Data Security

Linus Rundberg Streuli

Table of contents

1 Data Security

- Data security includes the planning, development and execution of security policies and procedures to provide “the four A’s”: proper Authentication, Authorization, Access, and Auditing of data and information assets. (See Section 4.6.1.)
- The goal of data security practices are to protect information assets in alignment with privacy and confidentiality regulations, contractual agreements, and business requirements.
- These requirements come from:
  - **Stakeholders**: Clients, patients, employees, students, citizens, suppliers and business partners all have privacy and confidentiality needs that must be recognized.
  - **Government regulations**: Some restrict access to information, others ensure openness, transparency and accountability.
  - **Proprietary business concerns**: If confidential, proprietary data is stolen, a business may lose competetive advantage.
  - **Legitimate access needs**: Business processes require individuals in certain roles to be able to access, use, and maintain data.
  - **Contractual obligations**: Some data is influenced by contractual and non-disclosure agreements.
- Effective data security policies and procedures ensure that the right people can use and update data in the right way, and that any other uses and updates are restricted.

2 Business Drivers

- The primary drivers of data security activities are risk reduction and business growth.
- Security is a valuable asset in its own right.

2.1 Risk Reduction

- Security organizations are often tasked with managing not only IT compliance requirements, but also policies, practices, data classifications and access authorization rules.
- Data security should be handled on an enterprise level. If different business units implement their own security solutions, costs may increase and the overall security levels may suffer.
- Data security begins with identifying which data requires protection. Steps include:
  - **Identify and classify sensitive data assets**: Different industries and organizations deal with different amounts and types of sensitive data.
  - **Locate sensitive data throughout the enterprise**: Security requirements may differ depending on where the sensitive data is stored.
  - **Determine how each asset needs to be protected**: Data type and technology influences what measures need to be taken.
  - **Identify how this information interacts with business processes**: Determine what accesses are necessary, and under what circumstances.
- Assess external (hackers, thieves) as well as internal (employees, processes) security threats.
  - Internal security incidents are often the results of missing or unenforced security controls, rather than malicious intent.

2.2 Security as an Asset

- One common approach to data security is through metadata.
- Creating a master repository of data classifications enables all parts of an organization to know the levels of protection required for different data sets and elements.
- Security-related metadata becomes a strategic asset, increasing the quality of transactions, reporting and business analysis, while reducing the costs associated with protection, regulation non-comformance, and data loss.
- One common tool for managing metadata is Apache Atlas, which we will take a look at later in this lecture.

2.3 Data Security vs. Cyber Security

- *Data* security aims to protect the confidentiality, integrity, and availability of data.
- It is a subset of the broader concept of *cyber* security, which involves protecting the entire digital environment from cyber threats.
- They are however closely linked, and a robust cyber security strategy involves data security measures, while effective data security contributes to the overall cyber security of an organization.

3 Goals and Principles

3.1 Goals

- Enabling appropriate access and preventing inappropriate access to enterprise data assets,
- Enabling compliance with regulations and policies for privacy, protection, and confidentiality, and
- Ensuring that stakeholder requirements for privacy and confidentiality are met.

3.2 Principles

- **Collaboration**: Data security involves IT security administration, data stewards/data governance, internal and external audit teams, and the legal department.
- **Enterprise approach**: Data security standards and policies must be applied consistently across the entire organization.
- **Proactive management**: Success in data security management depends on overcoming organizational or cultural bottlenecks such as traditional separation of responsibilities between information security, IT, data administration, and business stakeholders.
- **Clear accountability**: Roles and responsibilities must be clearly defined.
- **Metadata-driven**: Security classification for data elements is an essential part of data definitions.
- **Reduce risk by reducing exposure**: Minimize the amount of sensitive data, especially in non-production environments.

4 Essential Concepts

- Here follows a number of key terms related to information security.

4.1 Vulnerability

- A *vulnerability* is a weakness or defect in a system that allows it to be attacked - essentially a hole in an organization's defenses.
- Some vulnerabilities are called *exploits*.
- Examples include:
  - Computers with out-of-date security patches,
  - Web pages not protected with secure passwords,
  - Users not trained to ignore emails from unknown senders, and
  - Software vulnerable to commands that will give an attacker control of the system, such as SQL injections.

4.2 Threat

- A *threat* is a potential offensive action that could be taken against an organization.
- An occurence of a threat is also called an *attack surface*.
- Threats can be internal or external, and not always malicious - an uninformed insider can take offensive actions without even knowing it.
- Each threat should match to a capability that either prevents the threat or limits the damage it might cause.
- Examples of threats include:
  - Emails containing virus-infected attachments,
  - Denial-of-service attacks which overwhelm processes and leaves the organization unable to perform business transactions, and
  - Exploitation of known vulnerabilities.

4.3 Risk

- The term *risk* refers both to the possibility of loss, and to the thing or condition that poses the potential loss.
- Risk can be calculated for each possible threat using the following factors:
  - Probability that the threat will occur andits likely frequency,
  - The type and amount of damage created each occurence might cause, including damage to reputation,
  - The effect damage will have on revenue or business operations,
  - The cost to fix the damage after an occurrence,
  - The cost to prevent the threat, including by remediation of vulnerabilities, and
  - The goal or intent of the probable attacker.
- Risks can be prioritized either by potential severity, or by likelihood of occurrence.
- Prioritization of risk must be a formal process among the stakeholders.

4.4 Risk Classifications

- Risk classifications describe the sensitivity of the data and the likelihood that it might be sought after for malicious purposes.
- Classifications are used to determine which roles should have access to the data.
- Example classifications include:
  - **Critical Risk Data (CRD)**: Personal information aggressively sought for unauthorized use, by both internal and external parties.
  - **High Risk Data (HRD)**: Data actively sought for unauthorized used due to its potential direct financial value.
  - **Moderate Risk Data (MRD)**: Non-public information that has little tangible value to unauthorized parties, yet would likely have a negative effect on the organization if it was stolen.

4.5 Data Security Organization

- Depending on the size of the company, its security organization will differ.
- In all cases, data managers will need to be involved in data security efforts.
- However, in many organizations IT and data management lack standard procedures for collaboration and information sharing.
- An enterprise data model, categorizing and describing sensitive data, is an essential part of an effective data protection program.

4.6 Security Processes

- Data security requirements and procedures are categorized into four groups, know as the four A's: Authentication, Authorization, Access, and Auditing.
- A fifth category, Entitlement, has been added recently.
- The means to implementing policy and satifying the four A's are:
  - Information classification,
  - access rights,
  - role groups,
  - users, and
  - passwords.
- Security monitoring is also essential for proving the success of the other processes.

4.6.1 The Four A's

- **Authentication**: Validate users' access - via passwords or more stringent methods such as security tokens or biometrics.
- **Authorization**: Grant individuals privileges to specific views of data, appropriate to their role.
- **Access**: Enable individuals with authorization to access systems in a timely manner.
- **Audit**: Review security actions and user activity to ensure compliance with regulations and conformance with company policy and standards.
- **Entitlement**: An Entitlement is the sum total of all the data elements that are exposed to a user by a single access authorization decision. A responsible manager must decide that a person is "entitled" to access this information before an authorization request is generated.

4.6.2 Monitoring

- Systems should include montoring controls that detect unexpected events, including potential security violations.
- Some security systems will actively interrupt activities that do not follow specific access profiles, locking accounts or activities until security personnel have evaluated the situation.
- Other systems are passive, tracking changes and comparing trends against certain criteria, and sending reports to the person accountable for the data.

4.7 Data Integrity

- Data integrity is the state of data being protected from improper alteration, deletion, or addition.

4.8 Encryption

- Encryption is the process of translating plain text into complex codes to hide privileged information, verify complete transmissions, or verify a sender’s identity.
- Encrypted data cannot be read without the proper decryption key or algorithm, which is usually stored in a separate location.
- Some common methods of encryption include hash, symmetric (secret-key), and asymmetric (private-/public-key).
- Please note that this part of the lecture differs somewhat from the DAMA DM-BOK as there are some minor differences in the terminology used.

4.8.1 Hash

- Hash encryption uses algorithms to convert data into a mathematical representation.
- Strictly speaking, hashing is not an encryption technique but rather a way to get a unique representation of the information while keeping the content secret. It is generally a one way process.
- Common algorithms are Message Digest 5 (MD5) and Secure Hashing Algorithm (SHA).
- Hashing is commonly used to compare records, documents or senders, without revealing their actual contents or identities.

4.8.1.1 Hash Example

This is an example of hashing using the `hashlib` module from the Python standard library.

```
1 import hashlib
2
3 raw_string = b"Some data to be encrypted"
4 encrypted_data = hashlib.sha256(raw_string).hexdigest()
5
6 encrypted_data
```

```
'a007090e1d86b6e8c7a6fa4e730b99d0020e303968fffa93bd5c12b9764860d2'
```

4.8.2 Symmetric

- Symmetric-key encryption uses the same secret key for encrypting and decrypting data.
- Symmetric-key algorithms include Twofish, Serpent and the Advanced Encryption Standard AES (or Rijndael), which was developed by the U.S. National Institute of Standards and Technology, NIST.

4.8.2.1 Symmetric Example

This is an example of encrypting and decrypting data using the symmetric-key method. `cryptography` is a third-part Python library commonly used for encryption-related tasks.

```
1  from cryptography.fernet import Fernet
2
3  key = Fernet.generate_key()
4  f = Fernet(key)
5
6  raw_string = b"Some secret data"
7
8  encrypted_token = f.encrypt(raw_string)
9
10 encrypted_token
```

```
b'gAAAAABnHn-
YCx3rvGFvXnkOWy7u32R7ZmUb5VcZDBkp_SFw06YaVyVdJtclyp2AaswIm0bbLyXe7G9IxqmHR
```

```
1  f.decrypt(encrypted_token)
```

```
b'Some secret data'
```

4.8.3 Asymmetric

- In asymmetric encryption, the secret key is divided into two parts: one public key used to encrypt the data, and one private key used to decrypt the data.
- The public key can be shared while the private key must be kept secret.
- Common methods include Rivest-Shamin-Adelman (RSA) Key Exchange, and PGP (Pretty Good Privacy).

4.8.3.1 Asymmetric Example 1: Generate key pair

This is an example of generating a public/private key pair in Python using the `cryptography` library.

```
1  from cryptography.hazmat.primitives import hashes
2  from cryptography.hazmat.primitives.asymmetric import rsa, padding
3  from cryptography.exceptions import InvalidSignature
4
5  key_size = 2048
6  private_key = rsa.generate_private_key(
7      public_exponent=65537,
8      key_size=key_size
9  )
10 public_key = private_key.public_key()
```

4.8.3.2 Asymmetric Example 2: Encryption

We can then use our public key to encrypt data.

```
raw_string = b"Some data to be encrypted"

encrypted_data = public_key.encrypt(
    raw_string,
    padding=padding.OAEP(
        mgf=padding.MGF1(algorithm=hashes.SHA256()),
        algorithm=hashes.SHA256(),
        label=None
    )
)

encrypted_data.hex()
```

```
'56058f454c27f73082b21ecec10280840171ecd98f5b15dd5d4fafedd9ae929bf8e206afe
```

4.8.3.3 Asymmetric Example 3: Decryption

Finally, we use our private key to decrypt the data.

```
decrypted_data = private_key.decrypt(
    encrypted_data,
    padding=padding.OAEP(
        mgf=padding.MGF1(algorithm=hashes.SHA256()),
        algorithm=hashes.SHA256(),
        label=None
    )
)

decrypted_data
```

```
b'Some data to be encrypted'
```

4.9 Obfuscation/Masking

- Another method for making data less available is through obfuscation, or masking.
- Masking removes, shuffles, or otherwise changes the appearance of the data, without losing its meaning or relationships to other objects or systems.
- Obfuscation is convenient when displaying sensitive data on screens, or creating test sets from production data.
- There are two types of data masking: persistent and dynamic.

4.9.1 Persistent Data Masking

- Persistent data masking permanently and irreversibly alters the data.
- This type of masking is generally used in-between production and development or test environments.
- **In-flight persistent masking** occurs when the data is on its way between the source and the destination. This is very secure, as it leaves no intermediate file or database with unmasked data.
- **In-place persistent masking** is used when the source and destination are the same. The unmasked data is read from the source, masked, and then used to overwrite the original data. This method assumes that the data is in a place where it should not be, or that a copy exists in a secure location. It is also more riskful, and the in-flight method is generally preferable.

4.9.2 Dynamic Data Masking

- Dynamic data masking changes the appearance of data without making changes to the underlying data.
- This is useful when users need to access some sensitive data, but not all of it. For example, a database might store personal identity numbers using the YYYYMMDD-XXXX format, but only show the first eight digits in certain situations.

4.9.2.1 Masking Methods

- Masking methods include:
  - **Substitution**: Replace characters or whole values with those from a lookup or as a standard pattern.
  - **Shuffling**: Swap data elements of the same type within a record, or between rows.
  - **Temporal variance**: Move dates forwards or backwards a number of days, small enough to keep trends, but large enough to avoid identification.
  - **Value variance**: Apply a random value within a certain range.
  - **Nulling or deleting**: Remove data that should not be present.
  - **Randomization**: Replace characters or whole values with random characters.
  - **Encryption**: Use one of the methods described earlier to render values unrecognizable.
  - **Expression masking**: Change all values to the result of an expression. For example, change all entries in a free form database field (that might contain sensitive information) to be “This is a text comment.”
  - **Key masking**: Mask database key fields. Make sure that the masking process is unique and repeatable.

4.10 Network Security Terms

- Data security includes both data-at-rest and data-in-flight.
- For data to move between systems, a network is required.
- Data security needs to be part of a greater cyber security system.
- Here follows some terms and concepts related to network security.

4.10.1 Backdoor

- A *backdoor* refers to an overlooked hidden entrypoint into a computer system or application, allowing unauthorized access.
- Many backdoors are created by developers for maintenance purposes.
- Examples of backdoors include default passwords left unchanged, and vulnerabilities created by malicious software.
- All backdoors are a security risk.

4.10.2 Bot/Zombie

- A *bot*, or zombie, is a computer that has been taken over and is remotely controlled by an attacker.
- Bots are used to send large amounts of spam, perform denial-of-service attacks against legitimate businesses, and hosting fraudulent websites.
- A network of bots is called a *botnet*.
- According to the cybersecurity firm Thales, 32% of all internet traffic in 2023 was associated with “bad bots”.

4.10.3 Cookie

- A *cookie* is a small data file that a website installs on a computers hard drive to identify returning visitors and profile their preferences.
- The vast majority of cookies used by websites are perfectly harmless, but they are sometimes used by spyware.

4.10.4 Firewall

- A *firewall* is software and/or hardware that filters network traffic to protect a single computer or an entire network from unauthorized attempts to access or attack the system.
- Firewalls can scan both incoming and outgoing traffic for suspicious activity and prevent it from passing through.

4.10.5 Perimeter

- A *perimeter* is the boundary between an organization's environments and the outside world.
- This is typically where a firewall is placed.

4.10.6 DMZ

- Short for “demilitarized zone”, a DMZ is an area on the perimeter of an organization, between two firewalls.
- DMZ environments are used to temporarily store data moving between organizations.

4.10.7 Super User Account

- A *super user account* is an account that has administrator or root access to a system.
- The credentials for super user accounts should be tightly controlled by time, user ID, location, and other requirements.

4.10.8 Key Logger

- *Key loggers* are software used to record the keystrokes made by a user of a computer, and send them to another computer.
- Passwords, documents and other sensitive data may be captured in this fashion.
- Key loggers are often installed by malicious software or infected documents.

4.10.9 Penetration Testing

- *Penetration testing* is when an ethical hacker tries to break into a system from the outside to test its defenses.
- Results from penetration tests can be used to address vulnerabilities before releasing an application or deploying a system to production.
- All software contains potential vulnerabilities and should be periodically tested.
- When weaknesses are found, no blame should be applied - only security patches.

4.10.10 Virtual Private Network (VPN)

- A *virtual private network*, or VPN, use the unsecured internet to create a highly encrypted “tunnel” into an organization’s environment.
- This tunnel can then be used to securely communicate with an internal network from the outside world.

4.10.11 Virtual Private Cloud (VPC)

- A *virtual private cloud*, or VPC, is the cloud version of a VPN.
- It is a secure, isolated cloud within a public cloud.
- VPCs can be used together with VPC Service Controls to create perimeters within a cloud environment to further protect sensitive data.
- Devices within the perimeter typically has no connection to the outside cloud.

4.11 Types of Data Security

- Data security involves not only preventing inappropriate access, but also facilitating appropriate access to data.
- The important “least privilege” principle states that a user, process or program should only be allowed access to the information they need to perform legitimate business tasks.
- Without permission, no user should be able to see data or take any action within the system.

4.11.1 Facility Security

- Facility security is the first line of defense against bad actors.
- A locked data center with access only for authorized employees is a minimum for any facility that handles data.
- Employees must have the tools and training necessary to protect data in the facilities.

4.11.2 Device Security

- Mobile devices can be lost, stolen, and physically and electronically attacked, and so are inherently insecure.
- Such devices often contain documents and emails that can damage an organization if they are exposed.
- A plan to manage the security of mobile devices must be a part of an organization's overall security strategy.
- Device security standards include:
  - Access policies regarding connections using mobile devices,
  - Storage of data on portable devices such as laptops, DVDs or USB drives,
  - Data wiping and disposal of devices in compliance with data management policies,
  - Installation of anti-malware and encryption software, and
  - Awareness of security vulnerabilities.

4.11.3 Credential Security

- Each user is assigned credentials to use when obtaining access to a system, often a combination of a user ID and a password.
- To keep track of users, credentials and access policies, most organizations use some sort of identity management system.

4.11.3.1 Identity Management Systems

- It is often a good idea to implement a “single-sign-on” system, where a user logs in to a workstation and all authentication and authorization are executed through a reference to the enterprise user dictionary.
- This minimizes the number of user ids and passwords a user have to keep track of, stopping them from using unsafe methods to remember different passwords for different parts of a system.

4.11.3.2 User ID Standards for Email Systems

- To keep user IDs unique, most organizations use some combination of first and last names, and sometimes numbers, to create email or network IDs.
- It is generally discouraged to use employee ID numbers for email or network IDs, as these are data that shold not leave the organization.

4.11.3.3 Password Standards

- Every user account should be required to have a password set by the account owner, with a sufficient level of complexity as defined in the security standards.
- When creating new accounts, the temporary password must be set to expire immediately after the first login, and the user must choose a new password.
- How often users should change their passwords is under debate, as requiring users to change them too often might lead them to write them down.

4.11.3.4 A Recent Security… Mishap

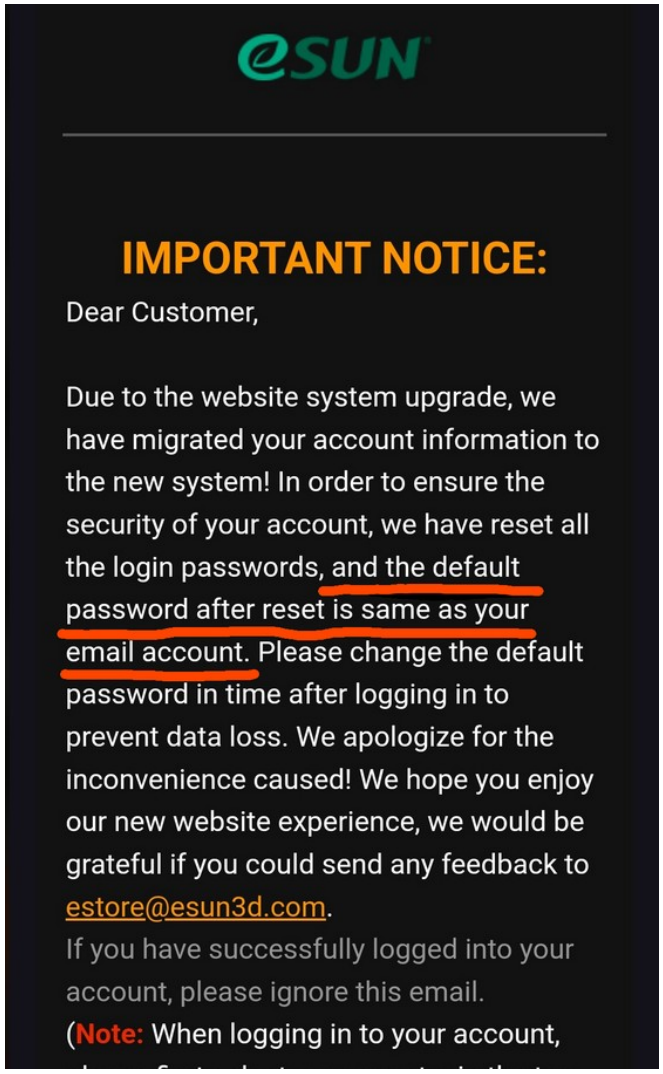

Figure 1: Screenshot of email sent to Esun customers. Source: reddit.com

- In October of 2024, customers of chinese 3d-printer filament supplier Esun received an email informing them of a website system upgrade.
- As part of the upgrade, all account passwords had been reset - to the email address connected to the account.
- This is of course a massive security failure.

4.11.3.5 Multiple Factor Identification

- Some systems require additional identification procedures.
- These include entering a temporary code into a connected device, using a hardware item the must be connected to the device logging in, or providing biometrics such as fingerprint, facial recognition, or retinal scans.
- All users with access to highly sensitive information should use two-factor identification when logging in.

4.11.4 Electronic Communication Security

- Users must understand the insecurities of sending information over email or other communications software, including social media and blogs.
- When data has been sent over these channels, the user no longer controls the data, and it can be forwarded to other people without the knowledge or consent of the original sender.

4.12 Types of Data Security Restrictions

- Two concepts drive security restrictions:
  - **Confidentiality levels**: Organizations determine which types of data should not be known outside the organization, or even within certain parts of the organization. Levels of confidentiality depend on who needs to know certain kinds of information.
  - **Regulation**: Regulatory categories are assigned based on external rules, such as laws, treaties, customs agreements, and industry regulations. Regulatory information is shared on an “allowed-to-know” basis.
- Any data set can only have one confidentiality level, established based on the most sensitive item in the data set.
- Regulations, on the other hand, are additive, and any data set may have multiple regulatory restrictions.
- These two concepts must form the basis of any user entitlement.

4.12.1 Confidential Data

- Confidentiality levels range from low to high. Typical classifications include:
  - **For general audiences**: Information available to everyone, including the public.
  - **Internal use only**: Information limited to employees or members, but with minimal risk if shared.
  - **Confidential**: Information that cannot be shared outside the organization without a non-disclosure agreement or similar in place.
  - **Restricted confidential**: Information limited to individuals performing certain roles with the “need to know”.
  - **Registered confidential**: Information so confidential that anyone accessing the information must sign a legal agreement to access the data and assume responsibility for its secrecy.

4.12.2 Regulated Data

- Each enterprise, of course, must develop regulatory categories that meet their own compliance needs.
- These categories should be combined into regulatory “families”, based on the similarity of their protective actions.
- Some sample regulatory families include:
  - **Personal Identifiable Information (PII)**:
    - PII includes any information that can personally identify an individual.
    - GDPR separates *personal data* such as name and personal identity numbers from *sensitive personal data* such as ethnicity, political opinions and health data. Sensitive personal data has a stronger protection in GDPR.
    - Integritetsskyddsmyndigheten (https://www.imy.se) is the swedish agency responsible for protecting person data.
  - **Financially Sensitive Data**: All financial information related to an organization, that has not yet been reported publicly.

4.12.2.1 GDPR

- The General Data Protection Regulation (GDPR) came into effect in May 2018.
- It should be fairly known and followed by most organizations by now.
- Article 5 of GDPR sets out seven principles that permeats the entire legislation. These are:
  - Lawfulness, fairness, and transparency,
  - Purpose limitation,
  - Data minimisation,
  - Accuracy,
  - Storage limitation,
  - Integrity and confidentiality, and
  - Accountability.

4.12.2.2 AI Act

- The EU AI Act will become applicable in the summer of 2025.
- The AI Act introduces a uniform framework across all EU countries, based on a forward-looking definition of AI and a risk-based approach:
  - **Minimal risk**: Most AI systems such as spam filters and AI-enabled video games face no obligation under the AI Act, but companies can voluntarily adopt additional codes of conduct.
  - **Specific transparency risk**: Systems like chatbots must clearly inform users that they are interacting with a machine, while certain AI-generated content must be labelled as such.
  - **High risk**: High-risk AI systems such as AI-based medical software or AI systems used for recruitment must comply with strict requirements, including risk-mitigation systems, high-quality of data sets, clear user information, human oversight, etc.
  - **Unacceptable risk**: For example, AI systems that allow “social scoring” by governments or companies are considered a clear threat to people’s fundamental rights and are therefore banned.

4.12.2.3 Data Act

- The EU Data Act will become applicable in September 2025.
- It does not regulate personal data, but rather enhances data sharing and enables a fair distribution of the value of data.
- Raw and pre-processed data that are readily available to a data holder as a result of the manufacturer’s technical design are subject to mandatory data-sharing obligations.
- Data Act might have a similar effect on overall data quality as GDPR had for the handling of personal data.

4.12.3 Industry or Contract-based Regulation

- Some industries have specfic standards for how to record, retain, and encrypt information.
- The most known example is the Payment Card Industry Data Security Standard (PCI-DSS) which applies to all organizations that accept or process payment cards.

4.13 System Security Risks

- System security risks include elements that can compromise a network or database.
- These threats allow legimitate employees to misuse information, either intentionally or accidentally, and enable malicious hacker access.

4.13.1 Abuse of Excessive Privilege

- The principle of least privilege should be applied when granting access to data.
- Users may be granted access to more data than they need, simply because it is challenging to manage user entitlements. As a result, many users receive generic default access privileges that exceed their job requirements.
- This leads to the risk of users intentionally or accidentally abuse their privileges.
- The solution is *query-level access control*, where accesses are not granted on database table level, but on specific columns and rows within the tables.
- The process of defining query-level access controls are time consuming and need to be reviewed and updated. Automated tools are commonly used, and cloud based services generally provide those.

4.13.2 Abuse of Legitimate Privilege

- Users may also abuse legitimate privileges for unauthorized purposes.
- An example may be a health care worker who has legitimate access to patient records, who decides to export the data and sell it, or use it for malicious purposes.
- Another example may be a teacher who exports student information from a secure location onto their laptop for easier access. Once this sensitive data exists on a device, it becomes vulnerable to theft and loss.
- A partial solution to this problem would be to not only enforce access policies based on user entitlement, but also based on time of day, device location and amount of data.
- This would avoid situations where employees export more data than they actually need, just to “save time”.

4.13.3 Unauthorized Privilege Elevation

- Attackers may take advantage of database platform software vulnerabilities to elevate access privileges.
- These vulnerabilities occur in stored procedures, built-in functions, and SQL statements.
- Privilege elevation exploits can be prevented by combining query level access controls with an intrusion prevention system, IPS. An IPS inspect database traffic and identifies patterns that correspond to know vulnerabilities.
- If a request is flagged by both the access control and IPS systems, an attack is almost certainly occuring.

4.13.4 Platform Intrusion Attacks

- An IPS is usually implemented alongside an intrusion detection system (IDS).
- The most common example of an intrusion protection system is a firewall, but as technology advances, this is no longer sufficient in itself.
- Database platform vendors release security patches to their software as new vulnerabilities are detected, but many organizations update their systems according to periodic maintenance cycles rather than as sson as a patch is released. This leaves the database unprotected until the update is applied.

4.13.5 SQL Injection Vulnerability

- SQL injection is when an attacker inserts (or “injects”) unauthorized database statements as part of an otherwise legal query.
- These statement can be used to elevate privileges or make unauthorized changes to the data.
- Always sanitize all inputs before passing them to the database.
- The XKCD webcomic by Randall Munroe once described the risks with SQL injections, as seen in Figure 2 on the next slide.

4.13.5.1 The Risks of SQL Injection

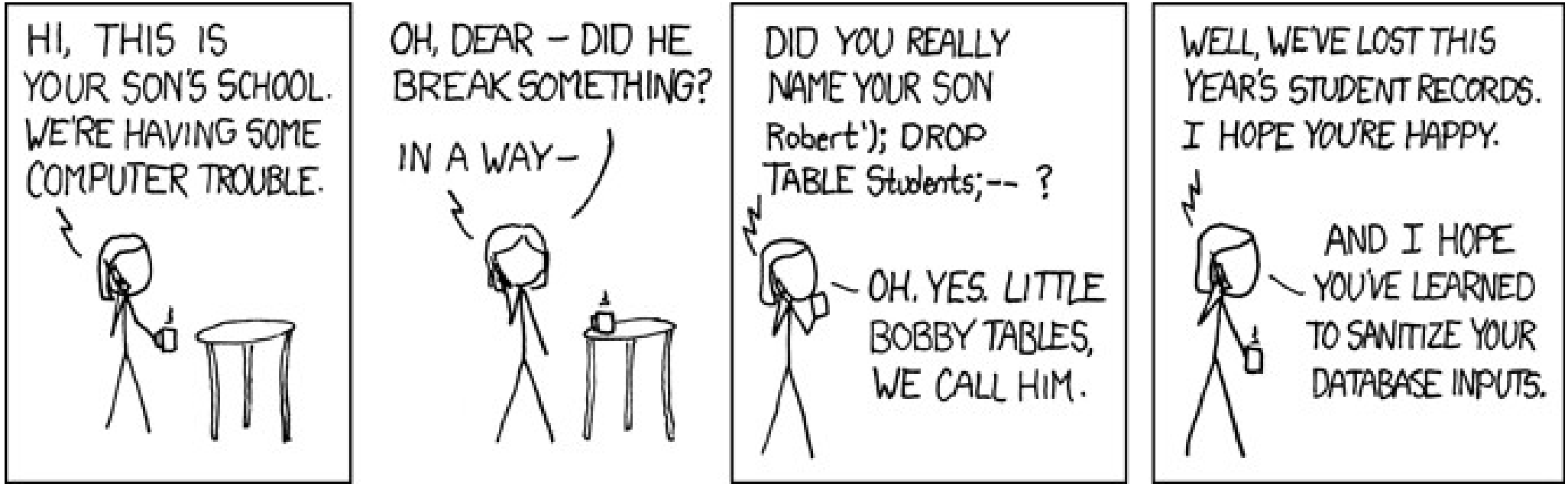

Figure 2: https://xkcd.com/327/

4.14 Social Threats to Security/Phishing

- *Social engineering* refers to how malicious hackers try to trick people into providing them with information or access. They use information about the organization to convince other employees that they have legitimate requests.
- *Phishing* refers to a phone call, instant message or email meant to lure recipients info giving out valuable or private information.

4.15 Malware

- *Malware* refers to any malicious software created to damage, change, or improperly access a computer or network.
- Viruses, worms, key loggers, and adware are all examples of malware.

5 Activities

- Activities for implementing security controls include:
  - Identifying requirements,
  - Assessing the current state,
  - Implementing security tools and processes, and
  - Auditing data security measures to ensure they are effective.

5.1 Identify Data Security Requirements

5.1.1 Business Requirements

- The business needs of an enterprise, its mission, strategy and size, as well as the industry it belongs to, define the degree of rigidity required for data security.
- Analyze business rules and processes to identify security touch points.

5.1.2 Regulatory Requirements

- Some regulations that have impact on data security have been mentioned earlier in this lecture.
- Create a central inventory of all relevant data regulations, and the data subject area (IT, Sales, HR, etc.) affected by each regulation. Add links to the corresponding security policies, and the controls implemented.
- This inventory should be in a format that is easiy updated, as regulations, policies, and data will change over time.
- Table 1 on the next slide shows a simple example of a regulation inventory table.

5.1.2.1 Regulation Inventory Table

Regulation	Subject Area Affected	Security Policy Links	Controls Implemented

Table 1: Sample Regulation Inventory Table

5.2 Define Data Security Policy

- Data security policies should be based on business and regulatory requirements.
- A policy is a statement of a selected course of action, and a high-level description of desired behavior to achieve a set of goals.
- For policies to have a measurable impact, they need to be auditable and audited.
- Defining security policies require collaboration between IT security administrators, security architects, data governance committees, data stewards, internal and external audit teams, and the legal department.
- Data stewards must also cooperate with business managers, who have the data expertise, to develop metadata catalogs and apply proper security classifications.

5.2.1 Security Policy Contents

- Different levels of policies include:
  - **Enterprise Security Policy**: Global policies for employee access to facilities and other assets, email standards, security access levels, and security breach reporting policies.
  - **IT Security Policy**: Directory structure standards, password policies, and an identity management framework.
  - **Data Security Policy**: Categories for individual applications, database roles, user groups, and information sensitivity.
- IT and data security policies are often combined. It is, however, preferable to separate them, as data security policies are more specific and require different controls and procedures.
- The data security policy should be approved and reviewed by the data governance council, and owned and managed by the data management executive.
- Employees need to understand and follow security policies. Compliance should be easier than non-compliance, and the reasons behind the policies should be cleary defined.

5.3 Define Data Security Standards

- Policies are guidelines for behavior.
- Standards supplement policies and provide additional detail on how to meet the intentions of the policies.
- One example could be a policy stating that passwords must follow guidelines for strong passwords, and a standard that describes what a strong password is. This policy is enforced through technology that prevents passwords from not meeting the standard.

5.3.1 Define Data Confidentiality Levels

- Store confidentiality classifications in the metadata directory that guide how users are granted access privileges.
- Any classification method should be clear and easy to use, and will contain a range of levels, such as described in Section 4.12.1 on slide 57.

5.3.2 Define Data Regulatory Categories

- As previously mentioned, there are many regulations around the world that are data-specific.
- It is often a good idea to group these together, such that all regulations that cover the handling of PII, for example, are managed by the same policy category.
- When confidentiality levels and regulatory categories are combined in the metadata repository, all employees handling data know the sensitivity of the data they are handling, transmitting, and authorizing.

5.3.3 Define Security Roles

- Data access control can be organized at an individual or group level.
- For smaller organizations, the individual level may be acceptable, but larger organizations will benefit from role-based access control.
- Role groups enable security administrators to define privileges by role, and grant these privileges by enrolling users in the appropriate role group.
- Try to assign each user to one group only, creating different user views if necessary.
- User identity and role group data should be managed centrally, and any changes tracked.
- There are two ways to define and organize roles: as a grid (starting from the data), and in a hierarchy (starting from the user).

5.3.3.1 Role Assignment Grid

- A grid can be useful for mapping out access roles for data, based on data confidentiality, regulations, and user functions.
- Table 2 on the next slide shows a simplified example.

5.3.3.2 Role Assigment Grid Example

	Confidentiality Level		
	General Audience	Client Confidential	Restricted Confidential
Not Regulated	Public User Role	Client Manager Role	Restricted Access Role
PII	Marketing Role	Client Marketing Role	HR Role
PCI	Financial Role	Client Financial Role	Restricted Financial Role

Table 2: Role Assignment Grid Example

5.3.3.3 Role Assignment Hierarchy

- The other way is to create group definitions at a workgroup or business unit level.
- A simple example is shown in Figure 3 on the next slide.

5.3.3.4 Role Assignment Hierarchy Example

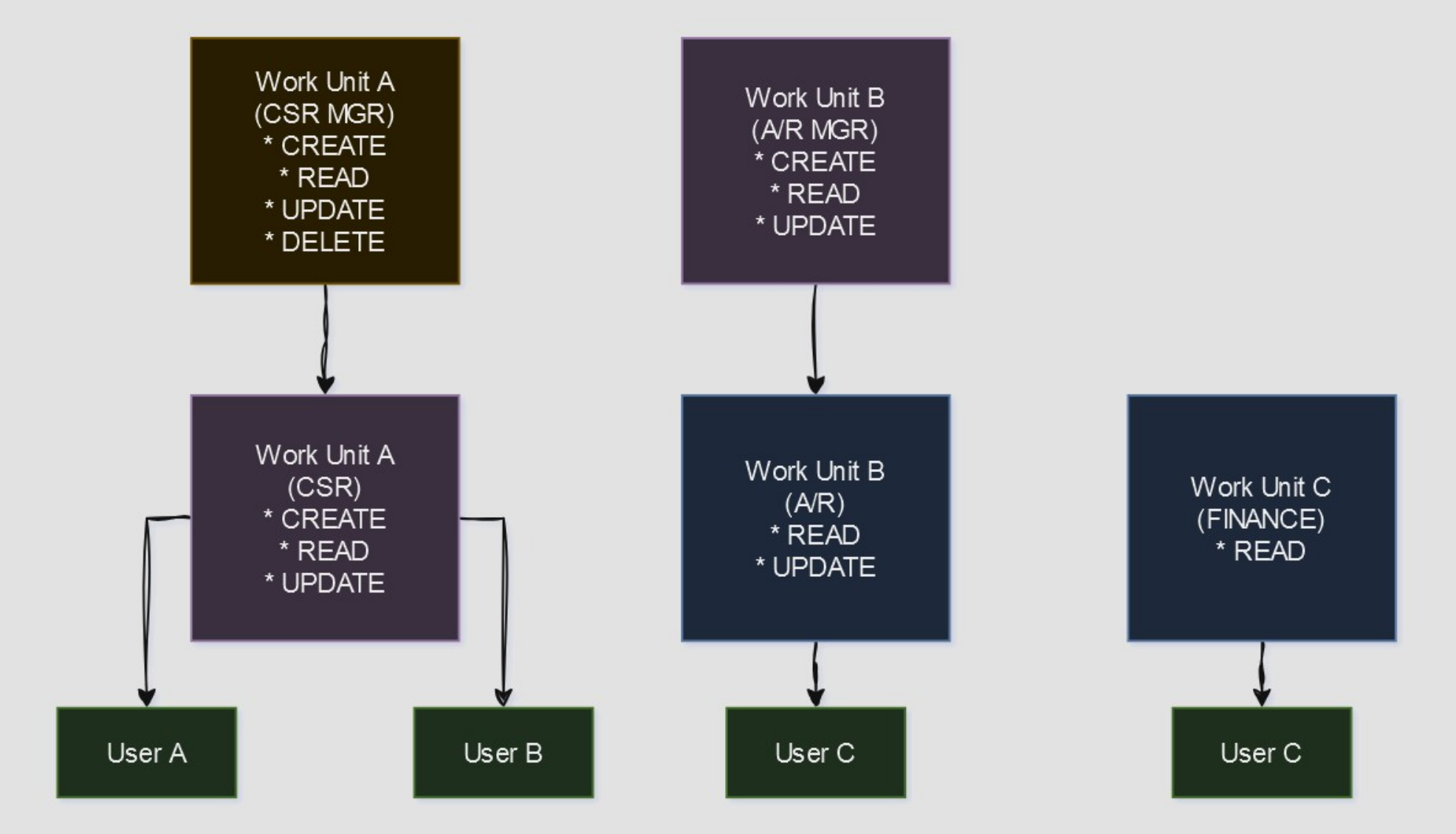

Figure 3: Role Assignment Hierarchy Example

5.3.4 Assess Current Security Risks

- Security risks include elements that can compromise a network and/or a database.
- Evaluate each system for:
  - The sensitivity of the data stored or in transit,
  - The requirements to protect that data, and
  - The current security protections in place.
- Document the findings as a baseline for future evaluations, as well as proof for compliance.
- Gaps must be remediated and the impact of improvements measured and monitored.

5.3.5 Implement Controls and Procedures

- Implementation and administration of data security policy is primarily the responsibility of security administrators, in coordination with data stewards and technical teams.
- Controls and procedures should at least cover:
  - How users gain and lose access to systems and/or applications,
  - How users are assigned to and removed from roles,
  - How privilege levels are monitored,
  - How requests for access changes are handled and monitored,
  - How data is classified according to confidentiality and applicable regulations, and
  - How data breaches are handled once detected.
- Some level of management must formally request, track, and approve all initial authorizations and subsequent changes to user and group authorizations.

5.3.6 Assign Confidentiality Levels

- Data stewards are responsible for assigning confidentiality levels for data, based on the organization’s classification scheme.
- The classification for documents and reports should be based on the highest level of confidentiality for any information found in the document.

5.3.7 Assign Regulatory Categories

- A classification scheme must be created, or adopted, to ensure that regulated data is handled in compliance.
- This scheme provides the foundation when responding to internal or external audits.

5.3.8 Manage and Maintain Data Security

- When all requirements, policies and procedures are in place, the main task is to ensure that security breaches do not occur, and if they do, to detect them as soon as possible.
- Continual monitoring and auditing of security systems is crucial to preserving data security.

5.3.8.1 Control Data Availability

- Managing user entitlements requires an enterprise data model where sensitive data is categorized.
- Data masking, or in some cases encryption, can protect data even if it is inadvertently exposed.
- Relational database views can be used to restrict access to certain rows and/or fields.

5.3.8.2 Monitor User Authentication and Access Behavior

- Monitoring entails a wide range of activities over a wide range of levels, from spanning across several systems down to certain data sets, users or roles.
- Monitoring can be automated, executed manually, or a mix between the two.
- Some automated monitoring should be part of any database deployment. Risks for unmonitored systems include:
  - **Regulatory risk**: Many regulations, GDPR included, require monitoring.
  - **Detection and recovery risk**: If there is a security breach, audit data can be used to find the violation as well as links to users involved, and may guide in repairing the system.
  - **Administrative and audit duties risk**: Users with administrative access to a database may be able to turn off monitoring to hide unauthorized activity. Auditing duties should preferably be separate from database administrators and database server support staff.

6 Tools

- The tools required for managing data security depends on the organization, the data architecture, and the data that is handled.
- Kinds of tools include:
  - Anti-virus software,
  - Employing HTTPS to encrypt data sent over the Internet,
  - Identity management technology,
  - Intrusion detection and prevention software,
  - Firewalls,
  - Data masking or encryption, and
  - Metadata tracking.

6.1 Metadata Tracking

6.1.1 Example: Apache Atlas

- https://atlas.apache.org/2.3.0/index#/
- One example of a data governance and metadata tracking tool is the open source Apache Atlas framework, developed and maintained by the Apache Software Foundation.
- Apache Atlas is closely integrated with the security framework Apache Ranger.
- You can find more information about how to use Apache Atlas on the link on the bottom of the slide.

6.1.2 Example: Microsoft Purview

- https://www.microsoft.com/en-us/security/business/microsoft-purview
- Another example of a data governance and metadata tracking tool is Microsoft Purview, a part of Microsofts Azure framework.
- You can find more information about how to use Purview on the link on the bottom of the slide.

7 Techniques

- Techniques for managing data security include:
  - A process for installing security patches as quickly as possible on all machines. Users should not be able to delay this update.
  - Using an enterprise data model to record security attributes in metadata,
  - Defining clear metrics for evaluating security measures,
  - Identifying security needs when planning new projects,
  - Implementing efficient search of encrypted data, and
  - Document sanitization, such as cleaning metadata from documents before sharing.

7.1 Metrics

- Focus on actionable metrics. It is easier to manage a few key metrics in organized groups, than pages of seemingly unrelated indicators.

7.1.1 Security Implementation Metrics

- Some examples of measurable metrics include:
  - Percentage of enterprise computers having the most recent security updates installed,
  - Percentage of computers having up-to-date antivirus software running,
  - Percentage of employees scoring more than 80% on annual security practices quiz,
  - Percentage of business processes successfully tested for disaster recovery, and
  - Percentage of audit findings that have been successfully resolved.

7.1.2 Security Awareness Metrics

- These areas should be considered when selecting appropriate security awareness metrics:
  - **Risk assessment findings** provide qualitative data to be fed back to appropriate business units, to make them more aware of their accountability,
  - **Risk events and profiles** identify unmanaged exposures that need correction,
  - **Formal feedback surveys and interviews** identify the level of security awareness in the organization, and
  - **Incident post mortems, lessons learned and victim interviews** provide a rich source of information on gaps in security awareness.

7.1.3 Data Protection Metrics

- When selecting data protection metrics, consider:
  - **Criticality ranking** of specific data types and information systems,
  - **Annualized loss expectancy** of mishaps, hacks, thefts or disasters related to data loss, compromise, or corruption.
  - **Risk of specific data losses** related to certain categories of regulated information,
  - **Threat assessments** performed based on the likelihood of an attack against certain valuable data resources, and
  - **Vulnerability assessments** of specific parts of the business process.

8 Implementation Guidelines

8.1 Readiness Assessment/Risk Assessment

- Data security is deeply connected with organizational culture. Building awareness and understanding of security requirements, policies and procedures is the best way to avoid data security breaches.
- Ways to improve awareness include:
  - Training,
  - Consistent policies,
  - Measure the benefits of security,
  - Set security requirements for vendors,
  - Build a sense of urgency, and
  - Ongoing communications.

8.2 Organization and Cultural Change

- One big challenge when developing data security policies is balancing risks with ease of access.
- Technical solutions must be in place, but in most organizations, the behavior of both management and employees will need to change if they are to successfully protect their data.
- Implementing data security measures without regard for the expectations of customers and employees can result in employee dissatifaction, customer dissatisfaction, and organizational risk.
- Well-planned and comprehensive security measures should make secure access easier for stakeholders.

9 Sources

- Sources not already mentioned in the lecture:
  - DAMA DM-BOK Chapter 7
  - Eryurek, et. al: Data Governance: The Definitive Guide (Chapter 7).
  - https://cryptography.io