armadillo

THE BRITISH ASSESSMENT BUREAU

ISO27001

Certification No. 200944

Considerations when embarking on a Data Governance Programme

May 2019

Considerations when embarking on a Data Governance programme

Where to Start?

The foundation of any successful data governance programme is a genuinely identified need, with support from senior leadership within the business. This need could be as clear cut as an imposed requirement from a regulator or as part of your legal obligations, or as complex as a need to regain control over a huge estate of aging and poorly sorted data in order to reduce the risk of a breach.

In either case, and certainly before tools and technologies are considered, you'll want to be sure about 'what good looks like' and you'll need a solid backing from senior leadership as implementing a data governance programme can have effects across the organisation and will need support from many different stakeholders.

The Pillars of Data Governance

Data Governance is often the driver "to take control" of data across an organisation. However, looking at data governance at the organisational level touches on many aspects of security. Gartner proposed the term 'Data Centric Audit and Protection' to refer to the programme of work needed to move away from a siloed approach of focusing on the security of applications which hold or process data to focusing on the data itself. Data which is important to the business rarely is created, used and retired all within the same system. More commonly, it is moved and transposed by users and other systems across the internal network and beyond, changing shape as it goes. A siloed approach is the worst of both worlds, it can create roadblocks for legitimate data access, while having blind spots around accidental and malicious use.

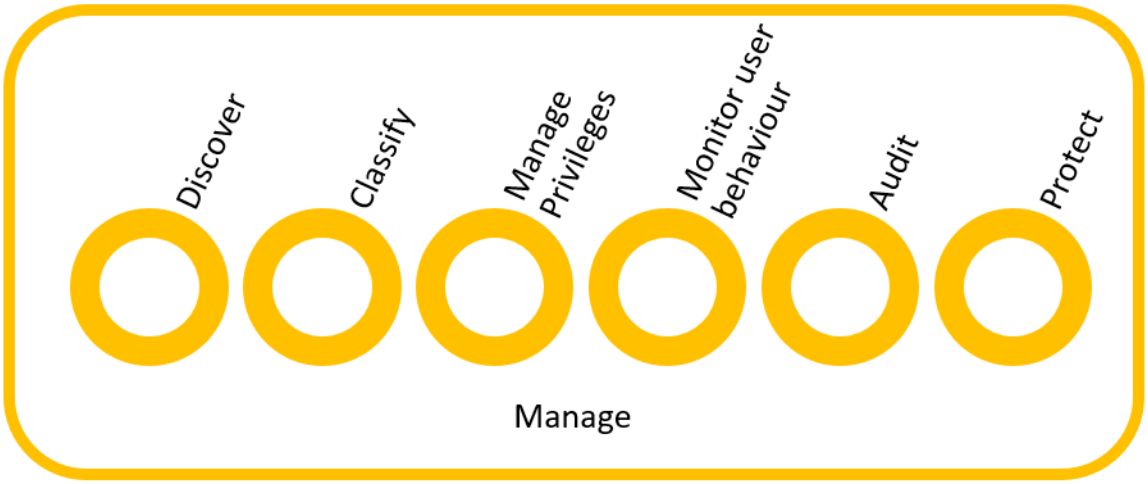

Figure 1 - Pillars of a DCAP Programme

Let's take these pillars one by one, with some suggestions around the approach taken when planning the programme, and any vendor evaluations.

Discover

Think about all the places that data could reside across an organisation. Are you focussing on shared drives and servers only, or do you need to get a grip on data stored locally on endpoints too? Unstructured data such as Microsoft Office files are usually the first thought, but what about information stored in databases – could there be information which is subject to a data subject access request under GDPR in there? What about cloud services?

Checklist

→ Ensure all data repository types have been identified and that the chosen platform can support them
→ Ask what support the tool has available to assist with discovering data repositories which you may not be aware of
→ Be sure to consider sanctioned cloud services, not just on-premises – and think about the direction of travel for organisations planning a cloud migration
→Do you need a process to cover GDPR DSARs?

Classify

Many organisations have a written classification policy – whereby documents or emails are classified as public, internal, sensitive etc. Not many organisations can say that all users could explain what each classification means or how the labels indicate that data should be managed. The classification policy should be granular enough to allow automated handling of data, but simple enough that users feel confident that they understand the classification of data that they create and manage.

Types of Classification

- **User-driven:** User driven classification is an evolution of the old manual process of 'labelling' a document with a classification in the headers and footers. Using a software tool, the person creating or updating a document or email can be prompted to select a classification from a pick list. As well as placing this classification onto the document as a visible marking, the classification can be added as meta data within the document. This can be read by other software tools such as Data Loss Prevention where policy decisions can be made based on the classification. The rationale behind user-driven classification is that the data owner typically has the context and knowledge required to properly label a document and therefore they should be empowered to make the decision on the classification. This also helps to involve users in data security, and 'live' the policy.

- **Automated:** Rather than rely on the user to classify the document, this method uses rules or machine learning to classify the document automatically. For example, searching for credit card numbers and then marking these documents as "Private and Confidential". The advantage of automated classification is that it can often detect misclassification by malicious users.
- **Hybrid:** a hybrid approach is a best practice, using both a user driven and automated classification together. This provides a 'trust but verify' approach which combines the best of both worlds – empowering the data owner but checking the classification against organisational policies.

Checklist

→ A written policy to define your categories is a vital foundation to build upon before considering technology
→ Determine if you have any existing tools providing classification – typically these may have existed in siloes within finance teams or legal teams, often using their own classification schema. Decide if these can be integrated or need to be replace with an organisation wide tool
→ Do your use cases lean towards a user-driven or automated approach or would you benefit from hybrid?

Manage Privileges

Once the location and types of data across the organisation are understood, you can start to understand how they are being accessed. Is that finance share accessible to just the finance team? Does Bob who used to work in accounts but now is in HR still have access and no-one realised? Do IT administrators have too broad a level of access, just because they can? How are user roles mapped to their need to know? Is there a need to enforce segregation of duties through data access?

Checklist

→ Do you know who your data owners are? If not, do you need an automated way to discover them?
→ Do you have any existing tools to check the permissions of files and folders for over-permissioned documents?
→ Do you want to be able to prevent classified documents from leaving the network? If so, you may need to consider integration of Data Loss Prevention technologies which will be important to your strategy.
→ Do you have pre-defined roles into which users are provisioned? These typically help to reduce management overhead instead of applying permissions directly to users

Monitor User Behaviour

Just because two hundred people have access to a file, it doesn't mean they all need it. There may be an opportunity to reduce the visibility of sensitive data and in doing so, reduce the risk. Consider how easy it is to create rules for who should be able to access what, or would machine learning help to infer rights and anomalies based on past history and what peers are doing? How would you like to be alerted to possible misuse, in real time or after the fact?

Checklist

→ Do you have a tool to monitor the use of documents and flag for anomalies? This could be either rules based (for example someone accessing 10,000 document per day) or Machine Learning based (someone accessing files outside of their team's usual pattern)

→ How are you going to respond to incidents generated by suspicious file access? Maybe you have an existing SIEM you need to ensure is supported by any new product?

→ Have you considered what you will do when examples of poor or malicious behaviour is detected? Are HR and legal teams fully briefed on the implications of implementing such a product?

Audit

Think back to the reasons why you embarked on this process – if it was due to a regulatory requirement then you better be sure that the programme you put in place can generate reports which quickly and simply demonstrate adherence to that regulation. If you are aligned to multiple requirements, can you report specifically on each?

Checklist

→ Ask for sample reports based on your specific regulatory framework and seek references if possible.

→ Does the tool support segregation of duties itself via role-based access?

→ Do you have a means of determining data which is past its retention period and/or not been accessed for a significant period. This stale data should be considered for secure deletion, as its value may be minimal, but it still poses a risk if it was to be part of a breach

Protect

Finally, you need to consider what will happen when invariably someone does something outside of the approved workflows. This could be a malicious insider try to exfiltrate data, or a well-meaning insider working around your business processes to meet an urgent client requirement. The level of protection desired is on a

continuum, and the desired balance between security and user experience will depend on the culture of your organisation. As well as active blocking of users from accessing or sending data, more passive approaches such as encryption, tokenisation, redaction or masking could be considered depending on the use case.

Checklist

→ Prioritise your requirements – it's unlikely you will be able to do everything at once
→ Blocking should be the end goal – but it will take some time to get there. Expect to make small, incremental steps through monitoring and warning before blocking is used
→ Think about realistic timeframes and seek sample implementation plans from implementation partners. Ask what resource will be needed from your organisation and start thinking early about who will contribute to a steering group
→ Consider the organisational culture and how much change could be borne and over what period. Where does your organisation sit on the continuum of usability vs security?

Manage

Arguably most importantly, to achieve value from the programme and truly break down the barriers between silos of data, there must be a way to manage all of these processes. These should be integrated into the day to day business operations of the organisation. It is important to consider this not just from an IT administrators' perspective but to include all stakeholders who will be involved in the process – end users as data owners, approving managers, auditors…

"Good" for your organisation might not mean you need to complete all of these pillars, or you may have priorities which mean you will only focus initially on a couple of these. In any case, embarking on a data governance programme is a journey which will take time and needs to be planned and executed in careful stages – it is not an overnight process.

Checklist

→ Make sure you don't end up moving from one siloed set of controls to a different set of siloed controls!
→ Depending on your use cases, one product may meet all your requirements. If so, thoroughly road test the interface for usability and ensure representation from the different use groups.
→ If multiple tools are required (including pre-existing ones) then you should consider if they need to be interconnected and how easy this will be. APIs are an easy response, but do you have the skills to implement this yourself or will you require professional services?

Adjacent Technology Areas

The following technology areas are closely related to Data Governance and may need to be considered in conjunction with it to complete the Data Governance journey:

User-driven Data Classification

If existing software is in place for user driven data classification (for example Boldon James Classifier, or Titus) it should be considered how this would work together with automated tools. For example, a conflict in classification between the two systems could indicate areas of the business where more training is required, or it may highlight an attempt to evade security controls in order to exfiltrate data. It should be investigated how the classifications can be mapped between the two solutions.

Data Loss Prevention

DLP tools may be required for the implementation of network-based blocking over email and web channels. Many DLP solutions can carry out automated classification of documents and emails as they leave the business, so if an existing toolset is in place, you may decide not to duplicate this in a data governance solution.

Identity and Access Governance

A well defined and managed process to manage the automated application, amendment and revocation of access rights to applications as employees join, move roles and leave the organisation can be a good way to enforce access to applications as well as data and may be a good consideration if it transpires than maintaining a secure permission set is a problem within the organisation.

Next Steps

Armadillo have experience with many of the vendors in this space and would be happy to broker any vendor conversations, demonstrations or even run an RFx on your behalf.

We recommend that we work together to define a concise set of requirements, based around your business requirements and from there we will be in a better position to recommend the best vendors specifically tailored to your requirements.