**Data Extraction**

**Data Protection Fortification**

|  |  |
| --- | --- |
| **Meta data** | |
| Date |  |
| Number of participants |  |
| Distribution of roles participating |  |

|  |  |  |
| --- | --- | --- |
| **1 - Discuss Data Source** | | |
| **General Questions** | | |
| ***Examples*** | Data source in focus |  |
| *Weather, prices, customer reviews* | What describes the data we get from this data source? |  |
| *What is the name of the provider? External to company?* | Who/What provides this data source? |  |
| *Fetching from API? Importing from database?* | How do you receive the data? |  |
| *Used in processing, aggregated, presented directly to end user, used in machine learning* | How is the data from this data source used in your services or products now, or will be in the future? |  |
| *Internal analysts, customers, customer support* | Who are the end-users for this data? |  |
| *Fetched when user triggers a service (clicks to view details about an order), data is downloaded by a CRON-job every night* | How often is data received from this data source? |  |
|  | What is the volume of the data received? |  |
| *Could the values be “anything”?* | Do you have an explicit schema or API-contract for data coming from this data source? |  |
| *Is it known who puts the data into this data source?* | How is the data coming from this data source generated? |  |
| **Questions for Security Implications** | | |
|  | How do you authenticate with this data source?  *Who is responsible for authenticating users trying to access this data source?* |  |
|  | Who has access to change the access URL used to connect with the data source?  *Where is this access URL stored? Is a change to this access URL logged?* |  |
| *Person Identifiable Information (PII), concerns business opportunities, data that could affect stock prices, data that could affect decision making processes* | How sensitive is the data in this data source? |  |
| *Analyst decisions, planned maintenance, conceal information or impact repudiation, company damage* | What could an attacker be interested in influencing through this data source? |  |
| *Denial of Service.*  *Users are denied of accessing a digital voting platform.*  ***Access to this platform*** *is only critical during the days where voting is open.* | What could the consequences be if the data source was no longer available, or parts of the data was missing?  *Are there certain times where the consequence would be greater?* |  |
|  | Who are the possible threat actors for this data source? |  |
| *Service availability monitoring (health checks). Users would discover unavailability or inconsistencies and report back to us* | Would you discover if the data from this data source is incorrect, or if the data source is unavailable?  *How would you report unavailability or inconsistencies to the suppliers of this data source?* |  |
|  | Other |  |

|  |  |
| --- | --- |
| **2 - Prioritization of Data Fields** | |
| **Estimate Value** | |
| Max value  Min value  Remaining values |  |
| **General discussion** |  |
| **Estimate Likelihood of Tampered Values** | |
| Max value  Min value  Remaining values |  |
| **General discussion** |  |
| **Prioritization Matrix** | |
| **Does the result shown in the matrix make sense?**  Should it be rearranged? |  |

|  |  |
| --- | --- |
| **3 – Identify Security Measures** | |
| **Evaluate security measures discussion**  Is the current handling of the data source sufficient?  In what ways can we fortify the security of handling the data source?  *Examples:*  *Validation (type-checking? min/max values? min/max length?)*  *Monitoring (anomaly detection, correlation, auditing)*  *Other measures that can decrease the risk?* |  |

|  |  |
| --- | --- |
| **Evaluation of Session** | |
| **Feedback from practitioners**  How did you feel like the session went? |  |