Do you DLP?
Maximising the business value of your Data Loss Prevention (DLP) solution

Data security

Roland Bastin
Partner
Advisory & Consulting
Deloitte

François Barret
Senior Manager
Advisory & Consulting
Deloitte

Data can be both an asset and a liability. As organisations grow, the volume and complexity of data required to support the business increases. All organisations store sensitive data that their customers, business partners, shareholders and the Board expect them to protect against theft, loss and misuse.

The intrinsic and contextual value of data and associated ownership risks vary throughout the data life cycle. The business value of information assets—gains on process and function performance, revenue and margin contribution—is a function of:

- Inherent value
- Contextual value
- Enterprise context
- Associated risk
- Cost of ownership

Data can be managed like any other enterprise asset, subject to the same net business value calculations balancing value, risk and total cost of ownership.

Figure 1 - Enterprise data lifecycle

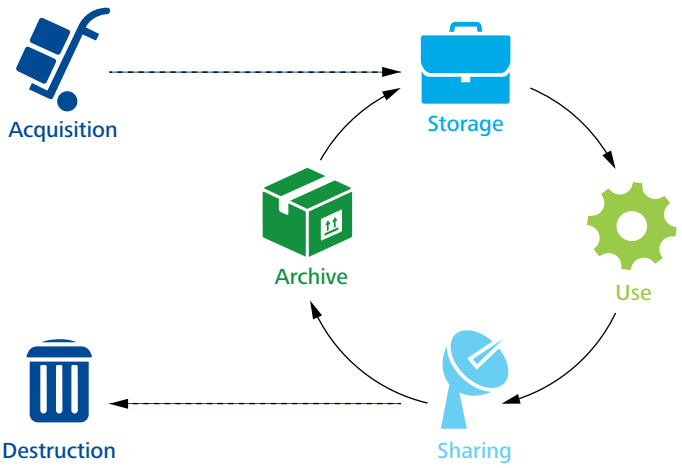


Figure 2 - Analysing data assets

Net business value	=	Data value	–	Data risk	–	Total cost of ownership
• Return on investment on managing data		Inherent value + Contextual value Data value		• Data in motion • Data at rest • Data retention versus liability exposure • Data leakage		• Managing data throughout its life cycle • Operating cost of supporting the infrastructure that handles the data - Facilities - Equipment - Staffing

When managed data and information has a negative net business value, the enterprise has several options, including:

- Increasing the value
- Reducing risk
- Discarding the data

Despite data management, high profile security breaches involving personal and corporate data continue.

What is data loss?

Data loss can be defined as the movement of an information asset from an intended state to an unintended, inappropriate or unauthorised state, representing a risk or a potentially negative impact to the organisation.

Data can be categorised using the following criteria:

1. **Form:**
   - Structured—hierarchical, relational, network: XML files, relational information (databases), files with detailed attributes, transactional information
   - Unstructured—free form (80% of potentially usable business information): email, blueprints, audio, video, images
2. **Type:**
   - Personal: credit card number, social security number, social insurance number, name and/or address, financial information, medical information, date of birth
   - Corporate: strategy, legal, intellectual property, intelligence information, financial information, sales information, marketing information
3. **The type of threat data is exposed to:**
   - Insider: disgruntled employee, ladder climber, petty ID thief, contractors, outsourcers, business partners/ vendors, fraudsters
   - Outsider: spies and industry espionage, gangs, ideologists, cyber terrorists, scammers (e.g. phisher), social engineer, script kiddies

Data loss can come in many forms, and may compromise various types of personal or corporate information. Data is being targeted by both internal and external groups.

A number of factors are driving organisations' data loss prevention needs: globalisation, varying regulations, varying customer expectations, customer privacy sensitivity, brand risk, advances in technology, mobile devices, advanced persistent threats (APT), extended enterprise, third party service provider risk, regulation and compliance (anti-money laundering, breach notification, PCI-DSS, GLBA, etc.) and data growth.

The data explosion

There has been massive growth in data volumes in recent years. Almost 3 trillion gigabytes of information was created and replicated as of 2012, compared to over 1 trillion in 2012 and 130 million in 2005. There are several factors driving this data growth and the associated challenges, including:

- **Globalisation:** *"70% of economic growth over the next decade will come from emerging markets, with China and India accounting for 40% of that growth"*[7]
- **Organisation:** *"40% projected growth in global data generated per year vs. 5% growth in global IT spending"*[8]
- **Consumerisation:**
  - *"On an aggregate, 56% of companies say yes to consumerisation and allow employees to use their personal devices for work-related activities"*[9]
  - *"31% of the mobile devices connecting to the corporate network are owned by the employees: 66% are laptops, 25% smartphones and 9% are tablets"*[10]

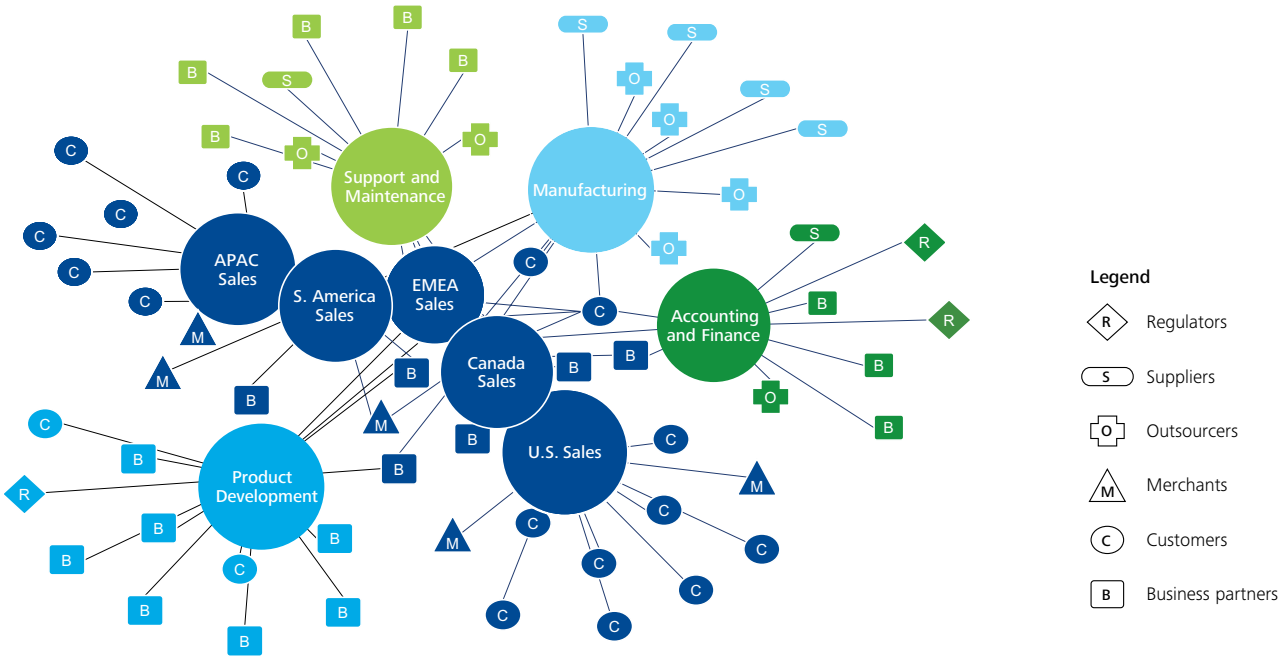
The rise in data volumes is forcing organisations to re-evaluate and refocus their information management practices to better integrate and leverage data in core business processes.

Sensitive data such as personal and financial information and intellectual property moves horizontally across organisational boundaries, including vertical business processes. Organisations commonly do not have a good understanding of the movement, proliferation and changes in their data leaving them susceptible to data loss.

Additionally, organisational boundaries are changing as enterprises become more virtual, blurring the distinction between internal and external. Perimeter-centric security often hinders business growth and brings a false sense of security when it comes to data protection.

7 World Economic Outlook Database, International Monetary Fund, UNWTO World Tourism Organisation
8 McKinsey Global Institute—Big data: The next frontier for innovation, competition, and productivity
9 Trend Micro Consumerization Report 2011
10 Trend Micro Consumerization Report 2011

Figure 3 - Blurred organisational boundaries



How data loss can happen to your organisation

Sensitive data can be lost or compromised in a number of intentional or unintentional ways, due to 'threat agents' (employees, users, hackers, etc.) acting in a malicious or innocent manner. Some common data loss scenarios are:

- Data in use (i.e. 'What is the agent doing with it?'):
  - Disgruntled employees copying files containing personal or confidential information to portable devices (e.g. flash drives)
  - Users printing sensitive data to equipment in common areas which can be accessed by others
- Data in motion (i.e. 'Where is the data going?'):
  - Users sending sensitive data to personal webmail accounts in order to work at home
  - Personal and confidential information being shared with third parties for valid business purposes using insecure transmission protocols
  - Malicious insiders transmitting personal and confidential information outside of an organisation's network
- Data at rest (i.e. 'Where is sensitive data located?'):
  - Business users innocently placing personal information in insecure storage locations where access is not administered by IT
  - Database administrators storing (unencrypted) backup copies of sensitive data in unapproved locations

The intrinsic and contextual value of data and associated ownership risks vary throughout the data life cycle

Data loss can come in many forms, and may compromise various types of personal or corporate information

Data loss proliferation

Data is growing at an exponential rate, as is the number of incidents in which data has been lost.

More than 1600 data loss incidents occurred[11] last year (See Figure 4).

Incidents involving digital media and hacking are most common[12] (Figure 5).

Data loss is occurring across industries, affecting organisations of varying sizes and different types of information assets[13] (Figure 6).

The variables to take into account when calculating the cost of a data loss incident are:

- Brand impact:
  - Media scrutiny
  - Loss of customers
  - Loss of business due to critical intellectual asset loss
- Regulatory impact:
  - Independent audit fees
  - Regulatory fines
- Financial impact:
  - Notification
  - Lost business
  - Response costs
  - Competitive disadvantage
- Operational impact:
  - Diversion of employees from strategic initiatives to work on damage limitation
  - Need to implement comprehensive (additional) security solutions

11, 12, 13 http://www.datalossdb.org/statistics

Figure 4 - Number of data loss incidents over time

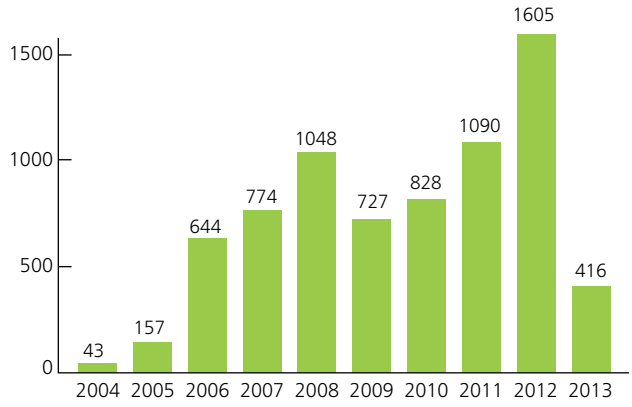


Figure 5 - Types of data loss incidents

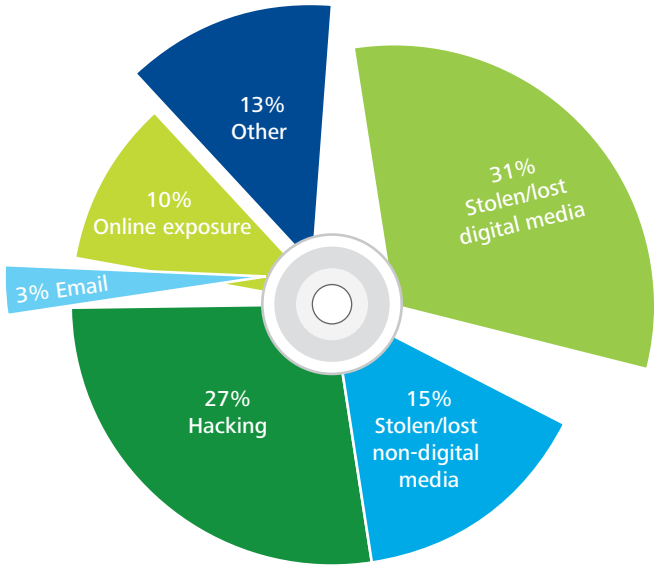
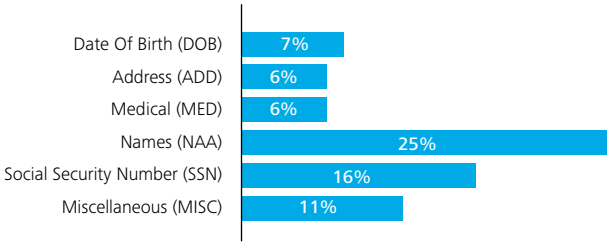


Figure 6 - Types of data loss



Moreover, a recent study by the Ponemon Institute[14] shows that the cost of data loss is steadily increasing.

Figure 7 - Average cost per record by cost activity

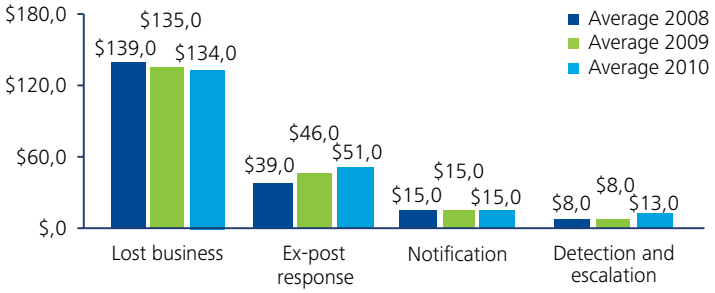


Figure 8 - Cost per record of direct and indirect costs

The cost to organisations occurs at each stage of the incident response life cycle—detection, notification, post-response—leading to the cost of lost business.

The cost of lost business has remained relatively stable last four years, and now averages US$135 per record compromised, or 63% of data breach costs

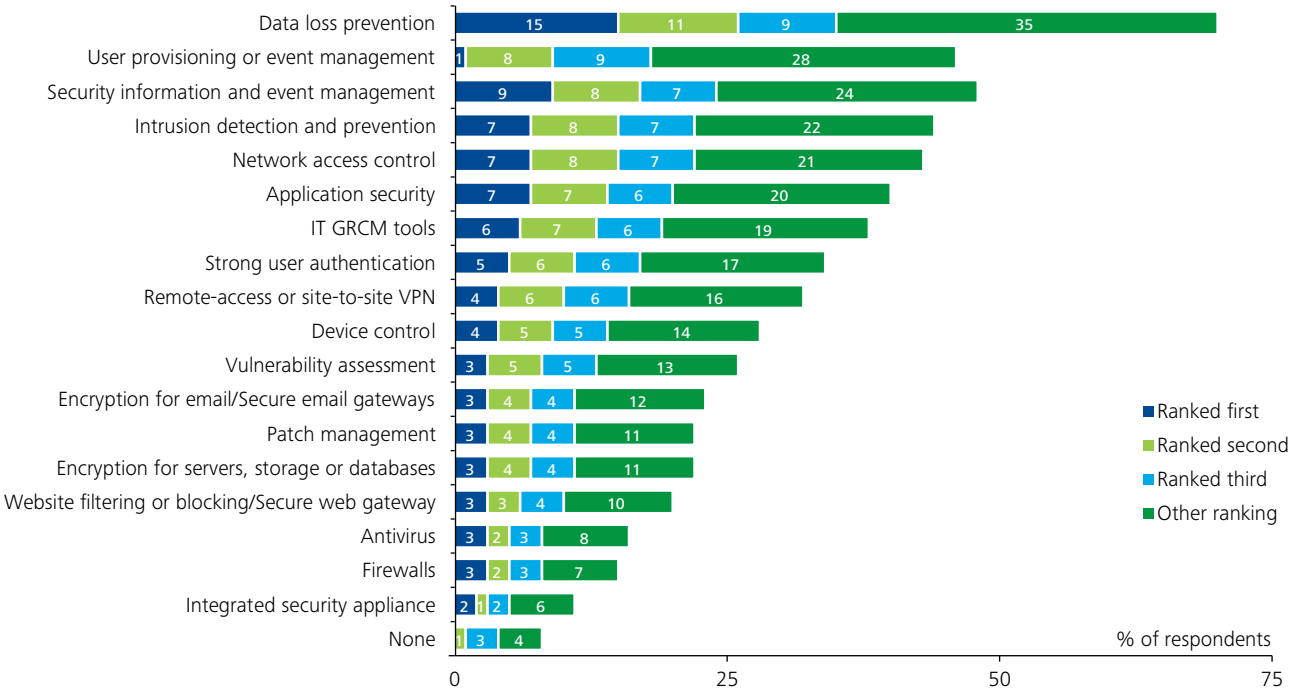
Data loss can have a significant impact on an organisation's bottom line, which is why organisations are increasingly turning to data protection measures in order to prevent data loss.

Data protection is a general term that encompasses a number of measures, including:

- **Data encryption**—this refers to a method of modifying data so that it is meaningless and unreadable in its encrypted form. It must also be reasonably secure, i.e. it must not be easy to decrypt without the proper key
- **Data obfuscation**—this is when data is rendered unusable by some means, but it is not considered a reliable form of encryption (obfuscating the data with a simple substitution cipher is not considered encryption)
  - Substitution, which replaces a value in the column with fictional data
  - Randomisation, which replaces the value with random data
  - Shuffling, which switches column values between records
  - Nullifying, which replaces column values with NULL
  - Skewing, which alters the numeric data by a random variance
  - Encryption/decryption, which employs reversible scrambling
- **Data masking** is a method of hiding sensitive data in a way that the clear text cannot be reconstructed from the displayed data. This is useful in situations where it is only necessary to display a portion of the data
- **Data generation** is a method of creating fictional data following certain patterns to completely replace the original data set with the intent of being fully displayed
- **Data redaction** is a method of locating unstructured data in the document, indexing it using OCR, and masking or obfuscating as appropriate
- **Data loss prevention**, which according to a recent Gartner survey , is the top priority for organisations implementing security technologies

14 http://www.symantec.com/about/news/resources/press_kits/detail.jsp?pkid=ponemon

Figure 9 - DLP implementation trends



What is DLP?

Data Loss Prevention (DLP) should be part of an overall information risk and data protection/privacy strategy. It starts with understanding what your assets are. Not all data can be protected equally—you must first understand what needs to be protected the most.

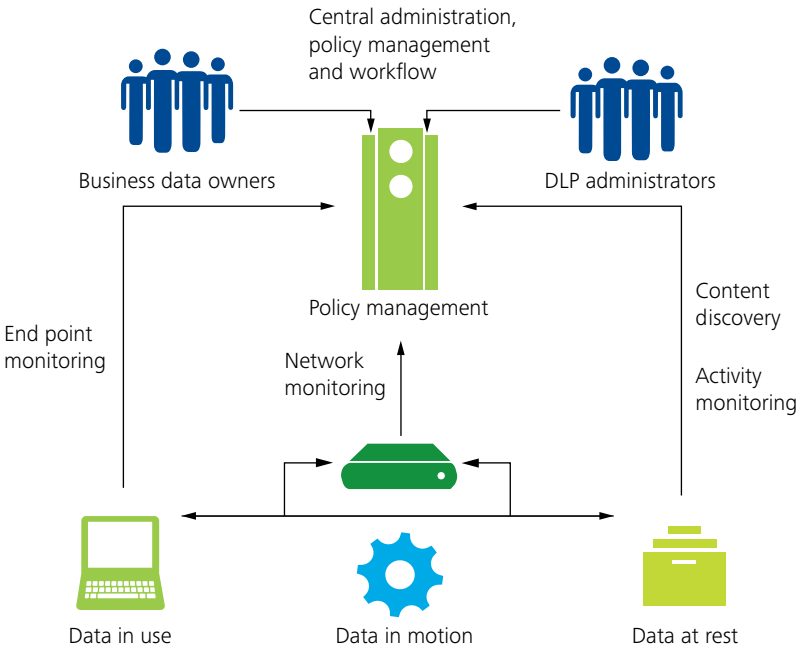
DLP involves tools that monitor, identify and protect electronic data as it moves to, from, and through an organisation. Typically, data can be described as being in a state of use, motion or rest:

- **Data in use:**
  - Monitor user interactions with data to identify, for example, attempts to transfer sensitive content to a USB drive and apply policy
  - Common controls include disabling Copy, Print, Print Screen, Open, Paste, Save, Save As, and Notification
- **Data in motion:**
  - Analyse data traffic over the network to identify sensitive content being sent via
  - email, IM, HTTP or FTP, and apply policy
  - Often requires integration with mail transfer agents, network components and other infrastructure
  - Common controls include Allow, Audit, Quarantine, Block, Encrypt and Notification
- **Data at rest:**
  - Scan and inspect enterprise data repositories to identify sensitive content and apply policy accordingly
  - Common controls include Encryption, Obfuscation, Quarantine, Deletion, and Notification

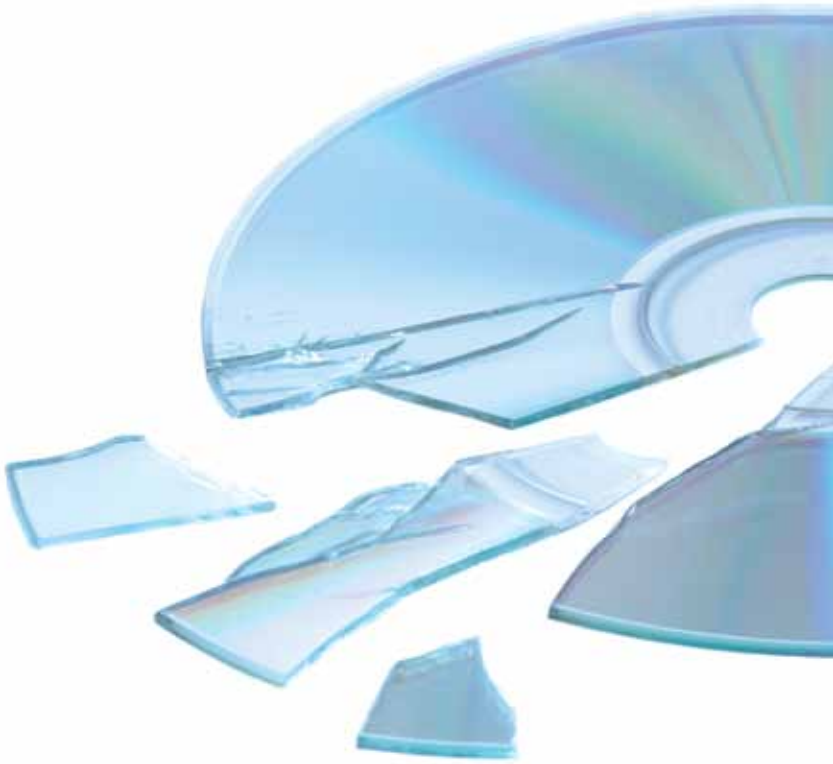
DLP tools typically consist of the following components:

- **Policy Management and Enforcement Servers:** a central platform for defining, deploying and implementing enterprise-wide DLP policies across various DLP components. Management servers are also used for incident response workflow management and reporting
- **End-point agents:** located within end-user devices such as desktops, laptops, etc. These agents discover and collect data on Data in Use activities performed on the device and are responsible for enforcing DLP policies on the device and reporting back to the Policy Management and Enforcement Server(s)
- **Network components:** can monitor network communications and restrict the flow of Data in Motion as necessary. Network components provide real-time monitoring and reporting of policy breaches to the Policy Management and Enforcement Server(s)
- **Discover components:** together with end-point agents, these components perform discovery activities for Data at Rest. Data discovery is based on the policies defined in the Policy Management and Enforcement Server(s)

Figure 10 - DLP solution conceptual model

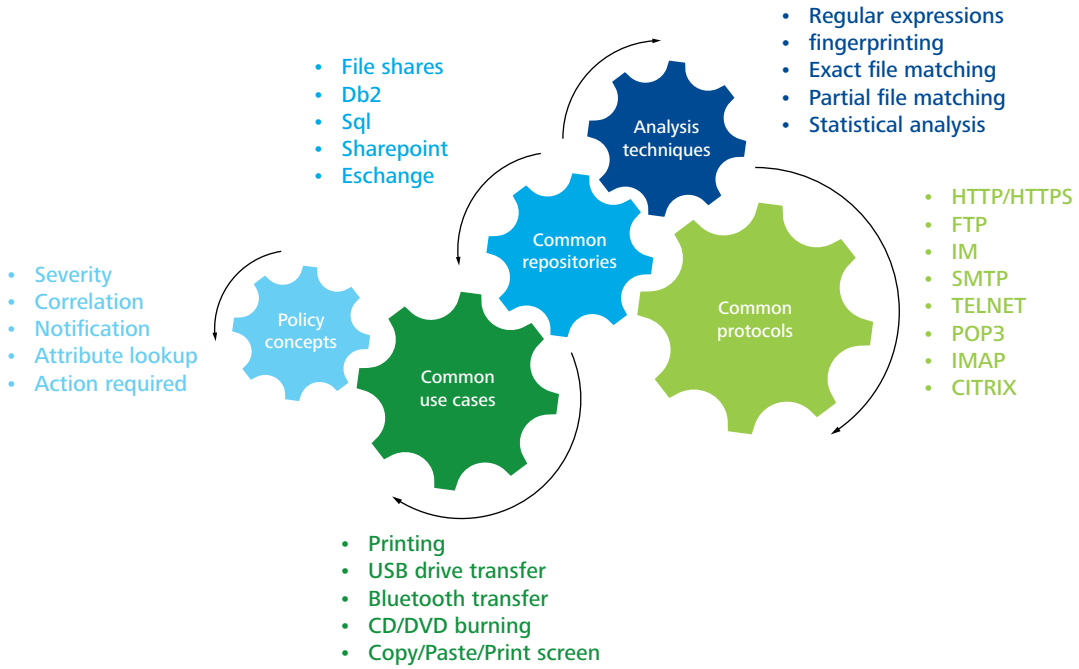


More than 1600 data loss incidents occurred last year



DLP tools vary significantly in their capabilities and have different strengths and weaknesses. However, there are some key capabilities and concepts that are generally applicable to most DLP tools, as summarised below:

Figure 11 - DLP key capabilities and concepts



Common DLP deployment challenges and their root causes

DLP solutions often do not achieve full business and data loss mitigation due to a number of common, but preventable challenges and root causes, including:

Challenges	Root causes
Business and IT sponsor frustration with the speed at which the solution becomes functional	• Lack of a DLP strategy provides no clear vision and direction for the solution • Poorly defined requirements cause work to be repeated, with a related cost • 'Big Bang' approach vs. proof of concept, pilot and phased implementation • DLP vendor marketing promises fail to materialise
Complaints from executive stakeholders that they don't understand the value the solution offers	• Poorly defined or lack of DLP metrics and success criteria • Inability to collect and report on metrics
Business community pushback due to a lack of communication or transparency	• Poorly defined or lack of a training, awareness and communications plan
Inability to correlate and report upon DLP and other types of security incidents and associated risks	• Lack of integration between DLP and Security Information and Event Management (SIEM) solutions • Lack of integration between DLP and Governance, Risk and Compliance (GRC) solutions
Advanced capabilities such as deleting, blocking, encrypting and quarantining are rarely implemented	• Lack of processes for business use case analysis and approval • Policies defined based on content vs. contextual analysis • Lack of processes for enabling efficient recovery of blocked or quarantined information • Lack of processes for managing encrypted messages/transmissions/files
Data in Use capabilities are rarely implemented, if at all	• Lack of processes for deployment and management of thousands of agents • Endpoint technology limitations or incompatibility with vendor solutions
Incidents are not responded to in a timely manner or at all, or all incidents are treated as "equal"	• Poorly defined or lack of incident severity levels and response workflows/ procedures • Roles and responsibilities not clearly defined • Insufficient training and resourcing of incident response team(s) • False positives caused by 'loosely' defined policies
High volumes of false positives lead to support team frustration, or legitimate business processes are blocked	• Lack of processes for business use case analysis and approval • Policies defined based on content vs. contextual analysis • Lack of sufficient testing and fine-tuning of policies over time before full-scale deployment
Sensitive personal and confidential information is consistently found in unanticipated/ undesirable locations and detected leaving the organisation's network	• Poorly defined or lack of data classification policy • Policies defined to monitor/search for minimal data elements and/or files • Lack of an inventory of network egress points, storage repositories and end points • Lack of business process re-engineering • Poor communication with business users regarding security expectations and their responsibilities • Poorly defined or lack of disciplinary measures and enforcement

Our approach

In our experience, a successful DLP solution/program must be approached holistically, focusing not just on the technology, but also on the people and processes needed to support and interface with the system(s). The approach we propose is as follows:

Governance

- DLP strategy
- DLP requirements
- Organisational structure
- Policies and procedures
- Training and awareness
- Metrics, monitoring and reporting

Process

- Business process analysis
- Incident response workflows
- Incident response plan
- Tuning and optimisation
- Policy change management
- Help desk procedures
- Business process re-engineering

Security integration

- Integration with enterprise security solutions

System implementation

- Hardware and software
- Egress points
- Storage repositories
- End points
- Policy configuration
- Access configuration

This approach integrates people, processes and technology. It allows DLP solutions to be aligned with business drivers and value

Key considerations for a successful approach

Below are some key considerations that should be taken into account as a first step towards a successful DLP tool selection and subsequent implementation:

Domain	Key considerations
General	• What information or data elements present the most risk? • What locations or business units present the most risk? • What are our mitigating controls? • How robust do we need our governance structure and incident response workflow to be to support our goals and mitigate our risks? • What type of resourcing do we need to support management of the tool and the incidents it generates on an ongoing basis?
Data at rest	• What types of data repository does the solution need to be able to scan? • What do we plan to do with the data once it is found?
Data in motion	• Do we care about outgoing transmissions only, or incoming and internal transmissions as well? • What protocols do we need to monitor and protect? • Do we need to block or encrypt traffic?
Data in use	• What platforms does the solution need to support? • What do we want the tool to accomplish when users are not on the network?

Conclusion

Approaching DLP in a more holistic manner and treating it as a program to drive organisational change, minimise business risk and realise full business value, as opposed to treating it as a technology "plug and play" type of solution, will bring some of the following key benefits:

- Clearly articulates the DLP program vision and strategy
- Helps prevent the cost of repeating work through clearly defined scope and requirements
- Demonstrate business value through 'quick wins'
- Maintains stakeholder support through clearly defined metrics and success criteria
- Helps to prevent business community and end-user outcry through well designed, planned and delivered training and communications
- Enables the use of advanced system capabilities that can help prevent significant legal, regulatory, compliance and brand issues
- Improves incident response capabilities, helping the organisation to respond more efficiently and effectively in the event of data loss
- Helps prevent business interruption through advanced search/monitor policy definition that consider not only content but context
- Facilities advanced incident correlation and reporting on governance, risk and compliance issues through integration with other security technologies