Bloor

Spotlight

Spotlight Paper by Bloor
Author **Daniel Howard**
Publish date **January 2022**

The ethics of sensitive data and Solix Technologies

What data qualifies as sensitive and how do you protect it? Most of all, why precisely should you care, at least beyond the bare minimum required by law?

Introduction

Sensitive data – and, more specifically, personal data and PII – is a hot topic. In the wake of GDPR, CCPA, and other recent data privacy regulations, there is a growing need to treat the sensitive data within your organisation with as much care as possible, not just to comply with the aforementioned regulations but also to satisfy an increasingly privacy-conscious public. In essence, there are both moral and practical reasons to handle your sensitive data ethically.

But what does this mean in practice? What data qualifies as sensitive and how do you protect it? Most of all, why precisely should you care, at least beyond the bare minimum required by law? It is these questions, and more, that we seek to answer in this paper. We begin by discussing the last of these questions before moving onwards to address the former: it will be easier to describe what you need to do if we've already established why you're doing it. We will finish by highlighting one particular vendor – Solix Technologies – that aims to provide solutions for many, if not all, of the issues raised.

“In essence, there are both moral and practical reasons to handle your sensitive data ethically.”

The incentives

> **The most pressing reason most organisations have to protect their sensitive data is a bevy of updated – and significantly more stringent – compliance mandates that have been put into place...**

Whither ethics?

First of all, as with many questions of ethics, there is a moral imperative here: few would disagree that treating personal information with due care, and respecting the privacy of your customers and employees, is simply the right thing to do. That said, we would be naïve if we thought this alone would be sufficient to convince businesses in general to act ethically. Fortunately, there are also significant material incentives for you to behave ethically in regards to sensitive data, the most prominent of which are described below.

Regulatory compliance

The most pressing reason most organisations have to protect their sensitive data is a bevy of updated – and significantly more stringent – compliance mandates that have been put into place by a number of countries, states and assorted international bodies over the last few years. The most well-known of these is GDPR (the EU's General Data Protection Regulation), but it is far from the only one: CCPA (the California Consumer Privacy Act), LGPD (Brazil's equivalent of the GPDR), and others are already out in force, with more likely to come in the near future. Industry-specific regulations, such as HIPAA, have been around for far longer, but are still relevant within those industries.

Regardless of which mandates apply to you – and bear in mind that if you want to trade internationally then GDPR at the very least probably will – breaking them can result in extremely heavy fines. For example, towards the tail end of 2020, British Airways was fined £20 million (equivalent to $26 million) for a data breach that occurred in 2018.

In the immediate aftermath of GDPR, there was a tendency for companies outside the EU to attempt to sidestep the regulation, at least temporarily, by simply refusing to serve European clients. This is a poor long-term solution, for several reasons: it cuts off a significant number of potential customers; it leaves you high and dry if regulations come into force that you can't avoid in the same way; and it sends a poor message to any customers you maintain outside Europe by effectively communicating that you are not looking after their personal data. Simply put, this solution is not sustainable. Like it or not, eventually you will have to figure out how to comply with modern data privacy regulations.

It's also worth bearing in mind that these regulations (GDPR, at least) tend to operate on a *"good faith"* basis: if your relevant regulating body sees you making a genuine best effort attempt to comply, they will likely treat you with more leniency than if, say, you were nakedly attempting to comply as minimally and technically as possible, rather than treating your customers and their personal information ethically and with consideration. In other words, obeying the letter of the law but not the spirit is unlikely to be looked on favourably.

Public relations

Following on from the prominence of GDPR, particularly in regards to news coverage, as well as a number of highly reported data breaches over the past several years (Cambridge Analytica comes to mind), the public is increasingly aware of their data rights and protective of their personal data generally. This means that failing to protect your customers' data doesn't just have a direct financial cost (in the form of fines) but also a significant reputational cost: if your company ends up in the news for failing to protect personal customer data, it's going to hurt your perception in the public eye. "*All publicity is good publicity*" does not hold in this case.

Moreover, many compliance regulations, including GDPR, give your customers some control over their personal data, usually in terms of the right to know what personal data you're storing as well as the right to have it removed from your system (in GPDR-ese, these are the *"right of access"* and the *"right to erasure"*, respectively). When your

customers action these rights (usually via DSARs, Data Subject Access Requests), the normal rules of providing a good customer experience apply: communicate openly, act promptly, and so on. Providing a difficult, duplicitous, or generally poor customer experience in relation to sensitive data is as deleterious to your word of mouth as any other.

Likewise, creating a service or website that technically complies with the regulations but in practice either tries to trick your customers into giving you their consent, or forces them to go through a laborious process to revoke it, is doing you no favours. Best case, you've delivered a horrible customer experience and retained a customer in spite of it; worst case, you've lost your credibility, and as a result, many customers.

When, not if

Data breaches are simply becoming a fact of life. If you are a large organisation, you will almost certainly suffer one at some point: it is a matter of when, not if. This means that you need to assume your system will be compromised in some way at some point and structure it accordingly, ideally so that your sensitive data will remain secure even if (when) after a breach, at least if it occurs at the most visible points of attack. Of course, you should still minimise your vulnerability to, and thus frequency of, data breaches, and if you're very lucky you won't have to suffer through one. But you shouldn't count on it, and if you do, it's at your own peril.

Finally, when a breach occurs, you're going to come out looking much better if you communicate with your users as quickly and with as much detail as possible, so that they can better take steps to protect themselves. You may also have a legal requirement to do exactly that. Honest and open communication can go a long way, and having processes in place to ensure that communication happens promptly during a crisis (which a breach certainly qualifies as) is important if you want to ensure it isn't overlooked or unduly delayed.

General applicability

So far, we have talked about sensitive data as if it was synonymous with personal data, whether that data belongs to your customers or your employees. However, this is not strictly the case. Sensitive data can also refer to data that is sensitive to your organisation itself: confidential documents, release plans, valuable IP, and so on. Moreover, the systems that are used to help ensure the security of your users' personal data can also be applied to this data. In other words, the practices you put in place to protect the rights of individuals can also protect your organisation. This is unlikely to (and arguably should not) be the driving force behind your treatment of sensitive data, but if nothing else it is a useful secondary benefit that can provide additional value.

> **"Sensitive data can also refer to data that is sensitive to your organisation itself..."**

The issues

> Sensitive data discovery can be exceedingly tricky over an entire enterprise worth of data, since sensitive data often exists pervasively but with poor visibility.

Discovery

The first step towards protecting the sensitive data you are storing is knowing what sensitive data you have and where you can find it. This sounds obvious, but the fact of the matter is that sensitive data inventory and discovery can be exceedingly tricky over an entire enterprise worth of data, since sensitive data often exists pervasively but with poor visibility. Discovery is impractical to do manually, and therefore you will want some way to automate this process. It's not uncommon, for instance, for discovery processes – once automated – to reveal sensitive data stored on data sources that you wouldn't even have thought to check. Note that manual intervention in this process is likely: since false negatives (read: sensitive data you've failed to discover, and is therefore left unprotected) are both legally and ethically unacceptable, false positives are necessarily common, and the usual solution is to have a human verify any discovery results (or at least the less likely classifications) before finalising them.

Compounding the issue is that defining what counts as sensitive data is much more squirrelly than you might expect, particularly since it will depend somewhat on the regulations that apply to you and the industry you're operating in. Most organisations are fairly comfortable with directly sensitive data: in other words, data that could be used to directly identify an individual. Its lesser-known sibling is indirectly sensitive data: data that can be used to re-identify anonymised data. To demonstrate the issue, researchers at the University of Texas were able to re-identify anonymised data released by Netflix with a 68% success rate by comparing it to movie reviews from a third-party website. To make matters worse, this sort of data is generally more difficult to find, especially using manual methods.

Adding further to this complexity, apparently simple methods of discovery are unreliable at the best of times. For instance, a standard and relatively primitive technique for discovering sensitive data is to look for relevant column names ("first name", "last name", "date of birth" and so on). But columns are often misnamed or poorly labelled. And what if human error means that sensitive data gets added to the wrong column? Clearly, more complex methods are needed. This will often include data matching, pattern matching, code matching, and more. What makes this even trickier is that no single discovery method is fool proof. This means that the only way to maximise the chance that you've eliminated all false negatives is to apply several of these methods at once. Of course, this can have a performance cost, and can lead to a large number of false positives. This means that keeping the rate of false positives down – crucially, without introducing any false negatives – is a priority. Discovery over different types of data is another consideration. Discovery on structured data is relatively well-supported; discovery on unstructured data, such as data held within documents, is less so. These are all elements you'll need to consider when building or licensing a discovery solution.

Protection

Once you've determined what it means for your data to be sensitive, and located it within your environment, you will need to protect it via anonymisation. You have several options for this, including static data masking (replace it with generated, non-sensitive data that preserves part or all of the structure and/or format of the original), dynamic data masking (the same, but in-flight rather than at rest), encryption or tokenisation (cryptographic security methods), and obfuscation (simply blanking out part or all of the data, commonly used for anonymising documents).

These methods all offer distinct advantages and drawbacks, and the most appropriate method(s) will depend on your situation and use case. For example, when and where do you need your data to be protected (at rest or in-flight)? Do you care about the data itself, or merely its shape (could you use masking to effectively get rid of sensitive information without impacting its intended use)?

Which properties of the data do you need to preserve under anonymisation (do you need to maintain referential integrity, or preserve the format of the original data, even while encrypted)? Should the anonymisation be reversible, or is it important that it isn't?

As a general rule, best practice would be to a) make sure you comply with all applicable regulation, b) ensure all sensitive data is protected as securely as possible, especially whenever it's moved, and c) allow as few people as possible to access (unprotected) sensitive data. You will most likely want to leverage a combination of the methods mentioned above to achieve all this in a way that is both comprehensive and performant.

Governance and consent management

It's not enough to protect all of the sensitive data that currently exists in your system: you need to protect all data that comes into it in the future as well. Data governance and automated policy management/enforcement can effectively be used as methods to both discover and protect sensitive as it enters your system on a continuous basis, as well as to continuously apply and enforce data protection policies to ensure that your customers' data remains safe and that you remain compliant.

You will also need to manage the lifecycle of sensitive data once it enters your system, most notably in the sense of managing data retention and removal. At least according to GDPR, you can only store personal data you've acquired its owner's consent to store, while you have a defined reason for storing it, and only for a limited time. This means that any personal data that enters your system will eventually need to leave it, either because you finish processing it, its natural lifespan elapses, or its owner explicitly withdraws their consent. Therefore, you will need processes in place to manage user consent, understand the lifecycle of your sensitive data, and keep track of data retention (including both why you're storing a given record and how long you've been storing it for) and removal of the data from your system once it expires (both naturally and on an ad hoc basis). Again, data governance platforms tend to be well-suited for this.

Additionally, you need to provide users with the tools to interact with their personal data, usually via DSARs, and make these tools easily and readily accessible. This means that you will need to be able to provide full visibility into a given user's data, as well as rectification and even deletion of that data, on an essentially at-will basis. This can pose some challenges, but for the third time data governance products usually provide support for this. Note that although legally you will usually have a month or so to acquiesce to these requests, it would behove you to at least respond promptly. Moreover, for the request for deletion in particular, you may want to opt to archive that data for a short time rather than erasing it immediately: it is not unheard of for customers to change their minds shortly after such a request.

Security and monitoring

Data breaches may be a fact of life, but you will still want capabilities in place to minimise the frequency and impact of their occurrence. The sooner you can detect and address them, the better. Accordingly, you will want to be able to monitor the sensitive data within your organisation so that you can see how your sensitive data is being accessed, by whom, and when. In particular, you will want a security monitoring capability that can detect and report anomalous – and what could be malicious – behaviour. This can provide you with advance warning of a data leak, possibly allowing you to prevent it, or at least respond to it quickly and appropriately.

You will also want to be able to control who has access to your sensitive data (whether anonymised or not) and will therefore need authentication and authorisation capabilities in place. This can (once again) be provided by data governance, in conjunction with tools such as Active Directory and LDAP.

“It's not enough to protect all of the sensitive data that currently exists in your system: you need to protect all data that comes into it in the future as well.”

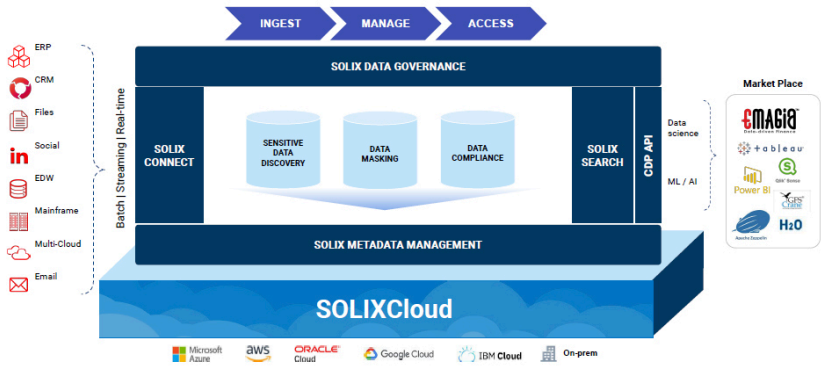
SOLIXCloud – Consumer Data Privacy Solution

What is it?

Solix Technologies speaks to the issues present within sensitive data via its product, SOLIXCloud, a multi-cloud enterprise data management platform. In particular, a subset of SOLIXCloud's capabilities, described as the SOLIXCloud Consumer Data Privacy Solution (or CDPS), has been designed specifically for this purpose.

CDPS contains the platform's solutions for sensitive data discovery, data masking, and data compliance and policy management. These are wrapped in a layer of data governance and metadata management, and bookended by extensive search and integration capabilities provided by Solix Search and Solix Connect, the latter of which provides connectors for hundreds of different data sources. Outgoing APIs are also available. The full architecture is shown in ***Figure 1***.

Figure 1 – SOLIXCloud Consumer Data Privacy Solution architecture



SOLIXCloud (and, as a result, CDPS) can be deployed in-cloud, on-prem or as part of a hybrid solution, across all major cloud providers. It operates across both structured and unstructured data, and at big data scale. This includes compliance, policy enforcement, and so on, and is extremely useful for ensuring that your good practices regarding sensitive data are all-encompassing, even on extremely unstructured types of data such as images or scanned documents.

Why should you care?

CDPS offers comprehensive capabilities for finding, protecting and governing your sensitive data. Its data discovery functionality, for instance, can scan, and subsequently 'map out', your enterprise data landscape, including both production and non-production environments and examining both metadata (such as column names) gathered during data ingestion and actual data (by looking for patterns in the data values). To reduce false positives, it uses sampling to help users decide what is and is not sensitive. It also comes pre-populated with discovery rules suitable for identifying PII, PCI, PHI, and other such sensitive data elements, and you can also create your own.

Its masking functionality is similarly substantial. It provides static data masking, both random as well as format-preserving encryption and tokenisation, and it can mask consistently, within single or multiple databases, and without losing referential integrity or moving your data outside of its original data source. In addition, dynamic masking is available for SolixCloud managed data. SolixCloud offers a variety of prefab masking rules and algorithms, some of which are generic (for example, a rule that produces a random sequence of letters) while others are specific to particular types of data, for instance date of birth or social security number. Custom algorithms are supported and can be written in Java, T-SQL or PL/SQL, as are masking algorithms for both structured and unstructured data (documents, in particular), the latter primarily via redaction. It also offers pre-packaged masking capabilities for Oracle and PeopleSoft application environments, provides masking support across most leading relational databases

as well as MongoDB and flat files, and has implemented its masking algorithms natively in Oracle PL/SQL.

The platform's further governance and compliance features include a data catalogue, a business glossary, data lineage, lifecycle management, and more. This allows you to classify and monitor your sensitive data as well as providing access to it via a secure environment equipped with role-based access. Notable features include policy management, which allows you to define enterprise-wide policies for automatically discovering and protecting your sensitive data, consent management, and a leading data archiving and retention management capability. The latter in particular gives you fine-grained, and automated, control over the lifecycle of your sensitive data, meaning that – for example – it will be assigned a lifespan and automatically archived when that lifespan is over (which is a requirement for many compliance regulations, most notably GDPR). At the same time, DSARs (Data Subject Access Requests) are supported, in terms of both producing all information pertaining to an individual and purging it from your systems on an essentially ad hoc basis. Solix has also implemented the concept of a *"legal hold"*, which allows you to freeze information that is currently under request to prevent it from being edited or deleted. Finally, Solix also provides a dedicated GDPR tool, which will scan your enterprise and produce a report on your sensitive data and its compliance with GDPR in specific.

"Solix Technologies speaks to the issues present within sensitive data via its product, SOLIXCloud, a multi-cloud enterprise data management platform. In particular, a subset of SOLIXCloud's capabilities, described as the SOLIXCloud Consumer Data Privacy Solution (or CDPS), has been designed specifically for this purpose."

Conclusion

It should not take much convincing for most companies to see the value in protecting their sensitive data: regulation in general and the GDPR in particular has seen to that. But we would caution that strict regulatory compliance is not the be all and end all when it comes to treating sensitive data ethically. People in general want their rights to be respected in earnest, are often savvy enough to know when that isn't the case, and are unafraid to react publicly. Accordingly, we implore you to treat any sensitive data in your care ethically – not just in a way that is technically considered compliant – and respect its owners right to privacy, not just for their sake, but for yours as well.

To that end, Solix is a very respectable solution for looking after the personal information that is in your care. It provides all of the core functionality you could want in a unified and integrated package, and its data retention, DSAR support, and lifecycle management capabilities, in particular, are highly developed. We would even call them market leading. In short, Solix and SOLIXCloud are a safe pair of hands for your sensitive data.

People in general want their rights to be respected in earnest, are often savvy enough to know when that isn't the case, and are unafraid to react publicly.

About the author

DANIEL HOWARD
Senior Analyst,
Information Management and DevOps

Daniel began his career in the IT industry relatively recently, in only 2014. Following the completion of his Masters in Mathematics at the University of Bath, he started working as a developer and tester at IPL (now part of Civica Group). His work there included all manner of software development and testing, usually in an Agile environment and usually to a high standard. In the summer of 2016, Daniel left IPL to work for Bloor Research as an analyst, and the rest is history.

Daniel works primarily in the data space, though he dabbles in development, testing, and DevOps. The former often (though far from always) involves working alongside his father, Philip Howard, while the latter allows him to leverage the technical expertise, insight and 'on-the-ground' perspective garnered through his old life as a developer to good effect.

Outside of work, Daniel enjoys latin and ballroom dancing, board games, skiing, cooking, and playing the guitar.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of ***Mutable*** business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

tel: **+44 (0)1494 291 992**
web: **www.Bloorresearch.com**
email: **info@Bloorresearch.com**