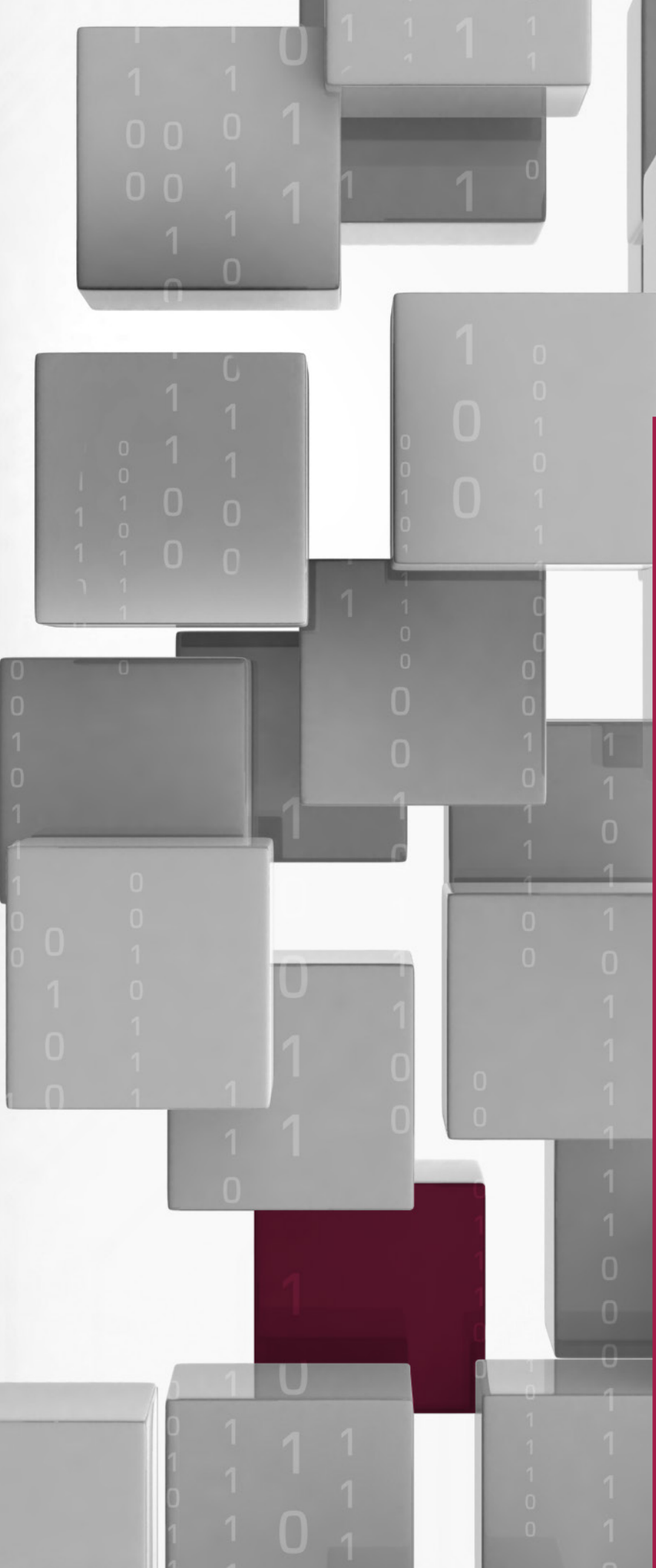


ISACA®
Trust in, and value from, information systems

PRIVACY & BIG DATA

AN ISACA WHITE PAPER AUGUST 2013

Improved decision making, faster time to market, better customer service and increased profits are just some of the benefits contributing to the explosion of big data implementation across enterprises of all sizes. The World Economic Forum describes the personal information garnered by big data as "the new 'oil'—a valuable resource of the 21st century." Big data analytics is the "new engine of economic and social value creation." Enterprises eager to reap the benefits of big data and its vast potential are recognizing their responsibility to protect the privacy of the personal data gathered and analyzed with big data. Risk and maintaining adequate mechanisms to govern and protect privacy need to be major areas of focus in any big data initiative. The comprehensive COBIT® 5 business framework for the governance of enterprise IT maintains a balance between realizing benefits and optimizing risk levels and resource use.

ISACA®

With more than 110,000 constituents in 180 countries, ISACA *(www.isaca.org)* helps business and IT leaders maximize value and manage risk related to information and technology. Founded in 1969, the nonprofit, independent ISACA is an advocate for professionals involved in information security, assurance, risk management and governance. These professionals rely on ISACA as the trusted source for information and technology knowledge, community, standards and certification. The association, which has 200 chapters worldwide, advances and validates business-critical skills and knowledge through the globally respected Certified Information Systems Auditor® (CISA®), Certified Information Security Manager® (CISM®), Certified in the Governance of Enterprise IT® (CGEIT®) and Certified in Risk and Information Systems Control™ (CRISC™) credentials. ISACA also developed and continually updates COBIT®, a business framework that helps enterprises in all industries and geographies govern and manage their information and technology.

Disclaimer

ISACA has designed and created *Privacy and Big Data* (the "Work") primarily as an educational resource for governance and assurance professionals. ISACA makes no claim that use of any of the Work will assure a successful outcome. The Work should not be considered inclusive of all proper information, procedures and tests or exclusive of other information, procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific information, procedure or test, governance and assurance professionals should apply their own professional judgment to the specific circumstances presented by the particular systems or information technology environment.

3701 Algonquin Road, Suite 1010
Rolling Meadows, IL 60008 USA

Phone: +1.847.253.1545

Fax: +1.847.253.1443

Email: *info@isaca.org*

Web site: *www.isaca.org*

Provide feedback:
www.isaca.org/privacy-and-big-data

Participate in the ISACA Knowledge Center:
www.isaca.org/knowledge-center

Follow ISACA on Twitter:
https://twitter.com/ISACANews

Join ISACA on LinkedIn: ISACA (Official)
http://linkd.in/ISACAOfficial

Like ISACA on Facebook:
www.facebook.com/ISACAHQ

Reservation of Rights

ACKNOWLEDGMENTS

ISACA Wishes to Recognize:

Project Development Team

Mario Bojilov
CISA, Meta Business Systems, Australia

Richard Chew
CISA, CISM, CGEIT, Emerald Management Group, USA

Francis Kaitano
CISA, CISM, CEN, New Zealand

Tichaona Zororo
CISA, CISM, CGEIT, CRISC, EGIT, South Africa

Expert Reviewers

Todd Atteberry
The Atteberry Group, USA

Goutama Bachtiar
Global Innovations and Technology Platform, Indonesia

Graciela Braga
CGEIT, Argentina

Girish Netke
CISA, A-N-G Computer Consultants, India

ISACA Board of Directors

Tony Hayes
CGEIT, AFCHSE, CHE, FACS, FCPA, FIIA, Queensland Government, Australia, International President

Allan Boardman
CISA, CISM, CGEIT, CRISC, ACA, CA (SA), CISSP, Morgan Stanley, UK, Vice President

Juan Luis Carselle
CISA, CGEIT, CRISC, Wal-Mart, Mexico, Vice President

Ramses Gallego
CISM, CGEIT, CCSK, CISSP, SCPM, Six Sigma Black Belt, Dell, Spain, Vice President

Theresa Grafenstine
CISA, CGEIT, CRISC, CGAP, CGMA, CIA, CPA, US House of Representatives, USA, Vice President

Vittal Raj
CISA, CISM, CGEIT, CFE. CIA, CISSP, FCA, Kumar & Raj, India, Vice President

Jeff Spivey
CRISC, CPP, PSP, Security Risk Management Inc., USA, Vice President

Marc Vael, Ph.D.
CISA, CISM, CGEIT, CRISC, CISSP, Valuendo, Belgium, Vice President

Gregory T. Grocholski
CISA, The Dow Chemical Co., USA, Past International President

Kenneth L. Vander Wal
CISA, CPA, Ernst & Young LLP (retired), USA, Past International President

Christos K. Dimitriadis, Ph.D.
CISA, CISM, CRISC, INTRALOT S.A., Greece, Director

Krysten McCabe
CISA, The Home Depot, USA, Director

Jo Stewart-Rattray
CISA, CISM, CGEIT, CRISC, CSEPS, BRM Holdich, Australia, Director

Knowledge Board

Christos K. Dimitriadis, Ph.D.
CISA, CISM, CRISC, INTRALOT S.A., Greece, Chairman

Rosemary M. Amato
CISA, CMA, CPA, Deloitte Touche Tohmatsu Ltd., The Netherlands

Steven A. Babb
CGEIT, CRISC, Betfair, UK

Thomas E. Borton
CISA, CISM, CRISC, CISSP, Cost Plus, USA

Phil J. Lageschulte
CGEIT, CPA, KPMG LLP, USA

Anthony P. Noble
CISA, Viacom, USA

Jamie Pasfield
CGEIT, ITIL V3, MSP, PRINCE2, Pfizer, UK

Guidance and Practices Committee

Phil J. Lageschulte
CGEIT, CPA, KPMG LLP, USA, Chairman

John Jasinski
CISA, CGEIT, ISO20K, ITIL Exp, SSBB, ITSMBP, USA

Yves Marcel Le Roux
CISM, CISSP, CA Technologies, France

Aureo Monteiro Tavares Da Silva
CISM, CGEIT, Brazil

Jotham Nyamari
CISA, CISSP, Deloitte, USA

James Seaman
CISM, RandomStorm, UK

Gurvinder Singh
CISA, CISM, CRISC, Australia

Siang Jun Julia Yeo
CISA, CPA (Australia), MasterCard Asia/Pacific Pte. Ltd., Singapore

Nikolaos Zacharopoulos
CISA, CISSP, DeutschePost–DHL, Germany

ISACA and IT Governance Institute® (ITGI®) Affiliates and Sponsors

Information Security Forum
Institute of Management Accountants Inc.
ISACA chapters
ITGI France
ITGI Japan
Norwich University
Socitum Performance Management Group
Solvay Brussels School of Economics and Management
Strategic Technology Management Institute (STMI) of the National University of Singapore
University of Antwerp Management School

ASIS International
Hewlett-Packard
IBM
Symantec Corp.

Introduction

Big data can be very powerful and have significant positive and negative impact on an enterprise. **Improved decision making, faster time to market, better customer service and increased profits are just some of the benefits contributing to the explosion of big data implementation across enterprises of all sizes.** Privacy breaches of big data can result in costly legal consequences for enterprises.

The definition of big data was given for the first time in a paper by Doug Laney.[1] He defined big data as data sets with three aspects that introduce specific processing challenges: volume, velocity and variety. Velocity is the speed at which data are created. This speed is increasing dramatically. During each minute in 2012, consumers spent US $272,000 on web shopping, and brands received 34,722 Facebook "likes." Variety refers to the diverse data types being processed. These have changed from simple files and relational databases to audio, video, sensor information, etc. Volume is the result of increasing variety and velocity. Today, enterprises are processing terabytes and petabytes of information.

In 2013, ISACA defined big data as data sets that are too large or too fast-changing to be analyzed using traditional relational or multidimensional database techniques or conventional software tools to capture, manage and process the data at a reasonable elapsed time. "Big data represents a trend in technology that is leading the way to a new approach in understanding the world and making business decisions."[2]

The World Economic Forum describes the personal information garnered by big data as "the new 'oil'—a valuable resource of the 21st century," and the analytics of this data as "the new engine of economic and social value creation."[3,4]

Enterprises eager to reap the benefits of big data and its vast potential are recognizing their responsibility to protect the privacy of the personal data gathered and analyzed with big data. Risk management and maintaining adequate mechanisms to govern and protect privacy need to be major areas of focus in any big data initiative. The comprehensive COBIT 5 business framework for the governance of enterprise IT maintains a balance between realizing benefits and optimizing risk levels and resource use. This framework can be applied very successfully to big data privacy requirements and challenges.

This white paper focuses on the impact of big data on privacy, privacy risk, big data privacy strategies, and governance of and assurance considerations for big data privacy.

---

[1] Laney, Doug; "3D Data Management: Controlling Data Volume, Velocity and Variety," gartner.com, 6 February 2001, *www.blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf*

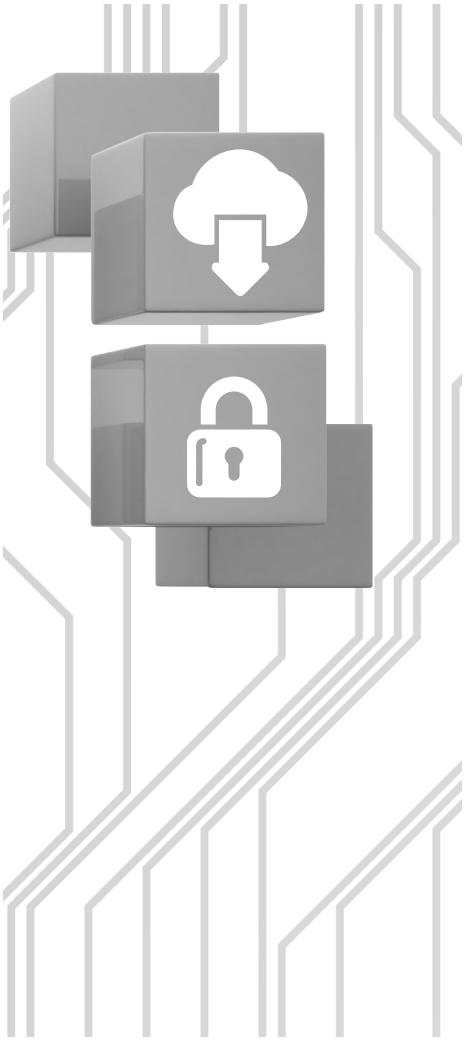
[2] ISACA.org, "Big Data: Impacts and Benefits," March 2013, *www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Big-Data-Impacts-and-Benefits.aspx*

[3] World Economic Forum, "Personal Data: The Emergence of a New Asset Class," January, 2011, *www.weforum.org/reports/personal-data-emergence-new-asset-class*

[4] World Economic Forum, "Unlocking the Value of Personal Data: From Collection to Usage," February 2013, *www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf*

Impacts of Big Data on Privacy

Across all industries, including banking, government, healthcare, media, energy and education, the exponential automation of business processes is widening the landscape of data consumption and analytics.



In the quest of enterprises to deliver acceptable and measurable big data return on investment (ROI), data that were stored in disparate online and offline repositories, in a variety of formats, are now available in digital format, ready to be correlated, aggregated and statistically analyzed in huge chunks of terabytes and petabytes in real time.

As data volume, data processing speed, data-type complexity, and privacy and security requirements continue to grow beyond expectations, enterprises are being forced to seek new ways to address legal, business and operational needs.

Big data is a significant force behind the rising number of enterprises that are making the decision to move data to the cloud and to use cloud-based analytics services and processing analytical databases, such as massively parallel processing (MPP) or symmetric multi-processing (SMP) analytical databases.

Big data has initiated discussions about international privacy and data protection laws. Currently, each region (European Union, USA, etc.), government and enterprise handles privacy and data protection in a different way. This geopolitical impact has forced enterprises to reconsider the way they handle and protect the privacy of individuals and the information collected about them and how enterprises implement their cloud-based big data solutions.

Big data adoption is also influencing how enterprises deliver IT projects. Most big data projects are technology- and data-intensive. The technology is complicated and the skills required to deliver are relatively scarce, which has resulted in project overruns and budget explosions.

The growth of big data has led to disparate repository storage of personally identifiable health records and credit card data details and transaction data. The storage and analysis of such data have increased the pressure on organizations to comply with data and privacy regulations, such as the Payment Card Industry Data Security Standard (PCI DSS), the 1998 UK Data Protection Act and the US Health Insurance Portability and Accountability Act (HIPAA). A pragmatic approach is required to address these big data compliance requirements.

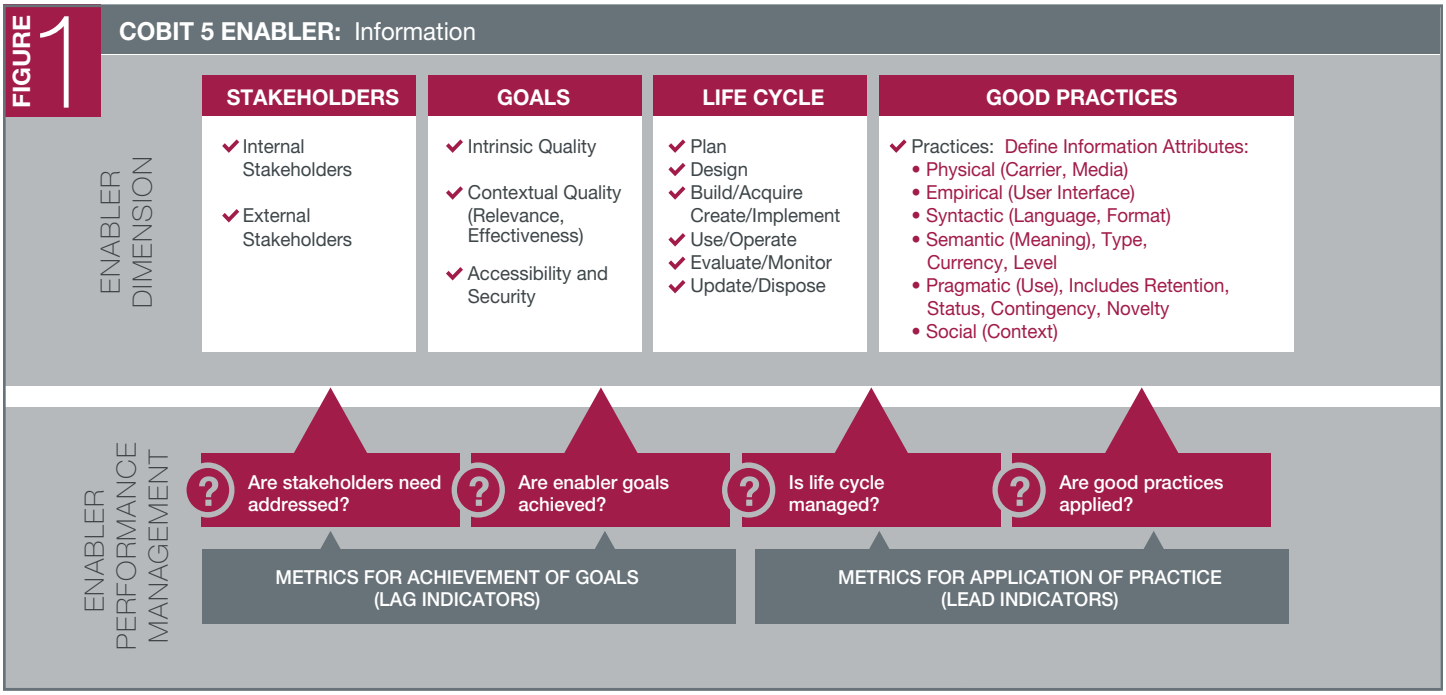
Dealing with data of this magnitude on a day-to-day basis and in real time presents a new frontier and the birth of data challenges, including the following:

- **Complex technological evolution**
- **Data privacy and integrity**
- **Security of the data at rest and in motion**
- **Availability and data system resilience (IT infrastructure)**
- **Incident response to and management of big data breaches**
- **Governance, risk and compliance**
- **Identity and access management**
- **Skill shortage in this domain**

Big data is inherently vulnerable to data and privacy breaches, making the matter of preventing them imminent. However, doing so has become challenging due to the following:

- **Increasingly complex IT environment**
- **Massive growth of transaction data volumes**
- **Explosion of new types of interaction data, such as social media and device data**
- **Use of insecure java-based frameworks, such as Apache™ Hadoop® and its programming paradigm MapReduce**
- **Insider and external threats**
- **Advanced persistent threat (APT)**

Data breaches in the enterprise are disturbingly common occurrences that are impacting companies and government agencies on a



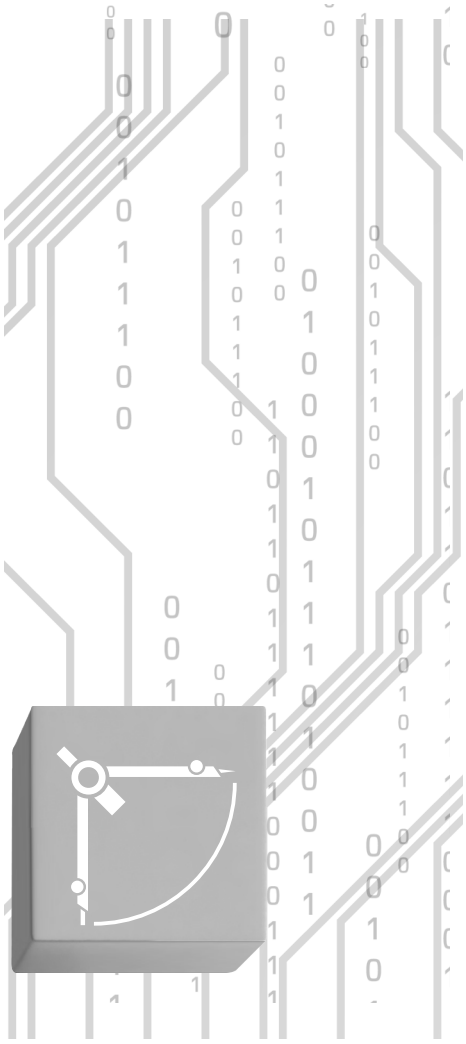
Source: COBIT® 5 for Information Security, ISACA, USA, 2012, figure 16

daily basis. In addition to the negative publicity, these breaches often have vastly deep effects on the business, including costs such as regulatory fines, litigation fees, consulting fees and the loss of customers. As a result, enterprises are in greater need of robust data privacy solutions to prevent breaches and enforce data security as data move from point to point and across borders.

As big data volumes grow, these new strategies need to be able to scale along with them. Enterprises need a robust data-privacy solution to prevent data breaches and enforce data security in a complex IT environment. The solution should empower enterprises to:

- **Identify all sensitive data**
- **Ensure that sensitive data are identified and secured**
- **Demonstrate compliance with all applicable laws and regulations**
- **Proactively monitor the data and IT environment**
- **React and respond faster to data or privacy breaches with incident management**

The COBIT 5 Information enabler life-cycle phases (plan, design, build/acquire, use/operate, monitor and dispose), shown in **figure 1**, allow enterprises to address these privacy solution features and streamline the governance, risk management and effective delivery of big data implementation projects.



Big Data Privacy Risk

Predictive analysis of big data is a tremendous tool when used and applied correctly. The collection of apparently unlinked data is harmless by itself, but its predictive powers can exceed expectations. For example, the health care industry sees value in big data as a predictor of patient behavior and future health. The Google Now™ service tracks a mobile device user's location, calendar events, search inquiries and personal preferences. This service predicts a user's information needs and displays the information on the user's mobile device. Using the personal data tracked with Google Now and payment card transactions, the health care industry predictive analysis can categorize patient lifestyles as healthy or unhealthy. The predictive analysis may categorize an individual incorrectly based on just one of the tracked parameters. Enterprises want to mitigate this type of risk with appropriate filters and cross checks.

Risk related to big data can be categorized as operational or information technology based. These risk categories can be mitigated with strong governance.

Operational risk encompasses external and internal factors that include geopolitical risk and the rush to satisfy the board and senior management who are eager to get ahead of the competition. Geopolitical risk, which is created by a country's policies, includes the European Union laws that restrict cross-border sharing and processing, privacy laws that prevent marketing to certain age groups and US privacy laws that prevent the labeling and sharing of personal, private and financial information, which can lead to identity theft and unauthorized transactions. Industry-specific legislation, such as the US HIPAA, can be very complex, and provisions assuming risk transfer may not be clearly documented or enforced. "Data itself does not create value or cause problems; its use does."[5]

Corporate chief information officers (CIOs) can be pressured by the board and senior management to iimplement big data to be able to compete before proper risk controls are applied. Ill-conceived application development controls can lead to data leakage and exposure of private data that are not intended to be seen by developers.

Methodologies such as Agile can support a controlled approach to risk while allowing flexibility. Mapping Agile to COBIT 5 can be an appropriate approach to governance, acquisition and development.

IT risk is business risk—specifically, the business risk associated with the use, ownership, operation, involvement, influence and adoption of IT within an enterprise.[6] IT risk occurs when safeguards are bypassed. For example, an enterprise may acquire software tools because technologists consider them scalable, but not necessarily because the tools meet the needs of those who plan to use them for business analytics. IT operations can be so focused on development and delivery that simple safeguards for capacity planning are overlooked and resources and data are not monitored or properly planned.

Enterprise policies need to ensure that employees keep stakeholder information confidential during and after employment. This risk is increasing especially given that information has become 21st century currency and data brokers are profiting from the sale of information—commonly referred to as data as a service.

[5] *Op cit* World Economic Forum, 2013

[6] ISACA.org, *The Risk IT Framework*, USA, 2009, *www.isaca.org/Knowledge-Center/Research/Documents/RiskIT-FW-18Nov09-Research.pdf*

Big Data Privacy Strategies

Various groups are beginning to address the need to establish norms for big data privacy. A World Economic Forum report recommends regulation that places restrictions on the use of personal data and the use of technology to empower individual privacy. Privacy can be embedded into the technology so that individuals have control of their own information.[7] The report suggests that future data systems could tag all collected data with code that indicates individuals' preferences for how their data are used.[8] The Software and Information Industry Association (SIIA) cautions against over legislation of big data and privacy and recommends that enterprises take the initiative to build privacy into big data policies. SIIA Senior Director David LeDuc expressed that enterprises can benefit from big data while protecting user privacy. For example, he recommends anonymizing consumer data as quickly as possible. Anonymizing permanently removes personal identifiers from data. The SIIA and other industry groups would like to see policy makers, consumer advocates and other stakeholders come together to create policy.[9]

By understanding global trends in privacy management, organizations can determine either a defensive or offensive strategy for big data privacy. A defensive strategy includes embracing the ability to use information and protect the privacy of individuals through various forms of encapsulation and tagging. Anonymization can support statistical trending and analysis and support individual privacy when required. An offensive strategy involves enforcing disclosure and reminding the consumer of the exchange of services and data. For example, this strategy would be implemented for services that obtain personal information in exchange for issuing coupons, exchanging free services, products as a *quid pro quo* of free magazine subscription, free flowers, free email or calendaring services.

Governance of Big Data Privacy

Big data gives enterprises the ability to access, aggregate and analyze ever-increasing amounts of data, including web pages, browsing habits, sensor signals, smartphone location trails and genomic information. Big data is an enormous opportunity to make information the leader of value creation, but without comprehensive principles, policies and frameworks, big data can generate enormous risk. Big data needs a governance framework that ensures trustworthy data practices.

Without proper governance, the same data that can be used to create value can be used to create intrusive and damaging outcomes and destructive decision making.

COBIT 5 enables enterprises to create optimal value from IT by maintaining balance between realizing benefits and optimizing risk levels and resources.

[7] *Op cit* World Economic Forum, 2013

[8] Lohr, Steve, "Big Data Is Opening Doors, but Maybe Too Many," *The New York Times*, 23 March 2013, *www.nytimes.com/2013/03/24/technology/big-data-and-a-renewed-debate-over-privacy.html?_r=1&*

[9] Software & Industry Information Association, "Data-Driven Innovation A Guide for Policymakers: Understanding and Enabling the Economic and Social Value of Data," May 2013, *www.siia.net*

Enterprises exist to create value for their stakeholders while optimizing risk and resource use. Delivering enterprise stakeholder value requires good governance and management of information and IT assets. Big data is an enterprise asset that fits naturally within the domain of COBIT 5 Principle 1—Meeting Stakeholder Needs.

COBIT 5 makes a clear distinction between governance and management, with governance responsibility at the enterprise board level. Big data stakeholder needs are ensured and maintained at a high level in the enterprise. The COBIT 5 RACI chart for one of the enterprise governance processes, in **figure 2**, shows that the board is accountable for critical initiatives, while the management responsibility lies with the chief executive officer (CEO) and CIO. The RACI chart assigns responsibility levels for process practices to roles and structures. The enterprise roles are shaded red and the IT roles are blue. The different levels of involvement are:

R (Responsible)—Takes the main operational stake in fulfilling the activity listed and creating the intended outcome

A (Accountable)—Is accountable for the success of the task

C (Consulted)—Provides input

I (Informed)—Receives information about achievements and/or deliverables of the task

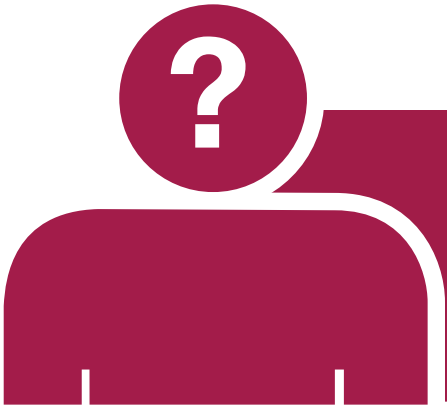
FIGURE 2

COBIT 5 EDM01:
Ensure Governance Framework Setting and Maintenance Process Role Assignments

EDM01 RACI Chart

Source: COBIT® 5: Enabling Processes, ISACA, USA, 2012, page 31

EDM01.03 Monitor the governance system	EDM01.02 Direct the governance system.	EDM01.01 Evaluate the governance system.	KEY GOVERNANCE PRACTICE
A	A	A	Board
R	R	R	Chief Executive Officer
C	C	C	Chief Financial Officer
C	C	C	Chief Operating Officer
R	R	R	Business Executives
I	I		Business Process Owners
R	R	R	Strategy Executive Committee
I	I		Steering (Programmes/Projects) Committee
I	I		Project Management Office
I	I		Value Management Office
C	C	C	Chief Risk Officer
I	I		Chief Information Security Officer
I	I	C	Architecture Board
I	I	C	Enterprise Risk Committee
I	I	C	Head Human Resources
C	C	C	Compliance
C	C	C	Audit
R	R	R	Chief Information Officer
C	C	C	Head Architect
I	I	C	Head Development
I	I	C	Head IT Operations
I	I		Head IT Administration
I	I		Service Manager
I	I		Information Security Manager
I	I		Business Continuity Manager
I	I		Privacy Officer



For proper big data privacy governance, the board and senior executives should ask the following questions:

The board and senior executives must embrace IT with leadership that ensures correct policies, processes and procedures, and an appropriate skill set.

Privacy within an enterprise comprises compliance with legal and regulatory requirements regarding data retention periods, cross-border regulations, privacy and intellectual property (IP). Big data privacy governance should ensure compliance, but at the same time enable an enterprise environment that effectively uses big data to create sustainable enterprise competitiveness.

- What principles, policies and frameworks are we going to establish to support the achievement of business strategy through big data?
- Can we trust our sources of big data?
- What structures and skills do we have to govern and manage IT?
- What structures and skills do we have to govern big data privacy?
- Do we have the right tools to meet our big data privacy requirements?
- How do we verify the authenticity of the data?
- Can we verify how the information will be used?
- What decision options do we have regarding big data privacy?
- What is the context for each decision?
- Can we simulate the decisions and understand the consequences?
- Will we record the consequences and use that information to improve our big data information gathering, context, analysis and decision-making processes?
- How will we protect our sources, our processes and our decisions from theft and corruption?
- Are we exploiting the insights we get from big data?
- What information are we collecting without exposing the enterprise to legal and regulatory battles?
- What actions are we taking that create trends that can be exploited by our rivals?
- What policies are in place to ensure that employees keep stakeholder information confidential during and after employment?

Assurance Considerations for Big Data Privacy

The main drivers for assurance include:

- **Providing interested parties with substantiated opinions on governance and management of enterprise IT according to assurance objectives**
- **Defining assurance objectives in line with enterprise objectives, thus maximizing the value of assurance initiatives**
- **Satisfying regulatory or contractual requirements for enterprises to provide assurance over their IT arrangements**

Assurance professionals should be part of the enterprise big-data initiative from its inception. To be able to provide valuable big data insights to the enterprise, assurance professionals need to have in-depth understanding of the business; the knowledge, as data scientists, to use big data tools such as Hadoop, the EMC® Greenplum® platform, Teradata® database software and analytic applications, HP™ Vertica™ analytics system and Palantir Technologies software; and the skills to be able to interpret the results and correctly explain them to stakeholders. Assurance professionals should keep well informed of new big data skills and terms and educate management and the audit team.

In addition to providing big data insights to management, assurance professionals should attest to the following:

- **Big data privacy and security solutions are implemented**
- **Sufficient big data privacy governance exists, such as:**
  - Data anonymization/sanitization or de-identification
  - Adequate, relevant, useful and current big data privacy policies, processes, procedures and supporting structures
  - Senior management buy-in and evidence of continuous commitment
  - Appropriate data destruction, comprehensive data management policy, clearly defined disposal ownership and accountability
  - Compliance with legal and regulatory data requirements
  - Continuous education and training of big data policies, processes and procedures

Conclusion

Big data is becoming increasingly pervasive, and enterprises need to find optimal ways to derive benefits from it. Privacy is one area requiring very close attention, since big data is centered inherently around the individual. The dependency has the potential to create significant negative consequences.

Big data is a valuable asset and, at the same time, a powerful tool with far reaching impacts. Consequently, big data initiatives need to have visibility at the board level and executive level sponsors.

The success of enterprises will depend on how they meet and deal with the various big data challenges and impacts, including privacy. **To harness value and deliver resilient and faster analytic solutions, enterprises must implement big data solutions using repeatable frameworks and processes coupled with a good governance and risk management framework.**