Dark data discovery and its importance in data protection

December 2019

What is dark data?

Dark data refers to data that has been collected, processed and stored by an organisation during the regular course of business, but which is yet to be tapped and analysed for any business purposes.[1] Dark data is generated by a company's CRM, ERP, SCADA, HTTP and Wi-Fi systems, which is not captured or left unanalysed. This type of unmanaged, uncategorised and unknown data is widespread across most industries.

Dark data is also referred to as unstructured data or dusty data. All modern modes of communication involve the storage/collection of dark data in one form or another.

What is dark data discovery?

Data discovery is the process of scanning through a large quantum of unstructured data or dark data to get a complete inventory of an organisation's data landscape. It helps an organisation to segregate relevant, necessary and actionable data.

As India works towards strengthening data privacy regulations such as the Personal Data Protection Bill, 2018, organisations must clearly understand the type of data they possess. Organisations find it difficult to manage dark data because it is stored across a distributed IT environment with no single owner.

Can dark data be personal data?

Personal data refers to information that can be used to identify a person.

Some of the metadata collected by organisations can be described as personal information and is therefore subject to privacy laws. Some examples of organisations using metadata as personal information are:


Employees personal records


Employees professional records


Employees financial records


Employees medical records

Privacy laws are applicable to dark data because by linking many different pieces of metadata together, a data analyst can develop a detailed understanding of a person's habits and likely actions in the future and use or misuse such information.

1 Source: https://www.gartner.com/en/information-technology/glossary/dark-data

Dark data discovery for privacy compliance

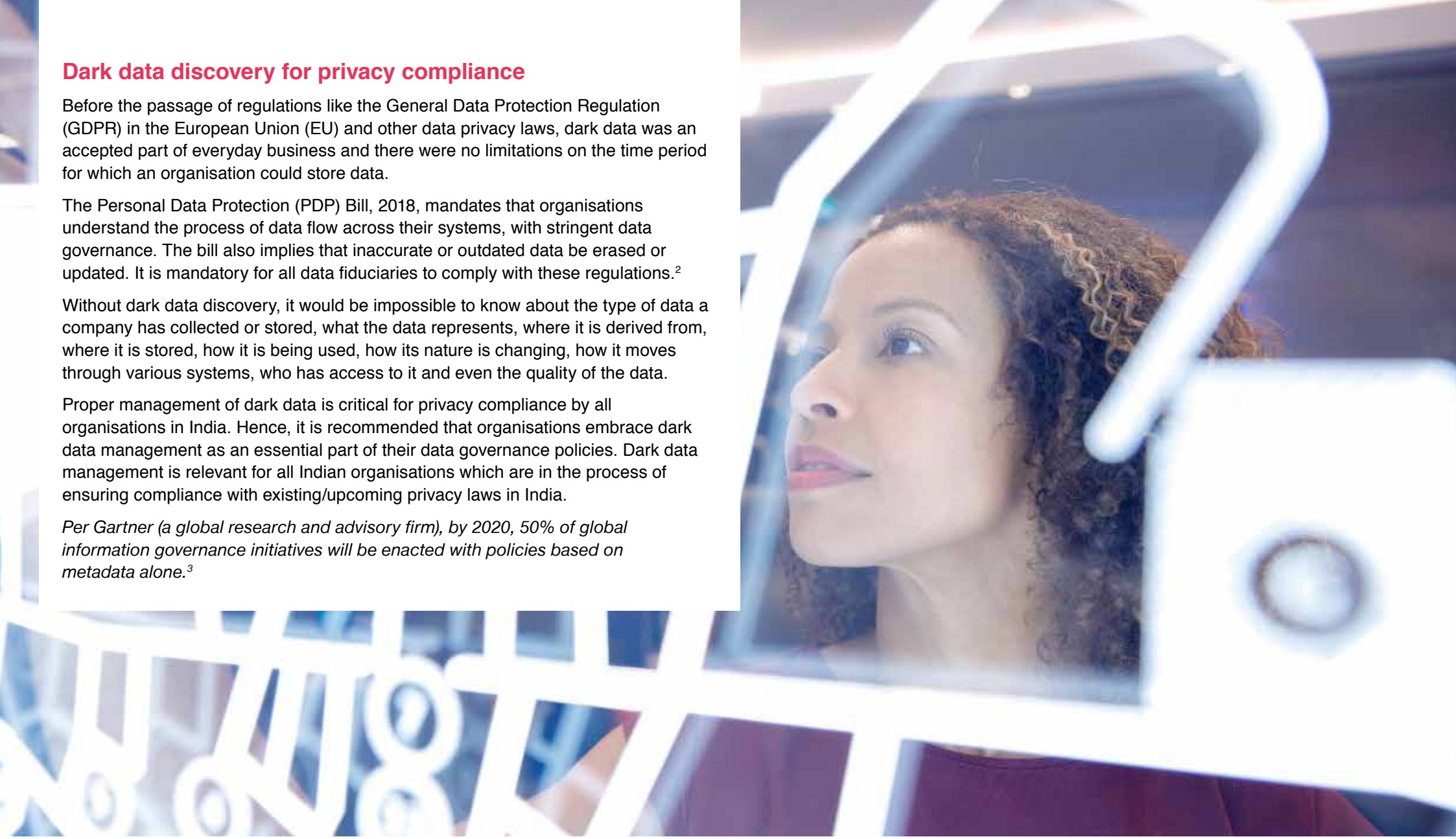
Before the passage of regulations like the General Data Protection Regulation (GDPR) in the European Union (EU) and other data privacy laws, dark data was an accepted part of everyday business and there were no limitations on the time period for which an organisation could store data.

The Personal Data Protection (PDP) Bill, 2018, mandates that organisations understand the process of data flow across their systems, with stringent data governance. The bill also implies that inaccurate or outdated data be erased or updated. It is mandatory for all data fiduciaries to comply with these regulations.[2]

Without dark data discovery, it would be impossible to know about the type of data a company has collected or stored, what the data represents, where it is derived from, where it is stored, how it is being used, how its nature is changing, how it moves through various systems, who has access to it and even the quality of the data.

Proper management of dark data is critical for privacy compliance by all organisations in India. Hence, it is recommended that organisations embrace dark data management as an essential part of their data governance policies. Dark data management is relevant for all Indian organisations which are in the process of ensuring compliance with existing/upcoming privacy laws in India.

Per Gartner (a global research and advisory firm), by 2020, 50% of global information governance initiatives will be enacted with policies based on metadata alone.[3]



2 Source: https://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf

3 Source: https://www.gartner.com/imagesrv/media-products/pdf/LogTrust/LogTrust-1-3F7HE3J.pdf

Optimising dark data for the upcoming PDP Bill

Here are five ways in which dark data management is likely to play a critical role in meeting the requirements for the upcoming legislation on personal data protection:

Right to access

The draft PDP Bill specifies that data principals (natural owners of data) have the right to access their personal data which is being processed by any organisation. An organisation would need to identify all possible data locations in advance in order to seamlessly use dark data to successfully locate the data required by its data principles.

Right to be forgotten

The data principal is empowered by the draft PDP Bill to demand the deletion of all his/her personal information from an organisation's records. Dark data can be effectively used by an organisation to exhaustively map and discover the whereabouts of personal data.

Data portability

In addition to the right by the data principal to be forgotten, the draft PDP Bill contains a provision on data portability. As per this, an organisation is required to transfer the personal information of any data principal to another organisation if the data principal so desires. Dark data management would enable organisations to effectively comply with this requirement.

Privacy by design

As per the draft PDP Bill, data storage systems must be designed with a strong focus on data protection rather than being viewed as an add-on feature. The sensitive nature of personal data items can be established using dark data, which can be used by organisations whilst designing their information systems.

Breach notification

The draft PDP Bill lays down specific timelines for organisations to provide information of a potential data breach to concerned data principals. Dark data enables an organisation to ascertain the incidence of a breach and its origin and timing.

Hence, dark data and its management are vital for an organisation that seeks to comply with the provisions of the draft PDP Bill and currently lacks robust metadata management procedures.

How can PwC help

Should an organisation become involved in a data-related lawsuit, dark data needs to be examined to see whether it contains any relevant information.

In enterprises worldwide, the ever-expanding data stores remain unstructured and unanalysed.

PwC's eDiscovery Services will search all possible data sources from a single compiled index to collect all relevant information, or copy, move, encrypt, secure, or erase the information completely and with forensic accuracy. This will also enable the client to locate and classify personal/business and regulated data quickly, thoroughly and in a scalable fashion.[4]

Using a wide array of data profiling techniques, PwC's eDiscovery platform generates a library of documentation describing the company's data assets and creates a metadata repository.

1. If any litigation or audit notice is issued from a regulatory body, an organisation can have compliance issues if it is not aware of what's in its data store, and whether it's being effectively protected. In such cases, our Cyber Security Services providing data protection impact assessments and Information Security Audit Services will help ensure compliance.
2. Clients' operating systems and browsers can be secured by installing privacy extensions to secure dark data.

4 According to International Data Group (IDG)

The benefits of dark data discovery are:

- generate insights for consumers and better business efficiency
- identify any possible links and connections between different data sets
- gain better understanding of customer feedback from web analytics
- identify new revenue streams for the organisation
- better-quality analytics
- reduced costs and risks
- improve company's privacy posture
- ensure information governance
- cater to data principal access requests with ease
- provide data inventory solutions.

Future of dark data

Dark data is said to be growing at a rate of 62% per year.[5] It is predicted that by 2022, 93% of the world's total data collected/stored will be dark data. Moreover, high annual growth rates, ever-increasing amounts of information consumed by organisations and continuously evolving data technologies are compelling organisations to efficiently manage their expanding volumes of data.

Also, the volume and severity of data regulations in India are expected to increase in the future, especially in terms of data privacy.

Dark data, if used properly, could result in economic benefits for organisations. By analysing this data, organisations can take information that was previously hidden or unknown and turn it into powerful insights, leading to new opportunities, reduced risk, and increased return-on-investment (ROI).

What should be done with dark data?

Some say that data should never be disposed of as storage is cheap, and the collected data might serve a purpose in the future. Data retention policies vary between different organisations.

Organisations are advised to proactively keep track of dark data which may prove to be valuable in the future. Delay in managing or abandoning such data may hamper future productivity or profitability of organisations.

5 https://www.pwc.in/assets/pdfs/consulting/forensic-services/dark-data-identification-and-remediation.pdf

About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 157 countries with over 276,000 people who are committed to delivering quality in assurance, advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.

In India, PwC has offices in these cities: Ahmedabad, Bengaluru, Bhopal, Chennai, Delhi NCR, Hyderabad, Kolkata, Mumbai, Pune and Raipur. For more information about PwC India's service offerings, visit www.pwc.in

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

Acknowledgements

This article has been researched and authored by Kush Wadhwa, Harbani Gill and Dhritimaan Shukla

Contact us

Dhritimaan Shukla
Partner and Privacy Leader
dhritimaan.shukla@pwc.com
+91-9899038326

Sonali Saraswat
Associate Director, Privacy
sonali.saraswat@pwc.com
+91-9620701515

PD/Decemeber 2019-M&C 3798