated player like IBM, technology project development ends up with prototypes and solutions that look more like the original model and source code and lacks orientation to people, process and touchpoints for applications on the ground. The resulting disappointing adoption rate in technology, lackluster outcome and performance is not uncommon.

Before the next book or business school case study on this story, the media, especially in the healthcare field, is already teeming with discussions of how data analytics, and more generally AI, can be used to transform the sector, as well as lessons of missteps in leveraging technology. Some lessons include:

Leading with too broad, general, scattered and ambitious an objective – While the purpose to help the healthcare industry to solve stubborn and complex challenges like cancer is respectable, IBM never clearly defined the questions to ask on where to take the technology. It was the proverbial letting data scientists or coders run loose while disconnected from the people who plough the domain and really under-

stand where the pains and opportunities lie. As reported, Watson's problem stemmed in part from IBM's emphasis on large, difficult initiatives, rather than narrower objectives. As it turns out, ML today is deployed regularly to deliver solutions superior to the more bespoke version, but mostly on specific well defined tasks.

Management's technology illiteracy – Watson was created to identify word patterns and predict correct answers to questions, and its AI is more along the branch of NLP. Indeed, the healthcare specialists who worked with IBM in the field were deeply impressed by Watson's performance on Jeopardy! They based their expectation on Watson as an omniscient master in knowledge management which was further inflated by the Watson marketing team far beyond the performance hype. Although the original Watson developer cautioned that winning a trivia game was a great distance from tackling data, information and expertise in healthcare, the management told the developer that he did not understand presumably the market potential of the technology. IBM might be holding the NLP tool as the proverbial hammer and seeing the world as a knowledge management problem while other solutions like robot process automation, value-based care solutions, and ML are required. At any rate, marketing hype eclipsed design thinking.

Failure to appreciate and connect with users and their experience – This goes beyond building trust from clinicians that Watson will not be the HAL computer in 2001: A Space Odyssey and replace the profession-

als. Watson's performance in a solution based mainly on NLP failed to generate insight and options that swayed the professional diagnostic and prescriptive orientation, let alone delivering a greater than 10X performance to overcome resistance to change. Solution providers need to understand how clinicians operate on a day-to-day basis. Clinicians generally spend a couple of hours with patients electronic health records (EHR) for every hour with patients, and have no need for another administrative burden to soak up their time.

Drowning in the data ocean – As it is clear from many established companies and start-up pouring resource to clean and sort the rising ocean of data, the Watson effort was frustrated by the complexity, absence in organization and formats, and gaps in the genetic and other healthcare and insurance databases. Data was commonly decentralized, unstructured, and often in handwriting. "A patient could be going to Dr. Smith in one hospital system to treat his tumor for lung cancer. But that doctor, and the EHR system that Watson was looking at, didn't know that the patient also went to the hospital two weeks ago for breathing trouble, because they might be on different data sets." [2]

Wanting in humility and agility – This seems more a lesson in corporate culture than data strategy, but then it is also where the proverbial buck stops. The problem may manifest in many forms, such as confirmation bias, authority bias, Dunning-Kruger effect, failure to ask the right questions, group-think, sunk-cost fallacy, to name a few. The absence or breakdown of a feedback loop to reach out to stakehold-

ers who sustain an ecosystem can easily mistake cures for side effects or symptoms for causes. The lack of a collective ability to harness opportunities from technology aggravates a problem when the organization lacks the humility and culture to embrace processes to experiment, learn from failing fast in order to course correct, or in other words, building design thinking into technology and product development. This problem's evil twin is readily recognizable by compliance folks when signs of product design and supply chain issues escalate into run-away liability disasters. This lesson reiterates the importance of integrating data analytics to connect with stakeholders on organizational purpose, strategy and processes.

IBM putting its proud history and marketing before on-the-ground outcomes was unfortunate because as the most visible AI project in clinical healthcare, it casts an oversized shadow on the emerging sector. Today the ecosystem and technology, even in NLP, has advanced far beyond the clinical AI in 2011 into which IBM's naiveté and ambition took Watson. Today, a majority of hospitals have an AI strategy in place, continuing to attract massive capital flows financing entrepreneurs with deeper appreciation of the healthcare sector and solutions. While fielding options in technology much more advanced than Watson, the healthcare sector, with hope, will be more inclined to put use cases and stakeholders' interest ahead of technology. Perhaps Tesla's humility in its approach to FSD, illustrated by its founder's comment about not deploying AI lightly, can serve as a contrast.

Business must appreciate that, just as coding regular expressions to generate outputs from inputs must be grounded on serving a purpose, algorithms honed from the interaction of inputs in and outputs from test data sets is no different, in addition to avoiding biases as well as conflating causation with correlation. Success will come from realizing that purpose throughout the ecosystem and working hand-in-glove with key stakeholders such as users of analytics and ultimate beneficiaries like customers. In other words, lead with strategies integrated with organizational purpose in the ecosystems, connected into stakeholders and specific business processes, for whom/which insights from the data infrastructure will need to serve.

From the top down, product and enterprise value propositions must substantiate the organizational reason to exist in an ecosystem which in turn drives business processes and guides insight from data analytics. A snapshot of the process may be as follows. The prime stakeholder (e.g., business leader) and domain expert (e.g., retail specialist) should specify desirable outcomes (informing the data scientist on labeling), score outcomes such as value and failure functions (calibrating weights to labels), prescribe performance KPIs toward the launch of the application to be informed by the analytics, and identify intended application users for their perspective in validation. Analysts and data scientists will need to explore whether potentially useful data exists and collaborate with an ML engineer to assess whether there is sufficient data and computing power to deploy ML on datasets as part of the analytics. Data engineers, domain experts and

data governance leads will need to ascertain whether data access is feasible and advise the decision maker on whether it is worthwhile and sustainable in the context of the business needs and strategy. Toward the back end, software engineers and QC engineers will need to work with others to prescribe performance criteria for a minimum viable product/application to dovetail into the realities surrounding the relevant business process, and build and maintain it according to specifications.

From the bottom up, how data infrastructure can support business processes and use cases informs business on requirements and actions to maintain competitive and sustainable (including compliance) in a functional feedback loop. The key players in data infrastructure and analytics function must cooperate to 1) find value propositions in the business and organization, 2) follow up on impact on relevant business processes, costs and risks, 3) build and test solutions and portfolios to serve these business opportunities and reduce costs and risks, 4) build the analytics solutions into strategy map to keep focus and balance on a sustainable ecosystem, and 5) refresh the feedback loop with measures to enhance implementation.[3] Just as the colleagues who mine and surf the data to gain insights for the business, data protection and other governance leads must also be in the room and pull their weight in this bottom-up process.

The next table illustrates the importance of integrating the data processes into the organization, augmenting strategy and other busi-

nesses process. Taking an example from each of the categories of organization purpose and strategy, business function, business processes, insight, analytical output, analytical engine, data type, and data source, the illustration tracks how the top-down process can serve stakeholders and domain experts to connect organizational strategies to relevant business processes, and seek critical insights from an entrepreneurial exploration of the data analytics supply chain.

Purpose/ Strategy	Lead Full Self Driving/Launch Robo Taxis	Smart Cities- Efficient energy & traffic flow	Optimize Customer Retention & Renewal	Cashflow Efficiency	Capital Risk Management	Data Protection	Keep Engaged People
Functions	Product Dev.	Technology	Sales & Market	Supply Chain	Finance	Governance	HR
Biz Processes	Monitor driving environment	Coordinating traffic lights	CRM	Inventory management	Credit extension	Managing data access	Candidates screening
Insight	Predict steering, turning, & accelerating,	Butterfly effect of timing of traffic turning across oncoming vehicles	Predict customer churn	Optimization of JIT deliveries upstream	Credit-worthiness	Optimization of corporate data needs and regulatory risk	Fitness for roles & culture of success
Analytical Output	Surrounding objects & vectors	Optimal Timing of traffic lights	Churn likelihood dashboard & customer flag alert	Supply chain dashboard upstream	Credit scoring	Dynamic multi-dimensional names & roles for access	Rank candidates by KPIs, shortlisting for interview
Analytical Engines	ML	Predictive modeling/ML	Predictive modeling	Prescriptive modeling & visualization	ML	Rules, workflow, monitoring for escalation	ML
Data	Video stream, sensor data	Video stream, municipal tabular data	Business tabular data	Business tabular data, sensors data, GPS data	Public and purchased data	Requests, authorization, exceptions	Video content, HR tabular data,
Data Source	Cameras	Cameras, other sensors, IoT, data warehouse	Data warehouse	Sensors, IoT, data warehouse	Third party API	Data warehouse, logs	Social media, corporate website, resumes

Integration of Business and Data Functions Across Use Cases

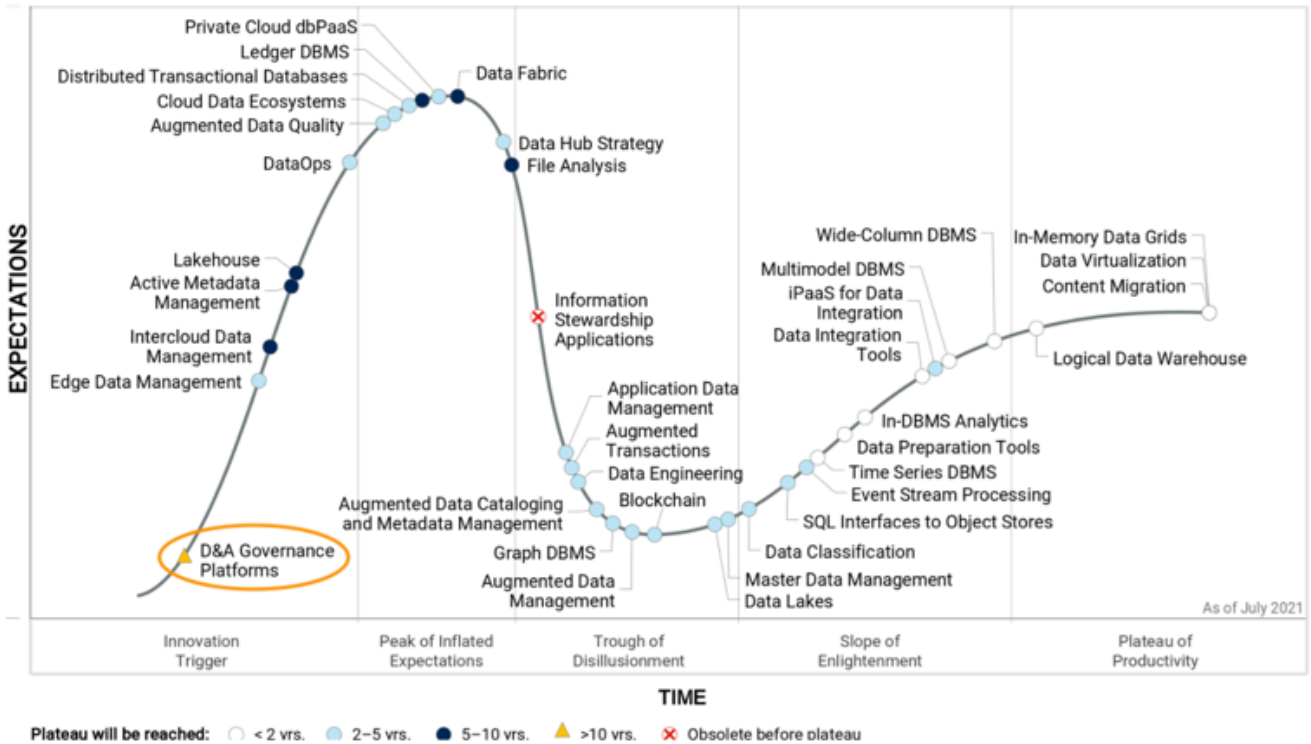
The development in data protection rules is a reminder that many stakeholders' vital interests underlying data and analytics governance differ from that of data management, with the latter driven mostly by business and, to most laymen, esoteric technology and markets. Nevertheless, they are not categorically conflicting, just as the erroneous approach many have taken toward business objectives and general

compliance. Data and analytics governance, when implemented as integral to data management and aligned with business strategy and operations, is complementary. It provides a framework to support decisions by further contextualizing decision rights, accountabilities and behaviors, akin to a RASIC protocol for the valuation, creation, consumption and control of data and analytics.

Unfortunately, the data protection function faces significant headwinds. Reminiscent of challenges in other compliance issues, different stakeholders and use cases will need to work out iteratively trade-offs in control, outcomes, agility and autonomy. Inevitably, organizations can navigate successfully only through cultural shifts toward harnessing the value of data and analytics governance. These challenges can delay the adoption of these data governance platforms and the process of their alignment with data and analytics initiatives and strategic business priorities. This is probably why the Gartner chart below ascertains that data and analytics governance will not arrive to be mainstream technology platform in at least ten years.

In the meantime, Gartner's recommendation also echoes the approach above to integrate data with business: "Design proofs of concept that will capitalize on the required critical technology capabilities. Identify the relevance of these technologies and their connection to business outcomes as a first step. Then look into their ability to support specific use cases (such as, risk management and compliance)." It also offers the lessons to ground the technology deployment on the

Hype Cycle for Data Management, 2021



Source: Gartner (July 2021)

organization's data and analytics governance strategy, control proliferation of tools and solutions, and leverage available market technology capabilities in end-to-end scenarios supported by emerging data and analytics governance platforms.[4]

First Steps in Data Protection

While organizations have yet to work the above playbook through their people and processes at both the strategic and operational levels, the emerging compliance demands will not tolerate in-decision or mere lip service to data protection. There is much the DP

function needs to do to fill the gap in and shape the intersection of the perspectives from business, legal and technology. From a technology perspective there are a number of vectors to inform the immediate call for action, but all are premised upon the need to establish a comprehensive data catalog.

Data catalog - To gain visibility of the data coursing through the organization's activities, companies should build up a comprehensive catalog. It is an organized inventory of data assets in the organization and creates a record of all of a company's data, its contents, location, and all other details about the data, including sensitivity levels.

Just as corporations go to great length carefully and correctly to keep track of fixed assets or finished and RIP inventory, including records of its composition, price (depending on the circumstance, cost or market value), source, and other key attributes, similarly data catalogs will lay the groundwork for data governance. While this is critical for compliance, it is increasingly table stakes to be in "ship shaped" and know exactly where some of the most valuable, and sensitive, corporate assets are.

In China, for taxation purposes, the government has begun to treat data in a fashion similar to the treatment of tangible goods through customs. Exporting certain data requires government approval, and it is more and more looking like a formal goods export transaction. Shanghai Data Exchange, opened on November 25, 2021, is China's

latest effort to facilitate data transactions between companies. Hence, keeping shipshape may help to monetize hidden asset value as well as avoid penalties and catastrophes.

To this end, a data catalog empowers organizations to leverage metadata to manage their data and enables data professionals to collect, access, organize, and enrich metadata to support data discovery and governance. For example, PII can be generated by combining multiple data sets, e.g., geo-location data from a device ID with addresses where device habitually spends the night and regularly visits. The design and operation of a data infrastructure and the flow of data and metadata should build in measures to anticipate and manage revealing of PII in the interaction of the parts of the data.

Recalling the side bar above on data taxonomy in Part 2, as the data industry is in the midst of creating, standardizing and systemizing metadata, the domain stakeholders on governance and data protection need to take leadership in discovering, developing and updating these metadata fields in the data infrastructure in ways that will qualify under emerging compliance standards. Further, the governance aspects in metadata should be integral to business use cases, how the organization explores, acquires access to, stores and maintains data, industry practices and, of course, specific regulatory stakeholders like bureaus for particular industries and consumers. In other words, the governance metadata should be triggered when any data is engaged and not operate in an insular system separate from the

rest of activities in the organization. In addition to cover existing data types, a data catalog should be elastic and robust to incorporate emerging metadata and related applications to connect governance to, and stay in sync with, other use cases.

On the foundation of a comprehensive, robust and elastic data catalog, companies will need to develop the following capability vectors to meet the data requirements that may apply from time to time.

Encryption and Data Masking - Ensure adequate capability to stop personally identifiable and other regulated data from being passed through the organization and to external parties without being properly protected. And target to execute this measure at or as close as possible to the source of the data.

Encryption works by encoding the original data with the help of sophisticated algorithms that convert it to unreadable text or ciphertext. A decryption key would be needed to revert to a readable format. Encryption is used to protect sensitive data, such as payment card information (PCI), PII, financial account numbers, and more. Data masking, also called data obfuscation, is a data security technique to hide original data using modified content. The main reason for applying masking to a data field is to protect data that is classified as PII, sensitive personal data, or commercially sensitive data. However, the data must remain usable for the purposes of undertaking valid test cycles. Data masking meets the requirements of most privacy laws such as

GDPR, PIPEDA and CCPA.

There are different techniques and approaches for data masking based on the type of data and the business requirement. In some cases, protected health information (PHI), PII or PCI data can simply be replaced with other characters and symbols resulting in only the sensitive data being anonymized and ensuing the data can be mined for value, leveraged for marketing or consumer analysis, and shared inside the organization. In other cases, where sensitive or secretive facts or financial figures are included in the data set, the data may need to be randomized or even partially deleted, potentially resulting in the loss of meaning of the dataset itself. In the case of masking an individual's name or address as part of a 360-view of an individual, typically the majority of the data value can be retained as the key demographic information remains intact. If a master key is assigned, the business can reconnect the individuals' records to their personal identification. One note of caution is that in some cases, third parties are able to join masked data with other data sets and derive PII.

Data Lineage or Traceability - Know where the data came from, where it went, and who has accessed it is critical to complying with many current and emerging privacy and data protection laws, e.g., the right to be forgotten. Data lineage is essentially a map of the data's journey through the organization. It typically includes the data's origin, the stops it made along the data pipeline, and an

explanation of how and why the data has moved over time. When properly implemented and operated, data lineage tools will provide the business with an understanding, recording, and visualizing of data as it flows from data sources to consumption points.

Organizations should incorporate traceability into data in systems and business processes. As the regulatory landscape becomes more complex and global, this capability will be essential in complying with differing requirements from multiple jurisdictions. Tracing data may also borrow from the expertise and capabilities such as tracking food and ingredients through supply chain, registering conflict diamonds, and artwork provenance. One day, perhaps blockchain technology can be deployed at scale in data protection and other use cases.

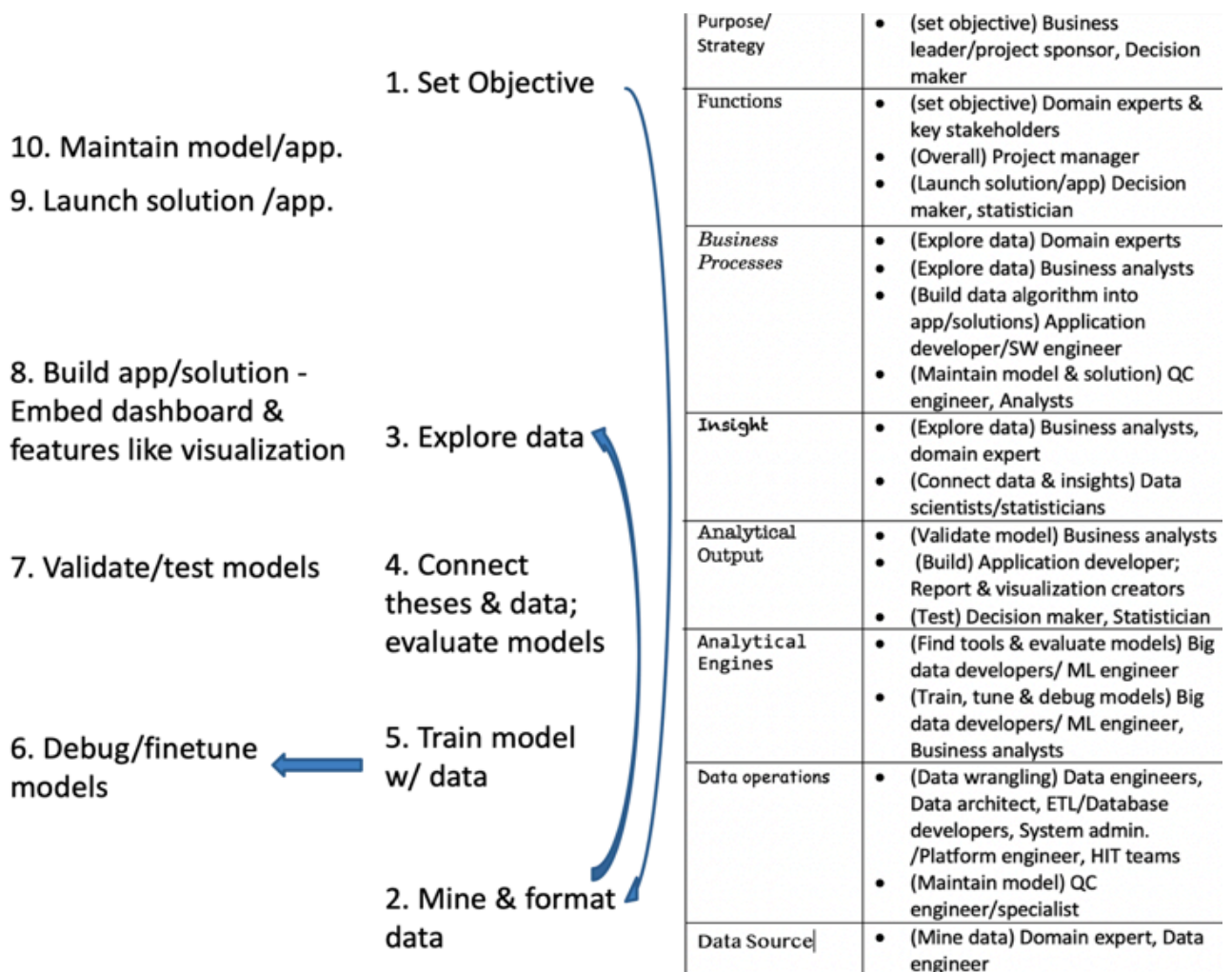
Access Protocol - Like an agile soccer team playing "one-touch" football, authorizations for data access inside an organization should be established based on need and not by default. Rather than granting access to data based on job title or department affiliation, the norm for data access authorization access should be explicitly requested and authorized. The design should distinguish data access protocol from business process decisions rights as the latter role may not even need to touch some data. The table above demonstrates a long runway to enhance data use cases in each business process, and opportunities to optimize data use protocols with some design thinking based on the experience of all use case stakeholders. Here again, Legal or other governance leads should work hand-in-glove with other

stakeholders in all business processes to button down each data use case across functions and from the operational to the strategic. Neither Legal nor IT can issue these protocols without thoroughly appreciating and engaging with the business processes and relevant stakeholders.[5]

On the foundation of a data catalog, three vectors above, once properly implemented as part of the business strategy and processes, supported by the right technologies and robustly maintained, should provide an elastic base case to support compliance with legal and regulatory requirements while equipping an organization with proper data analytics for strategic and operational business decisions.

Also, an effective data protection function cannot neglect to build a bench of data analytics talent to succeed in the era of data driven decision-making. Granted, the discussion above on trade-offs suggests that most companies will not be able to compete with the major players in data infrastructure and analytics, especially the Big Data platforms, as employers. Indeed, to wrangle the ever-advancing data streams and to modernize and maintain a data infrastructure, costs and risks, across strategic to operational needs as well as risk management, require a surprisingly large number of employees, contractors, and ecosystems of vendors and partners. Nevertheless, an organization that starts from a top-down approach described above will be more ready to conduct a realistic assessment of the current state and adjust at each round of the top-down and bottom-up data strategy

process. Evolving along virtuous cycles, it will invest in the trade-offs, and optimize the placement of resources in the data analytics supply chain and industry. Across roles and data functions illustrated in the table below, businesses may allocate their data resource internally and externally as part of their business strategy. However, strategic and operational excellence can only be built on a corporate culture of digital orientation, adequate data literacy and a proactive and engaging HR pipeline.



Finally, to help pull together some key concepts from all parts of the article, the following questions may serve as a useful checklist to inform data related operations. Any one reader is not likely to have the answers to questions in all categories, but a well resourced and grounded data analytics and protection team will likely get to the answers and will even figure the next questions that need to be addressed. A common thread runs through the perspectives of business, legal and technology highlights - that coordinated integration among functions and business processes is a critical success factor in solving for complex problems like data protection. These questions will also help to engage stakeholders and people with the knowledge and knowhow, from the macro to micro, to shape data protection fit for an enterprise.

On Business & Strategy:

- Where is your organization? Along the industry S-curves? In the BCG matrix of malleability and predictability? How does your organization come out from an analysis of the competitive landscape such as the Porter Five Forces?
- What is the purpose and reason of existence of your organization (as distinguished from measures and propaganda)? In the case of economic enterprises, strategy is generally beyond just a short-term path to profit, and increasingly covering broadening stakeholders in complex ecosystems.
- How is the business organized to serve this purpose? Is this aligned

in the organization through a strategy map covering key stakeholder outcomes, value propositions and operational objectives, related business initiatives and processes, and competence and resources required?

- Who are the stakeholders whose input and feedback are required to guide corporate strategies (alignment of stakeholder vital interests to drivers like sustainable enterprise profitability and robust cost-benefit management like accounting for externalities across sufficient activities scope and over business lifecycles)?
- Where and how do the stakeholder interests and corporate strategy manifest in the various business processes?
- How does the organization cover and rank the interests of all stakeholders and is organized to resolve conflicts? Is the assignment, ownership and responsibility clear?
- What is the JTBD for each of these stakeholders, including customers? Current as well as sustaining ones? Are there alternatives in fulfilling a particular JTBD?
- What type of inquiry is needed to unearth and appreciate the JTBDs? How and to what extent does data play a role in this inquiry (data to discover the JTBD)?
- Where may data exist/be generated and impact fulfillment of the JTBD of any stakeholder (data to help fulfill a JTBD)?
- In the case of an external stakeholders like customers, what is the competition landscape to discover their JTBDs? Does that differ when it concerns data?

On Legal/Governance:

- Is data protection an integrated part of overall enterprise compliance?
- To what extent is compliance integrated into achieving the overall organizational purpose, at both the strategic and operational level (recalling, e.g., the elements illustrated in the U.S. DOJ compliance guideline)?
- Has your organization deployed measures, such as circumstantial reminders, keyed in to detect behaviors and to “nudge” actors (beyond reciting traditional blackletter rules and punishments)?
- What are the primary policies of relevant data protection regulations? Competition, privacy or individual interest? Cause or opportunities to influence the public or industrial policies on this front?
- What is the level of transparency and clarity in the formulation, interpretation and enforcement of relevant regulations?
- Are there guidance or policy on data protection inside applicable industry or sectors? Or in adjacent space? How can you go about to configure measures to gauge regulatory standards such as sensitivity of data? How relevant and important is it for the organization to be capable to respond to certain demands, e.g., trace and track certain data and implement someone’s right to be forgotten?
- Are you building databases for emerging regulatory actions and integrating the same into the risk analysis?
- Getting into gear with the organization – Where is the legal/compliance function on engaging the rest of the organization in a joint

enterprise and ownership of problem solving rather than playing the traditional roles (and perpetuating the perception) of being the oracle, police and clean-up crew?

- Where are you on defining, collecting and tracking incidences of data protection failure modes? Is the rest of the organization on the same page?
- Are you connecting opportunities/upside with risks/downside? At the level of the organization, business, products, segment, as well as stakeholders?
- At the organization level and over measured timeframes, are you approaching opportunities and risks as a portfolio rather than isolated cases?
- Is the data protection function robust enough to be open to experiments and tests for risks and detect failure modes quickly?
- Is there a feedback loop in operation to fine-tune risk assessment, just like refining a data model to optimize false positives and negatives?
- Is risk analysis adequately quantitative and are you maintaining a pipeline of data to support it?
- Is your function staffed with T-shaped professionals? Are you seeking similar talent elsewhere in the organization to collaborate with?

On Technology:

- How has data proliferated historically in your business and do you have a perspective of what that looks like going forward?

- What is the scale of data that your business model has been experiencing? Do the challenges, costs and risks associated with this approach to data fit the business model? Do you need adjustments in the business model, the data model, or both?
- What are the key data value propositions inside the organization, which are closely integrated with major and strategic focus, and which incur the most compliance risk?
- Does the company own and operate a Customer Data Platform (CDP) and engage in Programmatic Advertising? If so, what are the functions served by this operation, such as sales & marketing, product development, supply chain?
- Do alternative data channels exist or can they be created?
- What kinds of disruptive forces are at work in your industry? Is Silicon Valley funding disruptive technologies and start-ups to take aim at your business model? What role does data play in their strategy?
- Are you familiar with the different kinds of structured, unstructured, and semi-structured data in your company? Can you name 3 examples of each kinds of data and its use by the organization?
- What level of data analytics sophistication is your company at today? Is the company keeping up the rapid software developments of the advanced analytics industry?
- What is your company's cloud strategy and how will this impact where and how data is stored, transmitted, and accessed?
- What are the key ongoing data management and advanced analytics projects in the company today and is there strategic focus and

roadmap going forward? How is data governance being incorporated into these projects and the overall data strategy?

- Do you operate a DTC e-commerce site? If yes, what first party data are you capturing, where does it go, and who is using it? Which partners are you sharing data with? Do you sell data? Are your products connected over IoT and streaming consumer data into your data center? If yes, what data are they streaming and where does it go?
- How stable is the data supply chain from a business perspective? Are there players in your data infrastructure gaining negotiation powers in the supply chain?
- Who is managing the licensing of your software and cloud services procurement? Are your end user license agreement (EULA) terms aligned with your business strategy and written to protect against the fast-changing software ecosystem? Does your cloud provider have the right to learn from your data and use patterns or metadata from it to train its own ML models? If so, have you bargained with the provider for their access?
- Has your company put a value on its data? Can you imagine a day when you track and trace data like you do WIP inventory? If data has value and it is moving across borders when would that be considered a taxable transaction? The challenge here is that data does not have intrinsic value but only in the context of the use case, but one should plan for paying for value in data which is on the drawing board everywhere.

- How would the organization apply a financial lens to its approach to data? E.g., Cost to acquire, process, maintain, archive data, factoring in escalating costs of compliance? Expensing versus amortizing the capital investment?
- What is the status of your company's Data Catalog?
- Are you effectively deploying data encryption and masking?
- How advanced is your company with data lineage?

#KennethTung #GlennMcCarthy #data #legal #inspection #AI #Big-Data #datamining #DataProtection #DataProtectionLaw #Compliance #DataCompliance #DataAnalytics #DataStrategy #LegalProfession #strategy #lawyers #BigData #BigDataAnalytics #ArtificialIntelligence #MachineLearning #legal #LegalProfessionals #LegalDepartment #LegalDesignThinking #LegalManagement #LawTechnology #Law-AndTechnology #LegalFunctionTransformation #DataStrategy #DigitalAdvertising #ProgrammaticAdvertising

Notes

1. S. Lohr, IBM Is Selling Off Watson Health to a Private Equity Firm, The New York Times, Jan. 21, 2022. See also J. Frownfelter, Why IBM Watson Health Could Never Live up to the Promises, Medcitynews.com, Apr.8, 2021; L. Vespoli, Where Watson Went Wrong, Medical Marketing & Media, Sep.8, 2021. Leveraging data analytics in healthcare is now a crowded field. Financial & strategic investors are paying tens of billions of dollars to acquire healthcare data, especially in the virtual health, EHRs and clinical workflow solutions fields. In 2018, Amazon, J.P.Morgan Chase, and Berkshire Hathaway announced formation of a joint venture, Haven, expected to disrupt the healthcare space. The venture shut down in 2021.
2. Ibid, quote from Where Watson Went Wrong.
3. J. Sun, Capture the Real Potential of China's Data and Analytics Market Growth, Gartner Local Briefing, Dec.6, 2020, p.55.
4. P. Russom & D. Feinberg, Hype Cycle for Data Management, 2021, Jul.27, 2021, <https://www.gartner.com/doc/reprints?id=1-27UXJCDD&ct=211102&st=sb>.
5. Most lawyers are not trained or in practice to mine evidence in daily work, but the best of litigators do, often through working to be somewhat of an expert in relevant domains.

About the Authors

Kenny Tung is General Counsel at Lex Sigma Ltd., where he served as the China advisor to a top U.S. PE fund, and the Asia Pacific advisor to one of the world's top auto components companies and continuing to advise companies in strategic projects and transactions in the region. Kenny also co-founded In-Gear Legalytics Ltd. which helps legal departments of world class companies and law firms to explore and design strategy and process optimization to facilitate transformation of legal services. Projects cover consulting, capability assessment, workshops to address longer term issues, but a common stream concerns the design and implementation of corporate legal strategies. Recently, Kenny undertook an additional role as the Senior Advisor to SSQ in Asia Pacific, facilitating business development and alliance for law firms and management of legal departments, focusing more on people aspects of the people-process-technology spectrum. Previously Mr Tung served as the Chief Legal



Counsel of Geely Holding (during which time the department received the top award for Best Asian & South Pacific Legal Department 2014 by International Legal Alliance Summit) and before that as general counsel in the region at PepsiCo, Goodyear, Honeywell and Kodak where he fielded a vast variety of issues and projects and drove efficiency projects/practices. In 1994, he came to China as a lawyer with Coudert Brothers and led major projects such as the Shanghai GM JV negotiation.

Glenn McCarthy is an entrepreneur and private investor with a primary focus on software and big data analytics companies in China & SE Asia. Glenn is currently the CEO, of Early Data Market Intelligence Co. Ltd., based in Shanghai, and sits on the boards of a number of early and growth stage companies. A native of Boston, Glenn has been living in Shanghai, China for 23 of the past 26 years. During the 1990's Glenn worked for General Electric, responsible for World-Wide Sourcing Programs in Europe, China, Asia, and then globally for the entire corporation



More by Kenny Tung

Kenny Tung regularly shares his expertise on Legal Business World. Click on one of these articles and you will be redirected to his publications.

Going In-house: A Changing Marketplace

"These are the times that try men's soul." : Thomas Paine, The Crisis, Dec.23, 1776...

If "Software Is Eating the World," Is Legal Service on the Menu?

* "Why Software Is Eating the World", Marc Andreessen, The Wall Street Journal, August 2...

How the challenges of cybersecurity reflect those in the legal...

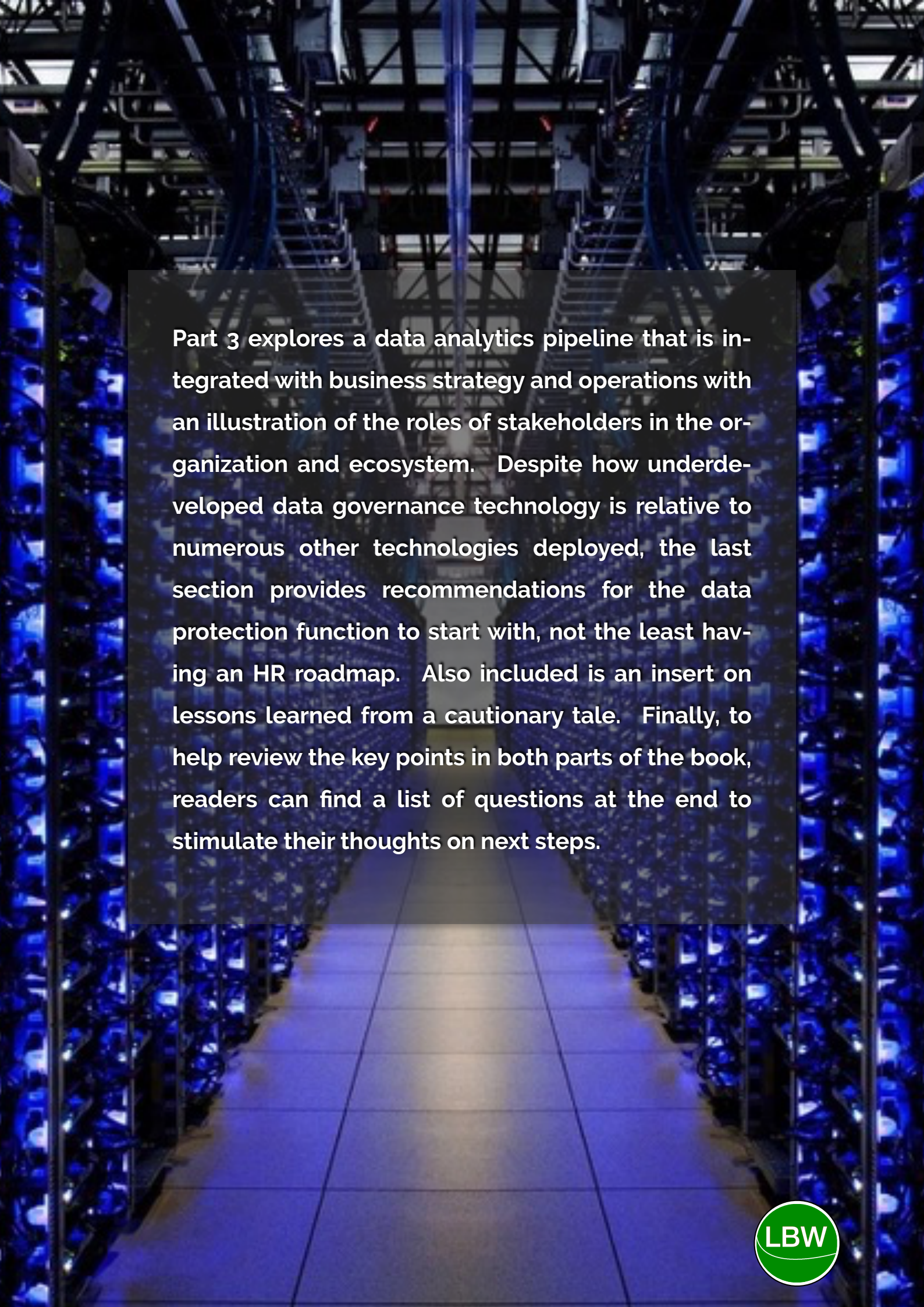
Cybersecurity used to be viewed as black magic. From a non-technical, user or...

Ecce Advocate - Reflections from Conversations in the...

The backward-looking, risk averse approach to the law, which is so common in...

Establishing a Legal Function Beyond Home Jurisdiction

Albeit less newsworthy today, news of surging and sustaining investments from...



Part 3 explores a data analytics pipeline that is integrated with business strategy and operations with an illustration of the roles of stakeholders in the organization and ecosystem. Despite how underdeveloped data governance technology is relative to numerous other technologies deployed, the last section provides recommendations for the data protection function to start with, not the least having an HR roadmap. Also included is an insert on lessons learned from a cautionary tale. Finally, to help review the key points in both parts of the book, readers can find a list of questions at the end to stimulate their thoughts on next steps.