www.pwc.in

Dark data
Identification and remediation

pwc

What is dark data?

Dark data refers to all information created, collected, processed and stored by organisations as part of their functioning. The data is stored without specific categorisation and serves no apparent business purpose. This data may be residing in handheld devices, local system drives, emails, network file shares, legacy paper files or cloud file sharing services and even structured databases such as document management systems. It is mostly obscure, with organisations having limited insight into its nature, potential risks and benefits, metadata details, storage details like location, encryption and access controls.

Current challenges

High annual growth rates, continuously evolving data technologies and increasing amounts of information consumed and created—all severely limit the ability of organisations to efficiently handle their burgeoning volumes of data.

There are significant costs involved in identifying and managing dark data due to its ambiguous and unstructured nature, and the lack of trained personnel to perform comprehensive in-depth analysis and handling and cleanup of the data.

Increase in volume and severity of data regulations across geographies, especially in terms of data privacy and other laws (GDPR).

Key risk areas

Legal and regulatory risk

An organisation's dark data could contain sensitive information covered under laws or regulatory constraints—for example, data relating to credit cards, bank accounts, and patient records. A leak of such data would result in serious legal and regulatory issues and financial liabilities.

Reputation risk

An organisation's lack of insight into the extent and nature of its dark data significantly increases the risk of data breach, and severely prohibits any approach to avoid or remedy the same, which, in turn, increases the associated reputational risk.

Intelligence risk

An organisation's dark data may be used to surmise its business operations, competitive advantages and strategies, important partnerships, joint ventures, proprietary information like algorithms or formulae and other sensitive data. This kind of exposure would significantly impact an organisation's business, relationships, market foothold, etc.

Opportunity costs

An organisation which does not invest in identifying, analysing and mining its dark data pool cannot tap into it for potential market intelligence, leads and opportunities. In the absence of dark data monitoring, these opportunities may be lost to a third party, which will make efforts to access and exploit the same.

Open-ended exposure

Dark data, by definition, is complex and expensive to extract and analyse as it contains unknown amounts of sensitive proprietary, legal or regulatory data. If an organisation has no comprehensive policy for the identification and handling of its dark data, there is no way for it to estimate the extent of exposure or risk that it may face from these dark data sets.

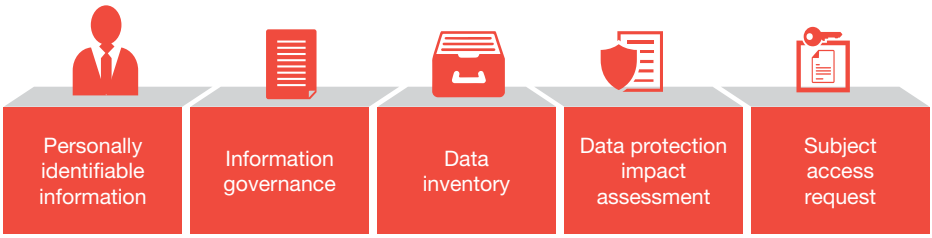
What can PwC do?

Identify personally identifiable information (PII)/business-related documents:

- Locate and classify personal/business and regulated data quickly, thoroughly and in a scalable fashion.
- Identify processes and sources of private data or documents to eliminate future problems.

Scalable retrieval and remediation: Search all possible data sources from a single compiled index to collect relevant information exhaustively, or copy, move, encrypt, secure, erase the information completely and with forensic accuracy.

Business benefits: Ensure that organisations meet their data privacy and other regulatory requirements. Mitigate risk of penalties and reputational damage.



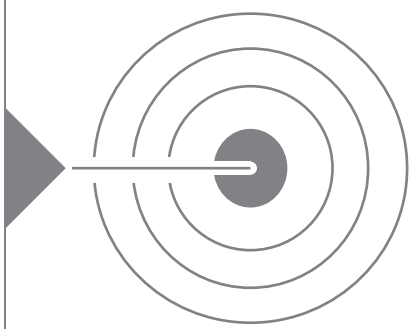
Our approach

Identifying the various dark data pools within your organisation.

Analysing data to understand metrics such as the data quantum, where data resides and its basic types (structured, unstructured, semi-structured).

Categorising data to begin understanding the quantity of each data type and the general nature of information included in those types such as format and age.

Classifying data based on how it will be handled—whether it will be archived/destroyed/analysed further. Once those decisions have been made, data can be grouped into different buckets for your team to review.



Benefits for the client

Speed and scale

Provide clients with the ability to comply with GDPR and incorporate principles in business processes with a minimal hardware footprint and within acceptable timeframes.

Operational efficiency

Understand and remediate threats to the business quickly, and locate, classify, and analyse sensitive and regulated data thoroughly.

Reduce risk

Identify the processes and sources of private data to protect it and eliminate future problems. Map risk content to location, owner, type, date, person or relationship.

Forensic accuracy

Ensure forensic accuracy and defensibility in reconstructing the extent and the timelines of incidents when reporting a breach and when searching data across your enterprise systems on the cloud or on premise.

About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 158 countries with more than 2,36,000 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, Chennai, Delhi NCR, Hyderabad, Kolkata, Mumbai and Pune. For more information about PwC India's service offerings, visit www.pwc.com/in

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

Contact us

Puneet Garkhel
Partner and Leader, Forensic Services
PwC India
M: +91 9820320181
E: puneet.garkhel@pwc.com

Dhritimaan Shukla
Partner, Forensic Services
PwC India
M: +91 98990 38326
E: dhritimaan.shukla@pwc.com

pwc.in

HS/Aug2018-14104