DATA MANAGEMENT, PRIVACY, AND SECURITY

Rebecca Wortzman and Lela Draganic– Big River Analytics

March 19th, 2021

Webinar Agenda

- Motivation
- 3 Key Takeaways
- Key Terms and Concepts to Understand
- The Data Management Lifecycle
  - Data Quality
  - Building Security and Privacy into your Data Management Lifecycle
- Privacy & Security: Strategy and Approach for Phase I

Appendix: Phases and Components to Privacy, Security, and Governance Programs

Topics in the Toolkit

- Concepts, Principles, and Practices for Data Security and Privacy
  - Plans, Policies, Procedures
- Considerations for IT Procurement
  - Total Cost of Ownership, System Requirements, Making a Business Base
- Data Quality
  - Documentation of metadata (i.e., information about a dataset)
  - Quality Assurance policies/programs
- Detailed treatment of Storage, Backup, Archiving, and Preservation

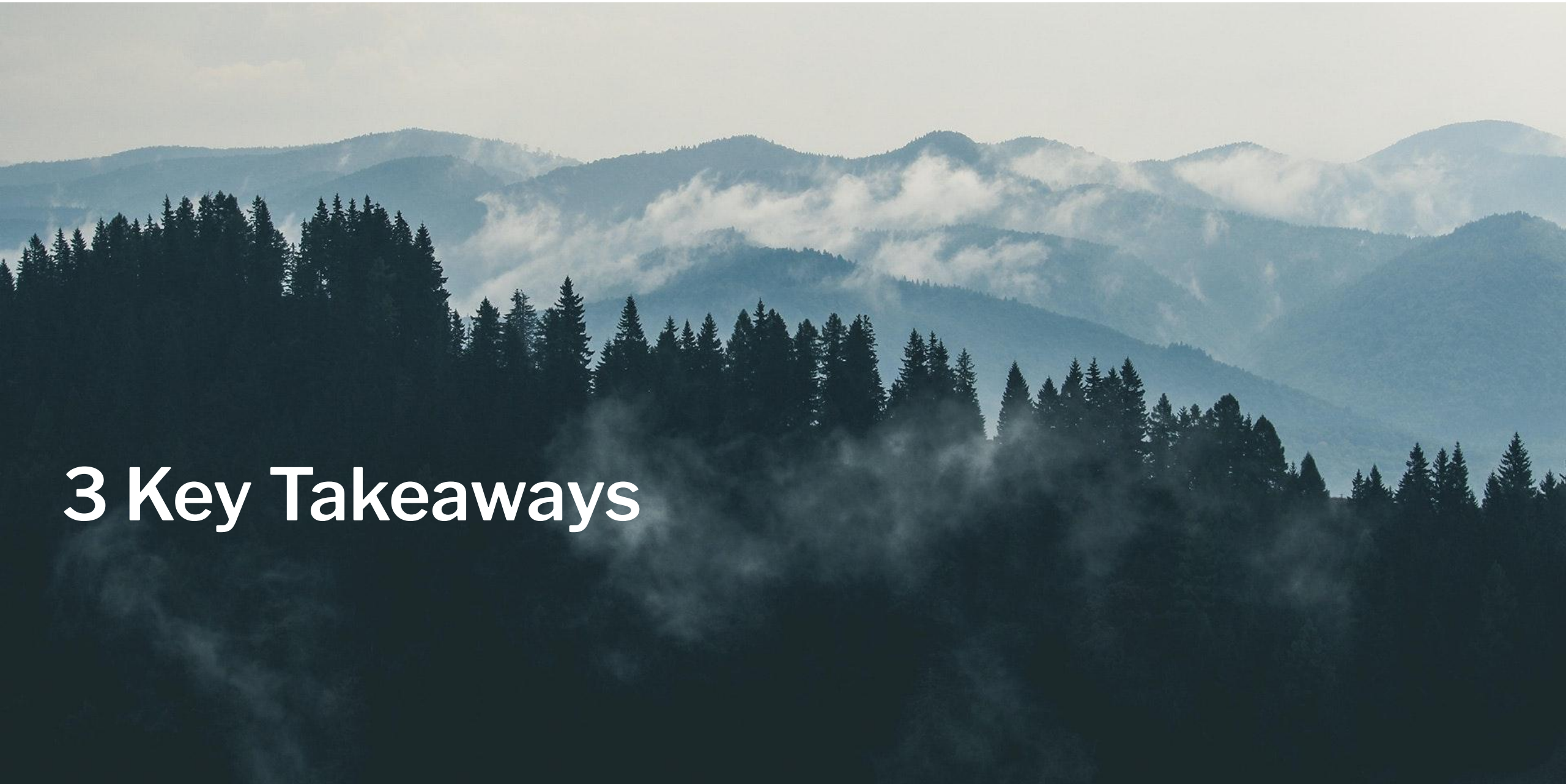
Motivation: Why are we here?

Data is not something we own.

When you protect personal data, you protect people.

This is a moral responsibility, on top of a legal one.

Strong privacy and security capacity can unlock new data avenues.



3 Key Takeaways

TAKEAWAY 1: Privacy and Security are a part of every aspect of data management.

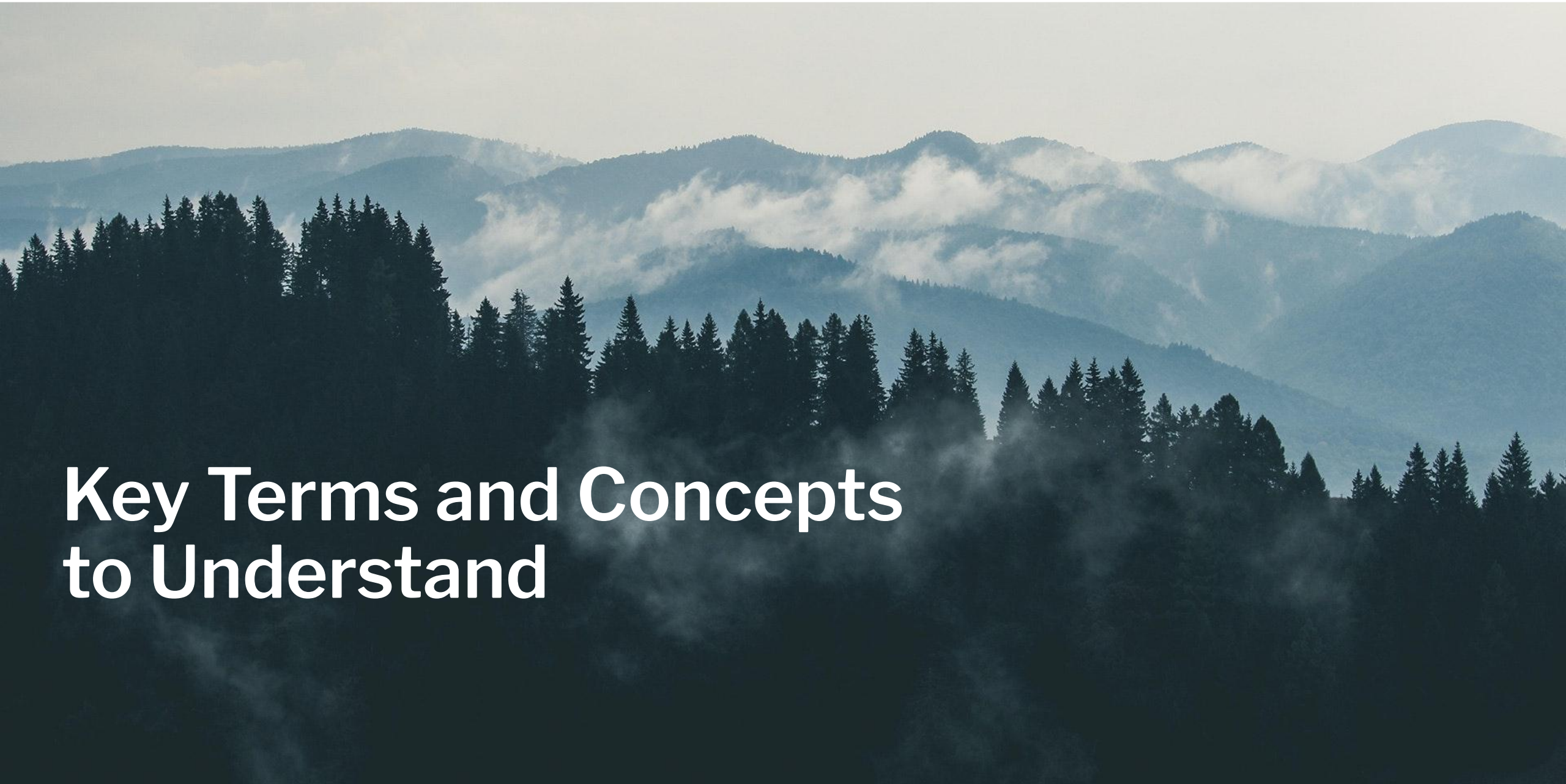
- Security and privacy **are not separate** from activities or stages of data management.
- They are a part of every decision you make from the moment you start thinking about collecting data, to conducting analysis, and ultimately storing it.

TAKEAWAY 2: Set realistic expectations for your privacy and security program — it will take time.

- Building cybersecurity and privacy capacity sounds intimidating and overwhelming to an organization of any size.
- This is a **long term initiative** that requires incremental change and a commitment to organizational change management.
- **Culture change** as much as this is about policy development.
- Reaching security maturity takes time and **your government cannot tackle everything overnight**.

TAKEAWAY 3: Data becomes sensitive when it has enough context.

- Data points, collected on their own, do not pose a privacy risk.
- **Context is key**. The more context there is around a data point, the more sensitive it becomes.



Key Terms and Concepts to Understand

Difference between key concepts and terms

Governance

IT

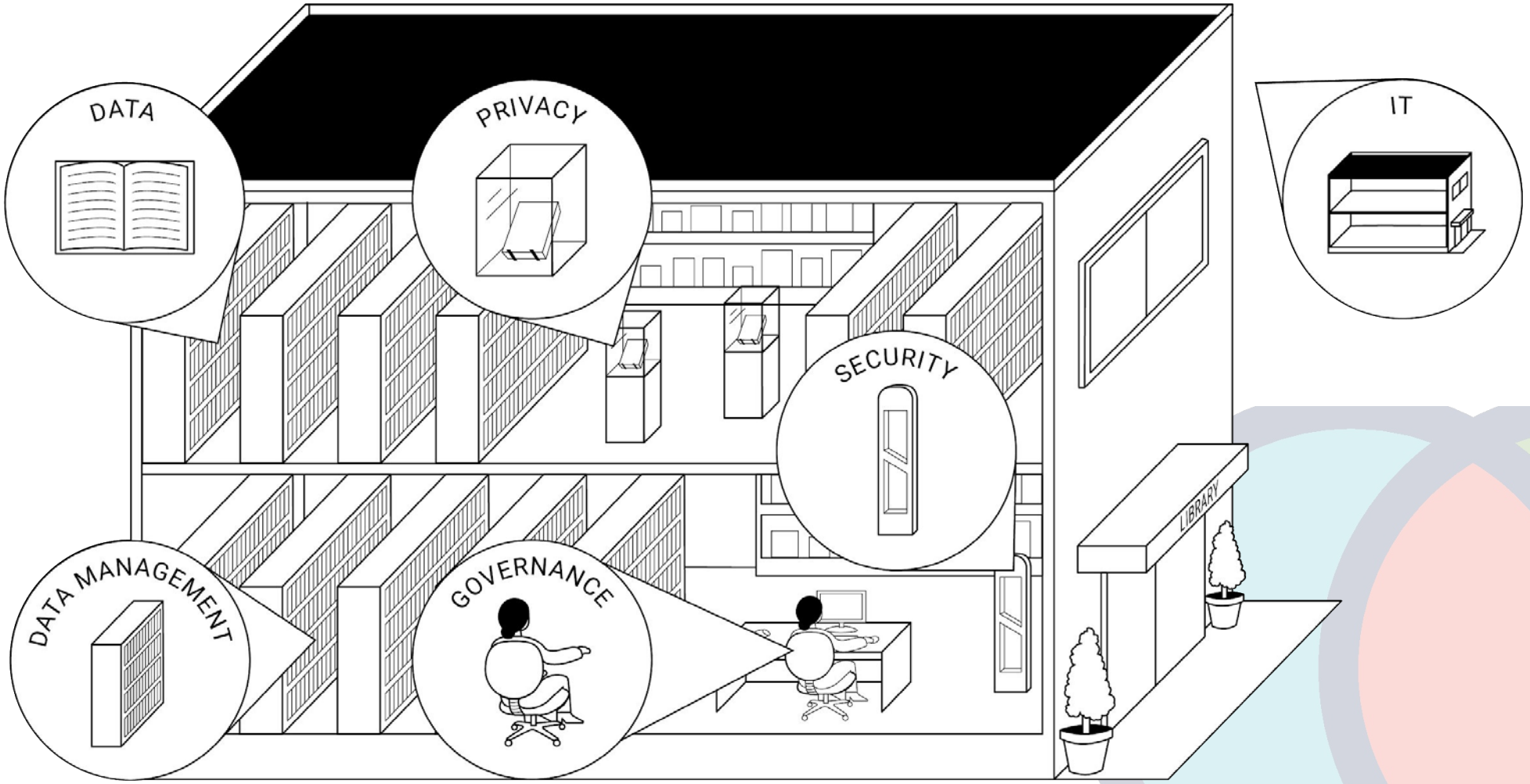
Cybersecurity

Privacy

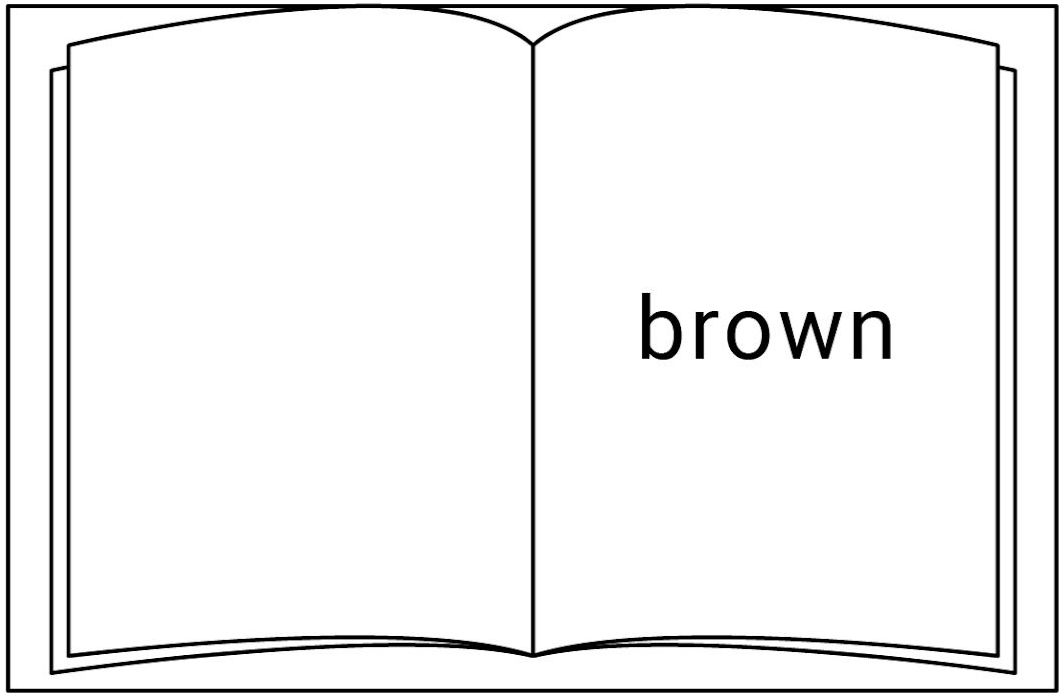
Data management

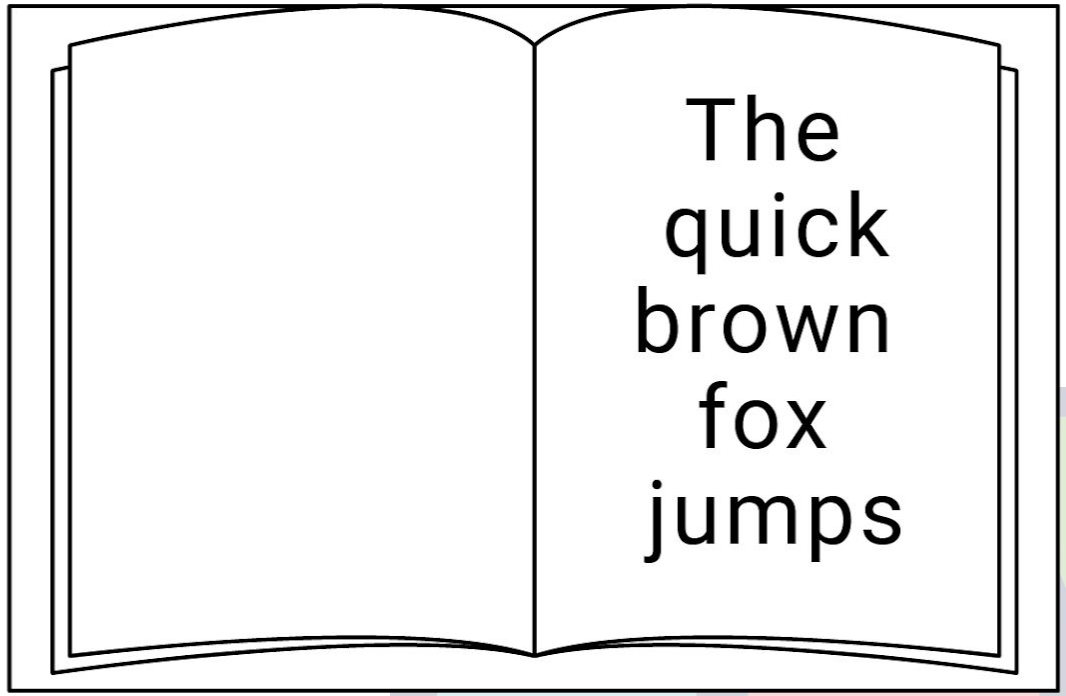
Data

Difference between key concepts and terms



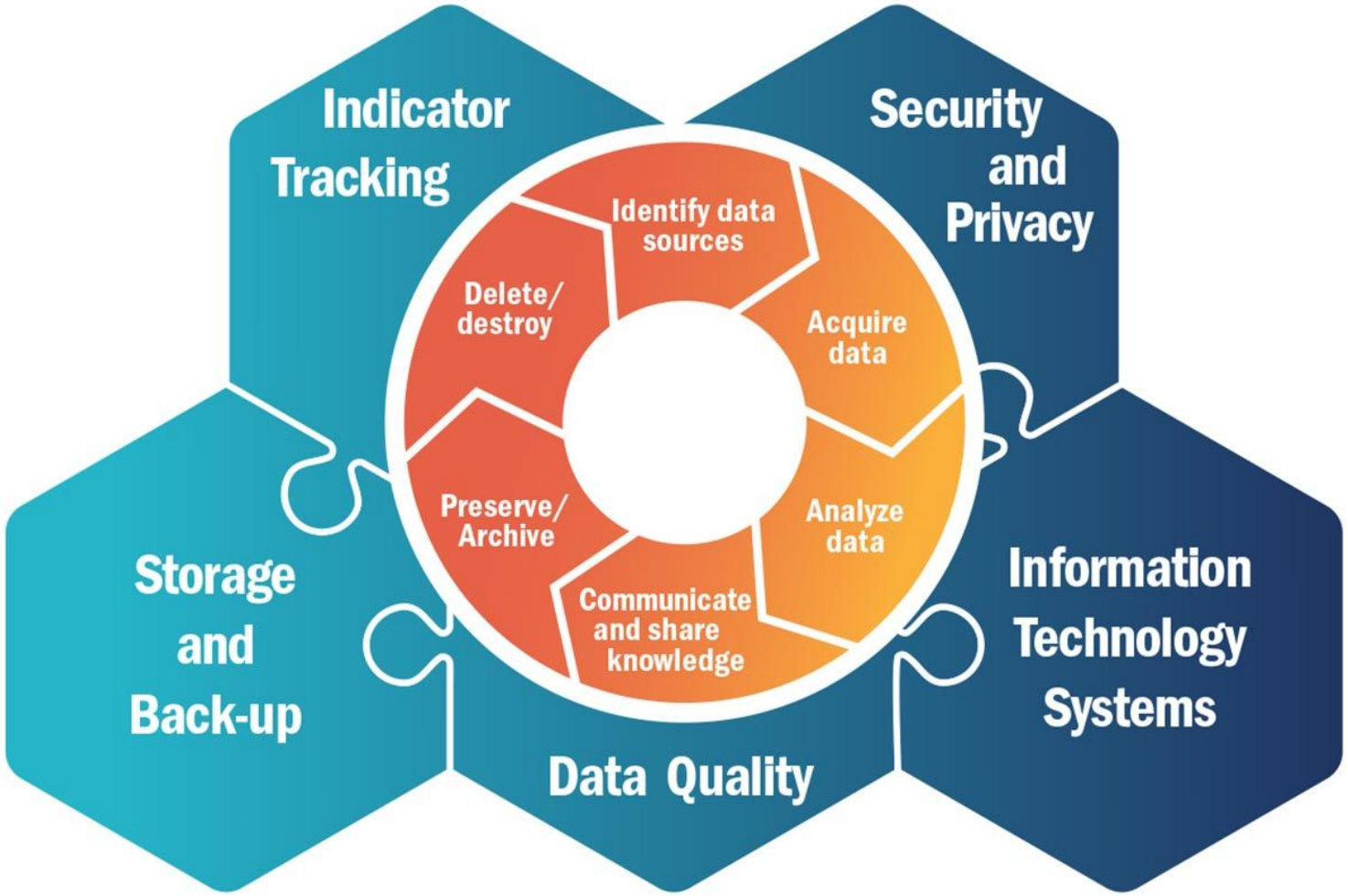
Understanding data





The Data Management Lifecycle

Building privacy and security into every phase of the data lifecycle



Dataset and System Quality

Data Management Lifecycle: Dataset and System Quality

Dataset and System Quality

- IT and System Design
- E.g. Single source of truth, appropriate coverage, sensible structure and storage

Quality Assurance and Best Practices

- Plans and processes to minimize and prevent errors

Data Management Lifecycle: Dataset and System Quality

Dataset Quality

- **Accuracy:** Is information factual?
- **Consistency:** Is information entered in the same format and way?
- **Completeness:** Is information collected for everyone for each field?
- **Uniqueness:** Are there duplications of records? Concepts within a dataset?
- **Timeliness:** Are data available, updated, or processed in a timely manner?
- **Relevance (Validity):** Does it measure what it's supposed to?

How To Build Privacy and Security Into Every Phase of the Data Lifecycle

Building privacy and security into every phase of the data lifecycle

Phase 1 and 2: Identify data sources and acquire data

- Am I collecting data in a secure way?
- Am I collecting consent from individuals?
- Am I recording consent appropriately?
- Do I really need to share/collect/store all this data or are only a few of these data points necessary?
- Can I anonymize this data or do I really need it fully readable?

Building privacy and security into every phase of the data lifecycle

Phase 3 and 4: Analyzing Data and Communicating and Sharing Knowledge

- Who truly needs access to this data and these files? Does your whole organization require access or can you limit it to a few people?
- What tools am I using to analyze this data? Are these custom, in-house developed tools our internal IT team is in charge of? Or vendor systems?
- Do we know if this is a zero-knowledge tool? Can the vendor read our information?
- Are we storing and sharing any sensitive data on movable media: USBs, external hard drives, CDs? Are we tracking these, with who and where they are?

Building privacy and security into every phase of the data lifecycle

Phase 3 and 4: Analyzing Data and Communicating and Sharing Knowledge

- Is all movable media protected by a password?
- Are we only communicating and sharing this data while on safe, encrypted networks? Are we using VPNs?
- Are we only accessing this data from password protected devices that have current antivirus software installed on them?

Building privacy and security into every phase of the data lifecycle

Phase 4: Communicating and Sharing Knowledge

- When sharing the collected data, how am I sharing it internally with my team? Am I emailing information back and forth or are we relying on one central storage where we have controlled access?
- Am I only sharing data and information others need? Am I utilizing access controls and providing only needed access: read-only, read-write, administrative?

Building privacy and security into every phase of the data lifecycle

Phase 5: Preserve and Archive

- What tool are we using to archive and preserve this data?
- Do we know how long we need to have this data archived, what is the data retention period?
- Do we need to preserve the full copy of the data or can we redact or obfuscate some of it before archiving?

Building privacy and security into every phase of the data lifecycle

Phase 5: Preserve and Archive

- Has the archive been tested for disaster recovery? How often are the tests conducted? Are backups of the archive distributed across different servers and geographical regions, in case of natural disasters etc.?
- Who has access to the archive? How do you keep track of who has access to it and how quickly can you revoke it?
- Are the archives encrypted at rest?

Building privacy and security into every phase of the data lifecycle

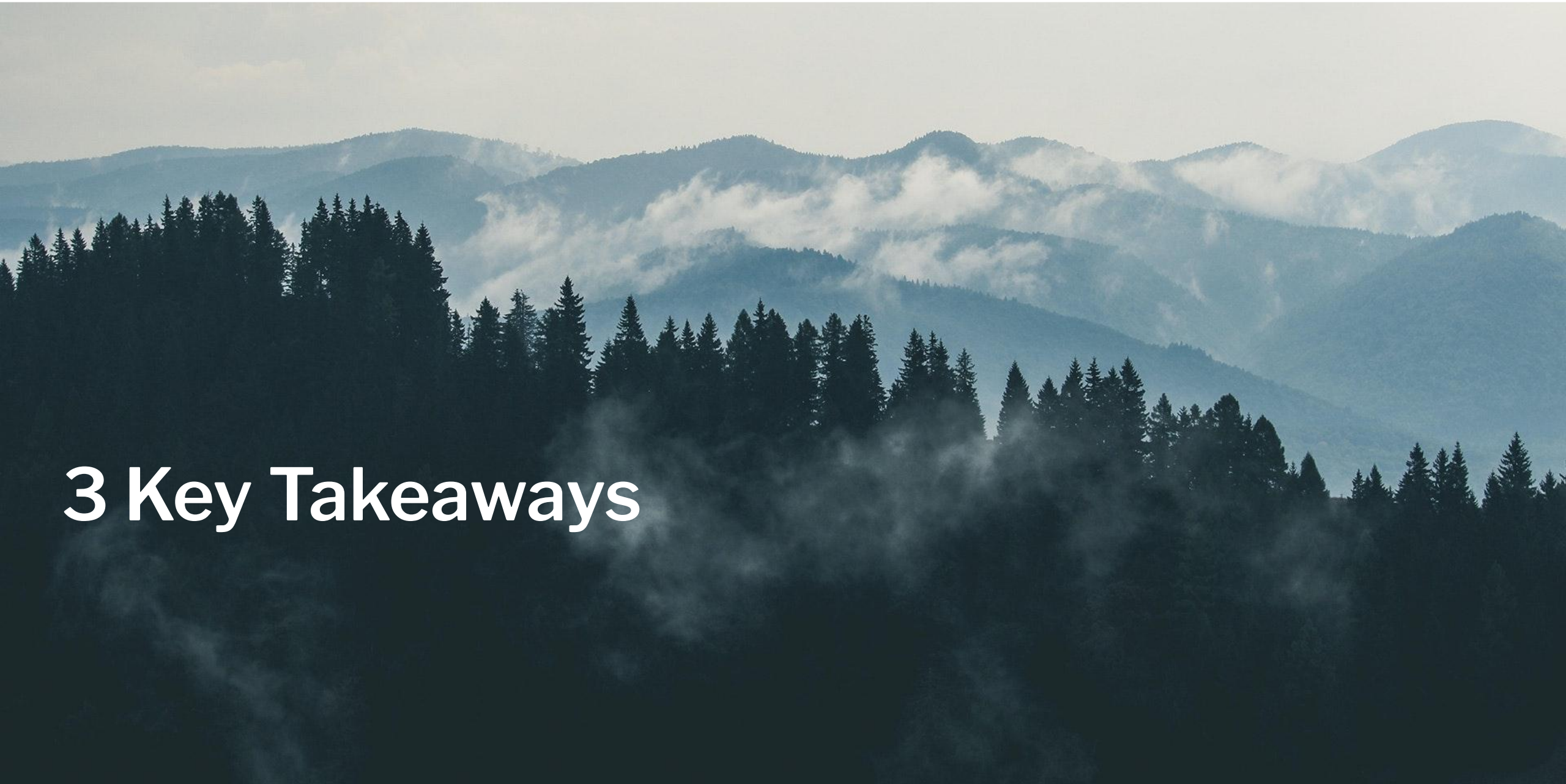
Phase 6: Delete/Destroy

- Do you know exactly where all your data and files are? Where has the data been copied? Have your team members made copies to their devices?
- How are you deleting this data from your archives or your devices? Will the data be impossible to retrieve and fully overwritten?

Where To Go From Here: Strategy and Approach for Phase I

Approach and Strategy for Phase I

1. **Get a lay of the land and create a map.** Identify all the platforms you are working with, that you are using to communicate with people, collect, handle or store data. The difference between collecting, handling and storing data.
2. Figure out **how the data is moving** between different applications and platforms.
3. Create a data dictionary and **identify and assess all the data you are working with**. Then within that identify which data is the most sensitive and which data types, when combined, reveal the most sensitive information about individuals, groups or organizations.
4. Review **who has or could have access** to this data. How does someone gain access to this data and the platforms where it's kept. Is this access attested to. Who takes that access away.
5. Setup your **Security and Privacy Governance structure.** Identify, employ or contract the necessary stakeholders that can guide this program for you.



3 Key Takeaways

TAKEAWAY 1: Privacy and Security are a part of every aspect of data management.

- Security and privacy **are not separate** from activities or stages of data management.
- They are a part of every decision you make from the moment you start thinking about collecting data, to conducting analysis, and ultimately storing it.

TAKEAWAY 2: Set realistic expectations for your privacy and security program — it will take time.

- Building cybersecurity and privacy capacity sounds intimidating and overwhelming to an organization of any size.
- This is a long term initiative that requires incremental change and a commitment to organizational change management.
- Culture change as much as this is about policy development.
- Reaching security maturity takes time and your organization cannot tackle everything overnight.

TAKEAWAY 3: Data becomes sensitive when it has enough context.

- Data points, collected on their own, do not pose a privacy risk.
- Context is key. The more context there is around a data point, the more sensitive it becomes.

Thank you!

Questions?

Appendix: Cybersecurity and Privacy Focus Areas, and Governance

Appendix: Security Focus Areas

What does a cyber security program entail? Controls are security measures that you have in place that can be grouped in the following focus areas:

- Access and identity management (who gains access how, who has access to what, when is that access taken away, how is it tracked)
- Third-party or vendor management (are the platforms safe, do you have a contract with them, do they have an obligation to let you know if there is an incident, what's their reputation)
- Network segmentation (do guests get access to your office network, do you have separate networks for guests and those working in-office, do you have firewall rules in place)
- Software defense (do all computers have antivirus on them)
- Hardware defense (can anyone just plug in a USB and exfiltrate data)
- Data transfer (how is data transferred, is there a VPN, is the email encrypted, do you get notifications when someone tries moving a large chunk of data out of an organizational account)
- Logging and monitoring (keeping logs of who has access to the information, what they did with the data, when)
- Disaster recovery (ensuring there are functional backups of the data, that have been tested and can be relied on in case of a natural disaster, complete system failure, loss of devices)
- Cyber Security Incident Response: having a guidebook in place in case an incident happens, knowing how to react, who to contact for support, who to notify, how to prevent further damage and how to collect evidence
- Education and awareness

Appendix: Privacy Focus Areas

What does a privacy program entail? A privacy program within any organization assumes that you are familiar, and have processes and procedures for the following:

- Regulation (what regulation and legislation do you need to be compliant with? What are the consequences or fines for non-compliance?)
- Disclosure (when and how do you need to disclose any breach or data leak? who do you need to disclose it to? how quickly?)
- Data classification system (categorizing data in line with the Restricted, Confidential, Internal, Public scale, or any other scale that works for you)
- Data dictionary (what data types you are working with and what the data classification of each data type is, as well as how the information was collected
- Consent records (have you informed individuals about the information you are collecting and why? have you recorded their consent? do you know how long you need to retain proof of consent?)
- Documentation (do you have privacy policy in place? if you have a website, do you have it published?)
- Information Access Request (do you know how you will respond if you receive an information access request, and you need to send an individual's data to them?)
- Data Agreements with Vendors (if you are using third party software and platforms in your organization do you have a data sharing agreement with those vendors?)
- Communication and training

Appendix: Privacy and Security Governance

What does a privacy and security governance structure look like

- **Chief Information Security Officer:** oversees the privacy and cybersecurity programs, where all subject matter experts and owners report to them.
- **Data stewards:** individuals who are familiar with the data being collected in the whole organization, or depending on the size, separate data stewards for each project or data source (community).
- **Privacy Counsel/Data Protection Officer:** either a lawyer or a data protection expert who is familiar with the relevant regulations and legislation and can advise your organization on how to respond to information access requests, report incidents, disclose breaches and draft policies.
- **Cybersecurity focus area owners:** individuals who are experts and can implement CyberSec controls (network security expert, system defence etc.). This can be one individual or a team of people, depending on the size and requirements of the organization.

Appendix: Phases of Data Security Maturity

Phase I: Knowing what you're working with. You're aware of how well educated and trained your team is, you know where your data is, what data sets you're working with, what solutions and platforms you're using and how they all interact. At this phase you would also identify and put in place your cybersecurity and privacy governance structure.

Phase II: You have assessed the security and privacy of all your tools, platforms and software. The risk assessment has shown you where you need to add cybersecurity controls, what documentation (e.g. privacy policy or incident response guidebook) you need to develop and socialize in your organization. You have not implemented these fixes yet at this stage, but have completed a detailed risk audit that tells you what your gaps are.

Phase III: You have started implementing the fixes that will close the cybersecurity gaps that were identified in Phase II. This phase can take a long time to complete, and that is ok. You just need to make sure to prioritize the critical and high risk vulnerabilities and close those first. The risk audit will tell you which issues you need to prioritize. At this stage you will be selecting new tools like a more aggressive antivirus, reviewing your firewall rules, setting up better access management policies, reviewing your data sharing agreements with third parties, adding encryption in transit and at rest, developing a detailed incident response guidebook etc.

Phase IV: Maintenance and improvement. Once you have covered all your bases and closed all the gaps, you now need to maintain your cybersecurity and privacy programs. This phase involves improving your governance structure to be more efficient, ensuring patches and updates are installed and applied in a timely manner, keeping up with the threat landscape and staying up to date with breaches or security news that impact your organization and having the capacity and resources to react quickly to these.