How to handle dark data compliance risk at your company

Slack and other consumer-grade productivity tools have been taking off in workplaces large and small — and data governance hasn't caught up.

Whether it's litigation, compliance with regulations like GDPR, or concerns about data breaches, legal teams need to account for new types of employee communication. And that's hard when work is happening across the latest messaging apps and SaaS products, which make data searchability and accessibility more complex.

Here's a quick look at the problem, followed by our suggestions for best practices at your company.

Problems

The increasing frequency of reported data breaches and expanding jurisdiction of new privacy laws are prompting conversations about dark data and risks at companies of all sizes, even small startups. Data risk discussions necessarily include the risk of a data breach, as well as preservation of data. Just two weeks ago it was reported that Jared Kushner used WhatsApp for official communications and screenshots of those messages for preservation, which commentators say complies with recordkeeping laws but raises questions about potential admissibility as evidence.

When it comes to data risk analysis, what's known as "dark data", especially if it qualifies as "personal data" or "personal information", is of particular concern for legal and compliance teams. It is not necessarily new, since organizations have dealt with "shadow IT" challenges for a long time. The current challenge relates to a confluence of the operational reality of bring-your-own-device policies and the proliferation of freely available software, like messaging apps and cloud storage services.

So what is dark data and why is it a topic of interest for legal and compliance teams? Dark data is ***data that is difficult to search or access***, or both. It is relevant to data risk management because searchability and accessibility of data are relevant to being able to find and secure it. Dark data can make it difficult or impossible for legal and compliance teams to find information when searching their

organization's files and documents. A common dark data example is chat files coming out of modern messaging apps, like Slack or WhatsApp. Unlike email messages, there is no chat message "standard", resulting in each company having their own file format (e.g. Slack uses a textual JSON schema, iMessage uses a binary SQLite database).

Use of tools that create dark data is already prevalent, since many are free. The pace of adoption for new tools that create dark data is fast and getting even faster, which means risk is being created everyday. Larger companies typically have a suite of approved tools for employees to use, but it's not uncommon for smaller companies and startups to pick up or change tools, like messaging apps, quickly. For startups, it may seem easier to accept dark data risk in favor of remaining agile, but, if ignored, the data risk snowball continues to roll downhill, making it increasingly more difficult to address as a company scales.

What are some specific risks and challenges related to dark data? These can range from producing evidence in litigation (e.g. ediscovery) to a company's ability to adhere to applicable regulations, such as data preservation and privacy laws. For example, GDPR presumes organizations know exactly what data they hold, making it important to understand if the company stores dark data (and where it is). It's also important to consider security risk, which boils down to the fact that an organization can't secure something if it doesn't know about it at all (or where it is).

Solutions

Mitigating risks presented by dark data can take several forms. Depending on the size of your organization, this may fall under the purview of a Legal, Security or Information Governance team at a larger company, or a Project Manager in a startup. Here are some actions you can take to mitigate risks around dark data.

1) Find out what data you have, and document where it is.

Understand what tools are in use across your organization, and their potential for creating or storing dark data. Without an understanding of what tools your internal stakeholders use to communicate, process and store data, you won't be able to assess risk related to dark data or establish that reasonable security controls are in place.

How?

- Take a business process approach to asking internal stakeholders questions about what kinds of data they create, access and store as part of their regular workflows. For example, find out what communication and productivity tools teams are using on a day-to-day basis, and how and where that data is stored. Are they approved tools, where the communications are searchable and accessible by the company? This is important to know. You may learn that a sales or marketing team stores personal information of EU prospects in a cloud-based note-taking tool, that your Legal team can't access!
- **Engage teams in the organization using a business process (or workflow) approach, and document what you learn. Here are some tools that can help:**

    - TrustArc Data Flow Manager: "[W]ith a data flow mapping tool, it is easy to standardize and operationalize the process of mapping all your customer and employee data flows. Create an up to date inventory of data collected, along with visual data flow maps of business processes."
    - BigId Data Mapping: "Data stewardship requires a big picture view of how data comes into an organization, how it gets processed and how it ultimately gets disposed. It will also benefit from a more detailed inventory that can be sliced and diced by data type, data subject, consent, calling application, system, country or even applicable regulation."
    - If a vendor solution is not in your budget, see what open source tools are available. For example, Everlaw created and then open-sourced a Google Sheet for this purpose, since many of the available tools came with a high cost. Larger companies likely have the budget to buy one of the many tools, such as those listed above designed for this purpose, but startups and smaller companies might be looking for a more accessible way to start answering questions about dark data and tools being used for business processes. The Everlaw Google Sheet was designed to capture business processes and tools that involve the use, storage or processing of personal information. It may also prove useful for startups looking to establish an information governance program.

2) Identify 3rd party and supply chain risk and come up with a plan to manage it.

Institute a process for third-party security and privacy vetting and risk management. A common topic discussed in legal and compliance circles, even at startups, is how they manage vendor security and privacy risk. Sending and receiving questionnaires aimed at facilitating a conversation around data processing, storage and security is now a common occurrence for both small and large

companies. Large companies typically have well-oiled vendor and procurement procedures, but now, startups are taking a hard look at their business partners, too.

How?

- There are several reference points and examples for organizations to use when it comes to either developing a security questionnaire to send to suppliers, or for developing an in-house checklist of standards that align with best practices. Whether an organization decides to pursue a questionnaire workflow, or embed security and data protection terms into contracts only, or both, will depend on specific threat models and risk tolerance.
- **Here are some examples of how an organization can operationalize either of these approaches:**

    - One example vendor security questionnaire is the Duke Law EDRM Project Security Audit Questionnaire, which "was designed primarily to help evaluate the security capabilities of cloud providers and third parties offering electronic discovery or managed services." This model questionnaire covers several regulatory frameworks, including Gramm-Leach-Bliley Act (GLBA) and PCI DSS. This template is useful because it can be customized to assign the weight (or importance) of each of the criteria, so that teams can emphasize what is most critical for their organization.
    - The Legal Cloud Computing Association (LCCA) website is another useful reference. LCCA published a set of standards aimed at helping legal and compliance organizations adopt cloud computing services. Their standards provide a framework that can serve as a checklist for an organization looking to meet the standard of "reasonable care" with regard to security, or, a starting point for developing a higher-level security questionnaire for a company's SaaS vendors.
    - Finally, Dropbox recently open sourced "the results of an experiment to improve vendor security assessments—directly codifying reasonable security requirements" into vendor contracts. They shared their model security legal terms and made them freely available on GitHub for anyone to use and modify.

In sum, dark data and how it interacts with business operations is an area where technology competence and understanding of how it is being created by organizations is increasingly important for legal and compliance teams, and for startups assessing data risk. Legal and compliance teams in both large and small organizations are paying attention to these risks, and looking to mitigate them by engaging their internal and external stakeholders and leveraging resources to assist them with data governance.