CORE CONCEPT SECURITY

Which Data Discovery Solution is Right For Your Business?

June 2019

CONTENTS

EXECUTIVE SUMMARY

While there are *many* uses for a data discovery tool, I have chosen GDPR as the primary impetus for getting one. No other regulation makes it such a necessity.

Even though GDPR has been enforced for just over a year now, many organisations are still unsure of where to even begin.

This is rather inexcusable given the enormous amount of free guidance out there. While yes, a lot of this 'free' advice is exceedingly self-serving, a little bit of homework would have provided everything needed to separate the good advice from the bad. Mostly anyway.

Just reading the GDPR through a couple of times, then turning to your supervisory authority for country specific guidance on implementation, would have enabled you to get started asking the right people the right questions to help *your* business. It's what I did.

It won't be very long before those who have done nothing will viewed in the same light as those who have been breached. Rightly so, because they are the ones who will be breached next.

So this white paper is not really for those who are just getting started on their GDPR journey. Trying to buy solutions first would suggest that you're throwing technology at a *business* problem. Instead, this is for people who have a reasonably good idea of what the steps to compliance are, have run an appropriate risk assessment, and are ready to pull the trigger on two of the most important steps in achieving compliance;

Data discovery and business process mapping.

Further, this is for organisations who have already performed some form of *manual* personal data to business process mapping and are now looking for confirmation that they have covered all their bases. There is not one data discover tool or even solution that will provide all of the business/legal context you need, but it's equally naïve to assume that *people* know where all of the data is. Departmental stakeholders will do their best to help, but you need to ***know***.

And that's the point of this whitepaper; *Reasonable* certainty.

If you're reading this because you heard that a data discovery tool will make you GDPR compliant, go back and start again because you missed the point. If you've already spoken to a few vendors, it's likely they were grossly exaggerating the 'deliverables'.

A data discovery tool can be an extraordinarily valuable asset across every facet of your business, if, and ONLY if it's used correctly. There are no shortcuts here.

Hopefully, by the end of this white paper, you will be on your way to answering 3 very important questions:

1. Do I even need a data discovery technology / solution?;
2. If yes, do I know exactly what it will be doing for me?; and
3. Can I ask the right people the right questions to select the right one?

ADDRESSING THE CHALLENGE

Grossly simplified, GDPR requires all organisation that process personal data (and I can't think of one that doesn't) to do so in a manner that is:

1. Lawful;
2. Fair;
3. Transparent
4. Specific;
5. Limited to what is necessary (processing and storage);
6. Accurate; and
7. Secure

There is no way to achieve this end unless you know:

1. where *all* the personal data is;
2. *exactly* what you're doing with it; and
3. how it's protected

Few organisations have any understanding of their data flow life cycles, let alone an understanding sufficient to *demonstrate* compliance to GDPR's numerous, and rather strict requirements.

For most organisations this means they will have to perform a data discovery and mapping exercise that should, ideally, take two forms:

1. First - Running a series of interviews and questionnaires with all unique departmental stakeholders (Sales, HR, Operations etc.) to manually map business processes and their corresponding data sources; and
2. Second - Running some form of data discovery solution to find data on end systems / databases / other file stores etc.

The interviews and questionnaires serve multiple purposes, all of which are critical for the success of a subsequent data discover exercise:

1. Each session should include a mini training exercise so that all departmental subject matter experts (SMEs) become increasingly familiar with GDPR;
2. SMEs now become 'departmental evangelists' to spread the word;
3. The GDPR project team becomes intimately familiar with how the business is run;
4. 'Hidden' or offline data stores can potentially be brought to light;
5. Data discovery output can be put into a proper business context

However, before getting to the data discover phase, there are not only a host of prerequisites that should be addressed, but a significant number of subsequent actions that should be thought through.

To put data discovery into an appropriate context, we must first determine what a start-to-finish GDPR compliance project looks like.

WHAT DOES A GDPR PROJECT LOOK LIKE?

Note: If you already have a compliance project in place, feel free to skip to the next section.

Every organisation is different, but regardless of your industry sector, size, or region, the below phase descriptions cover the majority of a GDPR project done well.

At a very high level, a project *should* contain all the following.

Step 1: Prerequisites[1]

1. READ IT! - There is not one person who does business with organisations 'established in the Union' to whom the GDPR does not apply. You are responsible in some way of ensuring other people's personal data is protected, and you should be aware of *your* rights when it comes to your *own* data;
2. Senior Leadership Buy-In - Like everything else, if the top people in an organisation are ignorant and/or ambivalent on a subject, nothing will happen. Your Board of Directors don't have to be GDPR experts themselves, but they had better take it seriously as they will ALWAYS be held liable if things go wrong. You can outsource a large part of the of GDPR *function* (even the DPO), but never the *accountability*;
3. Stakeholder Training - This can be as simple as a one-day engagement where an appropriate representative of each departmental vertical (HR, Sales, Operations etc.) undergo GDPR training bespoke to *their* business needs;
4. Designation of Project Ownership - If you already have a Governance function, this is easy, just give it to them. If you don't, you will have to assign the initial project to someone(s) with enough knowledge, influence, AND management support to be effective;
5. Wider Business Context - You must first decide on your objective: Regulatory compliance? Best in class? Best in the world? e.g. if you are functioning as a 'Processor', then your prospective client's policies and SLAs may go well beyond minimum legal requirements.

Step 2: Data Discovery[2]

Even if it's too late to perform the above steps, but definitely before you begin questioning prospective *legal* experts, you must know what personal data you've got, and what you're doing with it.

Under GDPR you are responsible for (amongst other things):

1. Determining your lawful basis for processing for each of your separate business processes (both internal and client facing);
2. Implementation of data subject rights in-line with 1. (erasure, portability etc.);

[1] See *GDPR Compliance Step-by-Step: Part 1 – The Prerequisites*
[2] See *GDPR Compliance Step-by-Step: Part 2 – Data Discovery*

3. Data minimisation (during collection and data retention);
4. Data confidentiality, integrity, and availability (all to defensible and demonstratable levels);
5. 'Legitimising' all transfers of, and responsibilities for, data to third parties / third countries

You must therefore determine what information you need to collect for the lawyers to get these needs met. It's actually quite straightforward.

Manual Process – (mandatory, and should be performed *first)*

A questionnaire and interviews can be used to collect the necessary information from each departmental subject matter expert(s). *Every* known business process should be examined. For example:

Sales Processes

- New Sales
- Up-Sell/Cross-Sell Existing Clients
- Account Maintenance

Finance Processes

- Recurring Billing
- Invoicing
- Payroll

HR Processes

- Recruiting & Onboarding
- Salary and Benefits
- Disciplinary

…and so on.

The completed questionnaire will look something like this:

Department: Sales
Sub-Department: Inside Sales
Sub-Sub-Department: Telesales
Business Process: New Client Acquisition

						DATA REPOSITORIES					
					Name:	Any CRM	Prospoects4U	Campaigns-R-Us	Email		
					Functional Responsibility:	Third Party (EU)	Third Party (EU)	Third Party (TC)	Internal (EU)		
					Data Format:	Application DB	Soft Copy (Editable)	Direct Connection (API)	Email		
Data Fields	Category of Individua	Category of Personal Da	Mandator	Responsibil	Data Type	X = Applicab	X = Applicab	X = Applicab	X = Applicab	Retentio	Comments
First Name(s)	Client - Prospective	Core Personal Data	Yes	Controller	IID	X	X	X	X	Indefinite	Data will be retained indefinitely to effect right to be erased.
Last Name	Client - Prospective	Core Personal Data	Yes	Controller	IID	X	X	X	X	Indefinite	Data will be retained indefinitely to effect right to be erased.
Title(s) (Mr, Mrs, Dr, etc)	Client - Prospective	Core Personal Data	No	Controller	IID	X	X	X	X	Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Gender	Client - Prospective	Core Personal Data	No	Controller	IID		X			Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Workplace Address	Client - Prospective	Contact Data	No	Controller	IID	X	X			Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Workplace Email	Client - Prospective	Contact Data	Yes	Controller	DID	X	X	X	X	Indefinite	Data will be retained indefinitely to effect right to be erased.
Workplace Phone Number	Client - Prospective	Contact Data	Yes	Controller	IID	X	X	X		Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Job Title (or Job Code, Function, etc)	Client - Prospective	Role Data	Yes	Controller	IID	X	X	X	X	Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Department (or General Ledger Code, Organizational Unit, etc.	Client - Prospective	Role Data	No	Controller	IID	X	X			Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Supervisor (or manager, reporting line etc.)	Client - Prospective	Role Data	No	Controller	IID		X			Indefinite	Data will be deleted if data subjkect opts-out of further marketing.
Private Mobile Phone Number	Client - Prospective	Contact Data	No	Controller	DID	X	X			Indefinite	Data will be deleted if data subjkect opts-out of further marketing.

It's very likely that the SME will have no idea how to find some, or even most, of this information, but that's OK, now you *know* what you don't know.

Once you have found as much information as possible, and assigned the tasks to find the rest, you can move to the next step.

[Note: For more information on conducting this exercise and for access to the Questionnaire Template and User Guide please go here: GDPR: Getting to the Lawful Basis for Processing]

Technology-Driven Process – Optional, but *highly* recommended

The above process will leave significant gaps, so it's critical that you cover all bases by looking for personal data where no-one knows it even exists. It's usually the stuff you don't know about that gets organisations into most trouble.

Step 3: Process Mapping[3]

If you have performed the data discovery exercises laid out in the previous section, you will now have a bunch of data with only limited context. For data to become *information*, you need to provide the appropriate context, which in GDPR terms, is in the form of a 'business process'.

It will look something like this:

DETAILED PROCESS NARRATIVE		
Department:	Sales	
Sub-Department:	Inside Sales	
Sub-Sub-Department:	Telesales	
Business Function/Process:	New Client Acquisition	
Affected Legal Entities:	Acme Ltd.	
Purpose of Processing:	New client acquisition based on qualified data received from a third party broker.	
Process Description:	Step 1: Qualified prospect data is received via email from Prospects4U in the form of an Excel spreadsheet; Step 2: Spreadsheet is formatted with appropriate field headers and imported into the Any CRM cloud-based application; Step 3: Data sent to US-based Campaigns-R-Us for additional email marketing campaign; Step 4: Clients contacted via email or phone by Acme employees, Any CRM is updated with the results of the contact.	
Categories of Individual:	Clients - Prospective	
Categories of Personal Data:	Core Personal Data, Contact Data, Role Data	
External Data Source(s):	**Name:**	**Category:**
	Prospects4U	Data Broker
External Data Recipient(s):	**Name:**	**Category:**
	Campaigns-R-Us	Marketing Campaign Service Provider
Third Country Transfers:	**Country/Region/Organisation:**	**Safeguards:**
	USA	[UNKNOWN]
Data Retention:	**Timeframe:**	**Justification:**
	Indefinite (see data inventory 'Comments')	All data will be retained indefinitely with consent, or deleted otherwise.
Security Measures:	Encryption in transit and storage, RBAC, pseudonymisation	
Automated Decision Making:	N/A	
Data Location(s):	UK, US	

[3] See *GDPR Compliance Step-by-Step: Part 3 - Process Mapping*

Step 4: Lawful Basis for Processing[4] [5]

While some scenarios would seem to be obvious; like doctors requiring personal data for vital interest, lawyers requiring personal data for legal reasons, or service providers requiring personal data to fulfil a contract, the devil is in the detail. Getting this wrong not only has a direct impact on your ability to demonstrate 'compliance', but you may also be implementing the wrong controls.

However, the lawful basis can have a significant impact on the technical and organisation measures you must put in place, so consideration must be given to things like:

- *Agree that it's the RIGHT decision* - Lawyers are only going to make decisions based on the facts / evidence provided in the Process Mapping step, they will likely have little insight into [or care about] the criticality of the business process in question. Or of the impact changes will have on the business. Legal decisions are almost always negotiable, make sure you are heard;
- *'Minimise' what's left (Data Categories)* - Data Minimisation is, by itself, one of the 7 Principles of GDPR, and the less data you have, the fewer things you have to do with it;
- *Consolidate what's left (Data Sources)* - Just because you need something, does not [necessarily] mean that you need several copies of it. You may only need ONE production copy of something (along with all requisite access and resilience obviously);
- *Shut down / amend the legacy data acceptance channels ("stop the bleeding")* - Now that you've worked out what you need to keep, stop the bad stuff coming in.

Getting to the lawful basis for processing is often confusing and difficult and involves the entire organisation in some way. Don't try to run a GDPR project in a silo.

With the above information and context, your data protection / privacy expert *should* be able to provide your legal basis for processing, note any caveats, and give you the next action items.

What you get back from them will look something like this for the sales process example above:

FOR [LEGAL TEAM]	
Lawful Basis for Processing:	Legitimate Interest (Article 6(1)(f))
Rights Available to Data Subject:	~~Article 7(3) - Right to Withdraw Consent~~ Article 15 - Right of Access by Data Subject Article 16 - Right to Rectification Article 17 - Right to Erasure Article 18 - Right to Restriction of Processing (absolute right when it pertains to marketing, no recourse) ~~Article 20 - Right to Data Portability~~ Article 21 - Right to Object
Additional Actions:	1. Perform due diligence on Prospects4U to ensure that data of any prospects was obtained fairly and lawfully; 2. Ensure Data Processing Agreements are in place with Any CRM, Prospects4U and Campaigns-R-Us; 3. Ensure Data Transfer Agreement is in place with Campaigns-R-Us, must include definition of Safeguards; 4. Ensure implementation of data subject includes onward transmission to third parties; 5. Perform a Legitimate Interest Assessment (LIA) to ensure an appropriate 'balance' is reached
Risks / Caveats:	Consent not required for B2B marketing unless (a) no opt-out provided or (b) opt-outs not actioned or (c) business is a sole trader or partnership.

[4] See *GDPR Compliance Step-by-Step: Part 4 - Lawful Basis For Processing*

[5] See *GDPR: Getting to the Lawful Basis for Processing* for instruction and samples

Step 5: Documentation[6]

Documentation is your evidence of compliance. It's as simple as that. Even if you're lucky enough not to have to maintain 'records of processing activities' (see Article 30(5)), you still must document everything else, including WHY you don't think you have to maintain records!

The word "*appropriate*" appears 115 times in the GDPR final text, and "*reasonable*" a further 23 times. It is therefore very clear that the determination of whether what you're doing meets the *intent* of the law, is just as important as meeting the *letter* of it.

Document everything, and have the results signed off at the highest levels of the organisation.

At a minimum you will require:

- *Policies* - Including those covering Data Protection / Privacy, Employee Privacy, Third Party / Third Country Transfers, Data Subject Rights, Information Security, Vendor Due Diligence and so on;
- *Personal Data Assets* - Steps 3 and 4 above should be documented in detail and all data stores recorded in an appropriate asset register;
- *Lawful basis for processing and corresponding data subject rights* - These must be clearly articulated and match the information contained within the Personal Data Assets;
- *Technical & Operational Security Measures* - If you haven't already documented your *entire* security program against a standard *like* ISO 27001 or NIST, you will need to;
- *Records of Processing Activities* - With a couple of caveats, if your organisation has fewer than 250 employees you do not need to record this, otherwise you will need to record everything your local supervisory authority requires of you[7]

Step 6: Operationalise[8]

If you don't build the necessary knowledge / processes into everyone's day jobs, your GDPR compliance program will falter. While data protection and privacy are everyone's responsibility, they cannot, and will not be at the forefront of everyone's mind as they work through an ordinary day.

You will also have pretty much thrown away everything you did on the first 5 steps.

These are the things that will need to be operationalised:

- *Governance - as it related to data security and data protection;*
- *Policies, Standards & Procedures - document management, review, ownership etc.;*
- *Employee On-Boarding / Awareness & Training - self-explanatory;*
- *Risk Management - there's no 'appropriate' without documentation risk;*
- *Asset Management - you can't manage what you don't know you have;*
- *Personal Data Life Cycles - the whole pint of this exercise;*
- *Vendor Due Diligence - they must all do what you're doing;*

[6] See *GDPR Compliance Step-by-Step: Part 5 - Documentation*
[7] The UK's ICO provides samples and guidance *here*
[8] See *GDPR Compliance Step-by-Step: Part 6 - Operationalise*

- *Incident Response / Breach Management - self-explanatory*

The whole point of a data discovery tool is to baseline your known-good processes to the point that any anomaly is detected as soon as possible. For example:

1. Why is a salesperson accessing financial data?
2. Why is personal data traversing the R&D network?
3. Why is unencrypted personal data being sent to a third party / third country?

There might be perfectly reasonable explanations for these things, but they should represent a *known* business process, not an anomaly against the *established* baselines.

ONE LAST THING BEFORE YOU START LOOKING

The specifics of the data discovery solution you go for are not determined by what you've seen available, and certainly not by a vendor, they are determined by the results of *your* risk assessment process.

No technology purchase should be considered before doing ALL of the following:

1. Conduct a Risk Assessment with Business Impact Analysis;
2. Perform a Gap Analysis comparing your risk to your *existing* mitigating controls;
3. Build a detailed list of the security *functions* required to fill the above gaps; and
4. Work out who is going to do *all* of the below:
    - Who is going to install it?;
    - Who is going to integrate it into the existing security operations?;
    - Who is going to maintain it (patching etc.)?;
    - Who is going to manage it (tuning etc.)
    - Who is going to monitor and respond to its output?; and
    - Who is going to measure its performance against the agree risk baseline(s)?

If none of the above can be handled by in-house personnel, then buying a fancy appliance is going to do you no good. You may have even *increased* your risk.

This is why you are not looking for *technology*, you are looking for a *solution*. Unless you can define what the solution should look like at a high level you will never ask the right questions.

The solution you end up with can be managed fully in-house, fully outsourced, or anywhere in between. Which is it?

You will also need to understand what the vendors mean by two phrases:

1. *Data Loss Prevention* - This was the traditional term for data discovery tools, but true DLP tools do not provide all of the functions we're looking for. They literally do what they say on the box, which is *preventing* certain types of data from being used inappropriately, accessed by unauthorised users, or leaving your network. Most of these tools pick one specific aspect of DLP (email, file-level data classification/encryption etc.), leaving you needing multiple solutions to obtain full coverage; and
2. *Business Intelligence* - It's now almost impossible to look for a data discovery solution that is not tying its deliverables to the phrase 'business intelligence' (BI). In fact, most see data discovery as a subset of BI. Generically BI is the process for "*analysing data and presenting actionable information to help executives, managers and other corporate end users make informed business decisions*" but no solution provider can tell you what BI is important to *your* business. So what *are* you getting?

Neither of the above points are a bad thing in and of themselves, it's just important to stay away from buzz-words or over-used terminology because they mean different things to different people. Stick to asking for what you need at the *functional* level to avoid confusion.

THE THINGS YOU NEED TO KNOW

The questions below are designed to be generic enough to suit almost any business, as well as potentially form some of the core questions of your Request for Information / Proposal (RFI / RFP);

1. *Do I need a permanent solution, or can a one-time, consultant-led, discovery exercise suffice for now?*

   It may well be that your business is small enough, and/or simple enough to only need a consultant led discovery exercise to help find, triage, and map your personal data assets. Once GDPR compliance is achieved, you'll likely need to run it again every year or so, or in times of significant change, but a permanent solution is just not warranted.

   If you do only require ad hoc scanning, you'll still need to ask all of the relevant questions below about the tool the *consultant* is using. Never assume they will have everything you need already covered.

2. *Can the solution perform discovery on the end-systems, databases, AND traffic 'on-the-wire'?*

   This is perhaps the most important question of all related to the functionality of a data discovery tool. And the one most overlooked. Some solutions ONLY search for data in databases, others only in static files, and others only 'on the wire' (data in motion). If the solution you choose cannot do all three, at least to some degree, then a lot of data could be missed, and a lot of the other functionality impossible (see below).

   A tool that only looks for data in specific formats is not necessarily a bad thing, you just have to know what you need, and what you're getting.

3. *Is the solution be installed locally on self-installed server(s), an appliance, or cloud-based?*

   Assuming you've decided on a permanent solution, what platform suits you best? Do you have the skill-set in-house to install, harden, and manage your own operating systems (stand-alone or virtual)? If no, then you are going to want a 'black box' / appliance which takes care of that for you (through a maintenance contract).

   What if the majority of your infrastructure is in the cloud? Does the solution work in Amazon, Microsoft *and* Google cloud offerings, or the one applicable to you?

   Some solutions will provide you a downloadable executable, others only come ready built, others provide both, and still others only provide solutions in the cloud. Which one is right, or do you actually need a hybrid solution instead?

4. *How will the solution manage encrypted files / databases / network protocols?*

   A data discovery solution that cannot decrypt databases or flat files prior to examination may not suffice, nor would solutions that cannot see into encrypted network protocols (TLS for example). The better solutions will integrate with whatever the decryption mechanism is for the relevant data format and provide a complete picture.

While the risk of *loss* of encrypted data is significantly less, you still have to worry about what and how you're *processing* it.

And what about email? Can the solution act as a data discovery 'gateway' to the most utilised communications tool in history?

5. *Does the solution cover both 'structured' and 'unstructured' data?*

   Structured Data *is comprised of clearly defined data types whose pattern makes them easily searchable (e.g. databases, spreadsheets);*

   Unstructured Data *is information that either does not have a pre-defined data model or is not organised in a pre-defined manner (e.g. Word docs, PDFs, emails etc.)*

   Some data discovery solutions make no attempt to examine unstructured data because of the significant complexity related to determining appropriate context, linguistic idiosyncrasies, slang, and numerous other factors. The false positive rates associated with the examination of unstructured data can often outweigh the benefits.

   It is therefore critical for any solution claiming to cover unstructured data be able to fully justify its position. Vendors throwing around phrases like 'artificial intelligence' (which does not even exist yet), 'machine learning', or 'neurolinguistics' had better to be able to back it up with hard facts.

   However, those that *can* have an enormous advantage over their more limited competitors.

6. *Can the solution learn / be taught organisation-specific parameters?*

   An extension of the previous question, some vendors will claim use of artificial intelligence, or machine learning techniques to get you to believe that their solution will actually get smarter and smarter all by itself as time goes by. As above, there is no such thing [yet] as artificial intelligence, and machine learning is often claimed when what's really happening is manual tuning.

   Regardless of the claims, what you *want* is a solution that can learn/be taught your organisation-specific idioms/idiosyncrasies so that the false positive rate drops to a manageable level. Your solution provider needs to explain, in detail, how they are going to accomplish this.

   If the solution also claims to map business processes (see below), then how these flows are recorded, baselined, and reported against will also need to be explained.

7. *Is the solution agent, agent-less, or both?*

   What is your greatest risk? Information held on a centralised database, or information processed on employee laptops / desktops? Agent-based data discovery has some significant advantages and disadvantages, they may also attract significant cost in both capital and resources. The results of your risk assessment should determine if 'end point' detection / protection is required.

8. *Can the solution perform network discovery e.g. (NMAP) and some semblance of asset management to guarantee ongoing coverage of systems?*

   Asset management is as key to data protection as it is to data security. Data is perhaps the most definitive asset in most businesses today. A data discovery tool that only looks for what's *currently* available as opposed to what *should* be available will again have limited value.

   The better data discovery tools will know exactly what physical asset (servers, desktops etc.), networks, and data stores (databases, file stores etc.) should be available for examination and anything missing should be flagged accordingly.

9. *Can the solution provide data-flow mapping?*

   Only available to data discovery tools that examine data 'on the wire', some solutions can map the flow of data from where it's stored, through which systems/networks it passes, even to the eventual external recipient (where packet header / URL information is available). This can to some extent automate the mapping the business process as well as form the baseline for all known-good processes.

   The best solutions can even apply the lawful bases for processing to these baselines and report against transgressions.

10. *Can the solution provide visualisation of data flow?*

    Direct extension of the question above, some solutions will actually attempt to map the flow of data into network diagrams. This can be very useful if you don't already have them, or if you want to begin questioning the efficiency / security of your current architecture.

    If you do already have comprehensive network diagrams, this can be a good sanity check.

11. *Can the solution be integrated with some form of compliance management tool?*

    A lot of organisations already have compliance solutions (Governance, Risk, and Compliance (GRC) for example), a data discovery tool can provide significant insight into what is happing within an environment. Therefore the output from data discovery tools should be easily 'parsible' into these systems, or better yet, already configured with APIs to the top providers.

12. *Is this solution available as self-managed, completely outsourced, or a hybrid?*

    Per the section above, before anything else happens, the questions of *who* is going to install manage, monitor etc. these solutions must be answered long before speaking to potential vendors.

    Some vendors give you an appliance and leave you to it, others will install it and tune it but no more, others offer a completely outsourced managed service. Some vendors can provide a hybrid solution where you can scale from completely internal to completely outsourced depending on your business needs at any given time.

Either way, any provider not offering a consulting wrapper to get their solution up and running effectively (even if it's outsourced) should be looked at with a little suspicion.

13. *Can the solution detect and alert against the top Digital Warning Signs?*

Digital Warning Signs (DWS) are powerful indicators that someone(s) internal to your network are either planning to do, or are already doing something that they shouldn't. They can include:

- Downloading or accessing substantial amounts of data;
- Accessing sensitive data not associated with their job function;
- Accessing data that's outside of their behavioural profile;
- Multiple requests for access to resources not associated with their job function;
- Using unauthorised storage devices;
- Network crawling and searches for sensitive data;
- Data hoarding, copying files from sensitive folders;
- Exfiltrating sensitive data outside of the organisation

No data discovery solution is going to be able to detect all of these things, nor can they completely baseline individual usage behaviours, but the more it *can* do the better.

14. *Can the solution support GDPR specific processes?*

This is where you must be especially careful. The GDPR is raising the emphasis on understanding your data like no other regulation before it. It is understandable and inevitable therefore that organisation are actually looking / asking for technologies that promise to handle GDPR specific processes.

Traditionally no data discovery tool is also going to provide an appropriate workflow for things like Data Subject Access Requests (DSARs), but that is exactly what some solution providers are now promising.

It is entirely possible that some of these offerings are sufficient for your needs, but until you have determined exactly what you need to accomplish to achieve full GDPR compliance, you should probably steer clear of this 'add-on' functionality.

Data discovery solutions are perfectly placed to *assist* in DSARs however, they can tell you exactly where all the data is!

15. *Does it only work with personal data?*

While we have focused almost exclusively on GDPR for the purposes of this white paper, the benefits of a data discovery solution can (and *should*!) go way beyond personal data. In fact, it's quite likely that personal data is not the most important data to your business.

From financial data (including payment data), to intellectual property, to R&D, buying a solution just for GDPR is not the best use of your budget.

Unfortunately, because of GDPR, data discovery vendors have 'retooled' their marketing to get their piece of the pie. I have no issues with vendors who built personal data into their product from the beginning, but many didn't, so what they have is basically smoke and mirrors.

As alluded to above, there is a great deal of personal data that is unstructured, the same goes for almost other types. It is critical to examine their search criteria/engine to ensure it covers all of your data assets.

16. *How long does it take to optimise?*

Something of a trick question, because the most you can ever hope for is continuous improvement. What constitutes personal data varies to such an extent that you will never be in a position where tuning of *some* sort is not required.

Any vendor who says you can just plug it in and it works is lying. Flat out lying.

Perfection is not 'reasonable', so don't strive for it unless you have unlimited time/money.

17. *How much does it cost?*

As a consultant, this is perhaps the most irritating question I am asked at the beginning of an engagement. Unless the client has *accurately* determined *precisely* what they need prior to contacting me, we are not ready to talk about money. Scoping comes first.

It really only matters that the deliverables meet your *defined* requirements, *including* the cost, so my job is to help you arrive at a mutually beneficial position or point you elsewhere.

The question should be; How much is the total cost of ownership for the specific deliverables I have requested?

Data discover vendors have many ways of charging:

1. By appliance;
2. By throughput;
3. By files examined;
4. By feature-set;
5. By month (for outsourced services)

So until you know exactly what you need, leave the money question until the very last.

First - narrow down the list of providers by comparing your required functions (determined at the risk assessment phase) to the vendor's list of product/service features;

Second - Run proofs of concept (or even a 'bake-off' if you have the resources) with the 3 most suitable vendors to see how each product / service works in reality;

Third – choose the vendor / functionality that gets you closest[9] to you desired goal, and THEN ask for pricing.

[9] Do not hold out for perfection, it does not exist.

SUMMARY

The technology behind data discovery and business process mapping has become increasingly more complex and accurate over the last few years, and the corresponding benefits to business of increasing value. With that has come the predictable increase in vendors trying to jump on the bandwagon. Most want to do a good job, some just want to sell stuff while the market continues to ripen.

However, few organisation require anything like what's on offer, certainly not until they have done a much better job of defining their business needs as a whole. The security, and data protection product industries are posterchildren for demand generation as opposed to fulfilling the real needs at hand.

Regulatory compliance too has been forcing organisations to get a better handle on their data, with GDPR being perhaps the most obvious and widely impacting. It follows therefore that vendors and service providers will use any acronym and buzz-phrase to draw attention to themselves.

Regardless of the technology in question, organisations must steer clear of the hype and focus only of the functionality they can define for *themselves* through robust risk management processes. It is entirely up to you to sift through to dross, or find the right help to do it for you.

It should be noted here that CCS is not an expert in data discovery technologies. At most, this white paper should be seen as the beginning of your journey towards choosing the right data discovery solution for your business.

As always, it is CCS's job to help you to ask the right people the right questions, the rest is up to you.

Best of luck!

ABOUT FROUD

David has almost 20 years of experience in areas of Information / Cybersecurity, including Regulatory Compliance, Secure Architecture Design, Governance Frameworks, Data Privacy & Protection, and FinTech.

As Project Lead for several Fortune / FTSE 'Enterprise Class' clients, David has performed hundreds of on-site security and compliance assessments for organisations globally.

Blog: http://www.davidfroud.com

Linkedin: http://www.linkedin.com/in/davidfroud

ABOUT CORE CONCEPT SECURITY

Core Concept Security (CCS) is an independent cybersecurity and data protection consulting practice based in the UK, but available globally.

The guiding principle behind all CCS's services is that security, while often difficult to achieve, has always been, and will always be, simple. There are no shortcuts to security, and there is nothing to be gained by just throwing money at it hoping the problems will go away. Technology will never fix what's broken, only people and process can.

The CCS approach is also simple; It's our job to help you ask the right questions, even if we aren't the ones who can actually answer them. You're hiring us, you're hiring everyone we know.

In the end, if your security program is not appropriate to your business needs it is a waste of your time and effort. Our commitment to our customers is to never settle for less than, or try to sell you anything *more* than, what you need.