KuppingerCole Report

WHITEPAPER

by Mike Small | April 2019

Big Data Analytics – Security and Compliance Challenges in 2019

There is an ever-increasing number of people, devices and sensors that generate, communicate and share data via the global internet. Analysing this data can help organizations to develop new products, improve their efficiency and effectiveness, as well as to make better decisions. This report describes the challenges of using Big Data in ways that are secure, compliant and ethical and how meeting these challenges requires a data centric approach to security.

by **Mike Small**
mike.small@kuppingercole.com
April 2019

Commissioned by comforte AG

Content

Related Research

Advisory Note: Big Data Security, Governance, Stewardship - 72565

Advisory Note: Maturity Level Matrix for GDPR Readiness - 72557

1 Introduction

There is an ever-increasing number of people, devices and sensors that generate, communicate and share data via the global internet. Analysing this data can help organizations to develop new products, improve their efficiency and effectiveness, as well as to make better decisions. However, there are increasing concerns over the trustworthiness of this data as well as security and compliance challenges over the way that it is used. This report describes the challenges of using Big Data in ways that are secure, compliant and ethical and how a data centric approach to security is essential to meeting these challenges.

Societal concerns over how data is being acquired and used are leading to increasingly tough regulations governing how organizations can acquire, store and use data. In addition to existing regulations such as HIPAA, recent examples are the EU GDPR (General Data Protection Regulation) and the CCPA (California Consumer Privacy Act of 2018). GDPR imposes a much tougher regulatory framework that affects organizations worldwide that hold or process personal data relating to residents in the European Union. In order to meet the obligations imposed by these laws, it is essential that organizations implement good data governance as well as data centric security controls over how they acquire, store, process, and analyse data.

Big Data is a key organizational asset and must be managed as such. Good information stewardship with data centric security provides a solution to these challenges. In the past there has been a tendency to view security as a technology issue. The KuppingerCole view is that this is wrong, and the KuppingerCole IT Paradigm takes an information / data centric view of security.

Information stewardship starts with good data governance at the organizational board level. The board must set clearly defined business objectives for the use of Big Data, together with the needs for compliance and the acceptable levels of risk. The responsibility for the data must be clearly defined and its lifecycle must be properly managed. To be able to demonstrate compliance it must be possible to audit the way in which it is acquired, analysed and secured as well as how its results are used.

Access controls are fundamental to ensuring that data is only accessed and used in ways that are authorized. Identity and access management are essential to control legitimate access but are not enough to protect against all risks. Encryption, tokenization, anonymization and pseudonymization provide important controls that help to implement information centric security. Importantly GDPR accepts the use of pseudonymisation as an approach to data protection by design and default as well as loosening controls over how pseudonymised data can be used.

Many of the data breaches have occurred because of the simplest controls were non-existent or were not implemented properly[1]. Don't let this happen to you. Organizations must implement best practices and data centric approach to secure data analytics infrastructure and adopt a privacy by design approach where personal data is involved. They must Implement controls to ensure that data moving to cloud services is properly protected. Where cloud services are used, organizations using them should require independent certification that they comply with the relevant laws and regulations.

[1] **https://ico.org.uk/action-weve-taken/enforcement/the-carphone-warehouse-ltd/**

2 Highlights

- An ever-increasing number of people, devices and sensors generate, communicate, share and access data through the internet. The analysis of this "Big Data" into "Smart Information" can help organizations to improve the efficiency and effectiveness by making better decisions.
- Societal concerns over the use of Big Data are leading to increasingly tough regulations governing how organizations can acquire, store and use data. In order to meet these compliance obligations, it is essential that organizations implement good data governance as well as data centric security controls over how the data is acquired, stored, processed and protected.
  - The infrastructure involved in the acquisition, storage and analysis of Big Data needs to be secured and this is often not the case.
  - The use of cloud services introduces new risks. Improperly set security controls can expose data on the Internet. Data moved to cloud services is often not protected.
  - Many IoT devices implement poor security practices with limited capabilities to resist cyber-attack and no capabilities for the defences to be upgraded. This could impact on the trustworthiness of the data analysed.
  - The use of MLS, Cognitive Systems and AI need training data and this introduces additional security and compliance risks.
  - There is often no clear ownership for Big Data and poor controls over its lifecycle.
- Big Data is a key organizational asset and must be managed as such. A Data Centric approach to the security and compliance of Big Data provides a sustainable approach that is independent of the tools and technologies used to analyse the data.
  - Information centric security puts the data as the central concern of the security objectives, policies, processes and technologies.
  - Information centric security starts with good data governance.
  - Big Data must be protected against unauthorized access and use. Encryption, tokenization, anonymisation and pseudonymisation are important controls to achieve this.
  - Pseudonymisation is encouraged to implement data protection by design and default, but the Data Controller needs to ensure the correct choice of tools.
  - Big Data must have an owner and its lifecycle must be properly managed from creation or acquisition through its use and disposal.
  - The infrastructure used to collect, store and analyse Big Data must be properly secured with care taken to remove vulnerabilities and implement best practices.
  - Encryption, anonymization and pseudonymization can help to ensure that data in transit and in cloud services is properly protected.
  - Where cloud services are used, organizations using them should require independent certification that they comply with the relevant laws and regulations.

3 From Big Data to Smart Information

An ever-increasing number of people, devices and sensors generate, communicate, share and access data through the internet. The analysis of this "Big Data" into "Smart Information" can help organizations to improve the efficiency and effectiveness by making better decisions.

Getting competitive advantage from data is not a new idea however, the volume of data now available and the way in which is being collected and analysed has led to increasing concerns[2]. As a result, there are a growing number of regulations over its collection, processing and use. Organization need to take care to ensure compliance with these regulations as well as to secure the data they use.

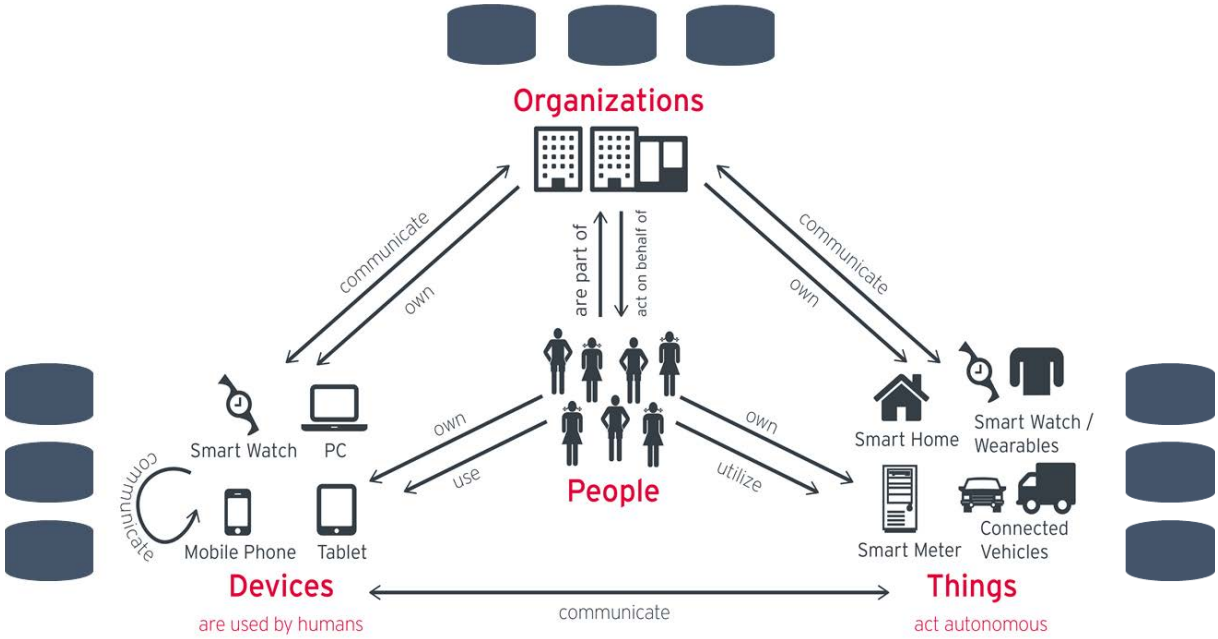

Figure 1: Big Data Everywhere

There are many sources of data as illustrated in Figure 1. On traditional devices such as PC's and servers, productivity applications generate business related data. An enormous volume of data is also generated for entertainment. For example, according to Ofcom[3] in the UK - "there are now more UK subscriptions to Netflix, Amazon and NOW TV than to 'traditional' pay TV services". Individuals and organizations also create large volumes of images and media for other purposes including advertising, self-publicity, advice and social communication. Embedded devices are also the source of data which is likely to increase dramatically as the deployment of the IoT (Internet of Things) evolves and matures.

Nearly everything in use today generates data; some of this data is created intentionally and some is inherent in the device's use. According to IDC[4], by 2025 the 175 Zeta Bytes (10^{21}) of data will have been created worldwide.

[2] **Findings recommendations and actions from ICO investigation into data analytics in political campaigns**
[3] **https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2018/streaming-overtakes-pay-tv**
[4] **IDC - Data Age 2025**

The smartphone is typical of today's smart technology which, as well as executing a primary function, includes a wide range of sensors that constantly monitor its performance. In addition to the data resulting from it direct use, readings from these sensors add to the accumulating amount of data available for analysis.

Big Data also includes the large amount of data that organizations have accumulated internally as well as that which comes from the infrastructure they control. Much of this is held in unstructured form like emails, word documents, spread sheets and presentation files. These are created in an ad hoc manner which creates a significant problem because it is hard for an organization to know what exists and where it is held.

Big Data extends beyond the enterprise and much of the potential value comes from being able to search for and utilize data from external sources. These sources include social media and publicly available data from government databases as well as other data that can be shared between organizations. Social computing like Facebook and Twitter provide large amounts of data that can be analysed to provide information on consumers' preferences and grumbles.

Smart Information is Big Data analysed to make it useful - for example to improve effectiveness, to help make better decisions and to accurately forecast likely outcomes.

However, this analysis creates its own challenges including the volume of data as well as the shortage of the skills needed to perform the analysis. Data scientists are in high demand, but there is a significant shortage in the industry. In addition, where each analysis is programmed individually it is very hard to adapt to the constantly evolving demands from the business. This creates a bottleneck.

The volume of data available together with the computing power provided by cloud services has led to the new forms of data analytics.

New analytics tools such as the Hadoop MapReduce framework have evolved to cope with the amounts of data that needs to be analysed. The computing power needed for this analysis has led to the use of cloud services which adds to the complexity of the security and compliance challenges. Use of cloud services can lead to a loss of control over the infrastructure and expose data. Where third parties are involved in the analysis process, this can also increase the risks of data being copied or misused.

This vast amount of data together with the difficulty of retaining skilled data analysts has encouraged the use of Machine Learning Systems (MLS). These have the potential to replace, or at least boost the productivity of, the hard to find skilled data scientists. However, their use exacerbates the security and compliance problems by needing training data and involving the use of cloud services to obtain the packaged tools, services and computer power needed.

4 Security and Compliance Challenges from Big Data Analytics

Big Data magnifies the security, compliance and governance challenges that apply for normal data as well as increasing the potential impact of data breaches.

Analysis of Big Data can identify individuals and their preferences more accurately, but often in a way that is not transparent to the individuals concerned. Big Data is often acquired from devices and through infrastructure that has not been designed, constructed and deployed with security in mind. The use of cloud services to store and analyse the data leads to additional challenges. In addition, there is often no clear ownership for Big Data and poor control over its lifecycle.

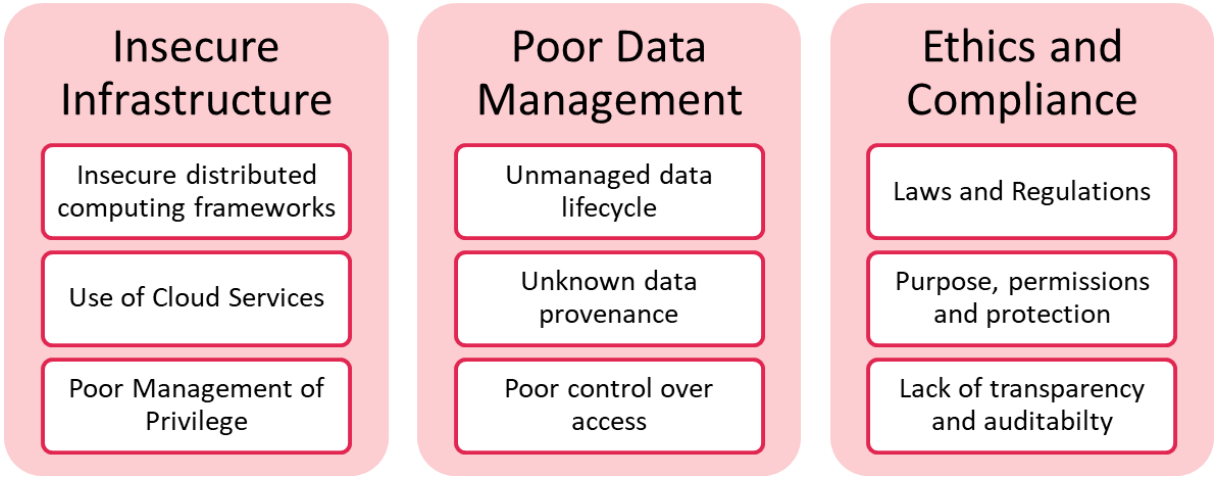

Figure 2: Big Data Challenges

Societal concerns over these challenges are leading to increasingly tough regulations governing how organizations can acquire, store and use data. In order to meet these compliance obligations, it is essential that organizations implement good data governance as well as data centric security controls over how the data is acquired, stored, processed and protected.

The impact of failures to secure big data can be very severe[5] both financially and to your brand.

[5] **Marriott Hack hits 500 million Starwood Guests**

Much of the Big Data originates from or can be attributed to individuals and hence its collection and use must comply with the obligations from privacy legislation. While this legislation varies in detail around the world some common basic principles apply. These include taking care of personal data and controlling how it is used. The way in which personal data is collected should be transparent to the data subject. The Data must be accurate and the organization using must ensure that the processing is legitimate.

The EU GDPR imposes a much tougher regulatory framework that affects organizations worldwide that hold or process personal data relating to residents in the European Union.

A good example of regulation is the EU GDPR[6] (General Data Protection Regulation). Under GDPR the definition of processing is very broad, it covers any operation that is performed on personal data or on sets of personal data. It includes everything from the initial collection of personal data through to its final deletion. Processing covers every operation on personal data including storage, alteration, retrieval, use, transmission, dissemination or otherwise making available.

The analysis of personal data is regulated by GDPR and it is up to the Data Controller to prove compliance.

GDPR defines a set of principles governing how personal data must be processed in Article 5. These principles require that both Data Controllers and Data Processors act lawfully, fairly and transparently. That personal data shall only be used for the purposes for which it was collected and must be relevant to the purpose while being the minimum necessary. It should be kept up to date and deleted when no longer necessary. Critically, the burden of proof to demonstrate compliance with these principles lies with the Data Controller.

Processing of personal data is only lawful if it satisfies one of the conditions set out in Article 6. These include use with explicit consent from the data subject for that use; for the performance of a contract or legal obligation; to protect the vital interests of the data subject; for a task in the public interest; or where processing is necessary for the legitimate interests of the controller.

The data subject has rights to access, correct and erase their personal data as well as to withdraw consent to its use.

[6] **EUR-Lex - 32016R0679 - EN - EUR-Lex**

GDPR sets out several rights that the data subject has in relation to their data. These rights include: the right to have confirmation from the Data Controller as to whether or not their personal data are being processed, and, where that is the case, access to their data; the right to the rectification of inaccuracies; the right to withdraw consent; and the right to have their personal data erased.

The sanctions for non-compliance are very severe with penalties of up to 4% of annual worldwide turnover.

GDPR sets out the sanctions to be applied to organizations that fail to comply with the regulation in Article 83. These depend upon the circumstances of the individual case and the degree of negligence involved. Two levels of penalty are defined: up to the larger of 2% of annual worldwide turnover or 10 Million Euro for internal matters; up to the larger of 4% of annual worldwide turnover or 20 Million Euros for breaching the principles, consent, subject rights or data transfers.

In order to comply with GDPR and other regulations as well as to ensure that the analysis results are trustworthy good data governance and secure processing are essential. While this may appear obvious it is not simple to achieve.

The infrastructure involved in the acquisition, storage and analysis of Big Data needs to be secure.

The technology that underlies the processing and analysis of Big Data was conceived to provide massively scalability rather than to enforce security controls. While this is not a new phenomenon in the IT industry there has not been enough time for the all inherent vulnerabilities and security weaknesses to become manifest.

For example: the Hadoop MapReduce framework is based on groups of standard computers which process data in parallel. Unless care is taken to remove known technical vulnerabilities, manage and secure the multiple administrator accounts as well as the content of the multiple file systems the framework is at risk from common cyber threats that could impact the confidentiality of the data as well as the integrity of the analysis.

Another Big Data technology is the so called “NoSQL” database. These databases provide for storage and retrieval of data which cannot be processed efficiently as relational tables. These data include video, voice, and relationship graphs. There are also “in memory databases” that provide optimized searching on relational tables by holding the data in memory and in columnar form. Many of these databases lack the normally accepted functionality needed to ensure secure processing such as control over administrator access to data as well as encryption.

The use of cloud services introduces new risks. Improperly set security controls can expose data on the Internet.

The volume of data that needs to be analysed and the computing power needed for this analysis has led to the use of cloud services. This adds to the security and compliance challenges through loss of control over where the data resides and potentially exposing the data on the Internet. It also makes unauthorised copying or misuse of data easier where third parties are involved in the analysis process. Where sensitive or regulated data is held in cloud services it must be protected against unauthorised access, copying and leakage.

Object databases such as AWS S3 provide a useful storage mechanism for large volumes of many kinds of data. However, unless correctly configured, this data can be exposed on the Internet. For example, according to a report from UpGuard[7] GoDaddy a major domain name registrar, was discovered to have files containing sensitive information stored in an unsecured S3 bucket. In response to this kind of problem, in November 2018 AWS rolled out new security features to prevent accidental S3 related misconfigurations – but organizations need to use these.

Many IoT devices implement poor security practices with limited capabilities to resist cyber-attack and no capabilities for the defences to be upgraded.

The IoT (Internet of Things) provides a rich source of Big Data. Active IoT devices include the vast number of sensors that are being deployed to monitor a wide range of parameters across many areas ranging from factories, buildings and even cities. These devices are capable of transmitting data without the explicit consent from the owner, or at time or in ways that the owner did not expect.

The use of MLS, Cognitive Systems and AI also introduce new risks.

MLS learning depends upon the use of large training datasets – to be useful these may contain regulated data and care is needed to ensure that this use is complies with regulatory obligations. For example, under which provision in Article 6 of GDPR is this use lawful?

In any case the training data must be properly secured with access strictly controlled. Since the training process may involve access by multiple groups of people including data scientists and subject matter experts, implementing these controls can be challenging. In addition, the training data may contain unintentional bias – which could potentially lead to claims of discrimination under various legal systems later. Furthermore, MLS systems are not able to explain their reasoning and so a large element of trust is needed to act upon the conclusions that they draw. Maintaining the integrity of the training data is essential to ensure this trust. In the case of disputes there could be difficulties in identifying who is legally responsible for conclusions drawn and actions taken.

[7] **Upguard - how configuration information for the world's largest domain name registrar was exposed online**

Big Data turns the classical information lifecycle on its head.

In the classical model data has a business owner who classifies it in terms of its business value and impact. The data is then created and used by business processes and is eventually deleted according to policies when no longer required to be retained. The provenance of the data is known, and its uses are largely predetermined. However, even this data is often not classified and is sometimes mishandled.

Many organizations now hold large quantities of unstructured data like emails, word documents, spread sheets and presentation files. This data is usually created in an ad hoc manner and has no formal owner or classification. Worse still this form of Big Data is often held on unstructured repositories like shared drives, SharePoint systems, cloud services - and is therefore highly mobile.

It is essential to ensure the trustworthiness of externally acquired data.

The creation of externally sourced Big Data may be outside the control of the organization using it. Therefore, its provenance may be doubtful and its ownership and consent for its use may be subject to dispute. The transmission of this data may not be properly secured to ensure that it has not been changed or leaked in transit and the responsibility for its protection may not be clear.

The volume of data and the computer power that is now available have widened the gap between what regulations and laws permit and what is technically possible.

Despite the many laws and regulations over the use of personal data there are still ethical concerns over the way in which Big Data is collected and analysed. The volume of data and the computer power that is now available have widened the gap between what regulations and laws permit and what is technically possible. This has changed the balance of power between individuals and the organizations that collect data. Organizations using Big Data need to be aware of these concerns and consider carefully how best to respond. Over time, organizations can expect that regulations will widen and strengthen, and need to ensure that they know what data they hold and use, where it came from and what justification they have to process it. This preparation will help to avoid future penalties.

5 Meeting the Security and Compliance Challenges

Big Data is a key organizational asset and must be managed as such. Good information stewardship with data centric security provides a solution to these challenges.

Information stewardship is not a new term; it has been in use since the 1990's and offers a consistent approach to managing the wide range of challenges where information is a key organizational asset. These challenges, which were described in the previous section, include the management of the complete information lifecycle from ownership to deletion as well as aspects like business value, data architecture, information quality, compliance and security.

Information centric security puts data as the central concern of the security policies, processes and technologies.

Good information stewardship for Big Data needs information centric, rather than technology centric, approach to security. In the KuppingerCole IT Paradigm, Information Security is a core discipline of IT. Hence the IT function must ensure the confidentiality, integrity and availability of corporate information. In the past there has been a tendency for organizations to view security as a technology issue. The KuppingerCole view is that this is wrong, and the KuppingerCole IT Paradigm takes an information / data centric view of security.

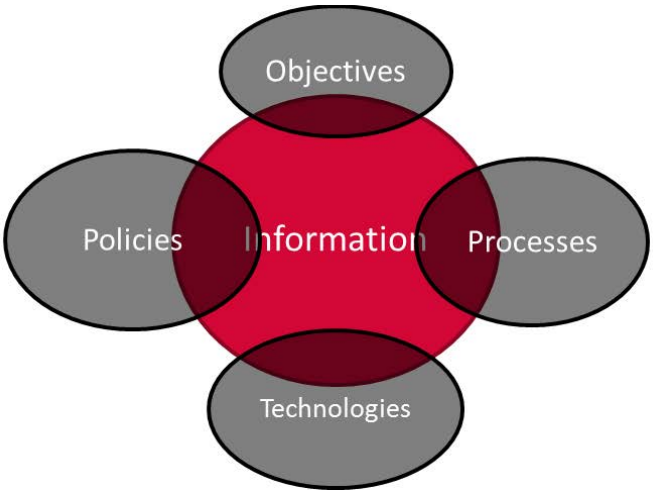

Figure 3: Information Centric Security

The basic objectives of information centric security are to ensure:

- Availability: individuals can access the Big Data and Smart Information they need to perform their business functions when and where they need it, and without delay.
- Integrity: individuals are only able to manipulate Big Data (create, change or delete) in ways that are authorized.
- Confidentiality: Big Data and Smart Information can only be accessed by authorized individuals and these are not able to pass data on to other individuals who are not authorized.
- Privacy and compliance: Big Data must be processed in a way that complies with laws and regulations and that regulated data is protected against leakage and misuse.

Information centric security starts with good data governance.

The distinction between governance and management is defined in COBIT 5[8]. Governance ensures that business needs are clearly defined, agreed and satisfied in an appropriate way. Governance sets the priorities and the way in which decisions are made; it monitors performance and compliance against the agreed objectives. Governance is distinct from management in that management plans, builds, runs and monitors activities in alignment with the direction set by the governance body to achieve the objectives.

For Big Data this means that:

- There must be clearly defined business objectives for the use of Big Data, the compliance objectives and the acceptable levels of risk must be set at board level;
- The responsibilities for Big Data must be clearly defined, and it must be possible to measure how well the business objectives have been met

Big Data should be protected against unauthorized access and use. Encryption, tokenization, anonymisation and pseudonymization are important to achieving this.

Access controls are fundamental to ensuring that data is only accessed and used in ways that are authorized. Identity and access management are essential to control legitimate access but are not enough to protect against all risks. Additional kinds of controls are needed to protect against illegitimate access, data breaches for example, and to ensure that the privacy of personal data is maintained when data is shared or held outside the organization or in cloud services.

Encryption, tokenization, anonymization and pseudonymization provide important controls. They are especially important where data can easily be copied or shared for example through cloud services.

[8] **http://www.isaca.org/cobit/pages/default.aspx**

Encryption protects data against some forms of unauthorized access but is only as strong as the control over the encryption keys. For example, classic symmetric encryption of data at rest does not protect the data during its use.

Homomorphic encryption overcomes this problem but with limitations and usually at a high computational cost. Rights Management using public / private keys strongly connected to individuals is an important control for unstructured data. DLP (Data Leak Prevention) technology and CASBs (Cloud Access Security Brokers) are also useful to help to control the movement of data outside of the organization.

Where data is being used for operational purposes it is usually necessary at some point for the (personal) data decrypted. However, where data is used for analysis this is not always the case, and anonymization and pseudonymization are methods that help here. Anonymizing data enables its analysis while making it unlikely that the identity of the individuals behind the data will be revealed. This must be done carefully and there are examples[9] of how by combining data sets it was possible to reveal the underlying identities.

Pseudonymisation is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. Provided that this additional information is kept separately and is properly secured the data cannot be attributed to individuals. Pseudonymisation is especially important because it is accepted by GDPR as an approach to data protection by design and default. Pseudonymization also allows operational data to be processed while it is still protected.

Tokenization, which replaces meaningful data with random tokens that obscure its content without access to a translation database, is an effective approach to pseudonymization.

Pseudonymisation is encouraged to implement data protection by design and default but the Data Controller needs to ensure the correct choice of tools.

GDPR Article 25 (and elsewhere) obliges the data controller to "implement appropriate technical and organisational measures, such as pseudonymisation" as an approach to data protection by design and data protection by default. Pseudonymisation is defined in Article 4 as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information..." with the additional proviso that the additional information is kept separate and well protected.

[9] **http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf**

In addition, Under Article 6 (4)(e), the Data Controller can take account of the existence of appropriate safeguards, which may include encryption or pseudonymisation, when considering whether processing for another purpose is compatible with the purpose for which the personal data were initially collected and the processing for another purpose. However, these provisos introduce an element of risk for the Data Controller relating to the reversibility of the process and protection of any additional information that could be used identify individuals from the pseudonymized data.

The lifecycle of Big Data must be properly managed from creation or acquisition through its use and disposal.

Many organizations are accumulating data without good reason or because it is more trouble to delete it than to keep it. This is especially true of the unstructured data in the form of emails, spreadsheets, documents and presentations. An organization should have a clear records management policy which identifies the information that must be kept for business purposes, why it is needed and for how long it must be retained. There should also be a clear policy for the deletion of information that is not retained for compliance or regulatory reasons.

In order to prevent the problems identified in the previous section, there should be a clear policy for data acquired externally that identifies the person within the organization that is responsible for the acquired data, its management and its lifecycle. This owner has responsibility for ensuring that that:

- The data is accurate, its provenance is certain, and its source is trustworthy;
- The purpose for its collection is clearly defined;
- The organization has the right to use the data;
- The data is only used in ways that are lawful, compliant and ethical;
- The data is held and processed securely to prevent unauthorized access;
- The data is protected using data centric technologies like encryption, tokenization, and methods such as anonymization and pseudonymization.

The infrastructure used to collect, store and analyse Big Data must be properly secured. A data centric approach provides independence from the technologies used.

Studies of the causes of data breaches[10] repeatedly find that most breaches could have been prevented using simple well-known controls. Standards and best practices should be used to ensure that common vulnerabilities are removed. Where new technologies are adopted the security aspects of these should be reviewed and appropriate controls implemented to mitigate potential weaknesses. Storing data in cloud services potentially exposes it to greater threats of leakage. Organizations must control which cloud services can be used to hold data and protect that data using data centric controls.

- Implement a vulnerability management program to detect and remove known vulnerabilities, implement and update anti-malware tools;
- Implement strong identity and access controls – especially around access to administrative privileges;
- Implement data centric security controls such as encryption, tokenization, anonymization and pseudonymization over sensitive and regulated data.
- Protect data in transit to ensure authenticity of source and destination as well as against interception;
- Implement privacy by design for systems and applications processing personal data;
- Control the use of cloud services using CASB technologies and protect data held in these services using a data centric approach;
- Choose the data centric encryption solution carefully to ensure that it is certified and meets your precise needs;
- Require independent assurance that sanctioned cloud services comply with relevant laws and regulations.

[10] **https://enterprise.verizon.com/resources/reports/DBIR_2018_Report_execsummary.pdf**

6 Recommendations

Take a Data Centric approach to the security and compliance of Big Data. This provides a sustainable approach that is independent of the tools and technologies used to analyze the data.

Smart Information obtained through the analysis of Big Data can provide tangible benefits to organizations. However, it also introduces new challenges around security, compliance and ethics that stem from increased complexity and loss of control. Organizations using Big Data need to manage the potential risks both to the organization and to those outside (for example to the data subjects) arising from this. These risks should be assessed and managed using best security practices, appropriate security controls enforced using the properly chosen tools.

A governance-based approach is needed to ensure that business objectives are clearly understood and to ensure that legal and regulatory compliance can be demonstrated in what is becoming an increasingly challenging regulatory environment. The only plausible approach to meet these challenges is for data centric security where controls follow the data.

There are three key pillars to this approach:

- The security and compliance controls must be data centric – they must secure the data to wherever it is stored and processed. To ensure compliance, these must include control over access as well as over how the data it is used. The data must be protected against accidental leakage and data theft to prevent costly data breaches. Encryption technology helps to protect data against some risks but is not generally effective during its processing. However, pseudonymization can protect data during analysis as well as during operational processing. Furthermore, pseudonymization provides an approach to privacy by design that is accepted by GDPR and additionally loosens the obligations over how pseudonymized data can be used.
- The lifecycle of Big Data must be properly managed. All data including Big Data must have an owner who is responsible for its classification, use and its complete lifecycle. Big Data that originates externally must be accurate, with known provenance and with verifiable and verified permissions for its use.
- The infrastructure for processing Big Data provides the foundation for its security and accurate analysis. A data centric approach to security provides independence of the specific technologies used. This must be acquired, built, run and managed using the same best practices and security disciplines as for other IT data processing. Adopt a privacy by design approach where personal data is involved. Using cloud services creates additional challenges by exposing the data to a wider range of threats. Control how data is moved to cloud services and ensure that it has additional data centric protection when held in these. Where cloud services are used require independent assurance that they comply with the relevant laws and regulations.

7 Copyright

The Future of Information Security – Today

KuppingerCole supports IT professionals with outstanding expertise in defining IT strategies and in relevant decision-making processes. As a leading analyst company, KuppingerCole provides first-hand vendor-neutral information. Our services allow you to feel comfortable and secure in taking decisions essential to your business.

KuppingerCole, founded in 2004, is a global Analyst Company headquartered in Europe focusing on Information Security and Identity and Access Management (IAM). KuppingerCole stands for expertise, thought leadership, outstanding practical relevance, and a vendor-neutral view on the information security market segments, covering all relevant aspects like: Identity and Access Management (IAM), Governance & Auditing Tools, Cloud and Virtualization Security, Information Protection, Mobile as well as Software Security, System and Network Security, Security Monitoring, Analytics & Reporting, Governance, and Organization & Policies.

For further information, please contact **clients@kuppingercole.com**

KuppingerCole Analysts AG
Wilhelmstr. 20-22
65185 Wiesbaden | Germany

Phone +49 (211) 23 70 77 – 0
Fax +49 (211) 23 70 77 – 11
www.kuppingercole.com