In alliance with
Linklaters

A step-by-step guide to assessing compromised data after a cyber incident

OCTOBER 2025

The most overlooked part of cyber incident response planning

Once a cyber incident has been contained, the most crucial part of any cyber incident response is a company's ability to rapidly understand the data that has been compromised and the impact on the business, individuals and others.

This phase is also typically the most *complex, costly, time consuming and stressful*. Delays or errors in data assessments and subsequent notifications can quickly attract the ire of affected customers, regulators, industry bodies and the media. They can also have significant regulatory implications.

And yet, data assessment and notification is also the most overlooked part of cyber incident planning. A company's approach to this task—including the methodology it will apply and the tools and specialists it may need to leverage—is hardly ever contemplated, let alone detailed in any useful way, in cyber incident response plans or playbooks.

This guide is designed to help you to start addressing that gap, and includes 13 key steps that cover what to do before, and after, an incident.

> *A 2025 global report found the average cost of assessing and responding to a data breach, and notifying those affected, was US$4.44 million.*[1]

Data assessments at a glance

By most estimates, unstructured data accounts for about 80-90% of the known data universe.[2]

The challenge

There is no one-size-fits-all approach to data assessments

Compromised data is rarely uniform, making it difficult to standardise the approach to data extraction and review.

Your methodology and tools will need to be adapted to the *purpose/s* of the assessment, the *volume* of data to be reviewed, and the *sources*, *format* and *structure* of that data.

There will always be a trade-off between time, accuracy and cost

Following a cyber incident, you will likely be under enormous pressure—from internal and external stakeholders—to rapidly and accurately assess compromised data and provide appropriate notifications to regulators, affected individuals and others about the nature of that data.

But expectations rarely reflect the *complexity* of this exercise, the *time* it takes to process and then assess the data, and the *trade-offs* that need to be made in order to speed up the review. It can often take weeks, if not months, to make an assessment and the pressure to accelerate the process frequently results in increased risks and errors in it.

Unstructured datasets

Assessing unstructured datasets (eg the contents of compromised email accounts, documents in document management platforms or network drives) is particularly challenging. A careful mix of algorithmic or AI tools, human oversight and review is normally required, given that these datasets typically include a disparate range of relevant data points, some of which are contextual.

What's more, reviewing this data to identify impacted individuals is more fraught than a typical document review or discovery exercise, because affected individuals need to be accurately identified, and then information about a specific individual—which often sits in multiple files or documents—needs to be accurately correlated to that individual. There is also typically a higher risk that false positives or negatives will be generated, which can lead to mistakes in regulatory and contractual notifications.

Key takeaways

Set expectations. Key stakeholders, including the broader incident response team, should all understand the data assessment process and associated timeframes. This can help these teams adapt their own plans and template communications. It is also important to set board and senior management expectations regarding the timing, cost and possible regulatory implications associated with this process.

Develop a data assessment and notification methodology. Although any template methodology will need to be adapted to the particular circumstances of the incident and the nature of the compromised dataset, having (in advance) an understanding of the data sources and record types that exist within the organisation and a review framework, will significantly reduce the risk of error and the duration of any review.

Compromised data assessments and notifications should be overseen by Legal. Data assessments are generally undertaken so that the business can be properly advised as to *whether* there needs to be a notification, *who* needs to be notified of a cyber incident (eg in accordance with privacy laws or applicable contractual obligations), *when* that notification should be issued and *what* it should contain. Legal should play a central role in coordinating the assessment and defining the necessary output, as well as in ensuring legal professional privilege is maintained where appropriate.

Know your tools and experts. Where possible, assess and pre-approve experts and tools, and refine your template methodology with them in advance of any incident. Your experts should be familiar with the tools that are available to assist in data assessments. Vetting and pre-approving experts will allow you to engage them quickly following a cyber incident. In addition, work with your IT team or experts to understand and enable built-in functionality within the systems you already use (eg Microsoft 365 or Google Workspace) to help with automatic detection and labelling of certain documents or personal information in the event of an incident.

Manage your digital footprint and maintain a robust data retention and deletion program. There's nothing quite like spending weeks or months (and incurring significant costs) assessing large volumes of data to remind you that *the best way to avoid or reduce the impact of a data breach is not to have the data in the first place*. Robust data governance practices and workflows will help minimise your digital footprint and your legal exposure.

Your 13-step guide

Before **an incident**...

STEP 1
Know your notification requirements

STEP 2
Enable functionality within your existing systems to detect, identify and tag personal information and data

STEP 3
Identify (and, if possible, engage) potential data assessment service providers and pre-approve their tools

STEP 4
Prepare a template data assessment methodology

In the wake of an incident...

STEP 5
Identify the compromised dataset

STEP 6
Tailor your methodology

STEP 7
Data processing and planning

STEP 8
Initial cull and preliminary assessment

STEP 9
Detailed review

STEP 10
Harm prioritisation

STEP 11
List generation

STEP 12
Notifcation

STEP 13
Record keeping and data management

Your 13-step guide

STEP 1

Know your notification requirements

ASK: *What are our regulatory and contractual notification requirements? Which of our datasets, if compromised, are likely to trigger those requirements?*

If you experience a data breach, you'll need to assess whether the compromise of that data might trigger any regulatory* or contractual notification requirements—each of which generally imposes different notification thresholds, and content and timing requirements.

Understanding these requirements *in advance* of any incident will improve your ability following an incident to quickly undertake your assessment and any required notifications.

**Depending on the circumstances of the incident and the nature of the company affected, notification could be required under the Privacy Act, SOCI Act, APRA's prudential standard CPS 234 (Information Security), CPS 230 (Operational Risk Management), the Corporations Act, Cyber Security Act and/or other sector-specific legislation (eg the Digital Health Agency in relation to potential or actual breaches affecting the My Health Record system).*

STEP 2

Enable functionality within your existing systems to detect, identify and tag personal information and data

ASK: *Are there any built-in tools within the systems we already use that could assist? Are these enabled?*

A number of commonly used tools, such as Microsoft 356 and Google Workspace, have data governance (or similar) modules that can assist with the automatic detection and classification of documents containing personal information (eg the ability to detect and label outgoing emails that contain driver's licence or passport numbers). Although unlikely to be sufficient on its own, leveraging this functionality may allow you to more quickly form a preliminary view as to what has been exposed, which may assist in the prioritisation of data for review.

STEP 3

Identify (and, if possible, engage) potential data assessment service providers and pre-approve their tools

ASK: *In what circumstances might we require additional tools or external expert assistance? Which organisations provide these and where are they located? How can we get assurance that they can do what they say they can, to the level we expect? Do they need to be pre-approved by our insurer and/or our internal IT security team?*

There may be circumstances in which the volume and/or complexity of data to be reviewed means your inhouse legal team does not have the availability or capability to review the compromised data in the time required (especially when they are already tied up on the broader incident response effort). What's more, traditional eDiscovery tools aren't fit for purpose for compromised data assessments—in order to conduct compromised data assessments efficiently and effectively at scale, organisations will generally need to leverage tools and workflows that are purpose-built to identify and link personal information to impacted individuals.

Data review providers (which include certain law firms and other data assessment vendors) can deploy resources, workflows and the right technologies to assist with the review, to inform the privacy and sensitivity review assessments and notification strategies developed by Legal.

Given the access these providers will have to compromised datasets, they should ideally be vetted in advance and pre-approved by your Cyber teams and (where required) your insurer.

They should also be engaged by internal or external legal (where the provider is not a law firm), for the purpose of facilitating advice to the business about the notification requirements applicable to it, so as not to waive privilege. Ideally, they should also be engaged *in advance* of any incident, to avoid having to negotiate arrangements in the wake of an incident, when negotiation leverage and time are limited.

Your 13-step guide

STEP 4

Prepare a template data assessment methodology

ASK: *What steps might we need to take, and what key decisions might we need to make, in reviewing any compromised data and making notifications?*

For the most part, notification regimes (whether regulatory or contractual) require that companies make notifications in respect of specific individuals, identifiers, data attributes or datasets. This means you will need to accurately identify affected individuals, their personal information that has been impacted, as well as other relevant attributes in the compromised dataset (including protected information or commercially sensitive information). You will also need to collate this information in a form that enables you to notify the relevant affected individuals or other third parties.

This is rarely straightforward—the complexity, duration and cost of this exercise are compounded where the compromised dataset includes *large volumes of unstructured or semi-structured data* (eg emails), or *difficult-to-read file types* (eg image-heavy files like pictures or scanned copies of identity documents).

Although some tradeoffs will almost always need to be made, regulators and contractual counterparties will generally still require that a comprehensive assessment is undertaken, irrespective of the complexity or cost.

Preparing a template data assessment methodology in advance of any incident will help you:

- *comply* with any data assessment and notification requirements
- *more readily explain* how you have approached data assessment and notification, should you need to provide an overview to regulators or contractual counterparties
- *educate internal stakeholders* (including senior management) about what is involved in undertaking these assessments and, in doing so, help to set expectations—in advance—around timeframes, costs and limitations. This can help *alleviate certain internal pressures* in an actual incident.

Note: Your template methodology will need to be adapted in the event of an actual data breach but given (1) the number of steps and decisions that need to be taken to sensibly assess the data and (2) that most of these can be outlined in advance, having a well-developed template will significantly reduce the margin of error and time required to consider these matters in the heat of the crisis.

Your data assessment methodology...

...should outline the steps to be taken to:

1. confirm the dataset in-scope for review and assessment
2. determine the parameters for review (and key markers to be identified—including customer identifiers that may enable easier identification of impacted individuals)
3. determine the nature and size of sampling to be undertaken to tailor the rule set / coding for detailed data analysis and to assist in culling (where appropriate) data that does not need to be reviewed
4. identify higher-risk personal information or other sensitive data and prioritise review of documents likely to contain this type of information
5. undertake the detailed data analysis (using both automated and manual review methods, as appropriate)
6. identify which individuals or regulators need to be notified to meet regulatory requirements
7. identify which individuals or other stakeholders require notification under contract
8. confirm contact information for affected individuals or other third parties
9. issue notifications
10. determine whether any follow-up notifications or other actions are required.

It should also contain a decision log (pre-populated as much as possible) to help you record the key decisions and judgement calls made in undertaking the review.

Tips!

- Your methodology should identify which attributes should simply be flagged (eg with a 'Yes' / 'No' to indicate their presence in the dataset) versus which should be extracted (eg contact details or, if credit card information is compromised, the last four digits and type of card).
- Remember—it is best to avoid creating new repositories of personal information unless there is a specific need to do so (eg to enable you to contact someone or to give them enough information to take steps to mitigate the risk, such as cancelling impacted cards), so the default option should be simply to flag it.

Your 13-step guide

STEP 5

Identify the compromised dataset

ASK: *Which data do we know, or should we assume, has been compromised?*

Your identification of the compromised dataset will likely be informed by a number of factors, including:

- the results of any forensic investigation—depending on the findings, you may also need to make certain assumptions regarding which data has been compromised; and
- whether any threat actor or other third party has published any data or otherwise provided it to you as proof that it has been compromised.

Tip! Identify any particularly sensitive data (eg suspicious matter reports) that may need to be carved out of the review and assessed separately (and not by certain third parties), as a consequence of more onerous statutory or other access restrictions.

STEP 6

Tailor methodology

ASK: *What do we know about this incident that may impact our methodology or the key decisions to be made?*

Your understanding of the compromised dataset and the circumstances of the incident should inform your approach to the data assessment. For example:

- **Timing:** depending on the regulatory frameworks applicable to your business, the specific nature of the incident and when it was first discovered, you may need to work to very compressed timeframes in order to meet regulatory notification deadlines. This will impact prioritisation and sequencing. For example, you may decide to prioritise an urgent '*pulse check*' of the dataset for the purpose of describing it to a regulator in an initial notification, with a more detailed review to follow. You may also need to identify and prioritise the notification of *highly vulnerable individuals*, particularly where their physical safety may be at risk , or prioritise review of datasets that you know have been published.
- **Nature of the dataset:** the tools and resources required to assess the dataset will depend on the volume of data and whether the data is structured or unstructured (or a combination of both).

Your methodology and decision log should continue to be updated to account for new information as the remaining steps progress. However, it is important to remember that additional effort and delay can arise from changes to agreed parts of the methodology, reversals of key decisions or the late-stage introduction of new review parameters.

STEP 7

Data processing and planning

ASK: *Do we have the internal capability to ingest and analyse the data? If not, have we vetted a service provider to assist with this process? Has the system(s) that provider proposes to use been approved by our IT security team? Does the system have any AI capabilities and AI-driven workflows to help identify personal information? What types of personal information does the system identify out of the box (eg credit card or passport numbers) and what might require manual review? How long will it take before the data and documents are available for analysis and review?*

Although the process of data ingestion and processing can be time consuming, your expert or data assessment service provider should be able to map out a timeline outlining when data will be available for assessment.

Your 13-step guide

STEP 8

Initial cull and preliminary assessment

ASK: *What steps can we take to reduce the dataset for review? What can we learn about the data from an initial scan and (where relevant) speaking to individuals who have created or who use the relevant files? How might this assist in helping us shape our approach to review or inform the techniques or tools we deploy (eg keywords, domains, AI models)? Are there any file types that will need to be reviewed differently?*

Interviewing the individuals closest to the affected files (eg the mailbox holder of an affected email account), and assessing the results of an initial scan, prior to commencing a detailed review of a dataset, can greatly improve efficiency and help to refine the tools and technical strategies to be used to assess the data. These steps can help to reveal things like:

- repositories containing higher-risk information for priority review
- repositories that do not contain personal information and can be excluded
- repositories containing high volumes of unstructured personal information
- files containing foreign language content
- files that require further processing to render them text searchable and, therefore, capable of detailed review (ie they require optical character recognition), eg audio / video files or scanned copies of identity documents
- files that are encrypted: depending on the level of encryption and the circumstances of the incident, you might determine that these files can be excluded from further review or decrypted at the processing stage if the passwords can be provided by individuals
- where differential processes need to be established based on file type: eg with the right tools, file types containing structured data (like .xls or .csv files) can usually be assessed via primarily algorithmic means. These types of files can be separated from those that will require a more nuanced combination of human and algorithmic review (eg PDFs, email text)
- whether it may be possible to focus on (or exclude) particular subsets of the data: eg if a mailbox holder is able to confirm (to your satisfaction) that certain inbox subfolders only contain benign data.

It is also often possible to significantly reduce the dataset for review by employing a range of preliminary culling strategies. These include identifying and removing duplicate documents, and correspondence from particular domains (eg @ newspapername), as well as documents that are clearly irrelevant to any data assessment (eg benign spam and marketing emails, or technology or administrative emails, such as reminders to install updates on company devices).

Tip! This preliminary assessment will also assist you to tailor your template methodology and confirm which attributes should be searched for as part of your review. It is important to plan early and settle the approach before the detailed review of documents commences. Revisiting documents already reviewed due to a new criterion being added to the assessments can push out timeframes and lead to increased costs.

STEP 9

Detailed review

ASK: *What level of review do we need to undertake of the refined dataset for us to understand who needs to be notified of the incident? What should the output of the review contain so that we can make the necessary notifications?*

Depending on the nature of the compromised dataset, this step is likely to involve a combination of:

- the application of *technical strategies and automated review*—for key identifiers that have been documented in the tailored methodology; and
- *manual review*—to validate the results of the automated review and to identify additional categories of personal information or other types of notifiable data that are not conducive to automated review (eg personal information in the nature of opinions about affected individuals or other important 'contextual' information).

Step 9 continued

More sophisticated systems now leverage AI to automatically detect personal information within documents that traditional searching may miss when used in isolation. In some cases, pre-built AI models can be supplemented by bespoke models that can be developed to target the specific types of personal information likely to be found within the particular compromised dataset (eg patient numbers and medical records). The precision of these types of AI models has improved considerably over time and can significantly reduce both the costs and time required for notification. However, in almost all cases, a manual review process is still required to validate and increase the accuracy of the results.

Tips!

- Because the purpose of the data assessment is generally to identify those individuals or other entities that should be notified of a cyber incident, results of the detailed review are often best presented on a per individual/entity basis (eg one line item for each affected individual/entity that flags the presence of each affected category of information, even if the relevant data points were originally spread across multiple documents). Just make sure that any 'merging' of similar or identical individuals or entities is undertaken carefully and with regard to more than one matching identifier—there may be more than two Charlie Smiths!
- You will need to determine whether contact information should be extracted from the dataset and included in the review output (where possible) or whether the business is likely to hold current contact information separately (eg in its customer relationship management (***CRM***)), or whether third-party contact information (eg a corporate customer or other partner associated with an individual) should be extracted instead.

STEP 10

Harm prioritisation

ASK: *Now that we know what information has been compromised and which individuals/entities have been impacted, specifically who must be notified of the incident (eg under privacy laws or contractual obligations)? Would we be better served by adopting a more inclusive approach to notification (eg issuing notifications even where not strictly required)?*

Once you know what categories of information are included in the dataset at a per individual/entity level, you can filter your list by data points to arrive at a refined subset of individuals/entities that will need to be notified of the incident.

How the appropriate data points are selected will depend on the work you undertook in Step 1 (to identify your notification requirements). For example:

- *if you are regulated by the Privacy Act*—you'll need to notify individuals for whom the exposure of that particular personal information is likely to result in 'serious harm'. This will require an assessment based on the circumstances of the incident, particular combinations of compromised personal information and what you know about the affected individuals; and
- *if you have contractual obligations to notify particular entities (eg customers) of cyber incidents affecting certain types of information*—you will need to assess whether the compromised information falls within the categories prescribed in the relevant contract.

Tip! You may decide to notify certain individuals or entities even where there is not necessarily a clear legal or contractual requirement to do so. Sometimes there can be good reasons for taking a more inclusive approach. These might include where:

- issuing a broader set of notifications would be faster and more cost effective than undertaking complex threshold assessments; or
- you expect to face a high volume of enquiries from those who were not notified of the incident but who have come to learn of it through other sources (eg colleagues or other customers) and now want to know whether they have been affected and why they were not informed directly.

Your 13-step guide

STEP 11

List generation

ASK: *What data techniques need to be used to effectively deduplicate the list without removing individuals with the same (or similar) names? What additional data fields need to be included in the list to assist with notification (eg customer numbers)?*

Once you have collated the data, it can still take some time to generate the list of individuals to be notified. A variety of techniques may need to be applied (eg fuzzy matching) to deduplicate the list without removing individuals with the same (or similar) names.

It will also be important to consider:

- which attributes should simply be flagged (eg with a 'Yes' / 'No' to indicate their presence in the dataset) vs extracted (eg contact details or, if credit card information is compromised, the last four digits and type of card); and
- what additional data fields may need to be included (eg a customer number) to assist with orderly notification.

STEP 12

Notification

ASK: *What channel will we use to notify individuals/entities? Do we have current contact information? If not, are there any third parties who have a closer relationship with the affected individuals/entities and whose involvement we could request? What are the privacy implications of that involvement?*

Email is a common means of communicating a data breach notification, although other channels may also be appropriate in certain circumstances (eg a phone call where the risk to a known individual or corporate customer has been assessed as high and imminent).

Often, however, current contact information is simply not available (whether in the affected dataset or in the business's CRM). This might be because your organisation does not have a direct relationship with the affected individual, or potentially because the information in question is very old and any associated contact information is now out of date.

In these cases, you may need to identify other avenues for notification. For example, does the affected individual have a direct (and relevant) relationship with one of your corporate customers or other partners? If so, and assuming you have extracted the details of this customer/partner during the detailed review stage, you may be able to work with that entity to have a notification issued on your behalf.

Tip! If notifications required by the Privacy Act cannot be issued directly to the relevant individuals, you may need to publish a general notification on your organisation's website and take other steps to publicise that statement (eg via physical notices in stores or publication via social media channels).

STEP 13

Record keeping and data management

ASK: *How long will we retain the compromised dataset with our service provider? Do we have a streamlined process to ensure access requests are handled appropriately? How easy is it to re-review or extract impacted data in order to respond to access requests? Have we extracted and saved tagged compromised data and notification lists? Are these appropriately secured?*

It is increasingly common for impacted individuals, other stakeholders (eg corporate customers) and regulators to seek further information following notification of a cyber incident. In some cases, impacted stakeholders may request access to, or deletion of, their information. Be prepared to respond to personal information access requests (made under the Privacy Act), general requests for information, and contractually entitled information or audit requests, or to delete data.

It's also important that you manage the ongoing security of the compromised dataset. Ensure that it is retained in accordance with your regulatory obligations and appropriately secured.

Key contacts

Valeska Bloch
Partner, Head of Cyber
T +61 2 9230 4030
Valeska.Bloch@allens.com.au

Isabelle Guyot
Partner
T +61 2 9230 4752
Isabelle.Guyot@allens.com.au

Lisa Kozaris
Chief Innovation & Legal Solutions Officer
T +61 3 9613 8944
Lisa.Kozaris@allens.com.au

David Rountree
Partner
T +61 7 3334 3368
David.Rountree@allens.com.au

Elizabeth Brown
Senior Associate
T +61 7 3334 3037
Elizabeth.Brown@allens.com.au

Maddison Ryan
Senior Associate
T +61 3 9613 8340
Maddison.Ryan@allens.com.au

Karan Mehta
Head of Legal Technology & Client Services
T +61 2 9230 4526
Karan.Mehta@allens.com.au

Peter Kirsopp
Legal Technology Senior Manager Integrated Legal Solutions, Legal Technology
T +61 3 9613 8243
Peter.Kirsopp@allens.com.au

Mary McEachern
Head of Review & Integrated Services
T +61 3 9613 8766
Mary.McEachern@allens.com.au

Our latest thinking on cyber

Defending from within: a guide to insider threat management

Why organisations must embed mental health and wellbeing support into cyber incident response planning

A step-by-step guide to assessing compromised data after a cyber incident

Establishing and preserving legal professional privilege in a cyber incident

Endnotes

1 See: Cost of a Data Breach Report 2023 | IBM.

2 See: AI Unleashes the Power of Unstructured Data | CIO
The Digitization of the World. From Edge to Core | IDC
What Is Unstructured Data And Why Is It So Important To Businesses? An Easy Explanation For Anyone | Forbes.